Journal of Official Statistics, Vol. 15, No. 1, 1998, pp. 115-124

Book and Software Reviews

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

Survey Measurement and Process Quality
Johnny Bull
Nonesponse in Household Sulveys
Paul J. Lavrakas
Statistical Reasoning and Methods
Katherine Lind
Dictionary of Statistics and Methodology
Karl R. Landis

L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin, eds. *Survey Measurement and Process Quality.* New York: Wiley, 1997. ISBN 0-471-16559-x. 768pp. 94.95 USD.

In 1995, an international conference on Survey Measurement and Process Quality was held in Bristol, United Kingdom. In addition to a large number of contributed papers, there were thirty-four invited papers which have been collected in the present volume. The papers are divided into five areas: Questionnaire Design, Data Collection, Post Survey Processing and Operations, Quality Assessment and Control, and Error Effects on Estimation, Analyses and Interpretation.

The papers are mainly of two types. The first type consists of articles that deal with a particular aspect of measurement error, from relatively general matters such as context effects (Wänke and Schwarz) or administration mode effects (Jobe et al.) to very specific topics, e.g., issues related to children as respondents (Scott) or differences in cognitive processing among various ethnic or racial groups (Johnson et al.). The second type are chapters which examine measurement and other nonsampling errors from the perspective of the survey as a production process.

Both sets of chapters make useful contributions to the literature or provide excellent summaries of recent findings and current practices. But the unique contribution of the volume is the papers that either treat the entire survey or some component of it as an integrated process. The framework for this approach to quality improvement in surveys is primarily presented in two chapters.

The first of these, is Dippo's Survey Measurement and Process Improvement: Concepts and Integration, the introduction to the section on Quality Assessment and Control. This chapter puts the process improvement approach in historical perspective, examines its basic concepts, and links the process to the goal of precise measurement in at least two important ways: first, in producing the measurements and second, in estimating the error associated with them. Regarding the first matter, Dippo points out that the application of this quality philosophy (pioneered by Deming and others) to improving survey measurement can take such forms as continuous quality improvement and recommended practices. On the second point, she notes that "All of the beautiful mathematics associated with estimating mean squared error or any variance or bias component assume a defined, reproducible process"

In their chapter on Continuous Quality Improvement in Statistical Agencies, Morganstein and Marker provide a very precise framework for understanding, analyzing and improving the production of survey measurements. As they explain, this framework "...moves quickly from focusing on the product (the survey results) to understanding and working with the survey processes and is a never-ending loop that returns to the needs of the survey client," whether internal or external to the organization. Their careful explication and detailed examples are a valuable guide for organizations considering this approach.

The book contains a number of good chapters that focus on particular aspects of surveys from the perspective of process improvement. In "Towards a Theory of Self-Administered Questionnaire Design," one of the most important chapters in the book, Jenkins and Dillman bring knowledge from several disciplines to bear on the specification of instrument design principles. The authors describe the interplay between the 'skills' respondents must call on to successfully complete a typical self-administered instrument, including, but not limited to language, visual perception, and motivation. The authors distill key factors from these areas into a very useful short list of design principles.

The Esposito and Rothgeb chapter, "Evaluating Survey Data: Making the Transition from Pretesting to Quality Assessment," covers nearly the full array of questionnaire testing and assessment tools. They maintain a useful distinction between the use of these methods during instrument development versus "post implementation quality assessment." They use the redesign of the U.S. Currrent Population Survey as a case study to demonstrate the integration of many of these methods into an efficient design and testing process.

The De Leeuw and Collins overview of Data Collection Methods and Survey Quality is a concise summary of many issues in the data collection process. The chapter places the new computer-assisted methods in the context of the overall data collection process. This serves nicely as a lead-in to Nicholls et al. who discuss the very important issue of "The Effect of New Data Collection Technologies on Survey Data Quality." Couper et al. and Blyth also address aspects of the effects of technological innovations on data collection. Any serious approach to quality improvement must be both cognizant of and have methodologies to evaluate these rapidly emerging tools.

