Book and Software Reviews

Joel Perlmann and Mary C. Waters. *The New Race Question. How the Census Counts Multiracial Individuals*. New York: Russell Sage Foundation, 2002. ISBN 0-87154-657-4. 389 pp. + refs. and index. 39.95 USD.

Scientific norms require that we specify guidelines and procedures followed to obtain a measurement so others can repeat our observations and judge the quality of information yielded by our measurement procedures. *The New Race Question* is not a book about scientific measurement, even though the methods and analysis appearing in several chapters are scientific in nature. As this book shows, race is not a theoretical concept driving a scientific measurement process. Because we cannot identify observable representations of a concept unless its meaning is clear, several chapters are geared to clarifying the mental imagery conveyed by the concept of race. Race is not just a subjective, social and cultural construct but has in the case of the United States, important legal and political effects.

The race question changes in the 2000 Census to capture emergent and multiracial identities. This change is in conflict with the American civil rights law that requires clear and nonoverlapping categories of race. At the same time, the classification based upon self-identification produces serious complexities when people select two or more races. The new format has led to the creation of sixty-three possible racial combinations. There are clear rules about how people of mixed race are to be counted for the purpose of civil rights monitoring and enforcement. Those people who select two or more races of which one is a minority race have to be allocated to the minority race. But this allocation rule is not a definitive answer.

This book is the outcome of the conference held at the Levy Economics Institute on September 22 and 23, 2000. In the introduction the editors point out the issues taken up

in the book: ''How did we come to measure race at all in the census and in our statistical systems? How was it decided which categories were identified as races and which ones were eligible for government protections and special legal statuses? What impact would allowing individuals to check more than one race have for the statistical system itself, for legal cases that use racial statistics, for other important research and policy development, and for average Americans' understanding of race and the government's role in measuring it?''

The book is divided into four parts. Part one comprises three chapters that explore directly what the efforts to count multiracials have shown. The first chapter describes the different processes implemented for the changes in the concept of race. For more than two centuries, the U.S. statistical system has classified each respondent into a single race. But now, in the Census of 2000, race has been operationalized as a multiple response question. Nevertheless, in Chapter 1, Farsley suggests that the concept ''multiracial'' will not enter popular discourse. He believes that most people will continue to assume that every American has a ''basic'' race. In Chapter 2, Harris considers the race question from a methodological point of view. He discusses different measurement methods, concluding that the way race is measured has a significant impact on estimates of the size of the multiracial population. Moreover he outlines several problems in respect of the race classification scheme. Chapter 3 studies the multiracial reality in California. Tafoya describes how diversity derives not only from immigration, but also from mixed racial and ethnic births.

In the second part, scholars debate how much this change in measurement will matter in the contexts that have used race counts – especially in civil rights law but also for understanding health, education, income, and other crucial topics regarding which public race counts have been common. Chapter 4 points out some of the issues that the Office of Management and Budget (OMB) directives may raise. Allocation rules have resulted from the disjunction between statistical policy and the laws for using racial data. Civil rights laws need single race categories, but the statistical system has to change to multiple race reporting. In Chapter 5 Harrison argues that this change could undermine the federal racial statistical system. He points to two essential aspects. Firstly, the category of combined races does not identify any recognised population group. Secondly, without good bridging techniques the ability of the federal statistical system to track changes in the welfare of racial groups in the U.S. society will be severely challenged as statisticians will be unable to tell whether the changes are attributable to real changes in the world or simply to a new way of measuring the group. Chapter 6 deals with the legal implications of a multiracial census. Persily indicates that most racial discrimination cases hinge not on population counts at the census level but rather on institutional data, such as the number of employees of a particular race.

In the third part of the book, the authors examine possible scenarios for the future taking into account that the rates of intermarriage among racial groups are high and mixed-race recognised. In Chapter 6 Snipp examines the American Indian case. American Indians have a long history of intermarriage and mixed-race people. Chapters 7 and 8 deal with population projections and they examine how the U.S. Census Bureau projections do not consider intermarriage effects.

