30th Annual Morris Hansen Lecture

The Evolution of the Use of Models in Survey Sampling

Wednesday, November 16, 2022
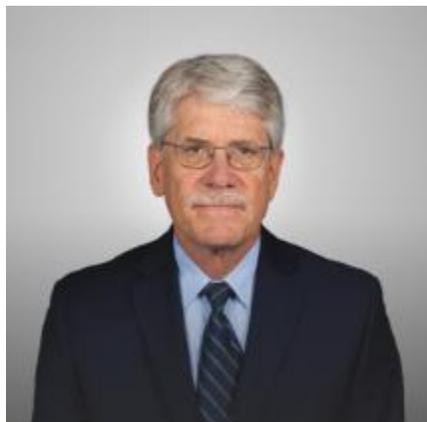3 – 5 p.m. EST

Register for Free

Jefferson Auditorium
USDA South Building
1400 Independence Ave. SW (between 12th and 14th Streets)
Smithsonian Metro Station (Blue, Orange and Silver Lines)
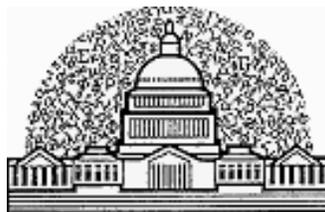*Registration Required*

Keynote Speaker: **Dr. Richard Valliant**
Research Professor Emeritus
University of Michigan, Ann Arbor and University of Maryland, College Park

Hosted by USDA's National Agricultural Statistics Service (NASS)
Sponsored by NASS, Washington Statistical Society, and Westat

Richard Valliant is a Research Professor Emeritus at the University of Michigan, Ann Arbor and the Joint Program for Survey Methodology at the University of Maryland, College Park. He received his PhD from Johns Hopkins University in Biostatistics and an MS in statistics from Cornell University. He has over 45 years of experience in survey sampling, estimation theory, and statistical computing.

He was formerly an Associate Director at Westat and a mathematical statistician with the Bureau of Labor Statistics (BLS). He has a range of applied experience in survey estimation and sample design on a variety of establishment, institutional, and household surveys, including the Consumer Price Index, the Current Population Survey, and other surveys done by BLS, the National Center for Education Statistics, the Consumer Product Safety Commission, the Department of Energy, and the National Agricultural Statistical Service among others. He is a Fellow of the American Statistical Association and an elected member of the International Statistical Institute. He was an associate editor of the *Journal of the American Statistical Association—Theory and Methods Section* (1989-1993) and the *Applications and Case Studies Section* (1996-1999), *Journal of Official Statistics* (2003-2010), and *Survey Methodology* (1996-2007). During his career, he has given invited lectures and taught courses in a number of other countries, including Canada, Chile, France, Germany, Italy, and Sweden.

Dr. Valliant is the co-author of three books: *Finite Population Sampling and Inference: A Prediction Approach* (2000) with A. Dorfman and R.M. Royall; *Survey Weights: A Step-by-step Guide to Calculation* (2018) with J.A. Dever; *Practical Tools for Designing and Weighting Survey Samples*, (2018, 2nd edition) with J.A. Dever and F. Kreuter. The first edition of the *Practical Tools* book was the winner of the 2020 Book Award from the American Association for Public Opinion Research. He is also the author of the R packages: *PracTools* and *svydiags*.

Dr. Valliant began his career at Westat where he worked with Morris Hansen, Joseph Waksberg, Ben Tepping, and other statisticians who were key contributors to early developments in the theoretical and practical applications of survey sampling. Following that, he was a student of Richard Royall who was one of the main proponents of the model-based approach to finite population inference. His own work has covered the design-based, model-based, and model-assisted approaches to sampling inference.

Abstract

The use of models in survey estimation has evolved over the last five (or more) decades. This talk will trace some of the developments over time and attempt to review some of the history. Consideration of models for estimating descriptive statistics began as early as the 1940s when Cochran and Jessen proposed linear regression estimators of means. These were early examples of model-assisted estimation since the properties of the Cochran-Jessen estimators were calculated with respect to a random sampling distribution. Model-thinking was used informally through the 1960s to form ratio and linear regression estimators that could in some applications reduce design variances.

