MAXIMIZING RESPONSE TO HOUSEHOLD INCOME ITEMS: RESPONDENT SELECTION[1]

John D. Hutcheson, Jr. and James E. Prather, Georgia State University

Household income estimates have become increasingly important, especially in applied research, yet it is also becoming more difficult to obtain income data from survey respondents. A number of techniques for handling missing observations in the analysis phase of a survey project have been suggested (Hutcheson and Prather, 1977); but few would argue that it is not preferable to prevent or minimize nonresponse in the field. A number of methods, including card-sorting techniques and probing procedures, have been employed to minimize nonresponse on income items. However, these techniques and procedures often require skilled interviewers and may not be feasible in applied situations in which resources are limited.

In this paper, we contend that the need for complex interviewing techniques may be minimized by carefully matching respondent selection procedures, interviewing techniques, and salient characteristics of the survey population. Using data from a recent survey, we identify characteristics that describe respondents who are most likely to be unable or unwilling to provide household income data and suggest respondent selection techniques that maximize potential response.

To maximize response to income items through selection of the respondent from the pool of potential respondents within a household, two possible reasons for non-response should be considered: 1) the unwillingness of the respondent to provide the information and 2) the inability of the respondent to provide the information. The researcher must seek to minimize both "refusals" and "don't knows." However, previous research exploring the correlates of non-response has often failed to treat these reasons separately. In most instances, "refusals" and "don't knows" are not distinguished and both unwilling and unable respondents are treated as one category of respondents, nonresponders.

In this study, the correlates of nonresponse identified in previous research are used to predict reasons for non-response -- either unwillingness to respond or inability to respond. While some research (Haberman and Elinson, 1967: 194; Spiers, Coder and Ono, 1971) has found that nonrespondents are likely to be male heads of household, Skelton (1963) demonstrated that female "subordinates" within households were also likely to be nonrespondents. This seemingly contradictory evidence, we believe, may be attributable to male heads of household being more likely to refuse and "subordinates" within the household being more likely to be unable to respond.

Respondents classified in higher skilled occupations and working full time have been shown to be likely nonrespondents (Ono and Miller, 1969), while nonresponders are also likely to be found in households with blue-collar and part-time workers. Again, we think that this contradiction is due to the combining of "don't knows" and "refusals" into one nonre-

sponse category, and we hypothesize that full-time, white-collar workers are more likely to be "refusers" and blue-collar and part-time workers are more likely to be unable to provide income data. Race, age and schooling have also been shown to be associated with nonresponse (Spiers, Coder and Ono, 1971) and we suspect that these attributes may be associated in different ways with the two reasons for nonresponse. Additionally, it seems likely that the presence of a number of full-time workers in a household would make it more difficult for the respondent to provide an answer to a family or household income question.

The data employed to test the above propositions were derived from a 1976 survey of a randomly selected sample of over 6,000 households in Atlanta and Fulton County, Georgia. Respondents were selected randomly from the pool of adults within each household and were asked to provide an answer to the following question: "About how much was your total household income, from all sources, for last year -- 1975 -- before taxes?" Full-time paid interviewers were used but most were inexperienced and no card-sorting or specialized probing techniques were employed. It is possible that some respondents masked unwillingness to respond with "don't knows," but the interviewers were required to undergo three days of training during which they were instructed to carefully distinguish between "refusals" and inability to provide income data.

Traditionally, heads of household have been thought to be the most reliable source of objective information about American households. Increasing numbers of single-person households and non-traditional families, as well as resistance to the designation of a household "head," however, have transformed what was formerly a researcher's maxim into what is now an interviewer's nightmare. For this reason, we believe that the designation of a head of household is no longer a very helpful procedure in respondent selection, and because of the many difficulties attempts to designate a head of household caused during the 1976 survey, "status" in the household is not included in the following analysis. The variables that are used in predicting reasons for nonresponse are those characteristics previously mentioned, with the exception of "status" in the household: the respondent's schooling (in years), whether or not the respondent was working full time or not, the number of persons in the household working full time, the respondent's age (in years), whether the respondent held a white-collar or blue-collar job, sex and race (White/Nonwhite).[2]

Multivariate discriminant analysis, using the above characteristics as discriminating (classifying) variables and responses to the income item ("refused," "don't know," and "income given") as categories was employed. This procedure identifies those characteristics which are most closely associated with different responses to the income item and permits the description of respondents who are most likely to be unwilling or unable to respond.

The interpretation of the classification function coefficients in Table 1 is similar to that of regression coefficients. Thus, controlling for the other variables in the analysis, the group of respondents unable to answer the income questions (the "don't know" classification) tended to have the fewest years of schooling, were least likely to work full time, the most likely to reside in households in which there were larger numbers of persons working full time, were likely to be younger, to hold blue-collar positions and to be female and Nonwhite. Respondents who had refused to answer the income question were more likely to have had the highest number of years of schooling, be the oldest, and were the least likely to be females and Non-whites. Respondents answering the income question were most likely to be those who were working full time. Almost without exception these findings support the hypotheses posited earlier.

The standardized discriminant function coefficients in Table 1 indicate the strength of the discriminating characteristics and the direction of their associations with the response categories. The analysis yielded two discriminant functions. Variables often thought to indicate socioeconomic status -- schooling, occupational category, and females not working full time[3]-- grouped logically in Function 1. Function 2 appears to delineate an antecedent factor which might be described as maturity. Age is associated with not working full time and years of schooling.

