Analysis of complex sample survey data using R

Joint Program in Survey Methodology
April 21 - May 02, 2025

Dates

Monday April 21, 2025 - Friday May 02, 2025.

Course objectives

The first objective is to explain what is hiddent behind the sampling weihts: Survey sampling data are the outcome of different random processes, from the random process that generated the characteristics of interest of the elements of a population, the sample selection process, as well as other perturbation processes as measurement error coarsening and non response. Analysis of survey data consists in taking into account all the processes that influenced the final outcome (the survey data), as well as all the relevant auxiliary information available to propose reliable estimators and variance estimators.

The second objective is to provide tools for inference, the course will focus mainly on the estimation of population totals and means from complex survey designs data under a model-assisted framework and weight adjustment methods, but will also provide an overview of likelihood based techniques for estimation of population model parameters.

The course will provide access to an interactive R shiny app that allows to compare different estimation techniques and help understand the assumptions that can guide the choice of one estimator against another. Real publicly available survey data from the Current Population Survey containing a series of different sampling weights will be used.

The R code to draw samples and analyzed survey data will be shared, and a review of the relevant R packages, including the packages sampling, and survey will be provided.

The instructor

Dr. Daniel Bonnéry is a Survey Statistician whose main domain of research is inference under informative selection. He has worked as a survey statistician and as a researcher at the University of Cambridge, at the Insee (The French National Statistical Institute), the French Ministry of Agriculture, and at the US Census Bureau, as well as a researcher on survey sampling for the Ensai (French National School of Statistics and Analysis of the Information), the Toulouse School of Economics, and at the Joint Program in Survey Methodology at the University of Maryland.

Contents

In details, the course will provide a review of the main concepts of survey sampling, for different inference paradigms:

- Basics of frequentist statistical inference: model, observations, estimation.

- Population models

    – Fixed population model for design-based inference

    – General population model for model-based inference

- Sample selection

    – Design, usual designs

    – Sample selection

    – Inclusion and double inclusion probabilities

- Other nuisance processes

    – Definition

    – Examples: non response, measurement error, and will focus on advanced inference

    techniques:

- What questions to ask when being handed survey data for analysis?

- Targets of inference: Population totals, means, population model parameters.

- Inference, parameter and variance estimation:

    – Model-assisted inference on the population characteristics, weight adjustment methods.

    – Likelihood based methods for population parameter estimation

Who should attend

Researchers and analysts who need to acquire any of the following:

- a global understanding of the random processes that generated the sample and the observations on the sample, that may or may not include the survey weights.

- a knowledge of the main weight adjustment techniques and their limitations.

- the ability to produce survey weighted estimates, apply different weight adjustement techniques, perform variance estimation of survey weighted estimators.

R, Rstudio

Registrants would find it beneficial to have a laptop with R (version ≥ 4.1.0) installed and Rstudio.

Course materials

Registrants will be provided with a course lecture notebook, as well as online access to lecture pdf slides, lectures, and R package containing code for illustration and interactive apps.

Course Schedule

Live sessions will be Friday April 25 and March 02, 3:00 PM to 4:30 PM EST.

The slides and talk are pre-recorded, spreading it out over 2 weeks.
Week 1, Monday – Thursday: The entire Week 1 recordings would be made available Online and participants could watch them at their own pace.
Week 1, Friday: A 1.5-hour live Online discussion, Q A with the instructor.
Week 2, following the same schedule as Week 1, (recorded lectures M-Th, live Online meeting Friday).