The Philippine Statistician

Vol. xxi, Nos. 3 and 4, 1972

A PILOT STUDY IN SAMPLING AND RESPONSE BIAS

Statistical Center Faculty*
University of the Philippines

I. INTRODUCTION

In the Philippines, specifically in the Greater Manila Area, studies based on sample surveys have increased tremendously in the recent years. Various academic, business and government groups have come to realize the indispensibility of the survey approach in conducting their researches in marketing studies, opinion polls and socio-politico studies. In studies like these sampling techniques are indispensible. Researchers have used different sample designs for similar studies or similar designs for completely different studies. As to the question of which design is the more efficient one, there seems to be some difficulty in venturing a categorical answer. Very few comprehensive studies, if there are any, have been made in this regard in the country. This paper, therefore, attempts to answer the question posed above, at least, partially. This study, however, was carried out only on a pilot basis. Further, ideas and findings emanating from this pilot study will, hopefully, be explored and extended in future larger-scale studies.

1.1. *Objectives of the Study*

This study aims to:

1) Compare the precision of several sample designs taking into consideration the rural-urban classification of the selected pilot places;

---

* This was a Center study where all the members of the faculty were involved. Members of the research staff and senior students were utilized in the field work. The processing, tabulations and computations were supervised by Mr. Ceferino Sinsioco, Assistant Professor of the Center, who also prepared the report with the cooperation of Dr. Cristina P. Parel, Professor and Dean of the Statistical Center and Miss Angeles Buenaventura, Assistant Professor.

2) Compare costs of several sampling schemes;

3) Make use of the data gathered for the purpose of objective (1) in extending the study to other areas of sampling;

4) Pinpoint certain areas of the study which might need further exploration and/or verification;

5) Generate data which will be of use to graduate students in the Statistical Center who are working on their theses or term papers; and

6) Examine biases caused by difference in time of interview and interviewers, etc.

The idea of carrying out this study arose from a desire to ascertain which sampling scheme would be most efficient at a given cost in rural and urban areas in the country. The study was formally started in July, 1971. By November, 1971, all interviews and computations were completed.

1.2 *Limitations*

As pointed out in the previous section, the project was contemplated to a pilot study. In this regard, San Miguel Area in Manila and the barrio of Lagundi, Plaridel Bulacan (see Appendices 1 and 2) were selected as pilot areas to represent the urband and rural sections, respectively.

Due to the financial constraint, considering that the funding of the project must come from the limited Statistical Center's research allocation, the time alloted to the operational aspects of the project was limited only to a period of five months.

In spite of the above limitations, however, the study is hoped to be of some use since, as a pilot study, the findings can serve as bases for further exploration and verification.

II. EFFICIENCY OF SURVEY DESIGNS

2.1. *The Urban Sector*

2.1.1. *Method of study*

To compare the precision of several sample designs, the following schemes were decided upon:

Type of Sampling	Frame Used	Sampling Fractions
A. Cluster Sampling		
1. By block	Map of San Miguel (Appendix 1)	1/4, 1/8
2. By precinct	Electoral precincts from COMELEC	1/4, 1/8
B. Systematic Sampling	The listing of households was done by graduate students of the Statistical Center	1/8, 1/16
C. Stratified Sampling		
By geographical location	Map of San Miguel	1/4, 1/8

After collecting the necessary data, using the various sampling methods and through the interview approach, certain variances and cost functions were then generated. Finally, the precisions of the different survey designs under consideration were compared.

2.1.2. *Variance and Cost Functions*

2.1.2.1. *Cost Functions* - Two cost functions are derived: one based on the block as the large unit, and the other, based on the electoral precinct as the large-unit. The table on the next page lists down the various information needed in the computation of certain constants in such functions.

Table 1. COMPUTATION OF INTERVIEWING COSTS, BY BLOCK AND BY PRECINCT

Interviewer	Imputed Daily Wage Rate	Block as the Cluster					Precinct as the Cluster				
		No. of days of Interview	No. of Schedules	Wages	Transp. Cost	Total Cost	No. of days of Interview	No. of Schedules	Wages	Transp. Cost	Total Cost
A	₱12.00	leave	—	—	—	—	18	73	₱216.00	₱18.60	₱234.60
B	16.00	19	84	₱304.00	₱21.00	₱325.00	25	77	400.00	21.60	421.60
C	17.50	19	99	332.50	22.80	355.30	15	56	262.50	24.00	286.50
D	13.00	19	81	247.00	22.80	269.80	15	44	195.00	24.00	219.00
TOTALS			264			₱950.10		250			₱1,161.70

Based on the Cost Function of the form

$$C = Co + Ce\sqrt{nm} + Cu\sqrt{n}$$

Where C = total costs incurred
Co = fixed costs
Ce = cost per element
Cu = cost per unit (cluster)
n = number of units (clusters)
m = number of households

The computed results are tabulated in Table 2.