Table 2 permits an assessment of the adequacy of the classifying or discriminating variables in predicting the responses to the income item. The F-Matrix shows that the "don't know" classification is clearly separable from the other two groups. In addition, even though the distinction between the "refused" group and the responding ("income given") group is not as clear, the difference is statistically significant. When actual responses are compared to responses which are predicted by the classifying variables, 61% of the "don't knows" and 42% of the "refusals" were correctly classified; but only 33% of the respondents who answered the income question were classified correctly.

This analysis clearly demonstrates that while respondent and household characteristics are less than accurate predictors of response or nonresponse, they are more adequate indicators of potential "refusals" and "don't knows." While attitudinal antecedents of refusal may make prediction on the basis of respondent and household characteristics somewhat hazardous, inability to provide income data ("don't knows"), by far the most frequent reason for nonresponse (72% of the nonresponses), is quite predictable.

What is suggested here is that attempting to predict nonresponse on the basis of easily discernible respondent and household characteristics may be futile. What can be done is to predict who is likely to be unable, and to a lesser degree who is likely to be unwilling, to provide income data. Using these predictions, the researcher wishing to maximize response to income items may employ respondent selection to avoid "high risk" respondents.

While recognizing that the limitations of the sample used here may limit generalization across cultural environments and that the purposes of some surveys may limit the researcher's latitude in respondent selection, we offer the following as a guide to respondent selection in surveys where income estimates are to be critical components of the product. First, it appears that respondents' inability to provide household income information is a more severe problem than refusals. In lower socioeconomic populations, especially in households with multiple workers, "don't knows" are most likely to occur. In such cases, male respondents who are working full time are most likely to be able to provide the desired information. Among higher socioeconomic populations, however, this selection strategy is likely to result in high proportions of refusals. The "best" respondent in higher socioeconomic areas may be a working female, but the risk of refusal is still relatively high. Finally, in both higher and lower socioeconomic areas, older persons who are not working full time are "high risk" respondents.

In general population surveys, if the objectives of the survey permit selectivity, researchers may consult existing sources of aggregate data, census data perhaps, to develop respondent selection instructions for different segments of the area to be included in the survey. Instructions should require that interviewers seek younger, working members of households; males in lower socioeconomic areas and females in higher socioeconomic areas. Failure to identify working females in higher socioeconomic areas will increase risk of refusals and may necessitate the collapsing of income categories, the use of card-sorting techniques or other procedures designed to decrease refusals. In lower socioeconomic areas such procedures would have little beneficial effect. Even so, in general population surveys, the researcher may find some comfort in the possibility that refusals in higher socioeconomic areas and "don't knows" in lower socioeconomic areas may have a tendency to offset each other when aggregated income estimates are derived.

We recognize that these suggestions offer no guarantee of response, no assurance of success. On the other hand, we do believe that their implementation would help avoid failure and increase response to household income items.

Notes

1. The authors are indebted to Robert E. Snow for his helpful comments on a draft of this paper.
2. Distributions of the interval variables were found to be both stable and normal.
3. Note that the coefficient signs for female and working full time are opposite in Function 1.

References

Haberman, Paul W. and Jack Elinson (1967) "Family Income Reported in Surveys: Husbands versus Wives." Journal of Marketing Research 4: 191-194.

Hutcheson, John D. and James E. Prather (1977) "Assessing the Effects of Missing Data." Proceedings of the Social Statistics Section of the American Statistical Association, Part I: 279-283.

Ono, Mitsuo and Herman P. Miller (1969) "Income Response in the Current Population Survey." Proceedings of the Social Statistics Section of the American Statistical Association: 277-288.

Skelton, Vincent C. (1963) "Patterns Behind 'Income Refusals'." Journal of Marketing 27: 38-41.

Spiers, Emmett, John Coder and Mitsuo Ono (1971) "Characteristics of Income Nonrespondents in the Current Population Survey." Proceedings of the Social Statistics Section of the American Statistical Association: 369-374.

TABLE 1

DISCRIMINANT ANALYSIS OF RESPONSE TO INCOME ITEM WITH RESPONDENT AND HOUSEHOLD CHARACTERISTICS
(N=6135)

CHARACTERISTICS	CLASSIFICATION FUNCTION COEFFICIENTS			STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS	
	REFUSED	DON'T KNOW	INCOME GIVEN	FUNCTION 1 (SES)	FUNCTION 2 (MATURITY)
RESPONDENT'S SCHOOLING (YEARS)	1.58	1.41	1.51	-.619	-.555
RESPONDENT WORKING FULL TIME	3.45	2.91	3.61	-.472	.507
# OF FULL-TIME WORKERS IN HOUSEHOLD	1.59	1.86	1.62	.348	-.065
RESPONDENT'S AGE (YEARS)	.34	.31	.32	-.297	-.757
RESPONDENT IN BLUE-COLLAR JOB	3.47	3.80	3.38	.262	-.261
RESPONDENT FEMALE	3.82	4.10	3.88	.165	.039
RESPONDENT NON-WHITE	.65	.91	.70	.160	.009
CONSTANT	-20.01	-17.02	-18.70		

TABLE 2

PERFORMANCE OF DISCRIMINANT FUNCTIONS IN PREDICTING RESPONSE TO INCOME ITEM

F-MATRIX

	REFUSED	DON'T KNOW
DON'T KNOW	45.5	
INCOME GIVEN	7.6	72.9

df = 7, 6126

ACTUAL GROUP	PREDICTED GROUP			N
	REFUSED	DON'T KNOW	INCOME GIVEN	
REFUSED	41.7%	29.2%	29.1%	702
DON'T KNOW	19.2%	60.7%	20.1%	1844
INCOME GIVEN	31.9%	35.4%	32.7%	3589

Percent of "Grouped" Cases Correctly Classified: 42.2%