

Estimating Proportions from Integrated Probability and Non-Probability Samples

V. Nekrašaitė-Liege^{1,2}, A. Čiginas^{1,3} and D. Krapavickaitė⁴

¹ State Data Agency, Statistics Lithuania, Lithuania

² Vilnius Gediminas Technical University, Lithuania
e-mail: Vilma.Nekrasaite-Liege@vilniustech.lt

³ Vilnius University, Lithuania
e-mail: Andrius.Ciginas@mif.vu.lt

⁴ Lithuanian Statistical Society, Lithuania
e-mail: Danute.Krapavickaite@gmail.com

Abstract

The estimation of finite population parameters by combining data from probability and non-probability samples is a prevalent research topic in survey sampling today. Various scenarios for data availability have been investigated in the literature. In some cases, the study variable is observed only in a non-probability sample and there is no possibility to combine samples at the unit level (Chen et al., 2020 and Yang et al., 2020). In other cases, both samples can be combined at the unit level and the study variable is observed either in non-probability sample (Kim and Haziza, 2014 and Kim and Wang, 2018) or both (Tam and Kim, 2018).

In this study, we focus on the situation where the study variable is available in both samples, and we aim to estimate the proportion using post-stratification and composite estimation methods.

We pay attention to the assessment of the variance for the estimator, taking into account not only the randomness of the probability sample but also the randomness of the non-probability sample. The influence of the non-probability sampling on the variance estimator is evaluated with respect to the distribution of estimated propensity scores.

Keywords: non-probability sample, post-stratification, propensity score, composite estimator.

References

- Chen, Y., Li, P. and Wu, C. (2020). Doubly robust inference with non-probability survey samples. *Journal of the American Statistical Association* **115:532** 2011-2021, DOI: 10.1080/01621459.2019.1677241.
- Kim, J. K. and Haziza, D. (2014). Doubly robust inference with missing data in survey sampling. *Stat. Sin.* **24**, 375-394.
- Kim, J. K. and Wang, Z. (2019). Sampling Techniques for Big Data Analysis. *International Statistical Review* **87**.
- Tam, S.-M. and Kim J.-K. (2018). Big data, selection bias and ethics – an official statistician’s perspective. *Statistical Journal of the IAOS* **34**, 577–588.
- Yang, S., Kim, J. K. and Song, R. (2020). Doubly robust inference when combining probability and non-probability samples with high-dimensional data. *J R Stat Soc Series B Stat Methodol.* **82(2)**, 445-465. doi: 10.1111/rssb.12354. Epub 2020 Jan 7. PMID: 33162780; PMCID: PMC7644042.