Conrad's chapter, "Using Expert Systems to Model and Improve Survey Classification Processes," gives both a very good overview of expert systems and clear examples of their application. He notes two measures for the evaluation of such systems: the production rate and the agreement rate. If these rates are equal to or better than those for manual coding and if the personnel costs, over time, are lower, the system is judged superior. Though he notes that this may be difficult to evaluate over a number of different surveys for which a particular expert system might be used. One wonders whether such systems might also have a role in improving the cost effectiveness of training coders and, perhaps, interviewers, who could gain a better sense whether elicited responses are codeable. The Granquist and Kovar chapter, "Editing Survey Data: How Much Is Enough," examines an often neglected part of the survey process and is one of the few papers to integrate cost considerations into their discussion.

In an insightful chapter, "Questionnaire Design: The Rocky Road from Concepts to Answers," Schwarz rightly notes that "...the research objectives of many studies are surprisingly ill-defined" and describes some common consequences of that lack of focus. He then provides a concise discussion of the standard survey response process model and its role in recently developed pretest methods. The chapter goes on to examine alternative data collection modes from the perspective of the cognitive tasks they present as well as factors stemming from social interaction between respondent and interviewer.

The chapter by Hox is concerned with "specific techniques to structure the process of going from theoretical concept to prototype survey question." In these chapters and elsewhere, the notion of an orderly progression from concepts and research questions to survey questions is advanced. Ideally, this may be the preferred approach.

In practice, the process of refining concepts, research questions and survey goals often does not become clear until designing the questionnaire. The careful researcher - who may not have done sufficient preliminary thinking - is often forced by the very process of writing (or selecting) questions to decide what particular questions really mean, which is to say how they will be understood by respondents. This suggests how the questions will (or can) be used in analysis, and whether they closely reflect the concepts of initial interest (which may lead to a re-examination of those concepts). Perhaps there is not as much wrong with this process as one first thinks. If many researchers, as it appears, do not follow the conventional wisdom of an orderly process, perhaps it is due, in part, to the proposed process being flawed. It is sometimes very difficult to clearly state survey goals and planned analyses; and difficult tasks will often lead to error. While not abandoning the conventional logic of survey development, it may be legitimate to recognize that many times a process in which the goals and concepts are *expected* to be refined during questionnaire design may be both more practical and more fruitful. Whichever view of the matter one takes, the way one goes about linking concepts and survey items is certainly a fundamental area of the survey process deserving close examination.

In general, quite a bit is known about which factors affect survey measurement and about how to reduce measurement error. But, at the end of the day, it is often difficult to know how good a job has been done – particularly if that assessment is to be stated in quantitative terms.

One aspect of this problem is the estimation of measurement error; another is understanding how measurement error affects different types of analysis. In "A Review of Measurement Error Effects on the Analysis of Survey Data," Biemer and Trewin develop two error models, one for continuous and one for binary data. They first examine (both correlated and uncorrelated) measurement error effects on a number of (simple random sampling) estimators; then turn to how measurement error may affect some commonlyused multivariate analyses.

Other chapters in this last section of the book deal with measurement error in several other circumstances, such as panel surveys (van de Pol and Langeheine), reinterviews (Bassi and Fabbris) and dietary intake surveys (Nusser et al.) Almost all of these chapters require a great deal of statistical sophistication to follow closely, but the main issues,

approaches and conclusions can often be followed by the general reader, leaving both the mathematical details and the application of these methods to the specialist.

Even a book as useful as this one, which runs nearly 800 pages, and represents the work of more than seventy authors, cannot fully encompass every aspect of the survey enterprise. Still, a few omissions deserve mention.

One curious shortcoming is the paucity of comprehensive discussions of costs. One of the most basic questions facing the practitioner is where to best allocate resources in the survey process for maximum benefit, however that benefit be measured. Any movement toward a methodology of survey process improvement has to address costs.

One component of the cost of process improvement that is easily overlooked is the expense necessary simply to document, even in a qualitative fashion, the different sources of non-sampling errors for a particular survey. (A U.S. Office of Management and Budget effort several years ago to require extensive documentation of nonsampling errors for federally-funded surveys, was, I believe, abandoned largely because the costs were likely to be prohibitive.)

Data collection and data processing are the stages that seem most amenable to the process improvement approach, if only because they are more formally organized processes to start with. A crucial part of data collection not given much attention is the management of the interviewing staff. This area is especially important considering that labor is typically the largest cost component of face-to-face and telephone surveys.

