SAMPLING ERROR ESTIMATES FOR PANEL STUDIES: IMPLICATIONS FOR STANDARDIZATION

Harold N. Organic, Brown University

Introduction

The Population Research Laboratory at Brown University has for a number of years conducted a series of annual sample surveys of the population of Rhode Island. The annual samples are related to one another within the framework of an overall panel design that permits continuing surveillance of the population and provides both cross-sectional and longitudinal data on community social structure, health needs and health care delivery.

While panel designs have long been known and often imaginatively used, they have generally been given short shrift by social researchers who have shown a marked preference for the cross-sectional model. The reasons for this are not far to seek when one recognizes that the cross-sectional model is free from certain difficulties inherent in the panel design. Chief among these, for our present purposes, is the attrition phenomenon that results in the loss of sample cases over time as respondents die, refuse continued cooperation at some point after the initial contact, or move their residence leaving no forwarding address for future contact. The bias introduced by this attrition has been difficult to measure, and consequently corrective steps hard to apply.

Nevertheless, the unique advantages of the panel design for providing time series data are manifest, and are assuming increasing importance in the view of many social researchers. This is especially true of, but by no means limited to those who deal with such applied problems as the provision of human services --- health care delivery being a specific case in point. In practical terms elected officials, administrators and interested professionals seek time series data for the planning, execution and evaluation of public programs and services. A research strategy directed toward continuing surveillance of a population and/or system and the recording of data thereon, implies some effective variant of the panel design.

The Sample Design

An experimental design for this purpose has been developed at the Population Research Laboratory and has been in operation since 1967. The purpose of this paper is to suggest a method for dealing with the attrition problem mentioned above. Three successive annual samples in Rhode Island were drawn independently from a frame constructed on the familiar clustered, stratified, area probability, multi-stage model [1], [2]. In each year approximately 1,100 household interviews were conducted, and follow-up interviews (usually by telephone) were taken at annual intervals thereafter. In all, 3345 respondents were enlisted in the three rounds of initial contacts in the Fall of 1967, '68 and '69. By the Fall of 1970 the first panel (Sample I) was three years old, and the others (Samples II and III) were two years and one year old, respectively. As would be expected, the oldest sample suffered the greatest loss (17.66 per cent), the second sample was intermediate (8.88 per cent), and the third sample sustained the smallest loss (6.01 per cent). These dropout rates were computed as a proportion of initial panelists who failed for any reason to give a follow-up interview in the Fall of 1970. The number of cases (N) in the initial samples and the number of dropouts by the Fall of 1970 are shown in the bottom row of columns 1, 3, 4, 6, 7, and 9 of Table 1.

The Problem

The fact of attrition brings to mind a number of questions:

1. how comparable were the initial panels;
2. has there been differential dropout by a given characteristic;
3. if yes, was the differential patterned;
4. was the difference between loyal and dropout components so great as to impair the "representativeness" of the residual loyal sub-sample;
5. if yes, at what point in the life of the panel did the departure from representativeness occur --- in the first, second...nth year;
6. is there available an efficient method to replenish losses, thus extending the representative utility of the panel; and finally
7. if yes, how can we test whether the replenishment cases are unbiased replacements for the losses previously sustained?

Appropriate statistics for estimating the sampling error of cluster samples have been available for some time. The calculations required for such estimates are tedious and time consuming, however, and it remained for the appearance of both large-scale electronic computers and the appropriate software to permit such calculations with reasonable economy of time and effort. Now that machines and programs for this purpose are at hand we should expect that sampling error estimates will appear together with the substantive findings published by researchers who employ the cluster sample design in their work. When this is fully realized we shall have come far toward achieving the standardization in procedure and reporting so important in this field.

The Data and Discussion

In attempting to answer questions 1 to 5 above we employed a recently developed program [3] to estimate the standard error of ratio differences for cluster samples. Ratio (percentage) differences between samples are expressed in standard error equivalents (SEE's) and shown in columns 10, 11, and 12 of Table 1. Thus, in comparing the initial panel Samples I and II for the age group under 35 (columns 1 and 4) we note a difference of 29.64 - 27.18 = 2.46 percent. One standard error for this comparison, as computed by the estimator program, was 1.85 percent, and the comparison is presented in column 10 as 1.33 SEE (2.46 / 1.85 = 1.33). All other between-sample comparisons for the characteristics selected were similarly evaluated and appear in columns 10, 11, and 12. A value of 2.00 SEE or greater is taken to indicate a significant difference between the sub-groups involved.

The first question asked above was --- how comparable are the initial panels? We note that for age, sex, and religion none of the 30 SEE's equals or exceeds 2.00. For these characteristics, then, we can conclude that the three panels are well matched. For the remaining characteristics, however, the matter is not so clear-cut as marital status, education, and total family income exhibit three, three, and five SEE's, respectively, that exceed 2.00. It is significant to note that 10 of the 11 SEE's exceeding 2.00 occur in columns 10 and 11, and that each of these columns involves Sample I.

