SAMPLING STUDIES IN THE SURVEY OF HOUSEHOLDS IN SELECTED BARRIOS IN THE PHILIPPINES*

by

Carmelita Cruz

Introduction

In 1958-59, the Statistical Center, University of the Philippines conducted special surveys in selected barrios in the country. The main objectives of these surveys are:

(1) To study the social, economic and demographic characteristics of the people in the selected barrios;

(2) To evolve appropriate schedules and to develop suitable concepts and procedures for investigating the above characteristics; and

(3) To study the applicability of different sampling techniques and to determine methods for increasing the efficiency of sample surveys.

Information on social, economic and demographic characteristics of all the households in these barrios was collected through trained investigators using well-designed schedules adopting standard definitions and procedures.

The survey data of two of these barrios, Pinambaran in Bulacan and Bungcalot in Batangas, provided the material for this study

The complete data for the two barrios, Pinambaran and Bungcalot, obtained in the survey were analyzed with respect to the important economic characteristics of the households and some of the demographic characteristics of the household members. The results are presented in this paper.

---

* Abstracted from the thesis submitted by the author to the Graduate School, U.P. in partial fulfillment for the degree of Master of Arts in Statistics, 1960.

SAMPLING STUDIES IN THE SURVEY OF HOUSEHOLDS IN SELECTED BARRIOS IN THE PHILIPPINES

Sampling:

For the study of sampling techniques and procedures for estimating some of these parameters different plans of sampling of households in each barrio were adopted. The samples were actually drawn from the list of households of these selected barrios and the different samples were drawn independently of one another. The several sampling plans studied are described below.

Plan A — Simple Random Sampling. In Pinambaran, three sampling fractions were used and sampling without replacement was carried out using a table of random numbers. These are indicated below:

A . 1 — 10% sample of households

A . 3 — 25% sample of households

A . 4 — 50% sample of households

Plan B — Systematic Sampling with Random Start. From the list of households arranged by the name of the heads of the households in a random manner, selection was made systematically with a random start. The different sampling fractions adopted are indicated below:

Pinambaran and Bungcalot

B . 1 — 10% sample of households

Pinambaran only

B . 2 Sample of 20 households

B . 3 — 25% sample households

B . 4 — 50% of sample households

Plan C — Stratified Random Sampling. Scheme I. Stratification was made according to the industry of the household. Three industry groups were formed: (1) Agriculture, (2) Manufacturing, (3) Others.

The size of the stratified sample was kept the same as in Plan A and the sampling within the strata was done as follows:

C . 1 — All the households in the barrio were arranged stratumwise; first the households were arranged in stratum one then those in stratum two and then the rest for the third stratum. Within each stratum the listing of households was in a random order. Starting with a random start between 1 and 10 in the first stratum and using a sampling interval of ten, sampling was done systematically for the whole list. This is the method followed in stratified sampling of households and is more convenient than separate sampling within each stratum. This gave a 10% sample from each stratum.

C . 1 — In Pinambaran, for comparison with unrestricted sampling with a size 20, a stratified sample of this size was also drawn with proportional allocation from the three strata. The selection was made systematically as under C . 1

Plan D — Stratified Random Sampling. Scheme II. In this scheme of stratification, the households in Pinambaran were also stratified with respect to the occupation of the head as well as of other members of a households. Six strata were under this scheme. Three different sampling fractions were used under this scheme. Three different sampling fractions were used with proportional allocation in each stratum. These are indicate below.

D . 1 — Sampling fraction 10% in each stratum

D . 2 — Sampling fraction 25% in each stratum

D . 4 — Sampling fraction 50% in each stratum

It is to be noted that in this study, stratified sampling with Neyman allocation was not separately examined because the values obtained using this method were practically the same as those under proportional allocation.

Since complete data for all the households in each of the two barrios were collected, the precision of estimates of some of the important characteristics using different methods of sampling of households could be examined in relation to the population values. Estimates were obtained for the following characteristics using each of these methods of sampling and compared with population values: average size of household, income and expenditure (total and the components) number of workers, relation between income and expenditure, and between household size and expenditure. The standard error of estimates were computed and studied so that the efficiency of different methods of sampling households could be compared and some general indication of relative advantages could be derived. Unbiased estimates of the standard error were also computed from the same data as is done in actual sample surveys, but since they are subject to sampling errors they will not quite be useful for comparing different sampling schemes. Coefficients of variation of the means were also computed.