Finally, there is little on the role of research staff training in the continuous process improvement paradigm. While most survey professionals will likely continue to be specialists, it would seem that under the process paradigm both managers and methodologists need to have a broader knowledge of the overall survey process.

Historically, survey practitioners are more likely to be specialists in particular niches than generalists. As in other fields there is good reason for specialization. The training and talents necessary for excellence in designing efficient sampling plans, writing good questionnaires, training interviewers and analyzing complex data sets are sometimes, but not often, found in individual researchers. Yet it is important that even specialists view the whole survey as a process rather than as independent elements, a view often fostered, perhaps unavoidably, by the organizational structure of most survey organizations. But, just as samples cannot be designed independently of the costs and practicalities of data collection, or questionnaires written with no eye on the planned analysis, a survey researcher who aspires to minimize total survey error by way of process improvement must approach the enterprise as an integrated operation.

Overall, this is an excellent compilation. It is probably more suitable for the professional than the student. While the generally excellent overview chapters that begin each section are suitable for advanced students, most of the chapters themselves require that the readers have a fair amount of knowledge.

Johnny Blair Survey Research Center Room 1103, Art-Sociology Bldg. University of Maryland College Park, MD 20742 U.S.A. Fax: 301 314 9070 This very important book is a must read for anyone who is seriously concerned about survey data quality. This includes those who plan and implement surveys, those who analyze survey data, and those who make policy and other types of decisions based on the findings of sample surveys. The book deserves a very wide audience because of the importance of the information it provides and the thinking it stimulates about the most serious problems now facing the field of survey research – Unit Nonresponse and the prospect of Nonresponse Error.

The book has limitations due to the particular approaches Groves and Couper chose to take in planning their research and also due to contraints placed on them by circumstances beyond their control. The primary limitation is the exclusive use of data from large scale government-funded surveys conducted only via the in-person mode and that all have extremely high response rates. Thus, their primary focus is on nonresponse in the in-person mode and the findings they present may not generalize to surveys with extensive nonresponse. The lack of a major database to investigate nonresponse error is another important limitation. These notwithstanding, the volume's most important contribution is that it provides a detailed road map that guides the careful thinker of survey methods to consider anew all that he or she thought was known about survey nonresponse. Whereas nonresponse has been addressed sporadically in the literature, it has never before received such careful and thorough theory-based interdisciplinary scrutiny. And never before has there been assembled original data about nonresponse on such a massive scale and breadth. Ultimately, the only unsatisfying aspect of the book is that it is not longer and more ambitious in its treatment of potential errors that may stem from nonresponse.

The authors explain their goal in their near-decade long investigation of nonresponse, which led to this book, as an effort to shed light on the question: "How important is nonresponse to surveys?" The foundation upon which they built their research and thinking is the premise that only by using "sound theories of human behavior" (mostly from the discipline of psychology) upon which to base the study of nonresponse can the field devise the most cost-effective procedures to increase response rates and to develop statistical adjustments for nonresponse error. A theme that is stressed throughout is the importance of thinking separately about noncontacts and refusals as the primary sources of nonresponse since: "Considering only the overall response rate ignores the possible counteracting biases of different types of nonresponse."

The massive original research, to which the authors appear to have devoted a noninconsequential portion of their relatively young lives for the past decade, is explained and extensively self-critiqued in Chapter 3. This one-of-a-kind information-base with many different types of data was developed with the cooperation of the U.S. Bureau of the Census. These data came primarily from the U.S. decennial census match nonresponse project which employed 1990 census data at the household level for tens of thousands of responders and nonresponders to large government surveys (e.g., the Current Population Survey and the National Crime Survey). (Of note, although most of these surveys use panel designs, the authors used only Wave 1 cases in their database.) Their information-base also included observational measures on the housing units of people who were respondents and nonrespondents to some other large surveys (e.g., the National Survey of Health and Stress). The authors supplemented these varied data with still other sources of information about nonresponse in large surveys, such as surveys of interviewers and focus groups with interviewers.