The final section of the book takes up the politics of race numbers. In Chapter 10

Jacobson points out how defining race combines scientific concerns about measurement with political concerns about measuring for political and administrative purposes. Chapter 11 stresses the danger of this combination. Chapter 12 describes how race came into the census and Chapter 13 examines more recent history, showing how political and administrative concerns have created the categories now in use. Chapter 14 contrasts the experiences of Brazil and the United States. In Chapter 15 Glazer stresses that the main racial divide comes from black-white inequalities. Chapters 16, 17, and 18 explore the possibility that the arrival of the multiple-race option will have profound effects on American social and political life. Finally, there is an appendix, called Chapter 19, where Tucker and his colleagues provide a detailed comparison of different methods of ''bridging'' between the old and new data, evaluating gains and losses of using any given method. In using bridging methods the intention is to reduce the effects of the changes in data collection. It is an interesting ''how to'' chapter.

The volume is a collection of essays that examine the ''new race question'' from a multidisciplinary perspective. The book index follows the preliminary conference. However, the length of the chapters is uneven. Chapter 10 by Jacobson and Chapter 11 by Sollors, are four pages each. Nevertheless the different approaches to the subject are well balanced. However, there is perhaps some overlap between the chapters. Occasionally, they repeat the same introductory information. But as in chess, departing from a similar position every chapter develops in different and interesting ways. The book is accessible to students, and also constitutes a valuable resource for those interested in civil rights law. In sum, this volume will be very useful for social statisticians, sociologists, and in general for those involved in social reporting.

Antonio Alaminos
University of Alicante
Departamento de Sociologia II
Edificio de Ciencias Sociales
Alicante Ap-99
Spain
Phone: 34 965903497
Fax: 34 965903495
Email: alaminos@ua.es
obets@ua.es

Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J.A. Little (eds). *Survey Nonresponse.* John Wiley & Sons, Inc., New York, 2002, ISBN 0-471-39627-3, 520 pp. (cloth) 89.95 USD.

In October 1999, an international conference on survey nonresponse was held in Portland, Oregon. This book, written to provide a review of the current state of the field in survey nonresponse, is an outcome of that conference. Given the reputations and experience of the editors and contributors (all leading academics and researchers in survey methodology, sampling, and statistics), the quality of the book is not surprising, and the selection of

chapters presented in this volume makes it a worthy addition to the Wiley series on survey methodology.

Survey Nonresponse contains 29 chapters divided into four sections: Perspectives on Nonresponse; Impacts of Survey Design on Nonresponse; Nonresponse in Diverse Types of Surveys; and Statistical Inference Accounting for Nonresponse. The first three sections (17 chapters) address the questions of why nonresponse occurs and how it might be reduced by survey design; the fourth is concerned with the problem of how to analyse data in the presence of nonresponse. While the papers in the latter section are clearly linked by their focus on statistical measures for dealing with survey nonresponse ''after the event,'' the boundaries of the other three sections are rather blurred. This means that readers interested in survey design, for example, will find something relevant in each of these three sections, not just in the section ostensibly devoted to this topic.

The book starts with an overview of survey nonresponse in survey design, data collection and analysis, written by the editors. The chapter provides the equivalent of an executive summary of the rest of the monograph. The rest of the first section includes chapters on international trends in household survey nonresponse, the development of nonresponse standards, and the effect of culture on survey nonresponse. The section concludes with two chapters specifically concerned with the decision processes that respondents go through when answering questions, and the implications of this for item nonresponse.

Part II contains six chapters, concerned with various aspects of survey design and how these affect nonresponse. It includes chapters on the influence of interviewer attitude and behaviour in different countries, the use of incentives and extended interviewer efforts to reduce nonresponse in household surveys, the effect of vigorous refusal conversion on unit and item nonresponse in a telephone survey, and the effect of visual design in self-completion questionnaires.

The section illustrates both a strength and a weakness of the book. On the one hand, it offers a variety of different perspectives, with examples from several countries. On the other hand, there is no strong unifying theme to the papers, and the level of detail ranges from a discussion of a particular aspect of questionnaire design, the presentation of skip instructions, to a comparison of different approaches to item nonresponse, including imputation, which might not have been out of place in the last section of the book. This produces a slightly uneven texture in the book, which is perhaps inevitable given the nature of the enterprise. However, if this is a weakness, it is only so in the sense that the chapters that are included emphasise the fact that there are a number of other interesting issues that might have been addressed. Furthermore, many readers will enjoy the somewhat eclectic nature of the book's contents and the sense of serendipity that this creates.

