Penalized spline propensity prediction imputation method with machine learning techniques

Hyonggin An
Korea University

Missing data lead to biased estimator and reduction of power in the study. A popular method for missing data is a model-based imputation. However, the method produced bias in the estimator when the imputation model is mis specified. As an alternative, a penalized spline propensity prediction (PSPP) method was proposed by Little and An (2004). In this method, if the response propensity is correctly predicted or the imputation model is correctly specified, it gives a consistent estimator. In this study, we propose that the response propensity is predicted by robust machine learning methods such as a random forest and a generalized boosting for more robustness in the PSPP method. This proposed method is compared with various imputation methods for the evaluation of its performance.