A Cumulative Count, General Approach to Item NonResponse

Michael Wood

Department of Sociology, Hunter College, CUNY; New York, NY

Abstract:

A complementary addition to existing perspectives on survey nonresponse is proposed: consideration of nonresponse across items, i.e. *cumlative item nonresponse.* Instead of considering nonresponse to individual items considered one at a time, a comprehensive view is taken of nonresponse over multiple items. Empirical modeling of cumlative nonresponse can be carried out using available GLM software (Poisson and negative binomial distribution functions). General models of nonresponse can be constructed compared to patterns of nonresponse for particular items. The approach holds out the promise of a better understanding of item nonresponse and item 'sensitivity,' as well as a better integration between the literatures on unit and item nonresponse.

(Due to space limitations the entire empirical analysis has been cut from the present paper. The author will be happy to provide the full version with text of survey items, tables and graphics.)

Keywords: Survey Nonresponse, Item Nonresponse, Poisson.

1. Introduction

Understanding of item nonresponse has been hampered by conceptual underdevelopment and ambiguous empirical results. "Item nonresponse" covers a wide variety of survey nonsubstantive responses, from "don't know" answers on opinion items, to refusal on objective items such as income. Research on item nonresponse has not generally produced theoretically consistent findings; item nonresponse theoretical perspectives for opinion items seem to have limited applicability to objective item nonresponse, and theories about opinion "nonattitudes" make different and opposed predictions than theories about opinion "satisficing" (Converse (1970), Krosnick (1991), Krosnick et al. (2002), see also Bishop (2005)). Furthermore, research on item nonresponse has not been well integrated with research on "unit" nonresponse.

This paper proposes a complementary addition to existing conceptual and empirical perspectives on nonresponse: consideration of the quality of response across items, i.e. *cumlative item nonre-sponse.* Instead of considering nonresponse to individual items considered one at a time, a comprehensive view is taken of nonresponse over multiple items. In this way an investigator can build a model of general nonresponse, and then compare it to patterns of nonresponse for particular items. Such models can be constructed using commonly available Generalized Linear Modeling software (Poisson and negative binomial distribution functions, especially). The approach holds the promise of a better understanding of item nonresponse and item 'sensitivity,' as well as a better integration between the historically separate literatures on *unit* vs. *item* nonresponse.

2. Essentials of a Cumlative Item Nonresponse Approach

Consider the issue of item nonresponse among respondents who have agreed to answer a surveyer's questions. Assume for the moment that 11 percent of respondents do not give a substantive answer to a particular question (i.e. the total of 'don't know' 'not applicable' or similar responses, including refusals). The survey analyst wants to know: did the respondents who gave no substantive answer differ from those who did give an answer to the question, realizing that that systematic differences between responders and nonresponders may vitiate results, and lead to wrong conclusions.

To investigate the possibility that nonresponders differ significantly from responders, the survey analyst typically uses tabular and multivariate techniques to determine if 'demographic' factors such as age and education distinguish between the two groups. Previous research of this kind has shown that nonresponse to a particular item may reflect that item's "sensitivity". It is known for example that questions about respondent sexual behavior, or financial situation are more sensitive –i.e. provoke more nonresponse– than questions about respondent gender, educational status, or political affiliation.

2.1 Limitations of Existing Approaches to Item Nonresponse

But there are several limitations of this approach. First of all, the number of missing cases is typically

small, and the variances high. The consequence is that it is difficult to statistically identify differences between responders and nonresponders. In other words, the survey analyst may find no significant differences, but have a strong suspicion that a differences do exist between responders and nonresponders.

Secondly, it may be that subtle differences among a) the particular survey, and b) the particular item, produce a different nonresponse outcome than expected from previous studies of item sensitivity. In other words, expected general relationships between respondent characteristics and item nonresponse may not be evident due to local variations –i.e. the mode of administration of the survey and and the particular wording of the question and nonresponse options.

Third, along these same lines, the operational definition of nonresponse varies, and studies have tended to show differences among "don't know" and "not applicable" nonresponse, as well as the case in which item wording includes no explicit nonresponse category. However, short of the realization that different nonresponse codings produce different outcomes, there have been few consistent findings about the effect of different nonresponse options. Indeed, along these lines, there are differing theoretical perspectives that predict opposite nonresponse outcomes (nonattitudes vs. satisfycing).

Finally, it can be observed that the study of "item nonresponse", defined as the study of nonresponse to particular items, has been conducted to a large extent independent of the more general case of "unit nonresponse." These two analytical concepts, though different, should be bridged through empirical study. But defining item nonresponse exclusively in terms of particular items tends to preclude such research.

2.2 Cumulating Item Nonresponse Addresses Limitations of Existing Approaches

The basic idea proposed here is to extend the definition of item nonresponse to include several items rather than just one. Cumulating nonresponse *across* items makes possible estimation of a *general* component of item nonresponse –heretofore a nonexistent analytic concept. By cumulating nonresponse across several items more information from each respondent is utilized, and a kind of 'smoothed' distribution of item nonresponse is obtained. In other words: a general pattern of nonresponse can be identified that may well differ from specific nonresponse

to a particular question.