In the first chapter the authors introduce a structure to compare the social exchange between interviewers and respondents with other types of conversations that happen between two people who essentially are strangers (e.g., telemarketing). The structure is used effectively to help explain that the decision to participate in a survey is most likely a stochastic process for most potential survey respondents. In Chapter 2 their conceptual model for survey participation is explained. The major components of this model are the influences on participation of (1) the larger social environment, (2) a particular survey's design, (3) the potential respondent's own characteristics and behavior patterns, (4) the interviewer's own characteristic and behavior, and (5) the uniqueness of the verbal and nonverbal interaction of these two individuals. Chapters 4–10 then provide detailed attention to various aspects of each of these components. Most chapters have a similar internal organization starting with a critical review of theory and the past research literature on a given component, then presenting original and detailed multivariate statistical findings from their own databases as pertains to the respective component and its attendant theory, and ending with practical advice for survey operations which follow from the research findings they present. Much of this advice is creative, yet sound, and the authors extend its applicability to other modes of surveying, in particular the telephone. Throughout their writing, Groves and Couper strive to stimulate new and needed research by acknowledging the considerable gaps that still remain before survey nonresponse is adequately understood.

For practitioners of survey research, by far the most important practical implication of their work is the importance the authors attach and the attention they devote to successful, creative "tailoring" by interviewers of the verbal and nonverbal behaviors they use to seek and gain the cooperation of household level gatekeepers and respondents themselves. Here, I share their belief that "while standardization at the measurement phase is a basic tenet of the scientific method, it has no proven value at the cooperation phase."

The final chapter takes a comprehensive look at how concern about nonresponse can (and should) be manifest in a multitude of practical decisions that are made in planning the operational aspects of a survey. This includes selection of the frame, choice of data collection mode, design of the questionnaire and other measures that might be gathered via observational means, hiring interviewers, and post survey adjustments for unit nonresponse. Ultimately, Groves and Couper challenge the practitioner (and the survey's funders) to make considered, non-arbitrary decisions about how to allocate finite resources so as to try to lessen the errors that may result from nonresponse.

They end with three major deductions that flow from their theoretical perspective on survey participation: (1) nonresponse rates of different subgroups (e.g., young adults, residents in rural areas, etc.) in the population may be affected by different factors (e.g., tailored introductory spiels); (2) survey nonresponse does not automatically lead to nonresponse error and the "mindless maximization of response rates has no theoretical support;" and (3) advances in effective post-survey adjustments will still be needed to deal with nonresponse error regardless of future progress that may be made in reducing survey nonresponse. Of additional note, the book is organized and written in a fashion that helps one skim those sections that have more details than necessary for a particular reader, yet provides enough depth on topics when one really needs it. This gives the book an appeal and value to multiple audiences as it makes it an easy process for the reader to learn what is there and what is not, in how much detail, and then to take away what seems to be enough information for one's needs at a given time and to know what has been left behind if a future information need arises.

In sum: this is a very important book that properly focuses the need for careful attention to the theory-based study of survey nonresponse and the need to pay more attention to it in the design and funding of sample surveys.

> Paul J. Lavrakas College of Social & Behavioral Sciences Derby Hall, Room 0126 154 North Oval Mall, Ohio State University Columbus OH 43210 U.S.A. Phone: (614)-292-6672 Fax: (614)-292-6673 email: lavrakas.1@osu.edu

R. Johnson and K. Tsui. Statistical Reasoning and Methods. New York: John Wiley & Sons, Inc., 1998. ISBN 0-471-04205-6. 75.95 USD.

I would recommend this introductory statistics text for a first semester undergraduate statistics course, despite some shortcomings. The authors promise in their preface to introduce basic statistical concepts and methods by presenting real-life, easy-to-understand examples and they do just that. Throughout the text, the authors use everyday situations to explain statistical concepts and their significance.

Ample examples and exercises provide students with the variety and repetition frequently necessary to master the basics of statistics. The use of real-life situations for examples and exercises also helps to demystify the use of statistics, which is particularly important for students in introductory statistics classes. An additional feature of this text is examples of syntax and output when using MINITAB software for statistical analysis. The examples provide detailed instructions for the syntax and keystrokes to run many basic statistical tests. Another particularly helpful feature for both students and instructors is the use of chapter objectives and key ideas. The chapter objectives provide a check list of concepts at the beginning of each chapter that one should understand after completing the chapter. At the end of each chapter, the key ideas provide a detailed review of all definitions, key concepts, and formulas discussed within the chapter.