We must assume that any sample of SEE's will itself be subject to sampling error, and that some SEE's will exceed the value of 2.00 merely by chance. Nevertheless, the concentration of significant SEE's in columns 10 and 11 should lead us to suspect that while Samples II and III appear to be comparable, there may be something unusual about Sample I. A large number of comparisons involving many additional characteristics (variables) would throw light on this point. If further investigation along these lines indicates that one sample is poorly matched to the others, the researcher may well review the sampling and field operations employed for that sample. It would also be possible that poor sample matches on a number of variables could be attributed to the errant variables themselves. Thus, the scheme of categories he chose may be the source of the difference, or a previously unsuspected departure from comparability in the wording or administration of the item. Not to be excluded from consideration is the possibility that the observed differences reflect a real change in the population under study --- income differences, for example, could result from a change in economic activity and salary/wage levels.

Here, as in the discussion that follows, the purpose is not to explain in detail the preliminary empirical findings, but rather to illustrate some of the ways that analysis of the data using this method can throw light on the researcher's problem of dealing with his material. In any case, the application of this scheme will not of itself releave the researcher of the task of close examination of the materials, but it may prove a useful tool.

The second and third questions asked --- has there been differential dropout by a given characteristics and if yes, was the differential patterned? --- are similar to the first, but involve examining within-sample differences. SEE's for loyal-dropout comparisons were calculated and are shown in columns 13, 14, and 15. Examining this section of Table 1 we note that neither sex nor religious preference exhibits differential attrition. The remaining characteristics, however, show evidence of attrition differentials of two sorts.

On the one hand age, and to a lesser extent marital status and total family income, exhibit patterned attrition; the case of educational attainment is less clear. To take up the characteristic of age, it appears that respondents aged 65 and over are strikingly more likely to be dropouts than any other age group. Given the mortality of man this is not a surprising finding, but it is interesting that the within-sample SEE's for this age group are 3.23, 2.66, and 1.68, respectively, for the three samples. In this instance both temporal and directional patterns can be discerned. The temporal pattern is evident from the fact that the SEE's vary monotonically with initial sample year. It will be recalled that at the time of the Fall 1970 follow-up interviews Samples I, II, and III were three, two, and one year(s) old, respectively. The direction of the differences is negative in all three cases, indicating a larger proportion of loyal than dropout respondents in each sample. The opposite is true for the age group 35 to 49. Suggestions of patterning (through SEE's are small) can be detected in the remaining two age groups. Here, as with other characteristics amenable to such manipulation, smaller interval sizes (10-year age groups, for example) could yield more detailed information about the fine-grain behavior of the variable.

The widowed and the aged are, to a considerable extent, overlapping categories. It is not surprising, therefore, that the widowed display a pattern of within-sample SEE's congruent to that seen among persons 65 years of age and over. The presently married, for similar reasons, appear to display attrition patterns similar to the younger age group. Examination of the table in this fashion also reveals total family income to behave in the same general way --- low income respondents having a greater tendency to drop out and high income respondents to remain loyal. As pointed out above, to the extent that these characteristics are related, to that extent the present analysis may be partially obscured. A more refined analysis will be required to evaluate the relative contribution to attrition made by each variable independent of the others.

If the overall study design employed the panel model solely for studying cohorts and their

TABLE 1

SELECTED SAMPLE CHARACTERISTICS: FREQUENCY DISTRIBUTIONS FOR EACH OF THREE SUCCESSIVE ANNUAL PANEL SAMPLES AND THEIR LOYAL-DROPOUT COMPONENTS IN PERCENT; AND BETWEEN AND WITHIN SAMPLE DIFFERENCES IN STANDARD ERRORS.