Part III of the book comprises five chapters, which examine aspects of nonresponse in mail surveys, business surveys, Web surveys, exit polls, and longitudinal household surveys, respectively. The chapters vary in terms of scope and content, but generally describe the characteristics of nonrespondents in the different types of surveys and discuss practical ways of reducing or dealing with the problem of nonresponse in these situations. An important issue raised in two of these chapters (though these are not the only chapters in the book in which it arises) is the possibility that increasing response rates may actually increase rather than decrease survey error.

The final section of the book is concerned with statistical approaches to nonresponse, and contains chapters describing weighting, imputation, substitution, and other statistical estimation procedures that address the problem of missing survey data. The chapter ends with a useful discussion of diagnostics for nonresponse adjustment methods. This is the most technical section of the book, likely to be of more interest to statisticians than survey researchers (though this is not to suggest that statisticians cannot be survey researchers, or vice versa). However, even in these technical chapters the basic concepts are described in a language that an interested reader with a limited statistics background could understand. This is another strength of the book.

Survey Nonresponse is, at the same time, a serious book presenting current thinking on an important issue in survey research and an interesting, readable book, with a strong applied orientation. This is a tribute to the editors and contributors. The book contains a mix of styles and perspectives but the chapters are uniformly thoughtful, rigorous and well-written. There is inevitably some overlap among the chapters, but this tends to add to rather than detract from the book because it reinforces its general theme, which is that only by blending social science and statistical approaches will survey nonresponse be understood and controlled.

There is no simple solution to the problem of survey nonresponse, and this book does not offer one. What it does do is to provide a conceptual framework for survey nonresponse and suggest some solutions to the problem for different types of surveys and at different stages of survey design and implementation. It then recognises that, despite researchers' best efforts, nonresponse will still occur, and presents various statistical procedures designed to reduce the effect of nonresponse on survey estimates. Thus *Survey Nonresponse* should be required reading for anyone with a serious interest in survey research, and, although it is not a textbook, it would be an excellent supplementary text in any graduate survey methods course.

Philip Gendall
Massey University
Department of Marketing
Private Bag 11 222
Palmerston North
New Zealand
Phone: 64 6 350 5582
Fax: 64 6 350 2260
Email: p.gendall@massey.ac.nz

Roderick J.A. Little and Donald B. Rubin. *Statistical Analysis With Missing Data*. Second Edition. Hoboken: Wiley Interscience, 2002. ISBN 0-471-18386-5. 368 pp. + refs and index. 94.95 USD.

Little and Rubin's first edition published in 1987 has become a standard reference for statisticians who regularly wrestle with missing data problems. While their work has broad potential applications, I am familiar with it in the context of sample surveys. The authors

review a wide range of methods for handling missing data, some of which have short-comings that the authors point out. The methods the authors favor have not found universal acceptance, but some, including multiple imputation, are more broadly used than they were a few years ago. One important contribution of their work is that it has helped make clear that the way we deal with missing data affects not only statistical estimates themselves, but the precision with which those estimates are measured.

The second edition contains three parts of five chapters each. A map in the Preface shows where materials from the 12 chapters of the first edition may be found in the second. In Part I, Overview and Basic Approaches, Chapter 1 provides an introduction to missing data problems and missing data methods, Chapter 2 covers issues of missing data in experiments, and Chapters 3 through 5 present complete case and available case approaches, single imputation and estimating imputation uncertainty, respectively. Multiple imputation is introduced in the last two sections of Chapter 5.

Part II, ''Likelihood – Based Approaches to the Analysis of Missing Data,'' begins in Chapter 6 with a review of the theory of likelihood estimation and Bayes' methods, covering estimation first with complete data and then with incomplete data. Chapter 7 discusses factored likelihood methods. Chapter 8, titled ''Maximum Likelihood for General Patterns of Missing Data: Introduction and Theory with Ignorable Non-Response,'' introduces the EM Algorithm. Chapter 9 covers large-sample inference, and Chapter 10 is devoted to multiple imputation. In Part III, ''Likelihood Based Approaches to the Analysis of Incomplete Data: Some Examples,'' the authors present applications in Chapters 10–14 for multivariate normal data, robust estimation, partially classified contingency tables and mixed normal and nonnormal data. The discussions in these four chapters are all based on the assumption that the missing data mechanism can be ignored. The final chapter in the book (Chapter 15) presents models for nonignorable missing data.