Although I appreciated the authors' rather laid-back, simplified approach to explaining statistical concepts that often frustrate even the most diligent student, I was disappointed in the ordering of chapters and concepts. One of the best statistical texts in my opinion is the classic Social Statistics by H. M. Blalock (1979). Although Blalock presents statistical material in a very dry manner, he does provide the best treatment of the differences between nominal, ordinal, interval, and rational data, which is essential to the understanding and proper use of various statistical tests. Johnson and Tsui's text only provides

minimal treatment of the various data types and fails to clearly demonstrate why certain statistical tests are only appropriate for specific types of data. Instructors who would use this text must be aware that additional emphasis must be made in the classroom concerning the differences between various types of data and the appropriate statistical tests for them. Finally, I think the textbook jumps at times from topic to topic with no clear path or stated association. In particular, I would suggest completing the discussion of probability and sampling prior to introducing the various types of data and statistical tests.

Overall, this textbook provides a good overview of basic statistics in a straightforward, understandable manner. The extensive use of real-life examples gives instructors an effective tool to make a somewhat intangible subject easier for students to relate to. The abundance of exercises throughout each chapter provides students with multiple opportunities to master key concepts. Instructors may want to reorder the chapters when teaching, though, to ensure students have a good understanding of the data types and sampling issues before moving on to the various statistical tests. Finally, for future versions of this text, I suggest the use of SPSS for the computer examples instead of MINITAB, as SPSS is much more widely used across all disciplines.

Reference

Blalock, H.M. (1979). Social Statistics (revised 2nd edition). New York: McGraw-Hill.

Katherine Lind Survey Research Laboratory University of Illinois at Chicago 412 S. Peoria, 6th fl., m/c 336 Chicago, IL 60607 U.S.A. Phone: 312-996-5308 Fax: 312-996-3358

W. Paul Vogt. *Dictionary of Statistics and Methodology*. Thousand Oaks, CA: Sage Publications, Inc., 1999. ISBN 0-7619-1273-8. 314pp. + references. 75 USD.

If you have ever been a student in statistics or research methodology in the social sciences, you will immediately appreciate the value of a dictionary that provides reliable information on the definitions of the statistical terms, concepts, and acronyms that seem so foreign when you first encounter them. If you have ever been an instructor in such a course, you will certainly understand the value of a dictionary that students can use as a summary of answers to frequently asked questions, since it summarizes the basic terms and concepts that roll off your tongue as a matter of course. Regular dictionaries are no help in this regard, since many of the terms common to statistics and social science research are not listed in regular dictionaries (e.g., ordinary least squares, maximum likelihood, heteroscedasticity). Definitions for statistical terms that are listed in regular dictionaries often exclude the statistical usage of the term (e.g., residuals, error) or they are written in statistical jargon that is all but unintelligible to the non-specialist (e.g., regression, correlation).

The Dictionary of Statistics and Methodology (DS&M) was written to provide students

and readers of social science research with a non-technical introduction to the jargon of statisticians and research scientists. Non-technical definitions can be very helpful in communicating a deeper level of understanding to students. They can also be useful to readers who want to review concepts and terminology that they confront only occasionally.

The author's experience as an instructor in statistics and methods is put to good use in the DS&M. Professor Vogt draws on this experience to provide many concise examples and excellent asides on topics that students often find confusing. Here is part of his explanation of the general linear model, "A common set of statistical assumptions upon which are based *regression, correlation,* and *analysis of variance.* ... The basic idea is that the relationship between a *dependent variable* and the *independent variables* can be expressed as a *linear equation.* ... You can get a practical 'feel' for the general linear model by spending a few hours calculating, by hand, *standard deviations, Pearson correlations, regression equations, t tests,* and *F tests.* You will quickly discover that many of the steps in these calculations are generally the same." (p122) All of the terminology which I have underlined in this definition is listed separately in the DS&M to explore this or any other series of interlinked terminology.