Table 2. COMPARISON OF COST PER ELEMENT (Ce) AND COST PER UNIT (Cu)

TYPE OF UNIT	Ce = Total Cost / Total No. of Schedules	Cu = (based only on the transp. costs in transferring from one cluster to another.)
Block	₱3.60	₱0.60) — roughly
Precinct	₱4.65	₱0.60)

The figures in the table indicate that using the precinct as the sampling unit would be more costly than that of the block. This is as expected since it takes more time locating the various elements within a precinct rather than within a block. It must also be pointed out here that cluster sampling by block was initiated ahead of that of sampling by precinct. The interviewers, therefore, were already exposed and more experienced, by the time they conducted the interviews by precinct. The difference in cost presented in the table is a conservative one.

2.1.2.2. *A Variance Function* - Mahalanobis (1944) and Hendricks (1944) developed general laws to predict how s^2_w changes with the size of the unit (Mi). Cochran (1963) presented a prediction model, useful in several agricultural surveys, which is of the form:

$$\text{Log } s^2_w = \log A + B \log M,$$

where s^2_w = within variance

M = size of the unit

A & B = unknown constants

The same model was tried in the current experimental study for the variable *income* and attribute attitude towards the *legalization of gambling*. In both cases, there is a significant relationship between log s^2_w and log M. In deriving the models, the large unit sizes considered were M = 3, 5, 10, 15, 30, and 50. The resuits are as follows:

Income : $\log s^2_w = 5.6505 + 0.0567 \log M$

$F = 25.07$ with 1, 4 d.f.

$\log s^2_w = - 0.4834 - 0.1395 \log M$

Legalization of Gambling: $F_w = 4.13$ with 1, 4 d.f.

With the use of the above prediction model, any size of the large-unit M can be studied.

Table 3 shows how the predicted values of s^2_w compared with the actual values.

Table 3. PREDICTED S^2_w VS. ACTUAL S^2_w

SIZE OF THE LARGE-UNIT M	Income Actual S^2_w (₱²)	Income Predicted S^2_w (₱²)	Legalization of Gambling Predicted S^2_w (₱²)	Legalization of Gambling Actual S^2_w (₱²)
5	497,066	489,930	0.2420	0.2461
15	523,034	521,420	0.2403	0.2209
30	540,345	542,310	0.2318	0.2063
50	543,808	548,260	0.1660	0.1962

Only one model was used in fitting. It is possible, however, that other models might yield better fits than the model which has been considered. Further studies might ascertain this.

2.1.3. *Comparison of Precisions of Several Sample Schemes* — Three were considered in this part of the study, namely: cluster sampling, systematic sampling, both the block and the precinct were considered. For stratified sampling, Neyman and equal allocations were examined. Geographical location and age - education stratification were considered. Sompling fractions of 1/4 and 1/8 were looked into.

Table 4 shows the precisions of the different sampling techniques used compare with each other.

Stratified sampling (by geographical location) and systematic sampling yielded variance estimates that are considerably lower than cluster sampling estimates. It must be noted, however, that the estimation of variances in systematic sampling is based on the assumption that the population is in random order, which is not necessarily true.

If, in fact, the population is randomly ordered, then, stratified random sampling should prove the more efficient sampling scheme.

2.2. *The Rural Sector*

2.2.1. *The Method of the Study* - The pilot area selected for this part of the study is the barrio of Lagundi, Plaridel, Bulacan. Lagundi is so situated such that one end of the barrio is towards the provincial highway and the other end is away from it. In fact, most of the barrios in the country seem to be positioned in the same way that Lagundi is. That is, one end is towards the more progressive poblacion. In this regard, two schemes of systematic sampling were studied. The first scheme involves taking systematic samples from units numbered from end (of the barrio) to end. In the second scheme, the units are numbered along the road sides.

To be explored in this study are:

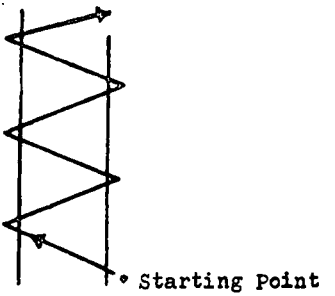
1) Comparison of precision between the two schemes

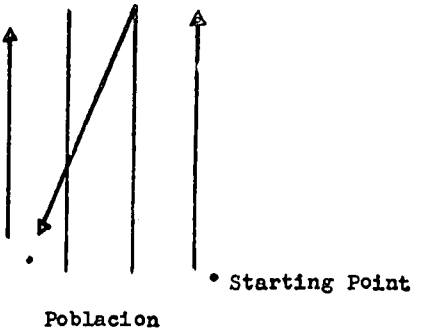
2) Comparison of precision between systematic, random and stratified designs

Table 4. COMPARISON OF PRECISION OF THE DIFFERENT SAMPLING SCHEMES

VARIABLE VARIATE	1/4 SAMPLING FRACTION					1/8 SAMPLING FRACTION				
	Cluster Block	Cluster Precincts	Stratified Geographical Neyman	Stratified Geographical Equal	Systematic (Assumption: Pop'n. in Random Order	Cluster Block	Cluster Precincts	Stratified Geographical Neyman	Stratified Geographical Equal	Systematic (Assumption: Pop'n. in Random Order)
Income	5,201.8363	3,077.6163	1,112.92	1,053.47	1,165.58	7,145.5777	18,940.570	2,875.32	2,336.71	2,252.7124
Expenditure	3,350.6710	1,452.3352	464.7468	534.30	384.65	1,219.0477	6,924.2525	1,137.9	1,044.0	713.0805
Household Size	0.2008	0.2532	0.0352	0.0576	0.0219	0.0154	0.0368	.07943	.11926	0.0483
Legalization of Gambling	0.0005	0.0021	0.0013	0.0021	0.0009	0.0020	0.0024	.00346	.00465	0.0013
Banning of Pornography	0.0409	0.0028	0.0010	0.0018	0.0005	0.0063	0.0641	.00223	.00398	0.0010

