Nonresponse_rates: theory-driven nonresponse propensity model

Xiaoxu Liu

2024-02-05

Sampling is a fundamental method to achieve accurate and timely estimates of crucial survey outcomes. Using sampling is usually more prior than using the full census. During the survey, researchers undertake significant control during the design stage, attempting to minimize errors. However, there are still uncertainties in data collection, introducing challenges such as unit and item nonresponse. If nonresponse isn't completely random(MCAR), complete case analysis becomes inefficient for losing relevant information. Thus, statistical adjustments become essential, and understanding the survey response mechanism is crucial for effective remediation strategies. In JSSAM publications, the article discusses the challenges and approaches in dealing with nonresponse in surveys, focusing on key aspects related to changes in response rates over time, building theory-driven nonresponse propensity models, and determining appropriate post-data treatment for nonresponse.

Overall, the declining trend in response rates across various modes (especially online surveys) raises concerns about the representativeness of survey samples, considering the impact on different study designs and countries. The article that introduces the concept of theory-driven nonresponse propensity models encourages surveys to enhance nonresponse adjustment, highlighting the importance of identifying predictive auxiliary variables; Then the article talks about the distinction between missing at PMAR and SMAR, and supposes different approaches, such as fractional imputation and a propensity-score-weighting approach under NMAR response mechanism. What's more, the final set of papers delves into post-data treatment for nonresponse, which evaluates general calibration models and alternative methods like post stratification and ranking.

One of the most basic issues is how nonresponse values will affect the research and how we can try to eliminate the errors by adjusting methods of data collection and filtration. To be specific, looking further into the adjustment part of the theory-driven nonresponse propensity models, we find that the chosen low response rates usually refer to systematic differences between respondents and nonrespondents on key estimates (Peytcheva and Groves 2009). To measure the existence and extent of these differences, we apply auxiliary variables that are available for all sample units (and key outcomes). Weighting adjustments and deviations are necessary for the measurement; besides, there is an article that utilizes the nonresponse propensity model at the individual level, which specifies the differences and potential ways to adjust data.

In the article "Level-of-Effort Paradata and Nonresponse Adjustment Models for a National Face-to-Face Survey", authors apply theory-driven nonresponse propensity models Wagner, et al. (2014) show that survey-related variables still affect a lot even if the data collection process and results are highly predictive. Its level-of-effort paradata contain accurate information such as the number and timing of attempts and if resistance occurs in a sampled case. Although this data measurement is useful, the authors conclude that "Whether level-of-effort paradata will predict survey variables" is not very predictive when there are other factors influences. Therefore, these variables affect a lot to the result even if the process of gathering and analyzing data is exactly correct.

But why not consider these auxiliary variables in adjustments comprehensively? Unfortunately, the auxiliary variables used in most surveys frequently do not satisfy the effective adjustment: significant associations with both nonresponse and the survey variables. The most possible solution is an approach to selecting weight variables that identify each factor, such as voting and volunteering, not usually considered. In short, effective nonresponse adjustments require a model that predicts well both nonresponse and the estimated quantity from the survey. As a result, the requirement for a nonresponse adjustment means satisfying both dimensions, which is relatively complex for people to build on a model.A more comprehensive model that includes most relevant key statistics will be helpful with predictions.

Some other researchers hold opposite opinions in contrast with Wagner. Amaya and Harring (2017) and Peytchev, Presser, and Zhang (2018) examine the association between civic engagement and survey participation. It is discovered that volunteers and voters are found to be more likely to participate, emphasizing the potential of these variables in predicting response behavior and improving adjustment methods. Beyond that, their models suggest that "not all routes to integration improve the probability of response." Some routes may even lower the extent of accuracy of predictability. These findings remain true regardless of the type of nonresponse that appears. In other words, too much relevant data variables are harmful to the accuracy of data analysis. What makes their ideas different with Wagner is that they believe the result of lower response rate itself is an implication of data relevance. Hence, not all data should be considered in modelling as we should pick the most important elements or we will be distracted by nuances. After finishing analyzing each perspective, we should recognize that their experiments are different based on different measurements, which may explain why their conclusions have conflicts.

When thinking over different opinions, we notice that the use of auxiliary variables is crucial to assess potential bias and implement post-data collection adjustments. The Imbalance Statistic proposed by S rndal and Lundquist serves as a monitoring tool to reduce sample imbalance through data collection interventions. The article emphasizes the effectiveness of the nonresponse adjustments when respondents are more representative, which implies a combined approach of data collection adaptations and post-survey adjustments will do better rather than simply applying each method. Certainly, we should also decide which method to apply under different conditions.

In conclusion, the article provides a comprehensive overview of the challenges proposed by nonresponse in surveys and the matching methods to address them. We still need ongoing research to adapt to changing surveys, improve prediction models, and enhance the effectiveness of post-data treatment for nonresponse. In building the related-theory nonresponse propensity model, we have different voices in applying factors, but it is certain that auxiliary variables undertake an important role in building model and nonresponse adjustments. The exploring process is hard for us to ensure whether it is right or not, so we are able to attempt to figure out the conflicts in the future based on the proved agreements before.