The second edition of the DS&M includes 600 more definitions than the first edition, but it is still a relatively slender volume at only 313 pages and about 2,000 definitions. Specific enhancements for the second edition include better coverage of terms from business research and economic research, better coverage of terms from qualitative research methods, and the incorporation of comments and suggestions from users of the first edition. Even though it is called a dictionary, it is designed more like a glossary with generous amounts of white space, a single column format, many detailed examples, and extensive cross-references. Students and non-specialists will be delighted to find this compendium of non-technical definitions for such insider terminology as artifact, asymptote, black box, cohort effects, secular trend, zero-sum game, and ecological fallacy. Readers will also be pleased to see that many definitions provide context for the use of a given term (e.g., the definition for recursive models includes a contrast with non-recursive models), many definitions include cross-references, and some definitions discuss the historical roots of the term defined.

The DS&M's coverage is most complete and its quality is at its highest when defining terms having to do with statistical research in psychology and education, which should not be too surprising given the author's background in these fields. Vogt's coverage of terminology having to do with the general linear model, correlational analyses, and experimental research designs is excellent. Unfortunately, despite the author's voluminous background research in preparing the DS&M – he reviewed well over 100 texts on methodology and statistics – it is uneven in its coverage and in the quality of the definitions. For example, the coverage of correlational analyses is sufficiently broad (e.g., regression, factor analysis, path analysis are all included in detail), but the definition of the central concept of ''correlation'' is not very thorough and includes no examples: ''The extent to which two or more things are related ('co-related') to one another. This usually is expressed as a correlation coefficient'' (p58). The remainder of the definition is a helpful aside on common pitfalls in confusing correlation with causation. On the other hand, the definition for ''degrees of freedom'' is excellent. This often-confusing concept receives a lengthy and detailed treatment, and includes a carefully explained example.

One important area of social science research that deserves better coverage in the DS&M is survey research. It includes definitions for sampling, instrument, respondent, pilot, pretest, nonresponse bias, CATI, order effects, and response rate, but many of the definitions for these terms provide only cursory treatment of their connotations in survey research. The coverage of survey research terms is also somewhat dated, since many of the terms that have received considerable attention in the literature on survey research methods over the last fifteen years are simply omitted. Here is a listing of some current survey research terms that are not included – incentives, self-administered questionnaire, nonsampling error, total survey error, context effects, confidentiality, informed consent, parental consent, quality control, monitoring, validation or verification, cognitive interviewing, random digit dial (RDD) sampling, Mitofsky-Waksberg RDD sampling, list-assisted RDD sampling, CAI, CASIC, CAPI, CADE, ACASI, usability testing, case management system, data transmission, AAPOR, and CASRO.

The DS&M also provides partial coverage of certain topics that seem to lie beyond the scope of a dictionary in statistics and methodology. For instance, some computer terminology is included (ASCII, bits, bytes, megabyte, gigabyte, UNIX, DOS), but most other computer terms are not (e.g., RAM, CPU, FAQ, Windows, World Wide Web, HTML, Internet, terminology on software, hardware, and peripherals). Some terms from social science theory are included (e.g., conflict theory, functionalism, cultural relativism, relativism, cybernetics), but the DS&M does not provide an exhaustive or representative listing of such terminology.

One particularly helpful feature of the DS&M is the inclusion of a brief bibliography titled Suggestions for Further Reading. The author has provided the reader with a carefully edited list of 50 books the author has found to be "especially clearly written, and/or widely cited 'classical' accounts." They are divided into three categories: Elementary Methodology and Statistics (appropriate for undergraduates), More Advanced Methodology and Statistics (more appropriate for graduate students), and Other Dictionaries and Reference Works.

In sum, the expanded DS&M will prove to be quite useful as a concise reference volume for students and other newcomers to social statistics and research methodology, especially those reading research in education and psychology. I am not aware of any other single volume that is specifically devoted to the definitions and connotations of the terms used in social science research. Anyone who needs to know what lemma, leptokurtic, or platykurtic mean; anyone who can no longer remember the distinction between a Guttman scale and a Likert scale; and anyone who does not know or remember what Lillefors Test, Lindquist Type I ANOVA, Arrow's Impossibility Theorem, Laspeyres Index, and Compertz curves are, will be happy to have access to the DS&M.

> Karl R. Landis, Research Triangle Institute 3040 Cornwallis Road P.O. Box 12194 Research Triangle Park, NC 27709-2194 U.S.A. Phone: (919) 541-7002 Fax: (919) 541-1261 email: klandis@rti.org