Some tables made on estimates of the means and the standard errors of the means and the sample estimates of standard errors of the means and coefficients of variation are given below.

Table 1

PINAMBARAN — HOUSEHOLD EXPENDITURE

Number of Households — 471

B — Systematic Sampling   C — Stratified Sampling

Particulars of Sample			Sample Mean (₱)	Standard Error of $\bar{Y}$ ₱		Coefficient of Variation of $\bar{Y}$ %	
Plan	Fraction	Size (n)	$\bar{Y}$	$\sigma(\bar{Y})$	$s(\bar{Y})$	$C(\bar{Y})$	$c(\bar{Y})$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total Household Expenditure: Population Mean $\bar{Y}$ = ₱95.29							
B. 1	10%	47	90.79	6.37	5.51	6.7	6.1
C. 1	10%	47	101.34	6.32	7.25	6.6	7.2
B. 2		20	91.94	10.07	11.95	10.6	13.0
C. 2		20	91.43	9.99	5.99	10.5	6.6

TABLE 1 (Cont.')

Food: Population Mean $\bar{Y}$ = ₱62.08

B. 1	10%	47	61.63	3.75	3.63	6.0	5.9
C. 1	10%	47	62.29	3.74	3.91	6.0	6.3
B. 2		20	56.84	5.93	6.64	9.6	11.7
C. 2		20	61.20	5.92	4.73	9.5	7.7

Housing: Population Mean $\bar{Y}$ = ₱5.37

B. 1	10%	47	3.98	0.71	0.44	13.2	11.0
C. 1	10%	47	5.54	0.70	0.73	13.1	13.3
B. 2		20	4.75	1.12	1.02	20.8	21.4
C. 2		20	5.70	1.11	0.72	20.7	12.6

TABLE 2

PINAMBARAN — HOUSEHOLD SIZE

Number of Households — 471

A — Simple Random Sampling
B — Systematic Sampling
C — Stratified Sampling I
D — Stratified Sampling II

Average Household Size in Barrio $\bar{Y}$ = 5.67

Particulars of Sample			Sample Mean (₱) $\bar{Y}$	Standard Error of Y (₱)		Coefficient of Variation of $\bar{Y}$ (%)	
Plan	Fraction	Size (n)		$\sigma(\bar{y})$	$s(\bar{Y})$	$C(\bar{Y})$	$c(\bar{y})$
A. 1	10%	47	6.19		0.39		6.3
B. 1	10%	47	5.72	0.35	0.37	6.1	6.5
C. 1	10%	47	6.04	0.34	0.37	6.1	6.2
D. 1	10%	47	5.92		0.39		6.6
B. 2		20	5.10	0.55	0.62	9.6	12.1
C. 2		20	4.65	0.54	0.48	9.6	10.4
A. 3	25%	118	5.81		0.19		3.2
B. 3	25%	118	5.70		0.20		3.6
D. 3	25%	118	5.58		0.18		3.2
A. 4	50%	236	5.57		0.12		2.1
B. 4	50%	236	5.67		0.12		2.1
D. 4	50%	236	5.57		0.11		2.0

A few broad conclusions regarding sampling of households were drawn from this study. We notice that the standard errors of estimates decrease generally with sample size in each of the three types of sampling considered: simple random sampling, systematic sampling with a random start and stratified random sampling. This is as expected from theory but we can use the magnitude of the actual standard error to decide beforehand how large should a sample of households be in order to obtain errors (or coefficient of variation) of the more important characteristics of a specified order of magnitude.

The study also showed that systematic sampling with a random start is equivalent to simple random sampling without replacement since in household listing, the order of listing the households is found to be quite random. The expression for standard errors for simple random sampling could be adopted instead of the more complicated expressions for systematic sampling. Thus, in practice, on any large survey, at the household stage, sampling could be done by systematic sampling with a random start; this could also be done within each stratum wherever stratification is adopted.

In the analysis made on some of the characteristics, it is clear that the type of stratification of households that was made did not increase the precision of the estimate appreciably. Better types of stratification could be used. Broadly, a discussion of the basis of information in the preliminary household list, between agricultural and non-agricultural households may be desirable as there would be differences in most economic characteristics. This is noticed from the analysis made in this thesis. Proportional allocation of sampling units to strata was found to be equivalent to Neyman Allocation and, therefore, the former can be used for sampling of households when the numbers of sample units within strata are small.

