sciendo

Journal of Official Statistics, Vol. 37, No. 4, 2021, pp. 1079–1081, http://dx.doi.org/10.2478/JOS-2021-0046

Book Review

Alina Matei[1]

Yves Tillé. *Sampling and Estimation from Finite Populations*. 2020 New York: Wiley, ISBN: 978-0-470-68205-0, 448 pages.

In 2020, Wiley published the book by Yves Tillé entitled 'Sampling and Estimation from Finite Populations'. This is the English version of the book 'Théorie des sondages: échantillonnage et estimation en populations finies' (second edition) originally published in French (Tillé 2019). The book is the outgrowth of 30 years of experience of the author in survey sampling research and practice. Yves Tillé authored more than 70 peer-reviewed journal articles and six books on survey sampling.

Featuring a broad range of topics, the book contains 16 chapters and offers both a classical and a modern view on sampling and estimation, from the history of survey sampling to nonresponse treatment. The book is broken down into two main parts; the first one is dedicated to sampling methods, and the second one to estimation problems. In the sampling part, simple and systematic sampling designs, stratified sampling, unequal probability sampling designs, balanced sampling, cluster and two-stage sampling are presented in detail. Additional topics on spatial sampling, sampling coordination, and multiple survey frames are more generally discussed and included in a single separate chapter. In the estimation part, estimators such as the ratio, the difference, the regression, the poststratified and the calibration estimators are fully covered.

Focusing on the sampling part, the book addresses two topics not fully discussed by the existing books; these are unequal probability sampling designs and balanced sampling. Yves Tillé provides a deep insight into unequal probability sampling designs that make use of first-order inclusion probabilities computed using auxiliary information. The strategy used to estimate a total is given by an unequal probability sampling design (without replacement) and the Horvitz-Thompson estimator. Chapter 5 presents a number of sampling schemes, including systematic sampling with unequal probabilities, Poisson sampling, the Rao-Sampford method, the Brewer method, as well as a few less popular sampling designs, such as order sampling, the pivotal method, and Deville's systematic sampling. All these are given together with variance estimation/approximation of the Horvitz-Thompson estimator. The entropy of a sampling design with unequal probabilities (Hájek 1981) is discussed in Subsection 5.3 (for equal probability sampling designs, see Subsection 3.11). This measure of randomness of a sampling design is then considered in connection with Poisson sampling and maximum entropy design (also known as conditional Poisson sampling). The choice of an optimal unequal probability sampling design is discussed, and the

[1] Institute of Statistics, University of Neuchâtel, Switzerland. Email: alina.matei@unine.ch

author concludes on page 110 that 'unfortunately, there is no ideal method'. Chapter 6 is dedicated to balanced sampling that (approximately) recovers totals of known auxiliary variables from the sample. It offers a major review on this sampling design, mainly focusing on the cube method introduced by Deville and Tillé (2004). The cube method is revisited in Chapter 8, where it is used to provide a sample that is both spread out geographically and balanced on auxiliary information; this method is called 'local cube method' (see Grafström and Tillé 2013).

Chapters 9, 10, and 11 focus on estimation and present design-based estimators that make use of auxiliary information. These are classical estimators such as the ratio, the regression and the postratified estimators. Chapter 12 offers a significant review on the calibration estimator (Deville and Särndal 1992). Together with the article of Särndal (2007), it currently represents one of the most important reviews on this topic from the design-based point of view. The chapter overviews the existing distances and calibration functions, discusses the main algorithm based on the Newton-Raphson procedure, and the use of the bounds in calibration. The chapter ends with a look at the generalized calibration, a very useful method to correct the bias due to non-ignorable nonresponse. This is briefly reconsidered in Chapter 16, which is dedicated to nonresponse treatment.

The book is mainly concerned with the design-based approach of inference. Chapter 13 copes, however, with model-based approach, where the inference is based on a superpopulation model conditionally on the selected sample. A regression model without intercept, relating the variable of interest to the set of auxiliary variables, that it is the usual model in survey sampling, is used as the superpopulation model. Model-based and design-based approaches are not presented as competitors. One finds in this chapter two topics advocated in the previous chapters: calibration estimator and balanced sampling. Model-based approach is shown to support an important design-based estimator, the calibration estimator, which is unbiased under the advocated model. On the other hand, the balanced sampling reduces the anticipated variance of the Horvitz-Thompson estimator of the population total, under a mixed approach, that is both model and design-based.

Chapter 14 on 'Estimation of Complex Parameters' and Chapter 15 on 'Variance Estimation by Linearization' investigate less common topics. Estimation of Lorenz curves, quantile share ratios and Gini indexes given in Chapter 14 are innovative topics in survey sampling books. The author provides an impressive work in Chapter 15 on variance estimation by linearization for complex statistics (such as logistic regression coefficients and Gini index), creating possibly the most extended review on this topic to date.

The book includes comprehensive introductory chapters, making it accessible to a broad audience, including survey statisticians, practitioners and researchers. Given that a set of exercises with summary solutions are available, the book is also an excellent support for advanced courses in survey sampling. Some chapters are more technical, and require more knowledge in survey sampling theory. Without any doubt, the book represents a salient contribution to survey sampling theory. I hope that it will soon be included in the list of the most influential books on survey sampling, such as Särndal et al. (1992) and Lohr (2019).

References

Deville, J.-C., and C.-E. Särndal. 1992. "Calibration estimators in survey sampling." J*ournal of the American Statistical Association* 87: 376–382. DOI: https://doi.org/10.1080/01621459.1992.10475217.

Deville, J.-C., and Y. Tillé. 2004. "Efficient balanced sampling: the cube method." *Biometrika* 91: 893–912. DOI: https://doi.org/10.1093/biomet/91.4.893.

Grafström, A., and Y. Tillé. 2013. "Doubly balanced spatial sampling with spreading and restitution of auxiliary totals." *Environmetrics* 14(2): 120–131. DOI: https://doi.org/10.1002/env.2194.

Hájek, J. 1981. *Sampling from a Finite Population*. Marcel Dekker, New York.

Lohr, S. 2019. *Sampling: Design and Analysis*. CRC Press, second edition.

Särndal, C.-E. 2007. "The calibration approach in survey theory and practice." *Survey Methodology* 33 (2): 99–119.

Särndal, C.-E., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. Springer Verlag, New York.

Tillé, Y. 2019. *Théorie des sondages: échantillonnage et estimation en populations finies*. Dunod, Paris, deuxième edition.