SELECTED CHARACTERISTIC		FREQUENCY DISTRIBUTIONS IN PERCENT FOR ANNUAL AND COMPONENT SAMPLES									SAMPLE DIFFERENCES IN STANDARD ERRORS					
		SAMPLE I			SAMPLE II			SAMPLE III			BETWEEN			WITHIN		
		INITIAL	LOYAL	DROPOUT	INITIAL	LOYAL	DROPOUT	INITIAL	LOYAL	DROPOUT	I-II (1-4)	I-III (1-7)	II-III (4-7)	I (2-3)	II (5-6)	III (8-9)
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
AGE:	Under 35	29.64	30.70	24.63	27.18	27.51	23.76	28.31	28.15	30.77	1.33	0.69	0.58	1.60	1.04	0.42
	35-49	30.17	31.47	24.12	29.90	31.08	17.82	27.01	28.25	7.69	0.14	1.62	1.27	2.20	3.27	4.95
	50-64	22.63	22.20	24.62	25.42	25.00	29.70	25.07	24.51	33.85	1.69	1.39	0.17	0.73	1.09	1.56
	65, +	17.56	15.63	26.63	17.50	16.41	28.72	19.61	19.09	27.69	0.04	1.39	1.47	3.23	2.66	1.68
SEX:	Male	41.61	41.27	43.22	43.18	43.05	44.55	42.09	41.83	46.15	0.89	0.66	0.27	0.48	0.29	0.63
	Female	58.39	58.73	56.78	56.82	56.95	55.45	57.91	58.17	53.85	0.89	0.66	0.27	0.48	0.29	0.63
M.S.:	Married	68.59	70.91	57.79	72.56	74.32	54.46	69.66	69.98	64.62	2.29	0.57	1.45	2.99	4.47	0.89
	Widowed	13.22	11.42	21.61	13.02	12.16	21.78	13.41	13.09	18.46	1.49	0.15	0.33	3.69	2.25	1.14
	Separated	3.11	2.58	5.53	3.25	2.90	6.93	2.50	2.17	7.69	0.24	0.78	1.21	1.67	1.68	1.60
	Divorced	4.17	3.99	5.02	2.64	2.51	3.96	4.72	4.72	4.62	2.04	0.62	2.38	0.64	0.83	0.04
	Never Married	10.91	11.10	10.05	8.53	8.11	12.87	9.71	10.04	4.61	1.90	0.93	0.89	0.43	1.48	2.17
REL.:	Roman Catholic	64.33	65.30	59.80	66.40	66.51	65.35	63.55	63.68	61.54	1.04	0.51	1.56	1.61	0.23	0.37
	Protestant	29.81	29.53	31.16	28.49	28.57	27.73	29.05	29.13	27.69	0.69	0.48	0.30	0.48	0.19	0.24
	Jewish	2.22	2.26	2.01	1.93	2.03	0.99	3.33	3.05	7.69	0.48	0.19	0.16	0.23	0.97	1.50
	None	2.66	0.54	1.51	2.64	0.48	0.99	3.52	0.49	0.00	0.04[a]	1.30[a]	1.28[a]	0.79[a]	0.50[a]	0.19[a]
	Not Ascertained	0.98	2.37	5.52	0.54	2.41	4.94	0.55	3.65	3.08	....[a]	....[a]	....[a]	....[a]	....[a]	....[a]
ED.:	Under 8 Years	14.11	13.47	17.08	14.87	13.51	28.71	14.62	14.07	23.08	0.52	0.38	0.16	1.42	3.62	1.52
	8-11 Years	41.36	40.30	46.24	36.41	35.91	41.59	34.87	35.24	29.23	2.47	3.24	0.72	1.98	1.16	1.00
	High School Grad.	30.60	31.47	26.63	32.98	34.27	19.80	33.95	34.25	29.23	1.24	1.51	0.51	1.49	3.78	0.84
	Some College	4.53	4.63	4.02	6.95	7.05	5.94	6.11	5.91	9.23	2.78	1.89	0.86	0.45	0.41	0.91
	College Grad., +	8.61	9.38	5.03	8.44	8.88	3.96	9.25	9.35	7.69	0.15	0.58	0.74	2.63[a]	2.80[a]	0.49[a]
	Not Ascertained	0.79	0.75	1.00	0.35	0.38	0.00	1.20	1.18	1.58	....[a]	....[a]	....[a]	....[a]	....[a]	....[a]
INC.:	Under $4500	27.15	24.57	39.20	22.51	20.56	42.58	22.66	21.86	35.38	2.36	2.35	0.08	4.04	3.67	2.47
	$4500 to 7499	29.72	30.50	26.13	27.53	27.22	30.69	25.62	25.59	26.15	1.07	2.13	1.34	1.31	0.74	0.11
	$7500 to 12,499	27.78	29.20	21.10	30.96	32.72	12.87	30.25	30.71	23.08	1.53	1.24	0.38	2.61	4.98	1.37
	$12,500, +	13.40	14.33	9.05	18.65	19.31	11.88	21.10	21.65	12.31	3.59[a]	4.91[a]	1.50[a]	2.35[a]	2.22[a]	2.41[a]
	Not Ascertained	1.95	1.40	4.52	0.35	0.19	1.98	0.37	0.19	3.08	....	....	....	....	....	....
	TOTALS	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00						
	(N)	(1127)	(928)	(199)	(1137)	(1036)	(101)	(1081)	(1016)	(65)						

[a] Standard errors not calculated for 'not ascertained' category.

experience over time, our last four questions would not be of crucial relevance. However, where an important purpose of the study is to monitor a representative sample of a defined population, the question of representativeness and how to cope with it can assume great significance. Using the case of age to illustrate the point, it is clear that as the total sample size grows through the addition of new annual samples there will be an upward shift in the age distribution. The cumulated sample of loyalists will grow older, and the annual addition of new respondents will not be sufficient to redress the balance. The problem becomes even more difficult as attrition introduces other biases and contributes to further departures of loyal respondents from representativeness of the universe under study.