3) Estimation of parameters from a single sample, when systematic sampling is used





2.2.2. *Comparison of Precision Between Schemes* (Schemes 1 and 2) and Between Sample Designs - Tables 5 and 6 give the computed results that allow comparison of the two schemes of systematic sampling and comparison of the three sampling techniques in question. Table 5 is based on systematic sample size $n = 10$, and the number of strata, $K = 10$. Table 6, on the other hand, is based on a systematic sample size of $n = 25$, and $K = 4$.

The results indicate that the stratified and the systematic designs definitely yield better precision than the random design. In most cases, though (10 out of 12 values), the "systematic" variance of the mean have values lower than those of the stratified sample.

There are no indications showing that one scheme is better than the other.

2.2.3. *Systematic Sampling: Estimation of Precision From A Single Sample* - Systematic sampling is often resorted to in view of some practical considerations. One problem often encountered in the use of systematic sampling in that of the estimation of the variance based on a single systematic sample. Some authors suggest the possibility of applying certain assumptions in order to estimate the variance. In this connection, some assumptions are made regarding the barrio of Lagundi. How these assumptions compare with each other are shown in Tables 7 and 8.

Table 5. COMPARISON OF PRECISION BY VARIABLES, BY SAMPLING TECHNIQUES AND BY SYSTEMATIC SCHEME OF SAMPLING FOR N = 10, K = 10

VARIABLE	ANOVA				Vy					
	Scheme 1		Scheme 2		Random		Systematic		Stratified	
	d.f.	M.S.	d.f.	M.S.	Scheme 1	Scheme 2	Scheme 1	Scheme 2	Scheme 1	Scheme 2
Income	9	36,536.51	9	44,970.60		40,297.36	3,447.11	2,245.71	3,288.29	4,047.35
	90	S^2 = 438,303.10	90	447,748.41	39,447.28					
Expenditure	9	19,278.69	9	17,595.06		23,827.24	1,231.83	1,065.28	1,735.08	1,583.55
	90	266,599.45	90	264,747.07	23,993.95					
HH Size	9	2.8120	9	7.93		29.06	0.62490	0.18960	0.2529	0.7137
	90	317.2500	90	322.98	28.55					

Table 6. COMPARISON OF PRECISION BY VARIABLES, BY SAMPLING TECHNIQUES AND BY SYSTEMATIC SCHEME OF SAMPLING FOR N = 25, K = 4

VARIABLE	ANOVA				Vy					
	Scheme 1		Scheme 2		Random		Systematic		Stratified	
	d.f.	M.S.	d.f.	M.S.	Scheme 1	Scheme 2	Scheme 1	Scheme 2	Scheme 1	Scheme 2
Income	24	S^2 = 24,777.02	24	32,745.48	5,578.62	5,826.01	153.8648	134.743	741.31	982.36
	75	S^2 = 185,954.03	75	194,200.30						
Expenditure	24	11,511.58	24	19,457.50	3,322.52	3,547.31	325.494	423.2835	345.35	583.72
	75	110,750.57	75	118,243.81						
HH Size	24	6.67	24	6.84	3.97	3.99	.1307	.1588	0.20	0.20
	75	132.39	75	132.90						

Table 7. COMPARISON BETWEEN THE POPULATION Vy sy AND SOME ESTIMATES OF Vy sy BY VARIABLE, AND BY SCHEME FOR N = 10

POPULATION VARIANCE AND ITS ESTIMATES	Income		Expenditure		Household Size	
	Scheme 1	Scheme 2	Scheme 1	Scheme 2	Scheme 1	Scheme 2
Vy sy (Parameter)	3,447.11	2,245.71	1,231.83	1,065.28	0.62490	0.18960
I Vy sy (assume	4,610.95	7,782.01	2,948.79	5,719.27	0.29	0.30
population is in random order)	1,730.13	2,980.03	985.16	560.34	0.24	0.48
II Vy sy (assume	4,846.68	3,086.14	3,811.03	3,122.30	0.37	0.20
stratification effect)	1,562.53	4,636.67	1,011.23	809.34	0.39	1.13
III Vy sy (assume	5,615.18	1,931.54	4,695.19	1,836.05	0.41	0.21
linear trend)	1,580.74	5,803.82	1,457.17	1,054.24	0.36	1.23

Note: 2 systematic samples are considered in each type of estimation procedure
* Variance estimates closest to the parameter value

