Journal of Official Statistics
Vol. 8. No. 2, 1992, pp. 241–246

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.

Särndal, C.-E., Swensson, B., and Wretman, J., Model Assisted Survey Sampling. Springer-Verlag, New York, 1992. ISBN 0-387-97528-4. xv + 694 pp. $49.00.

Survey sampling is a universally accepted approach for information gathering in academia, the private sector, and government. Therefore good textbooks are needed at all levels.

The third edition of Cochran's classical text *Sampling Techniques* appeared in 1977. In the same year, Cassel, Särndal, and Wretman (CSW in what follows for book and authors) gave a full coverage of the fundamental issues underlying the statistical inference theory approach to survey sampling, developed by Godambe and others in the 1960s. Theirs was the first textbook to systematically extend traditional survey sampling theory, using a statistical inference outlook. Now 15 years later, Särndal and Wretman are back with a new book on survey sampling, this time co-authored with Swensson (still a Swedish troika though, SSW for simplicity).

The book concentrates on *estimation theory*; we are not given a comprehensive theory for the entire survey operation. The objective is to give an up-to-date account, paying particular attention to recent developments. The reference list bears eloquent witness; there are some 300 items, a major part from the 1980s. The book adopts a unified perspective, that of the *randomization approach* with minimal assumptions concerning the population. Design-based inference ("fixed population, repeated sampling") is the declared principal theme of the book, i.e., inference about the finite population itself, its current state. SSW is thus different in structure from the comparison book CSW, where foundational aspects of inference theory were in focus. There is another fairly recent comparison book by Chaudhuri and Vos from 1988. In this book, besides inference theory, the issue is the search for a perfect selection scheme for sampling with unequal probabilities without replacement; in other words, there is little overlap.

The title says model *assisted*, not model-*based*. Statistical modeling is essential, SSW argue, both in the sampling process and in the choice of estimation technique. So another declared ambition is to present the systematic use of auxiliary information in survey sampling. An entire section of the book – Part II – is devoted to the topic of estimation through linear modeling; it contains much new material published in the 1980s. The use of a superpopulation model, important in model-based inference as

developed by Royall and others, is briefly reviewed by SSW, but not a central theme of the book.

The book is logically and cohesively constructed, with its material grouped into four main sections. Part I contains basic designs and estimation principles. There we find the familiar Horvitz-Thompson estimator, called the π-estimator by SSW. It is introduced in Chapter 2 and used repeatedly throughout the book. As was the case in CSW, considerable effort is expended in developing a consistent body of terminology and a flexible system of notation. Part II, as already mentioned, introduces linear models and auxiliary information. Part III provides an exposition of some additional (but not unimportant) material, such as domain and variance estimation. Part IV finally, provides a "broader view of errors in surveys," meaning essentially non-sampling errors.

The book is too long for a detailed review, though here are some of my impressions.

The section on the use of auxiliary information is impressive indeed, and in this area there is much recent progress to report. SSW take knowledge of the population scatter as a rationale justifying the use of regression estimates, not substantive knowledge of what is actually driving the data. Remember: this is the randomization approach, parsimonius on assumptions. We are assured, however, that estimation is robust even in the face of model-breakdown. New results on variance estimation are included in Section III. This important issue is gone into quite thoroughly. Small area sampling and estimation, are discussed within the framework of synthetic estimation. In Section IV, I was particularly pleased to see a whole chapter (44 pages) dealing with nonresponse and response meachanism models. However, problems occurring in surveys dealing with subjective and quasi-factual phenomena, as are frequent in private sector surveys, are not discussed.

Numerical illustrations and examples are taken from Canada and Sweden. The utility of the book is enhanced by numerous exercises with most answers in the back of the book.

I am uncertain of the audience for this book, e.g., where it fits into a department curriculum. The comprehensiveness of the text, exercises, and examples cause the book to be very long, 700 pages. In a normal undergraduate course, the instructor will have difficulties in finishing the book. SSW express the (ubiquitous) hope that "the book will be accessible to a wide audience," including undergraduate students. My own experience leaves me somewhat doubtful of this. I found the book tough going at certain points. It requires of the non-mathematically inclined reader great tolerance of mathematical notation, style, and arguments.

If I have a reservation, then it is the somewhat dry way of presenting the material: invariably numbered results with estimator, variance and variance estimator followed by numbered examples and remarks. Also I am not altogether comfortable with the SSW notation, but one will probably get used to it after a while.

In ending, however, I must admit to envious appreciation of the authors' achievement, a formidable work indeed! I can strongly recommend the book as a reference work and as the text for an advanced master's level or doctoral level course. It will provide the technical tools and the methodological approach necessary for undertaking original research. Another attractive feature: at 49 USD the book is a bargain!

References

Cassel, C.-M., Särndal, C.-E., and Wretman, J.H. (1977). Foundations of Inference in Survey Sampling. New York: John Wiley.

Chaudhuri, A. and Vos, J.W.E. (1988) Unified Theory and Strategies of Survey Sampling. Amsterdam: North-Holland.