The authors note that much has been written about missing data since the first edition of their text and they have expanded their coverage of many topics. The book's coverage of both single and multiple imputation methods has expanded, with over a chapter devoted to multiple imputation. In addition, the EM algorithm receives more thorough treatment, as do robust estimation and nonignorable missing data models. All in all, the 2nd edition has about 100 more pages of text than does the first.

The book is well organized and tightly written. It provides extensive references (the References Appendix is 15 pages long) and each chapter is followed by a variety of problems for use in courses. The authors present their own methods, as well as other approaches. They also give their views, sometimes critical, of these other methods. However, the authors in several places also point out where others disagree with the methods they propose.

The authors state in the Preface that the book is intended for applied statisticians. That may be its intent but I think it is only partially fulfilled. The term ''applied statistician'' is quite broad and a bit vague, so they and I may have different conception of who is an applied statistician. If by applied statistician one means a nonacademic who does statistical work but who spends little time on theoretical issues, then many of the sections of the text are written plainly enough for most such statisticians to use as a reference. However, other sections are quite dense and technical, and understanding then requires more background than being ''familiar with inferences based on likelihoods.'' I cannot recommend

this work as a ''cook book'' that can be applied without substantial prior training or experience in dealing with missing data.

The authors have done well in adding to their first edition. There have been many advances in this area and there will likely be more in the near future. For example, some statistical packages now provide procedures for multiple imputation; this perhaps will lead to the development of more practical procedures for using multiple imputation in the context of large data sets from surveys that employ complex sample design.

John Hall
Mathematica Policy Research, Inc.
Princeton, NJ 08540
U.S.A.
Phone: 609 275-2357
Fax: 609 799-0005
Email: jhall@mathematica-mpr.com

J.N.K. Rao. *Small Area Estimation*. New York: John Wiley & Sons, 2003. ISBN 0-471-41374-7, 344 pp., 99.99 USD.

The increasing demand for small area or domain estimators in recent years has motivated research on modelling and estimation of the small area parameters of interest. Professor Rao has contributed to the advance of this branch of research and now he adds another valuable contribution with this book. In general, Rao's book treats methods of inference applied to problems of small areas. These methods are principally point estimation and its respective measurements of variability. This book also considers model construction and some methods of model validation.

For a better understanding of the material contained in the book, it is essential to have a basic knowledge of statistical inference, linear mixed models, and sampling methods, as the author points out. Nevertheless the book contains sections where one finds a clear introduction to design-based estimation (Chapter 2), general linear mixed models (Chapter 6), empirical Bayes methods (Chapter 9), and hierarchical Bayes methods (Chapter 10). It is necessary to emphasize that in Section 10.2 one finds a clear introduction to Markov Chain Monte Carlo (MCMC) methods, which are very useful in modern Bayesian estimation.

The reader should be sure to read the introduction of Rao's book, since there he defines the concepts and terminology he uses in the rest of the book, such as small area and the direct and indirect estimation. Here some examples of small area estimation are presented. These examples will serve in later chapters to illustrate the usefulness of the methodology presented.

In Chapter 2, the author presents a review of design-based (direct) estimation, the Horvitz-Thompson estimator, the generalized regression estimator, and domain estimation with and without auxiliary information.

Demographic methods are the subject of Chapter 3. Here, traditional methods of indirect estimation in demography are reviewed. The author divides these methods into

symptomatic accounting techniques and regression symptomatic procedures, which are based on censuses and administrative records instead of sampling data.

In Chapter 4 the indirect estimation of domains on the basis of implicit models is presented. Here, synthetic estimation and composite estimation, a way of combining synthetic and direct estimation, are examined. The James-Stein method is presented as a special case of composite estimation methods.

