

Estimation methods for integrating probability and non-probability survey samples

B. Cobo¹, M. Rueda¹, S. Pasadas², L. Castro³ and R. Ferri¹

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

² Institute for Advanced Social Studies, Spain
e-mail: spasadas@iesa.csic.es

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

Abstract

In recent years, different methods have been provided to combine information from multiple data sources. We focus on the case of probability and non-probability samples that share the same questionnaire, combining both to maximize the efficiency of the estimates with the help of machine learning methods. We develop a new estimation method to integrate data from probability and non-probability samples, we evaluate the efficiency of the resulting estimates by comparing them with other strategies that have been used before. The application of this method to the second wave of the Survey on the impact of the COVID-19 pandemic in Spain allows us to conclude that the estimation method we propose is the best option to reduce the biases observed in our data.

Keywords: non-probabilistic surveys, machine learning techniques, propensity score matching, survey sampling.