How to Address Nonignorable Nonresponse in Data Collection?*

Quang Mai

March 7, 2024

Table of contents

Introduction

The editorial of "Special Virtual Issue on Nonresponse Rates and Nonresponse Adjustments" from the Journal of Survey Statistics and Methodology mainly talks about ways to go about unit and item nonresponse in data collection (Survey Statistics and Methodology 2019). To be specific, unit nonresponse occurs when a member of the data sample doesn't respond to the survey, while item nonresponse happens when a member of the data sample fails to respond to one or more survey items that they are eligible to answer.

In order to negate unit and item non-response, the editorial stresses on the importance of understanding "survey response mechanism" (Survey Statistics and Methodology 2019). This is crucial when current trends point towards decreasing response rates accross various data collection modes due to the prominence of web surveys. To address nonresponse bias, the editorial stresses identifying "auxiliary variables" that mediate both nonresponse and key survey variables (Survey Statistics and Methodology 2019). The editorial thus curates a list of papers that tackle the topic of nonresponse adjustments. From nonresponse propensity models, mutiple imputations (meaning replacing missing values with plausible numbers derived

*Data is at: https://github.com/ponolite/tutorial8.git from distributions of observed variables in the data set) to post-data collection adjustments (e.g. calibration weighting), the editorial highlights the need for non-response adjustments.

One paper in particular, "A Propensity-score-adjustment Method for Nonignorable Nonresponse" is interesting as it tackles 'nonignorable' nonresponse in surveys and how to address them in data collection (Minsun Kim Riddles 2016).

To note, this discussion paper was generated using the open-source statistical programming language R (R Core Team 2022).

Nonignorable Response

In large, the paper outlines how nonresponse is a significant issue in sample surveys, especially when participation rates in survey responses have been declining in general. While current methods to deal with nonresponse are numerous, ranging from weighting adjustments, regression weighting and more, these adjustment methods still assume nonresponse to be ignorable missing data, which is oftentimes untrue (Minsun Kim Riddles 2016). The paper thus proposes a new method to address nonignorable nonresponse data, namely "parametric model assumptions, or propensity score models, about the study variable among respondents only" (Minsun Kim Riddles 2016). In particular, to account for nonignorable nonresponse, the paper privileges using importance sampling on the distribution of the respondents to account for nonignorable nonresponse. In simple terms, this means focusing on analyzing observable responses from the data collection to extrapolate on nonignorable missing data. Overall then, the paper indicates the theoretical framework and presents results from the simulation studies and addresses on how this approach can apply to real-world data.

In short, propensity score model is a way to figure out the best guess for a dependent varaible's parameter. Thus, to account for nonignorable nonresponse in data collection, the paper proposes the theoretical framework that makes statisticians focus only on observable parts of the data which are the respondents and the study variable. By doing this, statisticians can then later extrapolate on parts that they can't account for, like nonignorable nonresponse. The paper especially highlights that, "the approach using the outcome model for respondent is more appealing" because it's more feasible to adjust for the unknowable nonignorable nonresponse (Minsun Kim Riddles 2016).

Overall, the paper remarks that propensity score adjustments can minimize the impact of nonignorable nonresponse. While focusing solely on modelling the characteristics of survey respondents, the results have proven that the model is robust even when "the model for respondents is misspecified" or they haven't responded to the survey (Minsun Kim Riddles 2016). The model works exceptionally well for categorical data. For continuous data, the model works well but might need to be refined based on the scope of the data as some continuous datasets can be difficult to manipulate. Still, the method and model, while efficient, can benefit from adding further calibration constraints or adjustments that make estimates more accurate by aligning them with known or expected values (Minsun Kim Riddles 2016).

References

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

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Survey Statistics, Journal of, and Methodology. 2019. "Special Virtual Issue on Nonresponse Rates and Nonresponse Adjustments." *Journal of Survey Statistics and Methodology*. https://academic.oup.com/jssam/pages/special-virtual-issue-on-nonresponse-rates-and-nonresponse-adjustments.