RESPONDENT SELECTION BIAS IN THE HAGEN-COLLIER APPROACH

Karol P. Krótki, Lorraine Porcellini, Temple University
Karol P. Krótki, Institute for Survey Research, 1601 North Broad Street, Room 502, Philadelphia, PA 19122

Key Words: respondent, selection, bias, Hagen-Collier

Introduction

When more than one household member is eligible for a survey it is necessary to devise a method that will assure that every eligible member has an equal chance of being selected. The random selection of a respondent in a household would seem at first to be a very simple operation. In reality, it poses several challenges largely due to the need to ensure that the interviewer does not introduce bias, intentionally or otherwise.

The most elaborate method for achieving this goal is known as the "Kish table" method which generates for each household a random number selection table which in turn indicates which respondent to choose depending on the size of the household, i.e., the number of eligible members (Kish, 1965). There is no argument that this method is guaranteed to ensure random selection. However, the disadvantages are, first, the screening form becomes more complicated; second, the application of the Kish tables requires more work of the interviewer; third, it is intrusive, since a complete household roster is required; and fourth, it lengthens the introduction to the interview which is the point during which most nonresponse takes place.

In light of these drawbacks, researchers have developed alternative strategies that are simpler to administer than the Kish tables but at the same time do not violate the basic requirement of randomness or absence of interviewer selection bias.

One such method is the last, or next, birthday method. Another is based on alphabetic ordering of members' first names. Both these methods are simple and quick to administer and do not require a complete household roster. However, the birthdate method requires the informant to be familiar with all household members' birthdates. The alphabetic ordering requires time to obtain a name for each member, order the list alphabetically, and select the correct sequence number. Both methods depend on the accurate reporting of information on the part of the informant.

Another method, known as the Hagen-Collier method, attempts to overcome these problems with a simple selection rule that consists of four equally probable selection rules: Youngest Male, Oldest Male, Youngest Female, Oldest Female. One of these four is randomly generated for each selected household. The interviewer prepares an abbreviated household roster asking only for first names of males and females separately and in order of age. The selection rule will indicate the respondent (Lavrakas, 1987).

The obvious weakness of the Hagen-Collier method is that "middle" males and females have zero probability of selection. That is, if a household has three or more males or three or more females, the members who are neither the oldest nor the youngest have no probability of selection.

The Research Question

The question to be addressed in this paper is the extent to which the Hagen-Collier method introduces bias. In particular, we have considered two specific questions:

1 - What is the proportion of all households with "middle" members? What is the proportion of all household members who have zero probabilities of selection?

2 - Is there any bias introduced by the excluded members? To what extent do they differ from other members in these households?

Our hypotheses are that, one, the proportion of the population excluded is extremely small and two, the bias introduced is negligible since there is no a priori reason for believing that these individuals differ from other household members.

Results

The results are based on data obtained from the 1990 US Census Bureau Public Use Master File (PUMS). Only adults (persons ages 18 and over) were included in the calculations. Those living in group quarters were excluded.

Hypothesis 1:

The left-hand portion of Table 1 shows the percentage of all households that contain three or more adults of the same sex, 18 years of age or over. It is the middle members of these households who would potentially be excluded using the Hagen-Collier method of respondent selection. "Middle" members are defined as adults in households with more than two adults all of whom are the same sex. Middle members are neither the youngest nor the oldest household member. Therefore, they have no chance of being selected using the Hagen-Collier method. Of the 91.9 million households that contain at least one adult, 3.9% contain potentially excluded middle members.

As can be seen from the right-hand portion of Table 1, the total number of potentially excluded members is 4.3 million which is 2.4% of the total population of 179 million persons 18 and over.

Census data indicate that the Hagen-Collier method excludes, relatively speaking, a very small proportion of the population. The proportion is sufficiently small that, even it were different from the rest, it would not be likely to influence overall results.

Hypothesis 2:

Middle members differ from the general population since they belong to larger households and possess traits that are characteristic of such households. It is illustrative to compare the excluded middle members to the rest of the population. The issue is, given a household with middle members, what is the effect of excluding middle members from the selection process. Over all such households, is there evidence that the excluded middle members differ from the rest of the members?

Table 2 addresses this issue. The first column contains results for all adults in households with no middle members. The second column contains results for non-middle adults in middle households. The third column contains results for the population of middle members.

As can be seen, middle members tend to be more male, more Black, more Hispanic, younger, less likely to be married, with higher education, and with higher employment rates than non-middle members. The most dramatic difference occurs for age. The proportion of middle members who are 65 or over is relatively lower than the corresponding proportions for non-middle adults in middle households as well as adults in non-middle households.

In spite of these differences, we recall that the proportion of the total population eliminated from the sampling process is very small (Table 3). The marginal differences are overwhelmed once the overall sample data are invoked for analysis purposes.

Future Research

The above results are compelling but they raise a few questions. For example, it is not clear what would happen if the target population were redefined as, for example, those aged 12-34, drug users, or women of childbearing age. In several cases, these results could be calculated precisely using census data as was done above. We postulate that the effect would depend on the extent to which the screening criteria are clustered by household. This could lead to a model which would relate the screening variable's homogeneity to the proportion of excluded middle members and the resulting bias.

References

Kish, Leslie. (1965). Survey Sampling. New York, NY:Wiley & Sons, Inc., pp. 398-401.

Lavrakas, Paul. (1987). Telephone Survey Methods. Newbury Park, CA: Sage Publications, pp.93-96.

Table 1

#HHLDs w ADULTS (POP ≥ 18)	91,888	TOTAL POPULATION OF ADULTS	178,599
#HHLDs w ≥ 3 ADULTS	3,600 (3.9%)	#"MIDDLE" ADULTS	4,298 (2.4%)

Table 2

	"Non-Middle" HHLDs	"Middle" HHLDs	
Characteristics	All Adults in "Non-Middle" HHLDs (n=166,315)	"Non-middle" Adults (n=7,998)	"Middle" Adults (n=4,286)
	%	%	%
MALES	47.1	56.2	57.3
BLACKS	10.1	18.4	20.3
HISPANICS	6.8	18.1	21.5
AGES 18-34	34.8	57.0	72.6
AGES 65+	17.2	8.9	2.3
MARRIED	62.7	30.6	17.5
< HIGH SCHOOL GRAD	23.7	35.3	29.6
COLLEGE GRAD	19.4	11.2	12.5
EMPLOYED	63.6	61.0	68.5

Table 3

	Population of All Adults (n=178,599)	Population with "Middle" Adults Excluded (n = 174,313)
	%	%
MALES	47.7	47.5
BLACKS	10.7	10.5
HISPANICS	7.6	7.3
AGES 18-34	36.7	35.9
AGES 65+	16.5	16.8
MARRIED	60.2	61.2
< HIGH SCHOOL GRAD	24.4	24.3
COLLEGE GRAD	18.9	19.0
EMPLOYED	63.6	63.5

NOTE: All totals have been rounded to thousands.