Sample Survey Theory and Methods: Past, Present, and Future Directions

J. N. K. Rao*
Carleton University, Ottawa, Canada, jrao34@rogers.com

Wayne A. Fuller
Iowa State University, Ames, USA, waf@iastate.edu

Abstract

A large part of sample survey theory has been motivated by practical problems encountered in the use of samples to obtain information. We discuss developments in sample survey theory and methods covering the past 100 years. The International Statistical Institute played a vital role in the development of survey sampling in the 1920s by creating a committee to report on the role of random sampling in estimating finite population totals. Neyman's 1934 landmark paper laid the theoretical foundation for the probability sampling approach to inference from survey samples. Classic sampling books by Cochran, Hansen, Hurwitz and Madow, Sukhatme, and Yates, which appeared in the early 1950s, expanded and elaborated the theory of probability sampling, emphasizing unbiasedness, the model-free features of the theory and developing designs to minimize variance for a fixed cost. These books were the foundation of survey practice for decades. During the period 1960-70, theoretical foundations of inference from survey data received attention, with the model-dependent approach suggested by Brewer and Royall generating considerable discussion. The introduction of general purpose statistical software led to the use of such software with survey data, which led to the design of methods specifically for complex survey data. At the same time, weighting methods, such as regression estimation and calibration, became practical. A bit later computer-intensive resampling methods also became practical for large scale survey samples. Improved computer power led to more sophisticated imputation for missing data, use of more auxiliary data, and more complex estimation procedures in general. Longitudinal surveys became more common and some treatment for measurement error beyond attempts to reduce the effect occurred. Design consistency replaced unbiasedness as the requirement for standard estimators. The most notable use of models was in the expanded use of small area estimation.
Future directions in research and methods will be influenced by budgets, response rates, timeliness, improved data collection devices, and availability of auxiliary data, some of which will be called "big data". Most importantly, survey taking will be impacted by changing cultural behavior and by a changing physical-technical environment.

Keywords: History of survey sampling, survey practice, survey inference.