

Official statistics on population parameters from a nonprobability sample or by integration with a probability sample

Danny Pfeffermann¹, Daniela Marella² and Donato Summa³

Abstract

Tightened budgets and constant decrease of response rates, due in part to increasing response burden in traditional surveys, have stimulated research into the use of non-probability sample data. In this paper, we consider the case where a large non-probability sample with observations on variables Y and X is available, but this sample is subjected to selection bias. An informative probability sample, with not missing at random non-response may also be available, but this sample only contains observations on X . The aim is to estimate the joint probabilities of (X, Y) and/or the total of the Y -values in the population which the samples are supposed to represent. We illustrate and compare the performance of several methods by use of simulations and by an application to real enterprise data from Italy.

Keywords: Empirical likelihood, NMAR nonresponse, Sample integration, Selection bias.

¹ University of Southampton and Hebrew University of Jerusalem: d.pfeffermann@soton.ac.uk.

² Sapienza University of Rome: daniela.marella@uniroma1.it

³ Italian National Institute of Statistics (Istat): donato.summa@istat.it