Table 8. COMPARISON BETWEEN THE POPULATION Vy sy AND SOME ESTIMATES OF Vy sy BY VARIABLE, AND BY SCHEME FOR N = 25

POPULATION VARIANCE AND ITS ESTIMATES	Income		Expenditure		Household Size	
	Scheme 1	Scheme 2	Scheme 1	Scheme 2	Scheme 1	Scheme 2
Vy sy (Parameter)	153.8648	134.743	325.494	423.2835	0.1307	0.1588
I Vy sy (assume	699.11	812.85	316.99	249.27	0.31	0.18
population is in random order)	1161.40	187.37	246.95	215.86	0.16	0.11
II Vy sy (assume	848.78	1172.17	548.05	264.77	0.35	0.20
stratification effect)	1478.70	184.83	413.23	134.26	0.15	0.12
III Vy sy (assume	935.00	752.96	385.20	217.89	0.38	0.29
linear trend)	1400.71	147.70	286.34	191.08	0.14	0.14

It is interesting to note that the first method of estimation (assumption of a population in random order) give the most number of variance estimates having values closest to the population variance (8 out of 12).

III. A STUDY ON RESPONSE BIAS

3.1. *Interviewer Effect*[1]

3.1.1. *The Analysis of Variance* - For the purpose of this study, two interviewers (A and B) were considered. To each interviewer were assigned 18 respondents. Each respondent was interviewed twice. To minimize the effect of time, the first and the second interviews were conducted within a period of just two weeks. The analysis of variance for the variables, income and household size, are shown in Tables 9 and 10.

Table 9. ANOVA — HOUSEHOLD INCOME

	d.f.	Sum of Squares	Mean Sum of Squares	F
(Interviewer) Columns (A & B)	1	268,766.64	268,766.64	1.21428
Row Time of Interview (1st & 2nd)	1	55,945.13	55,945.13	.25275
Interaction	1	15,341.71	15,341.71	.06931
Within Groups	68	15,050,976.40	221,337.88	

$F_{0.05} = 4.0$

Table 10. ANOVA — HOUSEHOLD SIZE

	d.f.	Sum of Squares	Mean Sum of Squares	F
(Interviewer) Columns (A & B)	1	.23	.23	.03066
Row Time of Interview (1st & 2nd)	1	56.89	56.89	7.58533
Interaction	1	2.70	2.70	.36000
Within Groups	68	509.80	7.50	

$F_{0.05} = 4.0$

[1] Reference, Mahalanobis, Experiments in Statistical Sampling in the Indian Statistical Institute, 1961.

The computed F's indicate the absence of both interviewer and time effects in household income but the presence of interviewer effect in household size.

3.1.2. *Correlation Analysis* - In the survey of San Miguel area, the listing phase was carried out by a group of not well-trained interviewers (students) while the survey itself was carried out by the well-trained interviewers of the Statistical Center. In the Lagundi survey, on the other hand, the listing was carried out by the trained interviewers while the actual survey was done by a mixed group of students, faculty, and staff. Using correlation analysis, differences in responses due to different sets of interviewers are studied. The following tables show the relationships between the responses to one group of trained interviewers and the responses to another group of untrained interviewers.

It might be of interest to note that while there is significant correlation in all cases considered, there are indications, however, showing that disparities in responses of the two groups are more evident in *Income - Lagundi* rather than *Income - San Miguel* and *Household Size - San Miguel* rather than *Household size - Lagundi.*

3.2. *Poll Issues: Interviewer Effect and Time Effect*

A respondents' opinion may possibly change depending on the type of interviewer, the length of time (survey period). and/or the type of poll issue under considerations. Thus, in this portion of the study, two issues are considered: gambling and pornography. The variate measured is the proportion of respondents which changed their views (responses) over the time periods under consideration ($t_1 = 2$ weeks, $t_2 = 3$ months). Two interviewers did the interviewing. The table below gives the various percentages of respondents who changed their opinions when interviewed for the second time.

Table 11. CORRELATION BETWEEN RESPONSES TO TWO GROUPS OF INTERVIEWERS
INCOME — LAGUNDI, PLARIDEL, BULACAN
(Survey by untrained interviewers)

LISTING BY THE TRAINED INTERVIEWERS	0-50 1	51-100 2	101-150 3	151-200 4	201-250 5	251-300 6	301-350 7	351-400 8	401-450 9	451-500 10	501-600 11	601-700 12	701-800 13	801-900 14
0-50	11	6	7	1	4	2	—	—	—	—	—	—	1	—
51-100	—	8	—	3	—	1	—	—	—	—	—	—	—	—
101-150	—	—	—	1	—	1	—	—	—	—	—	—	—	—
151-200	1	—	1	5	—	—	1	1	1	—	1	1	—	—
201-250	—	4	—	1	4	—	—	1	—	1	—	—	1	—
251-300	—	1	1	1	—	2	—	—	—	—	—	—	—	—
301-350	—	—	—	—	—	—	1	—	—	—	—	—	—	—
351-400	—	—	—	—	—	—	—	1	—	—	—	—	—	—
401-450	—	—	—	—	—	—	—	—	—	—	—	—	—	—
451-500	—	—	—	—	1	—	—	—	—	—	—	—	—	—
501-600	—	—	—	—	—	—	—	—	—	1	—	—	1	—
601-700	—	—	—	—	—	—	—	—	—	—	1	—	—	—
701-800	—	—	—	1	—	—	—	—	—	—	—	—	—	—
801-900	—	—	—	—	—	—	—	—	—	—	—	—	—	1
901-1000	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1000 & Over	—	—	—	—	—	—	—	—	—	—	—	—	—	—

