

On the Use of Auxiliary Variables in Multilevel Regression and Poststratification

Yajuan Si¹

¹ University of Michigan, USA
e-mail: yajuan@umich.edu

Abstract

Multilevel regression and poststratification (MRP) is a popular method for addressing selection bias in subgroup estimation, with broad applications across fields from social sciences to public health. In this paper, we examine the inferential validity of MRP in finite populations, exploring the impact of poststratification and model specification. The success of MRP relies heavily on the availability of auxiliary information that is strongly related to the outcome. To enhance the fitting performance of the outcome model, we recommend modeling the inclusion mechanisms conditionally on auxiliary variables and incorporating flexible functions of estimated inclusion probabilities as predictors in the mean structure. We present a statistical data integration framework that offers robust inferences for both probability and nonprobability surveys, addressing various challenges that arise in practical applications. Our simulation studies indicate the statistical validity of MRP, which involves a tradeoff between bias and variance, with greater benefits for subgroup estimates with small sample sizes, compared to alternative methods. We have applied our methods to the Adolescent Brain Cognitive Development (ABCD) Study, which collected information on children across 21 geographic locations in the U.S. to provide national representation, but is subject to selection bias as a nonprobability sample. We focus on the cognition measure of diverse groups of children in the ABCD study and show that the use of auxiliary variables affects the findings on cognitive performance.

Keywords: data integration, nonprobability sample, robust inference, model-based, selection/nonresponse bias

References

Y. Si (2023). On the Use of Auxiliary Variables in Multilevel Regression and Poststratification, under review, available at <https://arxiv.org/abs/2011.00360>.