VERSITA

Journal of Official Statistics, Vol. 30, No. 3, 2014, pp. 571–573, http://dx.doi.org/10.2478/JOS-2014-0037

Book Review

Joseph W. Sakshaug[1]

Raymond L. Chambers and Robert G. Clark. *An Introduction to Model-Based Survey Sampling with Applications*. 2012 New York, USA: Oxford University Press Inc., ISBN 978-0-19-856662-5, 265 pp., £41.99.

This book introduces readers to the fundamental concepts of the model-based approach to survey sampling. For those unfamiliar with the model-based approach, it is worthwhile to note that there are distinct differences between this approach and the more commonly taught design-based approach. In the design-based paradigm, survey inferences are obtained from estimators that are based on the assumption that samples are drawn repeatedly from a fixed and finite population. These design-based estimators lack any assumption regarding the structure of the population under study and therefore can be correctly applied in any setting. In contrast, the model-based approach to inference uses estimators that do not rely on repeated sampling properties, but rather take into account the properties of a specific population through the use of auxiliary information. The population is summarized through the use of a model and responses are assumed to be generated by a stochastic process. Model-based approaches to sampling and inference have received adequate attention in the literature. Some notable citations include Valliant (2009), Valliant et al. (2000), Brewer (1963), and Royall (1970). The book *An Introduction to Model-Based Survey Sampling with Applications* by Raymond L. Chambers and Robert G. Clark is a useful addition to this literature and should satisfy readers who are interesting in acquiring the theoretical and practical knowledge needed to carry out the model-based approach.

The book is divided into three parts. The first part (Chapters 1–7) covers the fundamental aspects of model-based survey inference. The second part (Chapters 8–10) introduces model-based survey methods that are robust to incorrectly specified models and outliers. Finally, the third part (Chapters 11–17) covers several modern applications of model-based survey inference.

Chapter 1 provides an introduction to survey sampling and introduces the relevant notation that is used throughout the book. Chapter 2 introduces the model-based approach and highlights important differences between this approach and other approaches to survey sampling. Chapter 3 considers the simplest possible model in which the survey population has no auxiliary variables (or the auxiliary variables are unrelated to the survey variables of interest). Several important questions are addressed, including how large the sample should be and how to carry out the sampling process. Chapter 4 builds on the

[1] Institute for Employment Research, Regensburger Straße 104, 90478 Nuremberg, Germany. Email: joe.sakshaug@iab.de

previous chapter by extending the model to account for stratified populations. Several stratification schemes are considered, including proportional and optimal allocation, equal aggregate size stratification, and multivariate stratification. Chapter 5 covers linear regression models for populations with a single auxiliary variable. Combining regression and stratification is also discussed. Chapter 6 deals with clustered populations in which the sampling process is carried out in two stages. The clustered population model is introduced for different optimal designs, including fixed sample size and fixed cost. Chapter 7 concludes the first part of the book by demonstrating how the population models presented in the earlier chapters are special cases of the general linear population model.

Chapter 8 demonstrates the robustness of the model-based approach. Specifically, the chapter shows that, under robust sampling designs, approximately unbiased inferences can be obtained even when the assumed model is incorrectly specified. Chapter 9 extends the idea of robustness to variance estimation under model misspecification. Chapter 10 describes strategies for making survey sampling techniques robust to extreme observations, or outliers. Nonparametric regression approaches are among those considered. Practical challenges of applying these outlier robust estimators are discussed at the end.

Chapter 11 departs from the linear population model and focuses on inferences based on non-linear population parameters, including population medians, quantiles, and ratios of two population means. Chapter 12 covers variance estimation for these complex statistics. Various techniques are considered, including random groups, balanced repeated replication, jackknife, and bootstrapping. Chapter 13 departs from the assumption used in previous chapters that there is only one survey variable of interest and deals with the reality that most surveys are multipurpose in nature and collect many Y variables. Switching from a univariate to a multivariate Y poses several design and estimation issues which are addressed here. Chapter 14 covers inference for population units belonging to particular domains. Situations in which domain membership is either known or unknown prior to data collection are both considered. Chapter 15 provides a thorough treatment of inference for small areas. Unit-level models for small areas are the focus of this chapter, which covers various methods including synthetic methods, methods based on random area effects, direct prediction methods, and the use of generalized linear mixed models. Chapter 16 presents methods for obtaining model-based inference about distributions and quantiles. These methods are considered under various survey designs, including stratification and clustering. An application of these methods is provided using data from a large-scale business survey. Lastly, Chapter 17 deals with the use of transformations to achieve linearity. A back transformation approach and a model calibration approach are described and empirical results of these approaches are provided.

In summary, the book covers a wide range of topics devoted to model-based survey sampling. I would highly recommend it for students and applied survey statisticians alike. I was particularly impressed with the book's ability to integrate both theoretical and applied concepts. While reading the book, I got a strong sense that it was written with the practitioner in mind. The book covers many practical survey applications, including an extensive chapter on small area estimation which is a topic growing in importance in survey research. Furthermore, the authors do a nice job of accommodating readers who may be less familiar with matrix algebra by waiting until Chapter 7 before introducing this

notation. If there is anything to criticize about the book, it would be the lack of a chapter devoted to longitudinal surveys which pose many design and estimation issues. Such surveys often have a rich amount of auxiliary information collected from prior waves and it would have been interesting to know how to utilize these data using model-based methods. Nevertheless, whether you are brand new to the topic of model-based survey sampling or are a seasoned user of such methods, this book should serve as a useful reference.

References

Brewer, K. 1963. "Ratio Estimation and Finite Populations: Some Results Deducible from the Assumption of an Underlying Stochastic Process." *Australian Journal of Statistics* 5: 93–105.

Royall, R.M. 1970. "On Finite Population Sampling Theory under Certain Linear Regression Models." *Biometrika* 57(2): 377–387.

Valliant, R. 2009. "Model-Based Prediction of Finite Population Totals." In *Chapter 13 of Handbook of Statistics 29: Sample Surveys: Design, Methods, and Applications*, edited by D. Pfeffermann and C.R. Rao,. Amsterdam: North Holland.

Valliant, R., A.H. Dorfman, and R.M. Royall. 2000. *Finite Population Sampling and Inference*. New York: Wiley.