Useful Post-data Treatments for Nonresponse

Yuean Wang

2024-02-03

Introduction

While surveys have been one of the most important methods to learn about how people live, think, or respond to different things, the nonresponses in the surveys has always been a problem that prevents people from getting reliable results. There're numerous reasons for why certain people would not like to answer certain questions, and such condition poses a great challenge to the statistians in getting an estimation with low bias and low variance.

Over the years, several different methods have been proposed and improved by different people, trying to compensate for the influence from nonresponses, and make the survey result more trustworthy.

In the next section, 3 different types of treatments for dealing with nonresponses are discussed, along with their advantages or disadvantages.

Different treatments

General calibration models

One of the most widely used method for handling nonresponses is to use other information to calibrate the result. Such models include generalized regression (GREG) estimator, post-stratification (a special case of GREG), ranking, and so on.

Han, et al. discussed the influences from different factors on the estimation of three different estimators (See Han and Valliant 2020):

1. GREG: $\hat{t}_{yreg} = \hat{t}_{y\pi} + (N - \hat{N})^T \hat{B}_r$
2. Post-stratification: $\hat{t}_{yPS} = \sum_{i=1}^{2} \sum_{j=1}^{2} N_{ij} \hat{\bar{y}}_i j$ (See Holt and Smith 1979)

---

*Code and data are available at: https://github.com/yueanchristiwang/Useful-Post-data-Treatments-for-Nonresponse

3. Ranking: $\hat{t}_{yrk} = \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k \in r_{ij}} w_k y_{ijk}$ (See Jean-Claude Deville and Sautory 1993)

While those estimators are shown to be generally equivlant in the absense of nonresponse by the authors, they will no longer be asymptotically equivalent in the presense of nonresponse errors.

The three different kinds of estimators are then used to estimate the main effect on the outcome Y of two categorical variables α_Y and β_Y, both with two levels. The data are randomly generated, to simulate simple random sampling (SRS) in the surveys.

In the simulations, the following factors are considered:

1. Variable scenarios: high R^2 in the model and no interaction, high R^2 in the model with interaction, and medium R^2 in the model with interaction.
2. Response model scenarios: different types or extents of interactions in the data.
3. Sample sizes: 200, 2000, and 8000.

By evaluating a series of metrics, including the relative bias, the empirical relative standard error, relative square root of mean square error, coverage rate of 95% confidence intervals, and bias ratio, the performances of those 3 types of estimators are compared using different combinations of the factors.

They found that while all three of them can reduce bias, the outcome model scenario is the dominant factor in choosing the best calibration estimator. If no interaction is in the model, GREG or ranking should be preferred. If additive interaction is present, post-stratification should be preferred, and otherwise, ranking should be preferred.

Furthermore, they also found that simply increasing the sample size does not improve the results, unless more meaningful covariates are incorporated. Because if an estimator is biased, even a confidence interval with high confidence level may fail to cover most of the true distribution.

The model specification also influences the performance of the estimators. As the interaction is strong, post-strafication will have advantage. However, in practice, ranking often performs better than post-stratification.

Therefore, it is important to analyze the metrics about the data carefully, and determine their characteristics, so that we can choose a better way to calibrate the estimation results.

Response mechanism from different sources

While we often assume that the nonresponses in the data occur at random (missing at random, MAR), we should be aware that the sources of nonresponses can be further categorized into two types: missing at random in the population (PMAR), and missing at random in the sample (SMAR), and they might not always hold at the same time (See Berg, Kim, and Skinner 2016).

Berg, et al. argues that while PMAR can be considered as a natural type of MAR, failing to hold SMAR can also lead to bias in the results (2016).

Berg, et al. analytically proved that SMAR can hold only when PMAR holds, with one of the additional conditions: either the sampling mechanism is non-informative, or the response mechanism is unrelated to response value or sample inclusion. Therefore, in order to make SMAR hold, one can choose to include special design information to ensure that the the sampling mechanism is non-informative (2016).

Therefore, fractional imputation (FI) is proposed to design consistent variance estimators for the imputation-based estimators. By comparing the performance of 3 FI methods (ordinary least squares, weighted least squares, augmented model method), the weighted estimator is found to have the best overall performance.

This work highlights the importance of the nonresponse from difference sources, and the importantance of holding both PMAR and SMAR to reduce the bias in the estimation results. It is important to make a good design of the surveys to make sure that both PMAR and SMAR can hold.

Propensity-score weighting

Propensity-score weighting is an approach for handling unit nonresponses, by increasing the sampling weights using their inverse response probabilities. In this way, the weight of each sample can be calibrated to better match their proportion in the population.

While propensity-score adjustment is a popular technique for handling nonresponses, it often relies on additional distributional assumptions like MAR. If MAR does not hold, it will be challenging to apply propensity-score weighting because the inverse response probabilities cannot be correctly acquired. Therefore, Riddles, et al. proposed an approach of maximum likelihood estimation based on the distributional assumptions of the observed part of sample. (2016), which can be easily verified from the data.

The method from Riddles et al. does not heavily rely on the distribution from the population, so, even using an inappropriate distrbutional assumption will not make a big difference to the results.

Conclusion

In conclusion, 3 different kinds of treatments to nonresponses are discussed, and each of them have their own advantages and limitations. The general calibration models are easy to calculate, but they need MAR to work well, and it requires detailed analysis on the data and responses to choose the best estimator; the fractional imputation focuses on satisfying SMAR in the presence of PMAR, but it cannot work if multiple sampling units are contained in

one cluster of data; the propensity-score weighting can work without additional distributional assumptions, but it does not satisfy the calibration constraints.

As none of the methods can replace the others, we should always be aware of their advantages and limitations, and choose the correct method in the corresponding scenarios.

References

Berg, Emily, Jae-Kwang Kim, and Chris Skinner. 2016. "Imputation Under Informative Sampling." *Journal of Survey Statistics and Methodology* 4 (4): 436–62. https://doi.org/10.1093/jssam/smw032.

Han, Daifeng, and Richard Valliant. 2020. "Effects of Outcome and Response Models on Single-Step Calibration Estimators." *Journal of Survey Statistics and Methodology* 9 (3): 574–97. https://doi.org/10.1093/jssam/smz057.

Holt, D., and T. M. F. Smith. 1979. "Post Stratification." *Journal of the Royal Statistical Society. Series A (General)* 142 (1): 33–46. http://www.jstor.org/stable/2344652.

Jean-Claude Deville, Carl-Erik Särndal, and Olivier Sautory. 1993. "Generalized Raking Procedures in Survey Sampling." *Journal of the American Statistical Association* 88 (423): 1013–20. https://doi.org/10.1080/01621459.1993.10476369.

Riddles, Minsun Kim, Jae Kwang Kim, and Jongho Im. 2016. "A Propensity-score-adjustment Method for Nonignorable Nonresponse." *Journal of Survey Statistics and Methodology* 4 (2): 215–45. https://doi.org/10.1093/jssam/smv047.