

Tutorial 8*

INF312: Worlds Become Data - Prof. Rohan Alexander

Luca Carnege

March 6, 2024

Please consider the Special Virtual Issue on Nonresponse Rates and Nonresponse Adjustments of the Journal of Survey Statistics and Methodology. Focus on one aspect of the editorial, and with reference to relevant literature, please discuss it in at least two pages. Use Quarto, and include an appropriate title, author, date, link to a GitHub repo, and citations. Submit a PDF.

In an editorial for the *Journal of Survey Statistics and Methodology*, the board covers an ever-present topic in statistics and data science: the pervasiveness of a lack of responses to surveys (“Special Virtual Issue on Nonresponse Rates and Nonresponse Adjustments,” n.d.). Having enough sample respondents to form a decent-sized sample is crucially important when doing studies of anything, since without a decent proportion of the population, you may not have a representative sample. The article discusses three papers that create propensity models and adjustment weights creation methods that try to explain what determines a person’s willingness to participate in a survey. The practical implication of these models and methods is that they can properly adjust estimates to account for the lack of responses in a survey.

For instance, Amaya and Harring’s (2017) paper examines this problem from the lens of the social integration hypothesis. The hypothesis posits that individuals who are more integrated into society (e.g. through civic engagement, etc.) are more likely to respond to a survey request than individuals who are socially isolated. They examine the components that make up social engagement; these could serve as indicators for the likelihood of a survey response. Looking at some “routes” of becoming more engaged (particularly civic engagement and contact with people) they regress survey response against these measures, while also testing if certain “routes” made respondents more or less likely to respond. Their models suggest that not all types of societal integration make someone more likely to respond to a survey. In fact, some factors, such as higher engagement with neighbours, was predictive of a lower likelihood of response. (Amaya and Harring 2017)

*Code and data are available at: <https://github.com/lcarnege/INF312/tree/main/W8%20-%20Hunt%20Data>

Wagner et. al (2014) explore something related yet completely different: level-of-effort paradata. These data denote how much effort or how much resistance there is to getting a sample for a survey. In the context of studying data from the Health and Retirement Study from the University of Michigan, they model how these paradata can be used to calculate adjustment factors for survey nonresponse. Some elements like an indicator of whether a polling target (a sampled unit) is located in a locked building or gated community or effort in the field are strongly associated with response. For other variables that aren't correlated with survey outcomes, however, including them in models would add unnecessary variability to weights and survey estimates, making results harder to generalize. Despite that, they conclude overall that certain paradata can be very useful in predicting the probability of a response in a survey (Wagner et al. 2014).

A third paper presented is Peytchev et al. (2018). In their paper, they first find that many auxiliary variables used in surveys for nonresponse adjustment are not “good”, statistically, since they have significant associations with both survey variables and nonresponse variables which leading to biased weights and estimates. They propose an approach to selecting weight variables that yields better, more unbiased adjustment weights for nonresponse. As an example of their new method's use, they analyze the 2012 US General Social Survey (GSS) to show that voting eligibility, voter turnout, volunteering and candidate choice met their statistical requirements for nonresponse bias reduction. Furthermore, they also found that doing nonresponse adjustments using their weighting technique yielded much improved estimates of GSS response rates as compared to traditional techniques (Peytchev, Presser, and Zhang 2018). However, as they say, more work needs to be done to inform their implementation and solidify it's applicability.

Overall, these three papers show interesting solutions to the problem of survey nonresponse. Their use in practice shows that researchers are still able to generalize results to population with decent levels of accuracy, despite not having extensive amounts of data. That said, it is important to note that these tools should be used more as a last-resort technique, rather than the standard tool, since they can be prone to error and bias; this perpetuates the problem of credibility in the social sciences. Better, more efficient data collection techniques should still be explored in order to gather more data, which is still the certifiable gold standard in collecting a representative sample of a population.

References

- Amaya, Ashley, and Jeffrey R. Harring. 2017. “Assessing the Effect of Social Integration on Unit Nonresponse in Household Surveys.” *Journal of Survey Statistics and Methodology*, April. <https://doi.org/10.1093/jssam/smx001>.
- Peytchev, Andy, Stanley Presser, and Mengmeng Zhang. 2018. “Improving Traditional Nonresponse Bias Adjustments: Combining Statistical Properties with Social Theory.” *Journal of Survey Statistics and Methodology* 6 (4): 491–515. <https://doi.org/10.1093/jssam/smx035>.
- “Special Virtual Issue on Nonresponse Rates and Nonresponse Adjustments.” n.d. *Oxford Academic*. <https://academic.oup.com/jssam/pages/special-virtual-issue-on-nonresponse-rates-and-nonresponse-adjustments>.
- Wagner, J., R. Valliant, F. Hubbard, and L. Jiang. 2014. “Level-of-Effort Paradata and Nonresponse Adjustment Models for a National Face-to-Face Survey.” *Journal of Survey Statistics and Methodology* 2 (4): 410–32. <https://doi.org/10.1093/jssam/smu012>.