Explicit models that consider the internal variation of the small area and the variation explained by auxiliary information are examined in Chapter 5. These models have been divided into two groups according to the available level of auxiliary information. Unit level models are used when auxiliary information is available for every population unit. Otherwise, if auxiliary information is only available at the level of the small area and not at the level of the population unit, area level models are applied. Univariate and multivariate cases of the two types of models are also considered. In addition, time series and cross-sectional versions of these models are examined. Furthermore, generalized linear mixed models for binary and count observations are reviewed.

Chapters 6, 7, and 8 are dedicated to empirical best linear unbiased prediction (EBLUP). Specifically, the theory of the EBLUP under general linear mixed models appears in Chapter 6. Here, the studied estimators are best linear unbiased prediction estimators, EBLUP estimators, maximum likelihood and restricted maximum likelihood estimators. A form for estimating the mean squared error (MSE) of the above estimators is also presented.

In Chapter 7, the author applies the EBLUP to both area level models and unit level models. The general linear mixed model with diagonal covariance structure is used in the discussion of these models. Again, the estimation of the parameters of interest is established, as well as the estimation of the respective MSE's. In the same chapter there is an interesting discussion of internal and external evaluation methods for verifying the validity of the selected model in the applications.

Several extensions of area level and unit level models are presented in Chapter 8. In the area level case, these models are: multivariate area level models, general linear mixed models with correlated sampling errors, time series and cross-sectional models, spatial models and multivariate nested error regression models. In the population unit level case, the models introduced are the random error variances linear models and the two-fold nested error regression model. Finally two-level models that integrate the use of unit level and area level covariates are reviewed.

Chapters 9 and 10 are dedicated to studying Bayesian methods as applied to small area estimation. Chapter 9 concentrates on the empirical Bayes (EB) approach and Chapter 10 on the hierarchical Bayes (HB) approach. In Chapter 9, the EB estimator is generally obtained from the expected value of the posterior density function of the small area parameters of interest by replacing the unknown model parameters with the estimations obtained previously. This estimation procedure is applied to the basic area level model, the linear mixed model and the models for binary data. In this chapter the empirical linear Bayes estimator and the constrained linear Bayes estimator are also presented.

In Chapter 10, the HB estimator is also generally obtained from the expected value of the posterior density function of the small area parameters of interest. The difference from Chapter 9 is that the posterior density function is directly obtained via Bayes Theorem. HB estimators are calculated for basic area level models, basic unit level models, general

ANOVA models, two-level models, time series and cross-sectional models and binary data models. As mentioned previously, this chapter contains an introduction to MCMC methods, which are necessary when the algebraic expression of the posterior density function of the parameters of interest is not available.

If the reader is interested mainly in the application of Bayesian methods applied to small area estimation problems, the most relevant chapters are 9 and 10. On the other hand, readers interested principally in the frequentist statistical methods used in this subject should concentrate on Chapters 4 to 8.

Furthermore if the reader is interested in how the methodology of small area estimation has been applied to real problems, he or she should follow up the examples that are given in the book. Some examples in the health area can be found in Chapter 4, where synthetic estimation methods are applied. Also, examples in this area are to be found in Chapter 9, where empirical Bayes estimation and empirical, linear Bayes estimation are used. In addition, Chapter 10 contains some applications of hierarchical Bayes methods and lognormal models in the field of health. In Chapter 4 there are also some examples concerning agriculture, where synthetic estimation is used. There are more examples concerning agriculture in Chapters 7 and 10. In Chapters 4, 5, and 7 there are examples in the socio-economic area, where composite estimation, the Fay-Harriot model, and empirical best linear unbiased prediction are utilized. In the same area, other examples are to be found in Chapter 10, where hierarchical Bayes estimation, basic area level models, unit level models and time series and cross-sectional models are applied.

It is necessary to mention that at the ends of Chapters 2, 3, 4, 6, 9, and 10 the proofs of some of the basic results presented in these chapters are given. Another interesting feature of this book is the large list of references on small area estimation and related topics.

This book is essential to any basic library on small area estimation. Both theoretical researchers and practitioners in this subject will certainly appreciate the themes treated in Rao's book.

José Elías Rodríguez
University of Guanajuato
School of Mathematics
Department of Statistics and Econometrics
Carlos III de Madrid University
Apartado Postal 402
Guanajuato Gto., 36000
México
Phone: 34 916 245 854
Fax: 34 916 249 849
Email: elias@cimat.mx