Book Reviews

Books for review are to be sent to the Book Review Editor Jaki S. McCarthy, USDA/NASS, Research and Development Division, Room 305, 3251 Old Lee Highway, Fairfax, VA 22030, U.S.A. Email: jaki_mccarthy@nass.usda.gov

Bradley Efron. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. New York: Cambridge University Press, 2010. ISBN 978-0-521-19249-1, 276 pp, $65USD.

This comprehensive journey leads us through the era of massive scientific production data and its proper statistical analysis. In order to understand the purpose of this book, the author divides the statistical discipline into three eras: the age of Quetelet, with huge census-level data sets; the period where scientists collected experimental data to be analyzed with the tools of the theory of optimal inference developed by Pearson, Neyman, Fisher, and Hotelling; and finally, the modern age characterized by millions of data coming from the big boom of microarray and genetics.

In some parts, the text is theoretical and an appreciation of multivariable calculus, matrix algebra, probability theory and mathematical statistics is essential for successful comprehension of the content. Nevertheless, the book is pleasant and reads smoothly. It is easy to follow the sequence of the content, especially because of the very useful and detailed examples, sometimes supported by results of simulations, provided to illustrate the statistical methods in each chapter. Of great interest is the manner in which the author incorporates the real biomedical examples and the development of statistical methodologies in the free software R. The reader may find both, data sets and R codes, on the web.

The first two chapters address the general problem of large scale inference. When the statistician faces a large mass of data, the traditional approach is maximum likelihood. Efron cites the work of Stein (1956) that shook the foundations of statistics with a result that claims that, for dimensions greater than two, the maximum likelihood estimator could be improved in terms of the mean squared error. Later, along with James, Stein (1961) developed an estimator which dominates the maximum likelihood estimator, for any choice of prior. Although the James-Stein estimator is better, the truth is that for some certain circumstances, it underestimates the parameter of interest. The first chapter emphasizes, as well as throughout the text, the development of problems, theoretical and practical, which support the readings. Using an example from microarray data, the second

chapter addresses the problem of hypothesis testing for large masses of data. The example deals with expression levels of genes where the aim is to know if a gene makes a difference in the development of prostate cancer, between controls and treatments. For this example, classical statistics would recommend the use of t-test; then adjusting by Bonferroni limits approach. However, this method underestimates the total number of nonnull genes.

The third chapter deals with hypothesis testing performed simultaneously. After making a brief review of Miller's (1981) book on simultaneous statistical inference, which is focused on controlling of Type I error, this chapter reviews and discusses some algorithms (stepwise and permutation) that are used to control this type of error and also introduces the family-wise error rate, defined as the probability of making at least one false rejection in a family of hypotheses. Bonferroni limits are a clear example of a control algorithm, which does not require the assumption of independence between the p-values. The Sidak procedure gives better results than the Bonferroni limits, but assumes independence. The Holm procedure exhibits a better performance, since the rejection regions are larger.

Chapter 4 reviews false discovery rate control. It deals with important questions such as the choice of control rates and independence assumptions, as well as some variations on the Benjamini-Hochberg (BH) algorithm, including a special subsection on significance testing of microarrays. The reader may find the empirical Bayes interpretation of the BH procedures useful. The next chapter provides a detailed explanation of local false discovery rates, where it is necessary to analyze critical values (such as $z = 1.96$) by their own terms. This chapter introduces the idea of Bayes' factors for a hypothesis testing procedure and presents Poisson regression estimates for the mixture density.

What I find most useful with this book is the role of the null distribution and its validity. In Chapter 6, this is the main issue and the author deals with four examples that show that sometimes it is inadequate to use the theoretical null distribution. Instead, the author claims that the statistician should use the empirical null distributions, which are estimated using each study's own data. Note that an incorrect choice of null distribution affects both local false discovery rates and its tail areas. In this sense, Chapter 7 derives formulas for a wide class of summary statistics in the estimation of the former distributions, and also discusses the nonnull distribution of z-values that can be well approximated by normality.