Within-sample differences between initial and loyal respondents have not been calculated in this analysis, although this could be done without difficulty. Unhappily, however, there are no clear-cut guidelines for determining how great a difference can be tolerated, so that such within-sample comparisons would not presently be of much help. The existence of a significant difference between loyal and dropout sub-groups indicates that there is some bias due to differential dropout. Whether or not this bias has an effect on the follow-up results depends upon the proportion of dropouts. In order for the follow-up results to be biased substantially there must be both a high dropout rate and a substantial difference between loyalists and dropouts. In most cases where there is a significant difference between these two sub-groups the frequency distributions for the loyalists are still very close to those of the initial sample.

An independent test of the representativeness of the 1970 follow-up is currently being made by comparing the distributions of characteristics with those in the distributed 1970 Census tapes. This will provide a test for the combined effects of sampling error, original non-response, and attrition. By comparing the 1970 follow-up data for the three samples with the 1970 Census we can answer the question of when the dropout rate begins to affect substantially the representativeness of the loyal respondents.

If substantial departures from representativeness are indicated there are two general strategies that might be employed to deal with the matter. One of these strategies would be to drop the affected panel(s) --- probably the oldest. The decision to drop panels would be guided by age of the panels and/or some selected level of dropout.

The second strategy (that might be employed in conjunction with the first) would be to replenish the eroded samples with new respondents. It would be important to insure the suitability of the new respondents for this purpose and to test whether new respondents were, indeed, well matched with the dropouts they will represent. At the present time (August 1971) an experimental study of this replenishment strategy is under way.

A sample of replenishment respondents for Sample III dropouts is being drawn, and they will be interviewed in the current Fall follow-up interview round. Sampling procedures employed in the original Sample III selection are to be repeated, with some modifications discussed below. Criteria for a replenishment interview are as follows:

a. residence at an address that appeared on the original sample selection list, if the current occupants were not living in the dwelling unit at the time of initial interviewing. An exception is made in the case of a household still remaining if that address did not yield an initial interview due to refusal. The purpose of this rule is to allow original refusals, and the occupants of previously vacant addresses, and <u>new</u> occupants of a vacated interview address to fall into the sample;

b. residence at a selected dwelling unit that was constructed since the original sample was drawn. Such new construction will be placed at the bottom of the original list and the original selection interval applied; and

c. residence in an initial sample household (family) still remaining at the original address, <u>if</u> the initial respondent has for any reason left the household permanently.

In all cases the original criteria for respondent eligibility and selection will be applied. The procedures outlined here will be applied to a randomly selected half of the original sample segments for purposes of economy. An examination of the characteristics of initial respondents from the selected and unselected segments has revealed no differences that could be detected, and it is believed that this 'split-halving' will not introduce bias into the replenishment sample. After weighting to account for the half-sample employed the replenishment sample will be compared with the dropout component (the 65 dropouts identified at the end of the Fall 1970 round plus those newly identified at the end of the current round), and the results reported.

<u>Summary</u>

The increasing need for the surveillance of sample populations in studies of social change and for the planning, execution and evaluation of public programs is presented. Whether for theoretical or for applied purposes, however, this need implies some form of panel design, and the problem of panel attrition and its consequences is discussed. The current panel study at the Population Research Laboratory at Brown University is described, and a method for measuring and coping with attrition is suggested. Preliminary findings and plans for further analysis are presented in the hope that a standardized method for measuring, dealing with, and reporting attrition can be adopted by social researchers who employ the panel survey design in their work.

Acknowledgments

The study reported here is supported by United States Public Health Service Grant HS-00246. The author wishes to thank his colleagues Professors Speare and Roa for their helpful comments and suggestions on an earlier version of this paper presented at the meeting of the American Statistical Association at Fort Collins, Colorado, August 1971.

References

[1] Organic, Harold N. and Goldstein, Sidney, "The Brown University, Rhode Island Population Research Laboratory: Its Purposes and Initial Progress," in The Community as an Epidemiological Laboratory: A Casebook of Community Studies, Irving I. Kessler and Morton L. Levin, eds., (Baltimore: The Johns Hopkins Press, 1970), pp. 212-232.

[2] Kish, Leslie, Survey Sampling, (New York: John Wiley & Sons, 1965).

[3] User's Manual for PSALMS Computer Program, mimeographed, no author indicated (Ann Arbor: Survey Research Center, The University of Michigan, 1971).