Sven Berg
Lund University
Lund
Sweden

Särndal, C.-E., Swensson, B., and Wretman, J., Model Assisted Survey Sampling. Springer-Verlag, New York, 1992. ISBN 0-387-97528-4. xv + 694 pp., $49.00.

This recently published book on sampling theory is refreshing in many ways. While reviewing it, I compared its material with some of the other textbooks on sampling; e.g., Cochran (1977), Hedayat and Sinha (1991), Raj (1968), and Sukhatme and Sukhatme (1970). I found that this book contains a fair amount of material not found in any of the other books. It contains more than the usual number of examples of how the theory is applied under practical constraints, including considerations such as non-response, measurement error, quality evaluation, coverage and frame errors, and so on. It has been written by three authors who have had extensive practical survey experience as well as having an excellent knowledge of the theory of sampling. A better title might have been "A Practitioner's Guide to the Theory of Sampling from Finite Populations."

This book has many uses. It can be used for a one- or two-semester university course in sampling. Each chapter contains problem exercises to reinforce the chapter's concepts. As well, some small finite populations based on real data are given in the appendices for use with some of the exercises. The book is an excellent reference for practising survey methodologists and should be read by anyone who wants a refresher course in recent developments in the theory and practice of sampling from finite populations.

I would recommend that this book be in the office of every survey methodologist. He/she will find it a very handy reference since most of the estimators used in real life problems are discussed. This includes ratio and regression estimators for totals, as well as more sophisticated estimators such as poststratification under general sampling plans. For the more specialized topics, which are not adequately covered due to space limitations, an excellent set of references is given.

The level of mathematical knowledge required to read this book is not very advanced, but to fully understand many of the proofs, the reader does need to have good mathematical abilities.

The main theme of this book is on effective use of auxiliary data for making inferences from finite populations. However, this can be misleading, since much of the book is also devoted to the simpler case where there are no auxiliary data. This case is discussed in detail in order to set the stage for the later sections.

The book has been divided into four main parts. These are entitled:

I. Principles of Estimation for Finite Populations and Important Sampling Designs (five chapters),
II. Estimation through Linear Modeling, Using Auxiliary Variables (three chapters),
III. Further Questions in Design and Analysis of Surveys (five chapters),
IV. A Broader View of Errors in Surveys (four chapters).

Part I covers all of the standard material found in any textbook on sampling. However, since space is limited, some of the topics are covered more briefly than I would have liked. For example, some of the alternative methods for sample selection were de-emphasized. I would have preferred to see a bit more discussion on designing surveys for several occasions, including some of the recent developments in collocated or co-ordinated sampling, which has been studied by Statistics Sweden and the Australian Bureau of Statistics.

I found the presentation easy to read. I particularly liked the first chapter. It gives an excellent overview of probability surveys and a good account of some of the non-sampling issues, including not only the usual types of non-sampling errors but also a mention of the use of administrative data and of privacy considerations. The discussion on sampling frames is excellent.

Unequal probability sampling and the Horvitz-Thompson estimator is introduced as early as Chapter 2. Recent developments in methods to draw unequal probability

samples are included in Chapter 3. Chapter 5 includes a discussion of methods of estimating variances or confidence intervals for non-linear statistics, including ratios, population variances and medians.

Part II introduces and exemplifies the main theme of the book. A unified approach to design-based estimation through the use of linear models is given. This is based on a number of papers in the literature by these authors and their colleagues. I am very pleased to see that these points are now available in a textbook format, as it widens the accessibility of this excellent theory to a broader audience. The authors' sensitivity to practicalities is apparent when they discuss the trade-offs between simplicity and efficiency gains.

Part III covers some material in much more depth than other sampling textbooks. The section on two-phase sampling with auxiliary data is particularly well covered. Domain estimation is also discussed comprehensively. Alternatives to variance estimation are covered briefly. Model-based optimal designs are discussed. There is a good introduction to data analysis from complex surveys.

Part IV covers aspects of practical problems which are not usually covered as comprehensively in other sampling textbooks. Sources of non-sampling errors are reviewed. Under "data processing," I would have preferred a better account of detection and treatment of outliers, but this field is still not well developed in the literature. The chapter on non-response contains one of the best discussions ever written in a textbook. Measurement error models are also introduced. The section on quality declarations seems out of place in this book since it does not go into sufficient depth to do it justice.

In summary, this would not only be an excellent textbook for the student of sampling theory, but it would also become an invaluable addition to your book collection, if you are involved with the theory or practice of survey sampling.

References

Cochran, W.G. (1977). Sampling Techniques, 3rd Edition. New York: John Wiley & Sons.

Hedayat, A.S., and Sinha, B.K. (1991). Design and Inference in Finite Population Sampling. New York: John Wiley & Sons.

Raj, D. (1968). Sampling Theory. New York: McGraw-Hill.

Sukhatme, P.V., and Sukhatme, B.V. (1970). Sampling Theory of Surveys with Applications. Ames, Iowa: Iowa State University Press.

David A. Binder
Statistics Canada
Ottawa, Ontario
Canada

Särndal, C.-E., Swensson, B., and Wretman, J., Model Assisted Survey Sampling. Springer-Verlag, New York, 1992. ISBN 0-387-97528-4. xv + 694 pp., $49.00.

