Journal of Official Statistics, Vol. 20, No. 2, 2004, pp. 411–415

Book and Software Reviews

Books for review are to be sent to the Book Review Editor Jaki S. McCarthy, USDA/NASS, Marketing and Information Services Office, Room 5305 South Building, 1400 Independence Ave., SW, Washington, DC 20250-2014, U.S.A.

Paul Knottnerus. *Sample Survey Theory: Some Pythagorean Perspectives.* New York: Springer-Verlag, 2003. ISBN 0-387-95407-4. 417 pp, EUR 74.95, 69.95 USD.

The book provides rather detailed information on methods for estimating variances and correlations between values of the target variable measured on cases selected using different sampling designs. I believe it is an excellent reference for persons involved in teaching or carrying out survey data analysis, as well as those more specifically concerned with writing statistical computing programs for estimating variances and correlations for data collected using the sampling designs. The manner in which material is presented allows the reader to compare different estimation methods.

Admittedly, the text is highly theoretical and an appreciation of calculus and matrix algebra is a requisite for successful comprehension of the content. Nevertheless, very useful examples and results of simulations are provided to illustrate proofs of various theorems.

Of great interest is the manner in which the author incorporates the ancient Pythagorean Theorem to provide solutions to optimization and estimation problems. This theorem traverses the boundaries of sample designs and estimation methods. The book shows that this ancient technique is still very relevant and applicable to modern-day survey sampling theory. Comprehension of the theorem application is facilitated by the graphical illustration of the components of equations solved thereby.

The emphases for estimation in the volume are population means, totals and variances. The Horvitz-Thompson (HT) estimator is commonly proposed and recommended as appropriate for sampling without replacement. The Hansen-Hurwitz estimator for population totals, on the other hand, is deemed suitable for sampling with replacement.

The author propounds variance estimators for the HT estimator of population total as derived by Horvitz-Thompson and by Sen (1953) and Yates and Grundy (1953). Both, expressed as functions of the inclusion probability, are compared in the text. After going through different methods of arriving at the variance estimators the author delves into the

matter of arriving at efficient estimators – minimum variance estimators. The volume details the importance of the sampling autocorrelation coefficient (ρ_z) in variance estimation for standard and complex sampling designs. There is also an exposition of different approaches to derivation of the coefficient.

Prior to the exposition of the main subject matter the reader is allowed a review of important elements of classical statistical distribution theory that are very helpful to appreciation of the remainder of the text. Thus, Chapter 2 – appropriately compiled to set the stage for that which follows – could serve as a very useful reference for introducing undergraduate statistics majors to basic statistical distribution theory as used in sample survey methods and other features of statistical data analysis.

The volume also introduces the role of the hypergeometric model in estimating the finite population correction (*fpc*) factor. The role of the *fpc* in adjusting the relevant variance estimates is also shown. Then *fpc* is required for adjusting estimates obtained through sampling without replacement. An *fpc* factor that makes use of ρ_z is introduced and applied to estimators throughout the monograph.

After dealing with estimators coming from standard sampling designs, the author deals with those based on more complex designs. Estimators based on multistage sampling are demonstrated through incorporation of selection probabilities and ρ_z as given for SRS. This follows from the notion that multistage sampling can be considered as SRS of clusters in stage one and of elements of clusters or of subpopulations in subsequent stages.

A commonality between stratified random sampling and cluster sampling is given as the basis of derivation of a variance estimator for the population total. Stratified sampling can be regarded as a two-stage cluster sampling. As such the variance estimator can be derived by combining the cases obtained by SRS in Stages 1 and 2. Thus the estimator is a function of the total number of persons selected in each stratum as well as the size of the stratum. This, again, allows extension of results from SRS to the case of multi-stage sampling.

Multisurvey sampling is introduced as a means of providing an improved set of estimators. The general restriction estimator is proposed and used to arrive at new values of estimators. Optimal values of estimators are obtained using recursive methods, which involve continuous updating of previous estimates until convergence. Examples of estimation of totals from contingency tables are also presented.

A fitting closure to an excellent piece of work is the exposition of the importance of weighting procedures. The different functions for the weights used to arrive at the estimators for population totals are presented. Their effect on the variance estimators as well as estimates of covariance are also explored. After providing the simple weighting procedures that flow from classical regression theory, other weighting procedures presented are the almost-minimum variance (AMR) procedure, the Lagrangian approach and the Kuhn-Tucker approach to the problem of negative weights. The Pythagorean regression equation for the residuals vector facilitates recursive use of the weighting formula to arrive at the AMR regression estimator for population total. This chapter, I believe, highlights the importance of appropriately accounting for the sampling design in arriving at estimates of parameters from survey data collection.

With this volume as a resource it would be interesting to investigate the methods used in studies and by statistical software algorithms to arrive at estimators for variances, means, population totals and correlation coefficients using data from various sampling designs.

The book opens up opportunities for additional research into application of the methods presented plus further examination and comparison of estimators using "real life" survey data.