$r = 0.56346$
$t = 6.13834$
$t_{0.05} = 1.99$

Table 12. CORRELATION BETWEEN RESPONSE TO TWO GROUPS OF INTERVIEWERS
INCOME — SAN MIGUEL AREA
(Listing By Students)

Survey By Trained Interviewers	0-50	51-100	101-150	151-200	201-250	251-300	301-350	351-400	401-450	451-500	501-600	601-700	701-800	801-900	901-1000	1001-1100	1101-1200	1201-1300	1301-1400	1401-1500	1501-1600	37
0- 50	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
51- 100	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
101- 150	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
151- 200	1	1	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	3
201- 250	—	—	1	3	1	1	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	8
251- 300	—	—	2	—	—	1	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	5
301- 350	—	—	—	—	1	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	3
351- 400	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
401- 450	—	—	—	—	1	—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	—	3
451- 500	—	—	—	—	—	—	—	—	—	2	1	—	—	—	—	—	—	—	—	—	—	3
501- 600	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
601- 700	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
701- 800	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
801- 900	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
901-1000	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	—	—	2
1001-1100	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1101-1200	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1
1201-1300	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
1301-1400	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1401-1500	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1501-1600	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1601-1700	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1701-1800	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1801-1900	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1901-2000	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

$r = 0.865$
$t = 10.2004$
$t_{0.05} = 2.03$

Table 13. CORRELATION BETWEEN RESPONSES TO TWO GROUPS OF INTERVIEWERS

HOUSEHOLD SIZE — LAGUNDI, PLARIDEL, BULACAN
(Survey by Untrained Interviewers)

LISTING BY TRAINED INTERVIEWERS	1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16
1 - 2	6	—	—	—	—	—	—	—
3 - 4	3	28	1	—	—	—	—	—
5 - 6	—	4	22	7	—	—	—	—
7 - 8	—	—	2	15	1	—	1	—
9 - 10	—	—	1	1	4	1	1	—
11 - 12	—	—	—	—	—	3	—	—
13 - 14	—	—	—	—	—	—	—	—
15 - 16	—	—	—	—	—	—	—	1

$r = 0.905$
$t = 21.548$
$t_{0.05} = 1.98$

Table 14. CORRELATION BETWEEN RESPONSES TO TWO GROUPS OF INTERVIEWERS

HOUSEHOLD SIZE — LAGUNDI, PLARIDEL, BULACAN
(Listing By Students)

SURVEY BY TRAINED INTERVIEWERS	1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16
1 - 2	—	—	—	—	—	—	—	—
3 - 4	2	3	—	—	1	—	—	—
5 - 6	—	1	10	—	—	1	—	—
7 - 8	1	—	1	8	1	—	—	—
9 - 10	—	—	1	—	4	—	—	—
11 - 12	—	—	—	—	1	—	—	1
13 - 14	—	—	—	—	—	—	—	1
15 - 16	—	—	—	—	—	—	—	—

$r = 0.898$
$t = 5.908$
$t_{0.05} = 2.03$

Table 15. PERCENTAGE OF RESPONDENTS WHO CHANGED THEIR OPINIONS DURING THE SECOND INTERVIEW

Interviewers A	Issue C	Length of Time Between the 1st & 2nd Interviews B	
		B_1 (2 weeks)	B_2 (3 months)
A_1	C_1 Gambling	33.33%	29.62%
	C_2 Porno-graphy	16.66%	18.51%
A_2	C_1 Gambling	31.57%	30.43%
	C_2 Porno-graphy	15.78%	17.39%

To analyze the above data, the analysis of variance approached discussed in a previous section is extended and performed on the transformed data.

The transformation $\phi = \tan^{-1}\sqrt{\frac{P}{1-P}}$ is applied on the above data

TABLE 16
ANOVA - POLL ISSUE, INTERVIEWER EFFECT, TIME EFFECT

A	1	0.5565	0.5565	2.2083
B	1	0.0171	0.0171	0.0678
C	1	184.0407	184.0407	730.3202***
AB	1	0.2556	0.2556	1.0143
BC	1	3.3784	3.3784	13.4063*
ERROR	2	0.5041	0.2520	

*/ Nearly significant
$F_{0.10} = 8.53$ with 1,2 df.

***/ Highly significant
$F_{0.01} = 98.49$ with 1, 2 d.f.

The overall results are shown in Table 16.

The analysis of variance computations show that changes in opinion depend much on the type of poll issue under consideration. There is also an interaction between the type of

issue and the length of time between the first and the second interviews.

IV. INCOME AND EXPENDITURES PATTERNS

4.1. *Income Distribution*

One of the most common observations in the study of the economics of income distribution is the fact that these distributions are humped to the left, or positively skewed. The income distributions of 117 household in the district of San Miguel, Manila and of 104 households in the barrior of Lagundi, Plaridel, Bulacan, show this characteristic. (Figure 1 and Tables 17 and 18).