In their spirited defense of randomization based sampling inference, Hansen, Madow, and Tepping (1983) concede a role for models when choosing among sampling designs and estimators. Despite this concession, there has been little movement to explicitly incorporate regression modeling into actual survey practice, at least not within major U.S. statistical agencies and survey organizations. The problem, I think, has been the absence of an authoritative textbook detailing the proper role of models within randomization theory. With the publication of *Model Assisted Survey Sampling*, this void has been splindidly filled.

Part I of *Model Assisted Survey Sampling* provides a very solid introduction to randomization based sampling theory, although two important topics – regression estimation and two-phase sampling are reserved for later discussion. The authors have decided to begin with the unbiased estimator of a population total under an *arbitrary* probability sampling design. This bold approach is not suited for the mathematically squeamish, despite claims to the contrary on the back cover. Nevertheless, it hammers home

the distinction between the radomization based approach to inference advocated in the text and the model dependent approach generally favored in other branches of statistics.

Part II introduces the regression estimator and the models that support its use. This section is not without flaws. Nevertheless, as a practical guide to the use of regression modeling within the randomization framework it is excellent. The text provides many examples of how such models are used in practice (not always knowingly). Moreover, the authors recommend a superior approach to variance estimation than that usually found in survey sampling textbooks. This is not surprising, since the approach comes from Särndal, Swensson, and Wretman (1989).

Part III has some good chapters and some weaker ones. The treatment of two-phase sampling (Chapter 9) is very good as is the discussion of optimal sampling designs (Chapter 12), although the latter could be better connected with the material in Part II. The text's approach to domain estimation (Chapter 10) is of some interest but is unduly limited by its reliance on randomization based inference. This is an area where even Hansen et al. acknowledge a potential role for models in inference.

The authors should have resisted the temptation to include chapters on replication methods and the analysis of survey data (Chapters 11 and 13), which are not among their areas of expertise. The interested reader, however, does receive a good set of references.

Part IV features useful chapters on nonsampling errors. Again, the treatment of nonresponse (Chapter 15), helpful as it will be to many readers, is hampered by the authors' reluctance to stray very far from randomization (or, more accurately, quasi-randomization) based inference. Fortunately, they manage to overcome this self-imposed constraint in their discussion of measurement errors (Chapter 16).

I found some, but not many, typographical errors. More distressing are inconsistencies across chapters. For example, the introduction of Chapter 11 refers to the "weighted residual variance estimator," a term found in Särndal, Swensson, and Wretman (1989) but not previously in the text. Section 7.12 discusses a class of "asymptotically design unbiased" estimators from the literature, but a different term, "asymptotically unbiased," was defined in Chapter 5. Moreover, estimators in this class are more than asymptotically unbiased, they are consistent, which should have been noted.

There are several topics I wish the authors had addressed. The first is the distinction Kish (1979) draws between the target and inferential population. This would be a welcome addition to the otherwise helpful first chapter.

The second topic is the Rao-Samford probability proportional to size (pps) selection scheme, especially given the emphasis in the text on arbitrary sampling designs. The third is the large population properties of certain estimation strategies, in particular, those involving Rao-Samford sampling or systematic pps sampling from a randomly ordered list. A discussion of this topic would provide a natural bridge from the treatment of pps sampling without replacement to pps sampling with replacement. Interestingly, the model based properties of the weighted residual method of variance estimation depends on the population being relatively large, a fact not mentioned in the text. This may explain why the weighted residual variance estimators the authors calculate in section 7.9.1 do not exhibit satisfying coverage properties (the sampling fraction is $100/281 \approx 0.36$, a value too large to ignore).

I would have liked the text to include a discussion of the use of random effects modeling in evaluating multistage sampling designs and in domain estimation. Random effects models *are* mentioned in the chapter on measurement errors, but, like some other important concepts (quasi-randomization for one), they are missing from the subject index.

The authors should take the criticisms of the last several paragraphs as advice for the second edition of what may very well become the standard text on survey sampling theory for years to come. I recommend

it for master's programs and for undergraduate programs that stress mathematics. I say this because there is an abundance of set notation and more than a few matrices. It is also an excellent self-study text for readers with advanced degrees in mathematics or statistics but little or no background in survey sampling. Even those well-versed in survey theory will want to have this book on their shelves as a handy reference. They may even learn a thing or two. I know I did.

References

Hansen, M.H., Madow, W.G., and Tepping, B.J. (1983). An Evaluation of Model-dependent and Probability-sampling Inferences in Sample Surveys. Journal of the American Statistical Association, 78, 776–793.

Kish, L. (1979). Populations in Survey Sampling. Survey Statistician, 1, 14–15.

Särndal, C.E., Swensson, B., and Wretman, J. (1989). The Weighted Residual Technique for Estimating the Variance of the General Regression Estimator of the Finite Population Total. Biometrika, 76, 527–537.

Phillip S. Kott
National Agricultural Statistics Service
Washington, DC
U.S.A.

Printed in June 1992.