An important characteristic of large scale studies is the implicit correlations within subjects and it is shown how correlation can drastically degrade estimation accuracy. Chapter 8 arrives with some answers to important questions concerning correlation. Specifically, it deals with row-wise and column-wise correlation of the data matrix and it provides an estimate of the root mean squared correlation, which is a key quantity in the expressions of the former chapters. Later, it discusses the independency assumptions of a set of microarrays in which correlation is inherent in the way genes interact at the DNA level. Also, advising that multivariate normality may be a risky assumption in large-scale studies, the author reviews the theory behind those ideas.

Stressed throughout Chapter nine are enrichment concern methods for making inferences on groups of cases rather than one at a time. The chapter begins with a cancer development study, for which just a single gene, from 10,100 genes, showed an association with a significant z-value. This way, by means of considering a set of cases, it is possible to identify those who are associated with the characteristic of interest. After a

brief review of randomization and permutation for the proper choice of a null hypothesis of enrichment testing, the author focuses on the power of considering scoring functions.

Chapter 10 presents proper techniques concerning the separation and combination of cases. As the author claims, omnibus combination may distort individual inferences in both directions: interesting cases may be hidden while uninteresting ones are enhanced. This chapter presents an extended model where the cases are divided into distinct classes, and the statistical background behind finer separation of classes. Finally, the book concludes with an approach to prediction through an empirical Bayes perspective. In that chapter, attached to all of the proposed models, one can find some valuable examples that illustrate the prediction procedures.

In conclusion, this book sets a new threshold in statistical practice. Based on the above, I believe this book is a very important tool for practitioners who are conducting research with large masses of data. On the other hand, I think this text should also become mandatory for the graduate students in biostatistics.

References

Miller, R.G. (1981). Simultaneous Statistical Inference (2nd edition). Springer Series in Statistics and Applied Probability. New York: Springer-Verlag.

Stein, C.M. (1956). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. Proceedings of the 3rd Berkeley Symposium in Mathematical Statistics and Probability, Vol I. Berkeley and Los Angeles: University of California Press.

James, W. and Stein, C. (1961). Estimation with Quadratic Loss. Proceedings of the 4th Berkeley Symposium in Mathematical Statistics and Probability, Vol I. Berkeley: University of California Press.

Hugo Andrés Gutiérrez
Facultad de Estadística,
Universidad Santo Tomás,
Carrera 9 No 51 – 11. Piso Tres,
Bogotá, Colombia.
Phone: (+57 1) 587 8797 Ext: 1450.
Email: hugogutierrez@usantotomas.edu.co

Martin Bulmer, Julie Gibbs, and Laura Hyman (eds). *Social Measurement through Social Surveys: An Applied Approach.* Surrey, England: Ashgate Publishing, 2010. ISBN 978-0-7546-7488-7, 226 pp, $39.95.

In *Social Measurement through Social Surveys*, Martin Bulmer, Julie Gibbs, and Laura Hyman have pulled together a collection of work that grapples with the continuing challenge of improving the measurement of social constructs, attitudes, and behaviors in social survey research. Although this challenge may be addressed in a variety of manners, in this edited volume Bulmer et al. provide a call to reinvestigate the underlying theories that inform these social measurements. This approach is the crux of *Social Measurement*,

and provides a novel contribution to the larger body of literature. As we tend to be fixated upon the reliability and diagnostic testing of such survey questions (very important in its own right), the editors' unique approach highlights the importance of considering the validity of a social construct that was originally intended to be captured by our survey questions. The result is a volume written by a number of methodologists and researchers addressing this call, many who have been involved in the creation and maintenance of the UK-based Survey Question Bank (SQB) over the past ten years.