No. of Household

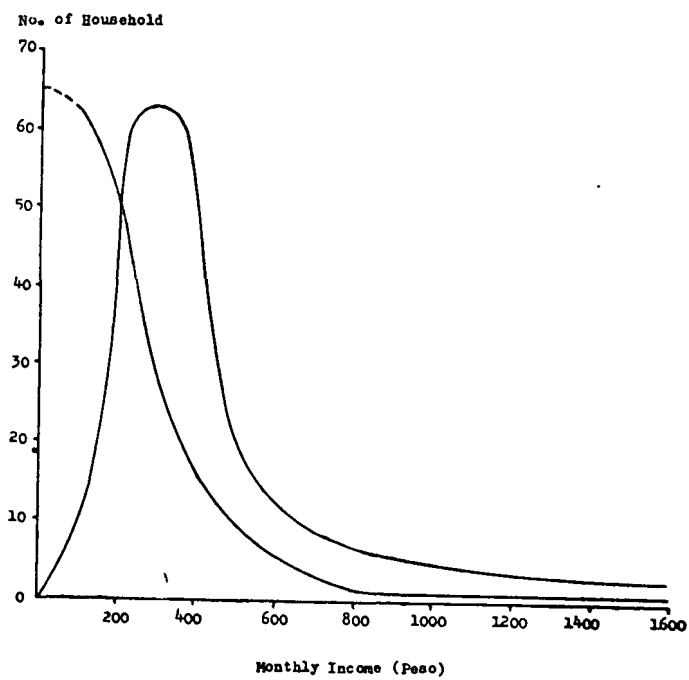


Figure 1

MONTHLY INCOME OF SAMPLE HOUSEHOLDS IN SAN MIGUEL, AND IN LAGUNDI
1971

Table 17. DISTRIBUTION OF INCOME OF SAMPLE HOUSEHOLDS IN SAN MIGUEL, MANILA, 1971

Monthly Income (Pesos)	No. of Households	Total Income	Cumulative No. of Households	% Dist'n Total Income
Below 200	9	1051	7.7	2.0
200 - 399	63	15882	61.6	32.0
400 - 599	20	9770	78.6	50.5
600 - 799	7	4653	84.6	59.4
800 - 999	6	5129	89.7	69.0
1000 - 1199	4	4060	93.2	76.8
1200 - 1599	5	7385	97.4	90.0
1600 - Over	3	4900	100.0	100.0
Total	117	52830		

Table 18. DISTRIBUTION OF INCOME OF SAMPLE HOUSEHOLDS IN LAGUNDI, PLARIDEL, BULACAN, 1971

Monthly Income (Pesos)	No. of Households	Total Income	Cumulative No. of Households	% Dist'n Total Income
Below 200	62	5724	59.6	27.9
200 - 399	29	7051	87.5	62.3
400 - 599	6	2769	93.3	75.8
600 - 799	6	4116	99.1	95.9
800 - 999	1	840	100.0	100.0
Total	104	20500		

The relative degrees of concentration of income distribution in the two samples may be compared through the use of the Lorenz curve. This curve is obtained when the cumulative proportion of income received is plotted against the cumulative percentage in income recipients from the lowest income levels upward. The diagonal lines represent perfect equality of incomes - at any given per centage of total income the proportion of income recipients, which men that every household earns the same income. Thus, the closer in the Lorenz curve to the diagonal, the less concentrated is the distribution of income. Figure 2 shows the Lorenz curves of income distribution for San Miguel and Lagundi. Comparison of the two curves show that income distribution is more concentrated in Lagundi.

issue and the length of time between the first and the second interviews.

IV. INCOME AND EXPENDITURES PATTERNS

4.1. *Income Distribution*

One of the most common observations in the study of the economics of income distribution is the fact that these distributions are humped to the left, or positively skewed. The income distributions of 117 household in the district of San Miguel, Manila and of 104 households in the barrior of Lagundi, Plaridel, Bulacan, show this characteristic. (Figure 1 and Tables 17 and 18).