Ratio and Regression Estimates

For each type of sampling described in the preceding section, ratio and regression estimates were obtained and compared with direct estimates with regard to precision.

Table 3

Pinambaran — Number of Males in Barrio Y
Total Number of Households — 471

B. 1 — Systematic Sampling with 10% Sampling Fraction

B. 2 — Systematic Sampling with Sample Size 20

C. 1 — Stratified Sampling I with 10% Sampling Fraction

Ratio Estimates from Total Number of Persons (X)

Population Y = 1,313

Sampling Method	Estimation Method	Estimate of Number of Males (Y)	Standard Error of $\hat{Y}$		Coefficient of Variation of Y (%)	
			$\sigma(\hat{y})$	$s(\hat{y})$	$c(\hat{y})$	$c(\hat{y})$
B. 1	Direct	1,323	99.25	92.78	7.6	7.0
	Ratio	1,310	63.17	66.20	4.8	5.1
C. 1	Direct	1,394	99.04	98.80	7.5	7.1
	Separate ratio	1,288	60.34	59.65	4.6	4.6
	Combined ratio	1,302	60.58	59.17	4.6	4.2
B. 1	Direct	1,201	156.85	165.39	13.1	13.8
	Ratio	1,335	99.87	100.21	7.6	7.5

Pinambaran — Average Household Food Expenditure ($\overline{Y}$)
Total Number of Households — 471

B. 1 — Systematic Sampling with 10% Sampling Fraction

B. 2. — Systematic Sampling with Sample Size 20

C. 1 — Stratified Sampling I with 10% Sampling Fraction

Ration and Regression Estimates from Total Household Expenditure (X)

Population Mean $\overline{Y}$ = ₱62.08

Sampling Method	Estimation Method	Estimated Mean $\overline{Y}$ (₱)	Standard Error of $\overline{Y}$ (₱)		Coef. of Var. of $\overline{Y}$ (%)	
			σ ($\bar{y}$)	s ($\bar{y}$)	C ($\bar{y}$)	c ($\bar{y}$)
B. 1	Direct	61.63	3.75	3.63	6.0	5.9
	Ratio	64.68	1.93	1.64	3.1	2.5
	Regression	64.30	1.75	1.57	2.8	2.4
C. 1	Direct	62.29	3.74	3.91	6.0	6.3
	Separate ratio	58.68	1.16	2.16	1.9	3.7
B. 1	Combined ratio	58.57	1.94	2.22	3.1	3.8
	Direct	56.84	5.93	6.64	9.6	11.7
	Ratio	58.91	3.06	3.45	4.9	5.9

These two tables provide comparison of ratio and regression estimates with direct estimates. Clearly, the first two methods are more precise than the last. Stratification appears to increase the precision of ratio estimates while the effect on direct estimates is negligible. Finally, we can say that from the analysis made on this study regarding methods of estimation of means, it can be said that wherever it is possible to find a variate closely correlated to the variate to be estimated, ratio and regression methods should be used as they give much higher precision in the estimates.

Range Methods of Measuring Precision of Sampling Techniques

In this study, the estimates of the precision of the sample means and totals using range were examined for the two important sampling techniques adopted for selection of households and compared with the root-mean-square estimates. The variances of three range estimates were obtained to compare their efficiency with the root-mean-square estimates.

In sampling without replacement from a population of size N, the variance of the mean $\bar{y}$ is given by

$$\frac{N - n'}{N} \cdot \frac{\sigma^2}{n'} \qquad (1)$$

Since $\frac{\bar{w}}{d_n}$ can be approximately used to estimate σ in the finite population, the estimate of the standard error of the mean will be taken as

$$\frac{\frac{N - n'}{N} \left(\frac{\bar{w}}{d_n}\right)^2}{n'} \qquad (2)$$

where n' is the sample size and n, the sub-sample size. The above method is applicable only on the assumption that the population is effectively infinite and nearly normal and, therefore, the observations are independent. The numerical study made in this section seems to suggest that the assumptions are fairly well satisfied.

The variance of this estimate is

$$\Sigma \ (di^2/Vi)n' \qquad (3)$$

It is clear that E ($\bar{w}$/dn) cannot be equal to σ exactly since the assumption mentioned above on which the value of d_n is based does not hold here. Therefore, the expression in (2) cannot be an unbiased estimate and the standard error of the mean; but it is likely that the bias will be small.

