

NON-PROBABILITY SURVEYS: A REVISION OF METHODS FOR INFERENCE

M. Rueda¹ and R. Ferri-García² and L. Castro-Martín³

¹ University of Granada, Spain
e-mail: mrueda@ugr.es

² University of Granada, Spain
e-mail: rferri@ugr.es

³ Andalusian School of Public Health, Spain
e-mail: luiscastro193@gmail.com

Abstract

Since the early XXth century, probability samples have been the standard procedure for the obtention of information from a population of interest, in those cases where a census was not feasible. However, the decreasing response rates and growing costs of traditional survey methods, which can guarantee probability sampling schemes to a certain extent, have favored the rise of non-probability samples to obtain information from a population of interest. Non-probability samples are often obtained from online surveys or using procedures aimed to collect large amounts of passive data. Despite their lower cost and immediacy, these samples entail a number of drawbacks, especially regarding their selection bias. This bias can be mitigated using design-based and model-based methods developed in literature. Design-based methods, such as Propensity Score Adjustment and Kernel Weighting, aim to estimate the probability of an individual of the population of being included in the non-probability sample, and use them to obtain weights or to match individuals with similar characteristics in an available probability sample. Model-based methods, such as Statistical Matching, also known as Mass Imputation, model-assisted or model-calibrated estimators, aim to predict the value of the target variable in a probability sample or a complete census of the population where the target variable has not been measured. In this session, we describe and compare the available methods for inference in non-probability samples, and explain how Machine Learning techniques could boost these methods. Finally, we give some recommendations on further research lines regarding estimation from non-probability samples.

Keywords: Non-probability sampling, Kernel Weighting, Propensity Score Adjustment, Statistical Matching, model-based estimators.