Novie Younger
University of West Indies
Tropical Medicine Research Institute, Sickle Cell Unit
Mona, Kingston 7
Jamaica
Phone: 876-977 0746
Fax: 876 927 2984
Email: nyounger@uwimona.edu.jm

Paul P. Biemer and Lars E. Lyberg. *Introduction to Survey Quality.* Hoboken, New Jersey: John Wiley and Sons, Inc. 2003, ISBN 0-471-19375-5, 376 pp. + refs and index, 94.95 USD.

This meticulous and comprehensive journey leads us through the survey process from planning and implementing a survey to managing the survey data. The title must be read broadly because this is a textbook for a survey research class rather than an advanced text on quality measurement and assessment. These authors are to be congratulated for incorporating many topics that we might not expect in a basic survey textbook. In many ways, it is geared to the practice of surveys so that students are better prepared for the real world of survey research. The authors have also targeted the personnel in a survey organization as an audience. Such personnel will gain a broader understanding of the elements that compete in the battle to maximize survey quality within budget constraints.

A regrettable reality is that the world changes over the long period needed to write a textbook and to bring it to press. The changes missed by this book are the growth of web surveys and the complications raised for RDD telephone surveys by new technologies such as caller identification, voice mail, and cell phones. While the text mentions web surveys in a list of other self-administered modes, it does not receive the attention it should in the discussions of survey modes. For example, usability is not in the book's index although usability testing is now standard in developing user-friendly web surveys. Instructors can easily remedy omissions with supplementary readings, and the next edition of the book can add such materials.

If I could change this book, I would give a more positive focus to the respondent's role. Biemer and Lyberg sometimes treat the respondent as the enemy who must be battled to provide quality responses or any response at all. For example, the authors characterize nonresponsive categories such as "don't know" and "no opinion" as "escape routes for respondents" (p. 113). On the contrary, the respondent needs to be treated with more respect and sympathy in order to gain cooperation. Testing with respondents may show that such categories are needed for questions that are inapplicable. A summary sentence in Chapter 4 reads, "As can be seen from this list of issues, the development of clearly understood and easily answered questions requires close collaboration between survey methodologist and survey researchers or subject matter experts." I would prefer that the

respondent be considered a crucial participant throughout the process of questionnaire design.

Another topic that needs a more positive focus is the impact of interviewers on the quality of survey data. For example, the first of four key areas for evaluating interviewer performance is "detection and prevention of falsified information," which is certainly the most negative issue. The usefulness of interviewer debriefings deserves a fuller discussion to show the partnership that is necessary to promote quality. While the authors include professional interviewers as a possible audience for their book, seasoned interviewers might find the tone too negative.

The ten chapters offer a wealth of interesting ideas and information, but the book is not necessarily composed for reading from front to back, especially for a survey course. Chapters 3 and 4 will make a much more interesting start for students than Chapters 1 and 2, which are written in a dry, technical manner. Chapter 1 will hold little initial interest for students as it defines surveys and quality. Chapter 2 is worse with its overview of types of errors with few examples to bring this chapter alive. An instructor should assign pieces of these chapters as supplemental reading with later chapters. One especially interesting illustration contrasts systematic and variable error. This illustration will be more compelling for students in the context of a later chapter.

Chapter 3 deals with nonresponse and coverage errors and shows many ways we can fail to reach our intended respondents. The diverse strategies for minimizing errors include Dillman's design principles for mail questionnaires, incentives for respondents, call-scheduling algorithms, and many more. Chapter 4 is a delightful chapter starting with measurement errors from questionnaires. This discussion includes many examples from establishment surveys. The rest of the chapter explores the response process model and conveys the fundamentals for understanding the respondent's interaction with a survey instrument.

Chapter 5 deals with survey instruments delivered by an interviewer. It includes an intelligent discussion of the debate over standardized and conversational interviewing. Instructors should select carefully among the many topics ranging from the highly technical such as design effects and to the very practical such as deciding on the length of interviewer training.

Chapter 6 reviews the strengths and weaknesses of typical survey modes of data collection and other data collection approaches such as diaries and direct observation. One of the highlights is the careful caveats for using administrative records as an alternative to data collection. Chapter 7 is the chapter that best represents the book's title. As it covers data processing issues, it successfully discusses measurable survey quality and the trade-offs with time and resources.

Chapter 8 focuses on evaluation methods for survey instruments. That chapter should be taught earlier, however, since it covers questionnaire testing. Chapter 9 covers sampling errors with enough formula derivations to put the most eager student to sleep. Instructors must make judicious selections from this chapter to keep students engaged.

As the final chapter, Chapter 10 covers several remaining topics. The most critical topic is balancing costs and survey quality. While this theme occurs throughout the book, it is well worth a final overview. The eclectic Chapter 10 includes two worthwhile reference

lists – the quality guidelines of the American Association for Public Opinion Research and the International Statistical Institute's Code of Ethics for Survey Workers.

This book sets a new threshold for topics to include in examining the survey process. Furthermore, helpful tables, illustrations, and boxed highlights make this an attractive choice for a textbook. Best of all, it includes intelligent debates and appropriate literature references. These insure that the beginning survey researcher will learn to make intelligent choices rather than merely follow prescribed formulas.

Fran Featherston
National Science Foundation
Arlington, VA 22230
U.S.A.
Phone: 703-292-4221
Fax: 703-292-9091
Email: ffeather@nsf.gov