After an introductory chapter that states the aforementioned goals, what proceeds is a series of ten chapters, each written by different authors and focusing on an individual social construct (or related grouping of constructs) and its measurement. This series begins in Chapter 2 with John Haskey's discussion of the issues faced when asking a basic set of survey questions on *families and households* in national surveys. True to the goal of the book, he begins his chapter with a discussion of the conceptualization of different types of families, how to define a family, and distinctions between families and households. This is followed by an example of survey questions asked to families in national surveys, a strategy on how to identify a household and collect key information on members of a family, and the importance of identifying relationships between family members. Next, Ian Brunton-Smith and Jonathan Allen focus on the measurement of *crime* in Chapter 3. Relative to other chapters in the book, they give little attention to the theoretical underpinnings behind measuring crime, but instead focus on a general overview of the different types of surveys used to measure crime in society (specifically victimization and self-report offending surveys), discussing strengths and weaknesses of each. The authors finish the chapter with a brief discussion of the measurement of crime against businesses.

In Chapter 4, Oliver Heath and Robert Johns discuss the social measurement of *political behaviors and attitudes*, giving attention to such problems as voter turnout, intended vs. actual vote choice, and the recall of one's voting behavior. The influence of political identification and issue salience are also covered using data estimates from the British Election Survey (BES). *Religious behavior* is the next social phenomenon that is covered (Chapter 5), where Peter Brierley provides a very brief discussion of the three aspects used by the sociologist Durkheim to define religion: affiliation, practice, and belief. The conversation regarding these theoretical constructs quickly ends there, as the author proceeds to provide a number of descriptive estimates and data sources as a set of practical references for the reader, accompanied with a brief historical background on religion in the UK. The challenges in measuring religion created by increased cultural diversity are also discussed. In Chapter 6, Eric Harrison covers the measurement of *social class*, a social construct that has a long history in the social sciences of being debated and discussed at a theoretical and conceptual level. Recognizing this, he begins with a succinct yet strong coverage of the theoretical and conceptual foundations that can be used to operationalize social class, and argues why measuring this construct is important in large-scale surveys. He then selects a particular conceptual framework (Erickson et al. 1979) from which to provide practical examples of measuring social class using survey questions on occupation, type of position, and supervision in the workplace.

Race and ethnicity are the next social measurement focused upon (Chapter 7). In this chapter, Martin Bulmer discusses the early history of racial/ethnic classifications in the UK, and provides an account of the introduction of measurement of race/ethnicity

in the 1991 Census of Population. Following this he covers some difficulties faced in the terminology used in capturing race/ethnicity, and measurement alternatives that have been used as opposed to asking a respondent to self-identify as being of a specific race/ethnicity. Chapter 8 follows, where Catherine H. Mercer and colleagues discuss the measurement of *sexual behavior*, a topic that has been gaining increasing attention in the survey methodology field. Here Mercer and colleagues justify the need for measuring sexual behavior in surveys, and note the issues that need to be considered when designing surveys and survey questions that yield the most accurate and unbiased results of sexual behavior. The authors give specific focus to first sexual experiences, sexuality, attraction and identity, number of sexual partners, characteristics of sexual partnerships, and use of condoms. Following this discussion, in Chapter 9 Nicola Shelton and Jennifer Mindell discuss the measurement of *health* in social surveys. As health is a wide-ranging and broad topic, the authors do not go into great detail, but instead provide a brief description of terminology related to health research, a brief history of measuring health on surveys in the UK, and justification of the importance of its measurement. They also provide a short overview of topics such as methods of data collection, level of analysis, and some suggestions for the future (a discussion relevant to novice researchers and methodologists).

In Chapter 10, Yaojun Li examines the measurement of *social capital* by providing a comprehensive theoretical discussion of the various conceptualizations of this concept, and how some of them would (or would not) be appropriate to use in survey research. After this assessment, Li uses Putnam's (2000) theoretical work as a guide and conducts an empirical study that examines the prediction of social capital (measured by trust, political efficacy, and voting behaviors) using various measures of civic participation, political activism, and informal volunteering. This approach is unique relative to the other chapters, as it provides the reader with a clear empirical example of how theoretical work on a particular social construct can be used to inform and guide the operationalization of an empirical study. In Chapter 11, Caroline Roberts describes various methods of measuring the nonobservable concept of *social attitudes* through self-reporting. She focuses not on specific attitudes, but rather on the different sources that introduce variability and error into measuring attitudes, including an individual's cognitive processes, questionnaire design, and willingness of respondents to be truthful with their responses. Finally, the book ends with a summary-style chapter by Bulmer providing some discussion on social measurement, which culminates with the suggestion of two strategies for promoting better social measurement: greater use of replication and more emphasis on longitudinal or panel design. Unfortunately, these suggestions are only briefly mentioned and not elaborated upon in any detail.

