

Workshop on Analysis of Survey Data

C. Boudreau, Ph.D.

Dept. of Statistics & Actuarial Science
University of Waterloo, Waterloo, Ontario

Abstract

This workshop on the Analysis of Survey Data is comprised of three modules/parts. The first one will be a general talk about why survey data differ from data collected by simple random sampling, and an introduction to survey methodology and concepts. In particular, we will discuss stratification, clustering, probability proportional to size sampling, and the Horvit-Thompson (HT) estimator (with its variance estimator).

The second part will focus on descriptive inference from survey data; in particular the use of the ratio estimator to estimate means and percentages/proportions of a finite population. Using SAS and Stata, we will illustrate how to estimate totals, means and percentages/proportions of a finite population, as well as their corresponding confidence intervals. These examples will utilize data from the International Tobacco Control (ITC) Policy Evaluation Project.

The third part will focus on analytical inference from survey data. Specifically, how does it differ from descriptive inference, its relationship to the concept of super-population, and how to perform such inference using pseudo-likelihood methodology. Examples of such inference using logistic regression and generalized estimating equations (GEE) routines in SAS, Stata and Sudaan (a specialized software for the analysis of data from complex surveys) will be discussed. These will be based on data from the ITC Project. Analytical inference for the Cox model will also be discussed and exemplified using Sudaan Proc Survival.