2.3 Item Nonresponse Posited to be the Outcome of Two Fundamental Components: General and Specific

Item nonresponse is posited to be the result of two fundamental processes: a general component associated with a respondent and the quality of his/her overall response to a survey, and a specific component associated with a respondent's specific reaction to a particular item. It is recognized at the outset that differences between the general and the particular will be common, and of interest. Concerning the issue of item sensitivity, this means that different factors may well come into play for a particular item vs the general level of nonresponse. Alternately, the same respondent characteristic may influence overall cooperation differently than response to a particular question. For example, a respondent from a higher socioeconomic status level might be expected to be reluctant to answer questions about finances. However, higher SES respondents have generally shown a higher level of cooperation and willingness to answer a surveyer's questions. In this important example, socioeconomic status is expected to increase nonresponse to a particular item, but to decrease *overall* nonresponse over multiple items. The present approach aims at providing empirical measures of general questionnaire nonresponse that afford comparison of the general to the particular.

Empirically this approach means that nonresponse is cumulated across items in a questionnaire. Such cumulation can be organized by type of question (e.g. objective vs attitudinal), item sensitivity (finances and sexual matters vs general opinions and beliefs), as well as by type of nonresponse option (e.g. don't know, not applicable, or refusal). The data produced by across-item cumulation of nonresponse is typically skewed (80 percent or more zero count) and can be modeled using GLM techniques with Poisson and negative binomial distributions.

3. Conclusions

An alternate approach to the study and analysis of item nonresponse, cumulating nonresponse across questions, shows both conceptual and empirical promise for addressing a number of issues in the study of nonresponse, as well as having value for the individual survey researcher and analyst. The approach, called here **cumlative item nonresponse**, provides a conceptual as well as an empirical bridge between *unit nonresponse* and *item nonresponse.*

- Conceptually, the approach grows out of recognition that unit and item nonresponse are related parts of the same phenomena. Cumlative item nonresponse variables can be built over an entire questionnaire, or over 'sensitive' questions only, or over behavioral report vs. attitude questions, for example. Conceptual bridging between unit and item nonresponse is thus achieved.

- Empirical levels of cumlative item nonresponse can be produced that support conceptual and theoretical questions. In the illustrative analyses above, 3 cumulative item nonresponse variables were constructed using objective financial report questions (length 3), subjective financial questions (length 33), and general attitude questions (length 18). The level of nonresponse to these cumulations varied in accordance with theoretical expectations about item sensitivity, from most sensitive (50 percent nonresponse for the objective financial items), to less sensitive (30 percent, and 19 percent, respectively).

- Furthermore, empirical predictors of cumlative item nonresponse can be identified. The illustrative analyses above showed that education and age had similar *general* effects across both 'sensitive' (money and income) and nonsensitive items: education lowered the odds of nonresponse, and age slightly increased them. Such empirical models can of course be compared across surveys, and compared to results of unit nonresponse research.

- Statistical modeling results from cumlative item nonresponse, when compared to models of *particular* item nonresponse, highlight differences and allow survey analysts to better understand both response *and* nonresponse for a particular survey. The financial example presented here showed that education had a different relationship with a particular survey question about household income than with three financial items overall. The inference from this difference was that slightly more higher education respondents had income item nonresponse than would be expected from the general (i.e. cumlative) case. Taken by itself, modeling of nonresponse for the income question showed only nonsignificant results.

- Since cumulative item nonresponse variables incorporate more information from each respondent, such cumlative variables may shed light on current issues in survey design, i.e. the effects of "don't know" vs. "refused" nonresponse categories for different respondents, and different questions. Cumulations can be constructed for only those who indicated "don't know", and compared to those who refused a question, for example (i.e. 'missing.'). Results for cumlative item nonresponse may provide more consistent results than previous research on individual item nonresponse.

- Finally, the approach has the practical advantages of being simple, easily understood by nonspecialists, and easy to carry out using conventional software (GLM, negative binomial).

References

Bishop, George F. (2005), *The Illusion of Public Opinion: Fact and Artifact in American Public Opinion Polls*, Rowman and Littlefield, Lanham, M.D.

Converse, Philip E. (1970), "Attitudes and Non-Attitudes: Continuation of a Dialogue, " Pp. 168-189 in E.R. Tufte (ed.) *The Quantitative Analysis of Social Problems*, Reading, Mass.: Addison-Wesley.

Krosnick, Jon A. (1991), "Response Strategies for Coping with the Cognitive Demands of Attitude Measure in Surveys," *Applied Cognitive Psychology*, **5**, 213-236.

Krosnick, Jon A., Holbrook, A.L., Berent, M.K., Carson, R.T., Hanemann, W. M., Kopp, R.J., Mitchell, R.C., Presser, S., Ruud, P.A., Smith, V.K., Moody, W.R., Green, M.C., and Conaway, M. (2002), "The Impact of "No Opinion" Response Options on Data Quality: Non-Attitude Reducion or an Invitation to Satisfice?", *Public Opinion Quarterly*, **66**, 371-403.