No. of Household

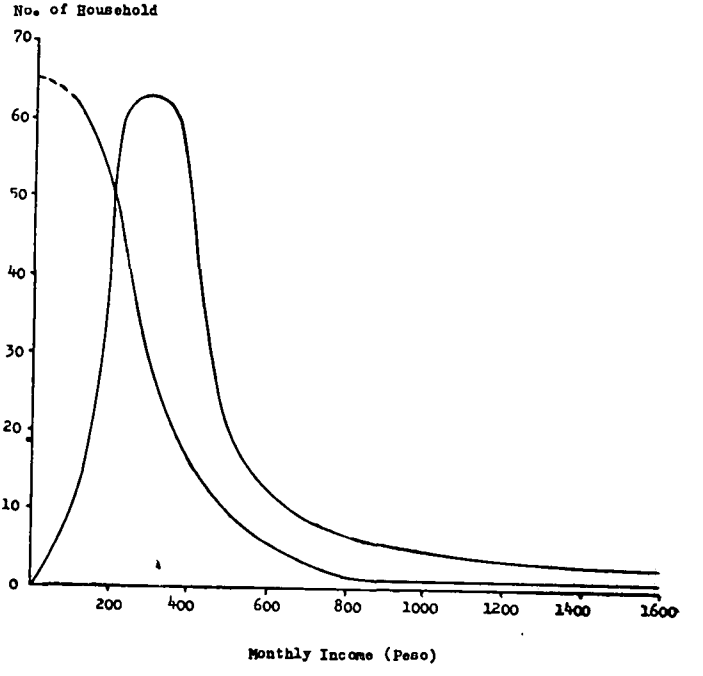


Figure 1

MONTHLY INCOME OF SAMPLE HOUSEHOLDS IN SAN MIGUEL, AND IN LAGUNDI
1971

Table 17. DISTRIBUTION OF INCOME OF SAMPLE HOUSEHOLDS IN SAN MIGUEL, MANILA, 1971

Monthly Income (Pesos)	No. of Households	Total Income	Cumulative No. of Households	% Dist'n Total Income
Below 200	9	1051	7.7	2.0
200 - 399	63	15882	61.6	32.0
400 - 599	20	9770	78.6	50.5
600 - 799	7	4653	84.6	59.4
800 - 999	6	5129	89.7	69.0
1000 - 1199	4	4060	93.2	76.8
1200 - 1599	5	7385	97.4	90.0
1600 - Over	3	4900	100.0	100.0
Total	117	52830		

Table 18. DISTRIBUTION OF INCOME OF SAMPLE HOUSEHOLDS IN LAGUNDI, PLARIDEL, BULACAN, 1971

Monthly Income (Pesos)	No. of Households	Total Income	Cumulative No. of Households	% Dist'n Total Income
Below 200	62	5724	59.6	27.9
200 - 399	29	7051	87.5	62.3
400 - 599	6	2769	93.3	75.8
600 - 799	6	4116	99.1	95.9
800 - 999	1	840	100.0	100.0
Total	104	20500		

The relative degrees of concentration of income distribution in the two samples may be compared through the use of the Lorenz curve. This curve is obtained when the cumulative proportion of income received is plotted against the cumulative percentage in income recipients from the lowest income levels upward. The diagonal lines represent perfect equality of incomes - at any given per centage of total income the proportion of income recipients, which men that every household earns the same income. Thus, the closer in the Lorenz curve to the diagonal, the less concentrated is the distribution of income. Figure 2 shows the Lorenz curves of income distribution for San Miguel and Lagundi. Comparison of the two curves show that income distribution is more concentrated in Lagundi.

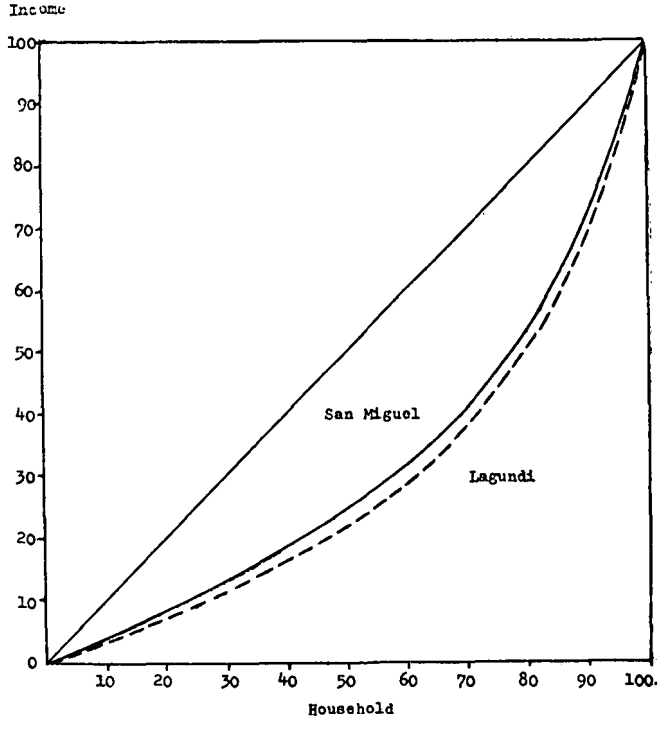


Figure 2

LORENZ CURVES OF INCOME DISTRIBUTION FOR LAGUNDI AND SAN MIGUEL
1971

Source: Last two columns of Tables 17 and 18

EXPENDITURE PATTERNS

Family budget studies have often shown certain regularities in the behaviour of family units with regard to the spending habits on various broad categories, such as expenditure on food and beverages, rent, utilities, clothing and miscellaneous items. The well-known Engel's law state that as income increases, the proportion of total expenditure spent on food and housing decreases. On the other hand, the proportion of total expenditure spent on clothing and miscellaneous items increases.