In a 1963 Australian Journal of Statistics paper, Brewer presented results for a ratio estimator that were entirely based on a superpopulation model. Royall (Biometrika 1970 and later papers) formalized the theory for a more general prediction approach using linear models. Since that time, the use of models is ubiquitous in the survey estimation literature and has

been extended to nonparametric, empirical likelihood, Bayesian, small area, machine learning, and other approaches. There remains a considerable gap between the more advanced techniques in the literature and the methods commonly used in practice.

In parallel to the model developments, the design-based, randomization approach was dominating official statistics in the United States largely due to the efforts of Morris Hansen and his colleagues. In 1937, Hansen and others at the Census Bureau designed a follow-on sample survey to a special census of the employed and partially employed because response to the census was incomplete and felt to be inaccurate. The sample estimates were judged to be more trustworthy than those of the census itself. This began Morris' career-long devotion to random sampling as the only trustworthy method for obtaining samples from finite populations and for making inferences.

Model-assisted estimation, as discussed in the 1992 book by Särndal, Swensson, and Wretman, is a type of compromise where models are used to construct estimators while a randomization distribution is used to compute properties like means and variances. This thinking has led to the popularity of doubly robust approaches where the goal is to have estimators with good properties with respect to both a randomization and a model distribution.

The field has now reached a troubling crossroads in which response rates to many types of surveys have plummeted and nonprobability datasets are touted as a way of obtaining reasonable quality data at low cost. Sophisticated model-based mathematical methods have been developed for estimation from nonprobability samples. In some applications, e.g., administrative data files that are incomplete due to late reporting, these methods may work well. However, in others the quality of nonprobability sample data is irremediably bad as illustrated by Kennedy in her 2022 Hansen lecture. In some situations, we are back in Morris' 1937 situation where standard approaches no longer work. Methods are needed to evaluate whether acceptable estimates can be made from the most suspect data sets. Nonetheless, nonprobability datasets are readily available now, and it is up to the statistical profession to develop good methods for using them.

Program

Jefferson Auditorium
South Building
United States Department of Agriculture
Thursday, November 16, 2022

Opening Remarks
Partha Lahiri (Program Chair)
Professor, University of Maryland, College Park

Keynote Speaker
Dr. Richard Valliant
Research Professor Emeritus
University of Michigan, Ann Arbor and University of Maryland, College Park

Discussants
Dr. Jay Breidt
Senior Fellow, NORC at the University of Chicago

Dr. Trivellore Raghunathan
Professor, University of Michigan, Ann Arbor

Hansen Lecture Committee
Partha Lahiri (Chair)
Brian Harris Kojetin (Past Chair)
Katherine J. Thompson (2023 Chair)
Carolina Franco (2024 Chair)
Keith Rust (Westat Representative)
Linda Young (NASS Representative)

Reception
Patio - Department of Agriculture
Jamie L. Whitten Building
(Across Independence Avenue)
5 - 6 p.m.

Morris Hansen Memorial Lecture Series

Morris Howard Hansen has been described as the most influential statistician in the development of survey methodology in the twentieth century. Early in his Census Bureau career he put together a staff to define the principal problems in the conduct of surveys, investigate these problems, and develop statistical methods to address them. Morris and his staff then widely distributed the results of their efforts, thus influencing statistical agencies all over the world. Generations of statistical students have learned from and been influenced by *Sample Survey Methods and Theory*, Volumes I and II by Hansen, Hurwitz, and Madow, which are commonly referred to by the authors' names instead of the title, and Hansen's other publications.

Morris was also known as an innovator and a leader in adapting electronic tools, such as computers and mark-reading sensors, to statistics. After his outstanding Census Bureau career, Morris joined Westat which was at the time a fairly small statistical research company. Morris again assembled a strong staff and expanded Westat's scope to take on large federal government statistical problems.

Morris also made outstanding contributions to professional organizations, serving as the president of both the American Statistical Association and the Institute of Mathematical Statistics and as the first president of the International Association of Survey Statisticians. He was elected to the National Academy of Sciences in 1976 and was an important member of many Academy committees and panels.

There have been many tributes to Morris since his death in 1990, such as memorial issues of both the *Journal of Official Statistics* and *Survey Methodology*. Westat issued a grant to the Washington Statistical Society to honor Morris with an annual lecture series. The series has been so successful in attracting top quality presentations on a wide variety of topics—in keeping with Morris' broad interests—that Westat has added to the original grant.

Sponsored by:

Washington Statistical Society

Westat®

AGRICULTURE COUNTS

30th Morris Hansen Lecture Series