In the case of stratified sampling without replacement in each stratum with proportional allocation, the finite multiplier may also be used for getting the estimate of the variances of $\bar{y}$. Thus, the standard error of $\bar{y}$ is given by

$$s(\bar{y}) = \sqrt{\frac{1}{N} \Sigma N_h - n_h \frac{(\bar{w}_h / d\, n_h)^2}{n_h}} \qquad (4)$$

Suppose, it is possible to divide the sample within each stratum into sub-samples of the same size n. If the mean of the ranges of the m_h sub-samples in the h th stratum is denoted by $\bar{w}_h$, then the following expression of the standard error of $\bar{y}$ may be suggested as an alternative to that in (4).

$$S(\bar{y}) = \frac{\bar{w}}{dn \sqrt{n'}} \qquad (5)$$

where

$$\bar{w} = \frac{\Sigma m_h \bar{w}_h}{m}$$

$$m = \Sigma m_h \text{ and } n' = \Sigma n_h$$

The finite multiplier may also be used in this case. Thus

$$s(\bar{y}) = \sqrt{\frac{N - n'}{N}} \; \frac{\bar{w}}{d_n \sqrt{n'}} \qquad (6)$$

In (5) the $\bar{w}$ is simply the mean of the ranges of all the sub-samples in the various strata. If there were no stratificaton and n' observations were drawn and randomly divided into m groups of n each, the mean of the ranges in all the samples is likely to have an expected value larger than the expected value of $\bar{w}$ in (4). For, since in stratification, units which are more homogenous with respect to the characteristics studied are grouped together, the mean range of the m_h sub-samples of the same size in an unstratified sample. It is, therefore, likely that $\bar{w}/dn \; n'$ should be a good estimate of the standard deviation of a mean in a stratified sample. The simplicity and ease of estimation are its special attractions.

The variance of the estimate is obtained easily, thus:

$$V\left(\frac{\bar{w}}{dn\sqrt{n'}}\right) = \frac{1}{d_n^2 \, n'} \left[\frac{\Sigma \; m_h^2 \, V(\bar{w})}{m^2}\right] \qquad (7)$$

$$= \frac{V_n}{d_n^2 \, n' m^2} \Sigma \; m_h \, \sigma_h^2$$

The different range estimates of the standard errors of $\bar{y}$ are shown in the following tables along with the corresponding root-mean-square estimates. In Table 6, the range method, using equation (2,) was applied to systematic sampling B. 1 which is equivalent to simple random sampling without replacement. In Table 7 the two range methods, using equations 4 and 6, were applied to the stratified sample C. 1 drawn without replacement. In this table we let Range — A denote the estimates obtained by using equation (6). R. M. S. denote the root-mean-square estimates.

Table 6

Bungcalot — Household Expenditure Items
Number of households — 264; sample size: n' = 27; size of groups: n = 9; Number of groups: m = 3

B. 1 — Systematic sampling with 10% sampling fraction

Estimator	Estimate of	Variance of estimate	
		$V(s\bar{y})$	$v(s\bar{y})$
Total Expenditure	= 16.64		
Range	14.69	4.8244	5.9263
R. M.S.	15.13	3.7620	4.9030
Food Expenditure	= 6.26		
Range	5.89	1.0735	0.9533
R. M.S.	5.45	0.8371	0.6357
Fuel and Lighting	= 0.53		
Range	0.37	0.0076	0.0038
R. M. S.	0.34	0.0060	0.0025

Table 7

Pinambaran — Household Expenditure Items
Number of households — 471

C. 1 — Statistical sampling with 10% sampling fraction

Estimator		Estimate of ($\bar{y}$)
Total Expenditure:	= 6.32	
Range — A		7.48
Range — B		7.46
R. M. S.		7.25
Food Expenditure:	= 3.74	
Range — A		4.23
Range — B		4.19
R. M. S.		3.91
Fuel and Lighting Expenditure	= 0.24	
Range — A		0.16
Range — B		0.15
R. M. S.		0.18

The range methods of estimation of standard errors of estimates provide very quick and simple methods when compared with the usual methods. Although some assumptions are involved in using existing tables of means and variances of range, the results obtained can be considered to be fairly valid. Comparison of range estimates with those of root-mean-square estimates shows that their efficiency is only slightly less. The methods suggested in the case of stratified samples are expected to be more useful. Extension of range methods to estimating stage variances in multi-stage sample appears to be possible and this could then facilitate better designing of sample surveys.