Table 19. PERCENTAGE DISTRIBUTION OF EXPENDITURES OF THE 117 SAMPLE HOUSEHOLDS FROM THE DISTRICT OF SAN MIGUEL, 1971

Level Income	Food and Beverages	Rent	Utilities	Clothing	Miscellaneous	Total
Below 200	55.4	11.1	7.5	5.8	20.2	100.0
200 - 399	56.5	14.3	7.2	3.3	18.6	100.0
400 - 599	52.5	19.2	7.4	5.7	15.2	100.0
600 - 799	51.5	17.5	6.0	2.0	17.0	100.0
800 - 999	45.4	15.7	6.5	6.4	26.0	100.0
1000 - 1199	36.3	10.6	7.3	3.5	42.2	100.0
1200 - 1599	52.2	13.5	4.6	5.5	24.2	100.0
1600 & Over	27.3	19.1	9.7	12.4	31.4	100.0

Broadly speaking, Table 19 shows a decrease of the proportions spent on food and beverages and an increase in the proportions spent on miscellaneous items as income increases. The proportions spent on rent, utilities, and clothing, however, do not show any perceptible trend. An exception is the expenditure on clothing and utilities in the rural area which increases with income, if the one household in the 800-999 bracket is excluded. This could be either due to sampling fluctuations, or to the influence of other factors, such as household size, average age of household member, etc., which were not taken into consideration. It may be noted, however, that the proportion spent by the urban households in the highest income level (12.4%) is significantly higher than the proportion spent on this item by the households in the other income levels.

The effect of household size on the spending patterns of households may be included in the analysis by various methods. One method is to express income and expenditure data on a per capita basis by dividing each household's variables by the number of household members. This procedure is sometimes refined by expressing all variables per equivalent adult instead of per person. The equivalent adult scales give different weights to each persons expenditures depending on his sex and stage of maturity.

Regression methods may also be used to analyze the influence of household size on the proportion of total expenditure spent on particular expenditure items. For example, the following relationship may be estimated:

$$\log E = L + B \log L + \sigma \log N + e$$

where E = food expenditure
Y = total expenditure
N = household size

The use of the logarithmic relationships will result in the estimation of coefficients which represent elasticities. For example, the value of B is the elasticity of demand for food and σ is the expenditure elasticity of household size. These two coefficients are both expected to be less than unity. B is expected to be less than unity in support of Engle's law about the decline in the proportion of total expenditure spent on food as income (and total expenditures) increase.

The elasticity coefficient σ is expected to be less than unity, due to the influence of economies of scale. As household size increases, the percentage increase in food expenditures is less than the percentage increase in household size.

Other so-called Engel functions may be estimated for the other expenditure items.

The various types of analysis described above will be presented in a subsequent study for all expenditure items for both the rural and urban sectors of this study.

TABLE 20: X^2 — Analysis: Rural - Urban comparison of Responses to certain Poll Issues

	Rural-Urban Responses to Poll Issues					Rural-Urban Percentage of Non-Response				
		Level of Significance (%)					Level of Significance (%)			
POLL ISSUE	X2	1	5	10	Insignificant	X2	1	5	10	Insignificant
Pornography	16.4547	***						**		
Medium of Instruction	5.6651		**							xxx
Graft & Corruption	32.5335	***							*	
Gambling	31.3266	***								xxx
Bringing about change	2.0194				***					xxx
Rallies & Demonstration	0.1520				***					***

V. RURAL-URBAN COMPARISON OF RESPONSES TO POLL ISSUES

Six poll issues are considered in this study. These issues are all included in the overall questionnaire. For comparison purposes, 104 households (complete enumeration) were all used to represent the rural area and a systematic sample of 117 households were taken to represent the urban area. Chi-square analysis is then applied to study the responses.

Table 20 shows how the rural respondents of Lagundi differ from their urban counterparts in as far as poll issues are concerned. In 4 out of six poll issues, they differ. The two issues where both sectors concur in their respective opinions are "bringing about change" and "rallies and demonstrations". Both issues deal with activism.

The right-hand portion of the table, on the other hand, gives certain indications that the incidence of non-responses is the same in both the rural and the urban areas (4 out of 6 show no difference).

VI. SUMMARY AND CONCLUSION

The previous discussions regarding the comparison made on several sample designs indicated that a certain design is relatively more efficient than the others; that a certain design is less costly than the others; and that designs applicable or practical to use in urban areas are not necessarily applicable or practical to use in the rural areas.

Studies on response bias showed that the amount of bias depends on the type of variate being measured (type of poll issues or type of variables). The length of the survey period interacts with the said bias. Another possible source of bias is the place where the interview is taken.

Income and expenditure patterns in the rural area (Lagundi) and the urban area (San Miguel) differ from each other. The income distribution is more concentrated in Lagundi than in San Miguel. The expenditure patterns of various expenditure items for the San Miguel area confirm Engel's law (as income increases, the proportion of total expenditure spent on food and housing decreases, while the pro-

portions of the total expenditure spent on clothing and miscellaneous items increase).

While the proportions of non-responses to the total number of respondents are the same in Lagundi and in San Miguel, the nature of responses to certain poll issues, on the other hand, differ.

The above findings and observations, it must be pointed out, are true only for the pilot areas selected, San Miguel as the urban place and Lagundi, as the rural place. The results, however, can be used as baselines for further larger-scale studies. It must further be pointed out that such findings and observations are based only on specific methods and approaches used in the study. There is therefore a great possibility of exploring other methods and approaches in the future.