The Survey Statistician, 2024, Vol. 90, 38-39

Book and Software Review

Sampling and Estimation from Finite Populations

Angelo Moretti[1]

[1] Department of Methodology and Statistics, Utrecht University, The Netherlands, a.moretti@uu.nl

Abstract

The book *Sampling and Estimation from Finite Populations* by Yves Tillé was published by Wiley in 2020 in the series Probability and Statistics. The book constitutes an important reference on survey sampling and estimation for a wide range of scholars, users, and practitioners, focusing on the 'representation side' of the Total Survey Error (TSE) framework (Groves and Lyberg, 2010). The book contains 16 chapters presenting both classical and modern approaches to survey sampling and estimation. The main focus of the book is on design-based inference; however, there is also some discussion on model-based techniques and design-based estimators that use auxiliary information in survey sampling and estimation. Importantly, it covers also issues on nonresponse, which, nowadays is a great deal in data collection studies.

Keywords: sampling, complex survey design, design-based inference, variance.

The book *Sampling and Estimation from Finite Populations* by Yves Tillé, published by Wiley in 2020, provides a comprehensive overview of modern survey sampling and estimation problems, and the author importantly defines the discipline as a 'living discipline'. The book has a technical nature, however, sometimes, some demonstrations are omitted; in this case, the author provides bibliographical reference for these, which can be helpful to interested readers. The main focus of the book is on the design-based approach; given that, as Tillé stresses, that is consistent and ethically acceptable to public statistics.

After two introductory chapters, i.e., Chapters 1 and 2, where an historical overview is given and the concepts of population, sample and estimation are introduced, the book covers two areas of the TSE framework, i.e., sampling and estimation issues. Various sampling designs including simple and systematic sampling, stratified sampling, unequal probability sampling designs, balanced sampling, cluster and two-stage sampling are extensively presented and discussed. Furthermore, topics such as spatial sampling, sampling coordination, and multiple survey frames are addressed. A comprehensive discussion on some important estimators is presented, i.e., ratio, difference, regression, poststratified, and calibration estimators. Each chapter contains exercises, and their solutions are provided in Chapter 17.

Chapter 3 is about simple designs, such simple random sampling, Bernoulli sampling, and systematic sampling. It also focuses on specific problems such as sample size determination and entropy calculation. Chapter 4 introduces the problem of stratification, which allows the use of auxiliary information in the sampling design. Different types of allocation are discussed and the notion of optimality is applied to estimators other than totals. Other issues, such as taking into account for costs and power allocation are presented. Chapter 5 is about unequal probabilities

designs, i.e., systematic sampling with unequal probabilities, Poisson sampling, the Rao-Sampford and the Brewer methods. Other methods are also presented and discussed in detail. The problems of variance estimation and entropy are also presented. Chapter 6 is on balanced sampling, which draws samples whose expansion estimators are equal to or at least very close to the population totals for one or more auxiliary variables. Here, different methods are presented, putting an emphasis to the Cube method. Variance approximation and estimation procedures are presented. The chapter concludes with special cases and practical aspects of this sampling design**.** Chapter 7 is about clustering and two-stage designs, where auxiliary information may be used to improve the organisation of the survey. More advanced issues are discussed here, such as self-weighting two-stage design and multi-stage designs. Chapter 8 presents a wide range of problems, e.g., among the other problems, spatial sampling, modifications of stratified sampling, coordination in repeated surveys, multiple frames survey and adaptive sampling designs. Overall, this aspect is particularly innovative as these topics are not always covered in a basic sampling textbook.

Design-based estimators that use auxiliary information are presented in Chapter 9 (ratio estimator), Chapter 10 (post-stratification and calibration), and Chapter 11 (regression estimator). The important problem of calibration is treated in Chapter 12. A wide range of approaches are presented here, focusing also on the problem of generalised calibration and its application in practise. Furthermore, the author provides the reader with references to various software (among others R, SAS, SPSS) through which it is possible to develop calibration estimators.

Although the book focuses on design-based approaches mainly, Chapter 13 introduces the model-based inference problem. Here, some model failures issues are discussed. Interestingly, the chapter concludes with a robust approach by looking at the inference from both the design and the model. The issue of estimation of complex population parameters is presented in chapter 14. For example, the problems of estimating the covariance, and the Gini index are presented, as well as quantile estimation.

Chapter 15 is about an important challenge, i.e., variance estimation. Here, many different linearization methods are proposed, which constitutes a great review of approaches. Chapter 16 is about non-response and how to treat it. Reweighting and imputation procedures are discussed and, the link between regression imputation and reweighting is presented. The topic of nonresponse is often overlooked in similar books on survey sampling, however, it is a crucial issue in nowadays data collection studies, especially in social surveys where nonresponse rates are dramatically increasing. Importantly, in this chapter, the author considers both target parameter estimation problems but also variance estimation and provide software suggestions to deal with this usually underestimated problem.

In conclusion, the book begins with detailed introductory chapters, ensuring it can be understood by a wide range of readers, including statisticians, professionals, and researchers alike. With a collection of exercises featuring clear, synthetic solutions, it serves as a valuable resource for advanced survey sampling courses. Certain more advanced chapters delve deeper into technical aspects, demanding a solid understanding of survey sampling theory. To highlight is the extensive bibliography, often updated with recent years, presented in this book to allow the reader to delve into topics of their own interest. Undoubtedly, this book stands out as a significant contribution in the field of survey sampling theory.

References

Groves, R.M. and Lyberg, L., 2010. Total survey error: Past, present, and future. Public opinion quarterly, 74(5), pp.849-879.