The greatest contribution of Bulmer, Gibbs, and Hyman's *Social Measurement through Social Surveys* is its call for using theoretical and conceptual work to better guide the operationalization of our social measures. Even when recognized, this emphasis can easily be forgone while in the process of handling the details and logistics of creating a survey, and this volume serves as a reminder to researchers and methodologists that they should not lose sight of theory. In a number of the chapters, the authors visibly took this call to heart and provided clear examples of the benefits of emphasizing theory, particularly those authors writing on the measurement of families and households, social class, sexual

behavior, and social capital. Unfortunately, I did not see this as the case with other chapters, specifically those discussing crime, religious behavior, and health. Regardless, all of these chapters used a variety of UK-based surveys and individual survey questions throughout, providing readers with "real-life" examples to help demonstrate the various points being made throughout each chapter.

In sum, this volume provides insight on social measurement, and provides numerous examples using real data and survey questions. Its information may not be novel for those with experience in recognizing and applying theory to their research, but depending on the concept of interest, this book could serve as a welcome introduction to basic considerations for measuring specific social constructs. Regardless, the limited space of each chapter prevents the book from being a comprehensive guide to measuring any one topic, and readers will need to pursue additional sources to help better design theoretically-informed survey questions. While new information presented in this book may be limited for some experienced survey researchers, this volume would be most useful for students or researchers who are new to survey research, where it could begin to provide insight into social measurement in social surveys.

References

Erickson, R., Goldthorpe, J., and Portocarero, L. (1979). Intergenerational Class Mobility in Three Western European Societies. British Journal of Sociology, 30, 415–441.

Putnam, R. (2000). Bowling Alone: The Collapse and Revival of American Communities. New York: Simon and Schuster.

Brian W. Ward
National Center for Health Statistics,
3311 Toledo Road, Room 2330
Hyattsville, MD 20782
U.S.A.
Voice: 1-301-458-4568.
Email: bwward@cdc.gov

Estimation, testing, and prediction blend in this framework, producing opportunities for new methodologies of increased power. New difficulties also arise, easily leading to flawed inferences. This book takes a careful look at both the promise and pitfalls of large-scale statistical inference, with particular attention to false discovery rates, the most successful of the new statistical techniques. Published September 20th 2010 by Cambridge University Press (first published August 31st 2010). More Details ISBN. Start your review of Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Write a review. Jan 31, 2011 David Bickel rated it it was amazing. cambridge university press Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, SaÌƒo Paulo, Delhi, Dubai, Tokyo, Mexico City Cambridge University Press The Edinburgh Building, Cambridge CB2 8RU, UK Published in the United States of America by Cambridge University Press, New York www.cambridge.org Information on this title: www.cambridge.org/9780521192491 c B. Efron 2010 . This publication is in copyright. We will be using empirical Bayes ideas for estimation, testing, and prediction, beginning here with their path-breaking appearance in the James–Stein formulation. Empirical Bayes theory blurs the distinction between estimation and testing as well as between fre1. Empirical Bayes and the James–Stein Estimator. 2. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing and Prediction. Bradley Efron Stanford University. Foreword. At the risk of drastic oversimplication, the history of statistics as a recog-nized discipline can be divided into three eras 3 The era of scientic mass production, in which new technologies typi-ed by the microarray allow a single team of scientists to produce data sets of a size Quetelet would envy. Empirical Bayes blurs the line between testing and estimation as well as between frequentism and Bayesianism. Much of what follows is an attempt to say how well we can estimate a test-ing procedure, for example how accurately can a null distribution be esti-. Foreword.