



Bayesian post-stratification models using multilevel penalized spline regression

Qixuan Chen*

Columbia University, New York, USA – qc2138@columbia.edu

Yajuan Si

University of Wisconsin-Madison, Wisconsin, USA – ysi@wisc.edu

Post-stratification is a standard approach to account for differences between survey samples and their target population by using auxiliary variables known in the population. However, classical post-stratification estimator can have unacceptably high variance because of sparse or empty post-stratification cells. Instead, we proposed a multilevel penalized-spline post-stratification model, in which in the first level of model a distinct mean and variance are assumed in each post-stratum, and in the second level of model the cell mean and variance is further assumed to follow a normal distribution with the mean as a spline function of sample weights. The multilevel model not only facilitates the estimation of cell mean and variance in sparse or empty cells but also gains efficiency in the survey estimates. Our simulation studies showed that the Bayesian modeling outperforms the classical weighting approach with smaller mean squared error, better confidence coverage and shorter interval width. We applied our proposed method to estimate the prevalence of mental health illness in a survey sample of Ohio Army National Guard Service members.

Keywords: penalized spline; post-stratification; raking; small-area estimation.