NATIONAL SAMPLE SURVEY OF HOUSEHOLDS
Report on the Sampling Designs and Estimation Procedures

by

Domingo C. Alonzo*

INTRODUCTION

The object of this paper is to give a brief account of the design and estimating procedure to be used in the national sample survey of households[1] which is being conducted by the National Economic Council. This is one of the most recent surveys in which a sound basis of sampling design has been developed to provide information far more rapidly and economically than complete enumeration.

The purpose of a sample survey is to gain information about the characteristics of a certain population based on sample. By means of the basic techniques of sample design this purpose is achieved at minimum cost for given accuracy or vice versa.

It will be convenient if some basic principles of surveys based on sampling techniques may be briefly explained at this stage. The totality of all possible values that could be selected in a complete enumeration is known as the *population* under study. Before a sample survey is begun, it is important to have a clear definition of the population, and the units from which the selection is to be made. The list of units from which the sample is selected is known as the sampling *frame*. Having decided on the population, the next problem is to determine the size of sample; how to draw the sample; and what are the ultimate sampling units.

The sampling method for any particular investigation may be designed to suit its particular needs. Whatever method is adopted for selecting the sample, the first thing to remember is that the sample shall be representative of the population. This may be achieved by assigning to each unit in the popula-

---

*Chief Statistician, OSCS, National Economic Council and Professorial Lecturer of Statistics, The Statistical Center, University of the Philippines.

tion an equal change or probability of being included in the sample. It is a known fact that the larger the sample is, the more likely it is to reproduce salient characteristics of the population. But still the magnitude of the error or discrepancy will also depend on the method of sampling adopted. In brief, the art of sampling is focused towards choosing sampling methods appropriate for selection from the populations concerned, and it is advisable to adopt such sampling methods that will yield reasonable accuracy for practical purposes without being too difficult to execute and also will give convenience in tabulation.

The sampling method adopted in the national sample survey of households is multistage. In this method a number of units are selected at random; within each selected unit a number of sub-units are next selected at random; and within each selected sub-unit a number of sub-units are finally selected. This is called a three-stage sampling because in this methods the selection is carried out not in one single stage but in three different stages. The coverage of the process of selection of sampling units is repeated over a smaller and smaller region.

To mention some of the advantages of multistage.

i) It enables existing natural divisions (municipalities, barrios, etc.) of the population to be utilized as units at various stages.

ii) It permits the concentration of field work at areas scattered over the population. The supervisors can devote most of their time to supervising the enumeration instead of travelling a great deal to hunt for the enumerators.

iii) The construction of the second-stage frame need only be carried out for those first-stage units which are actually included in the sample and similarly for the third stage frame.

Although multistage sampling is less efficient than other simple sampling methods the cost saved in listing and travelling is more than enough to cover the inclusion of more units to reach the desired accuracy. It should be added that multistage sampling is particularly valuable in surveys of a country like the Philippines where no frame exists which is sufficiently detailed and accurate for the subdivision of the population into reasonable small sampling units.
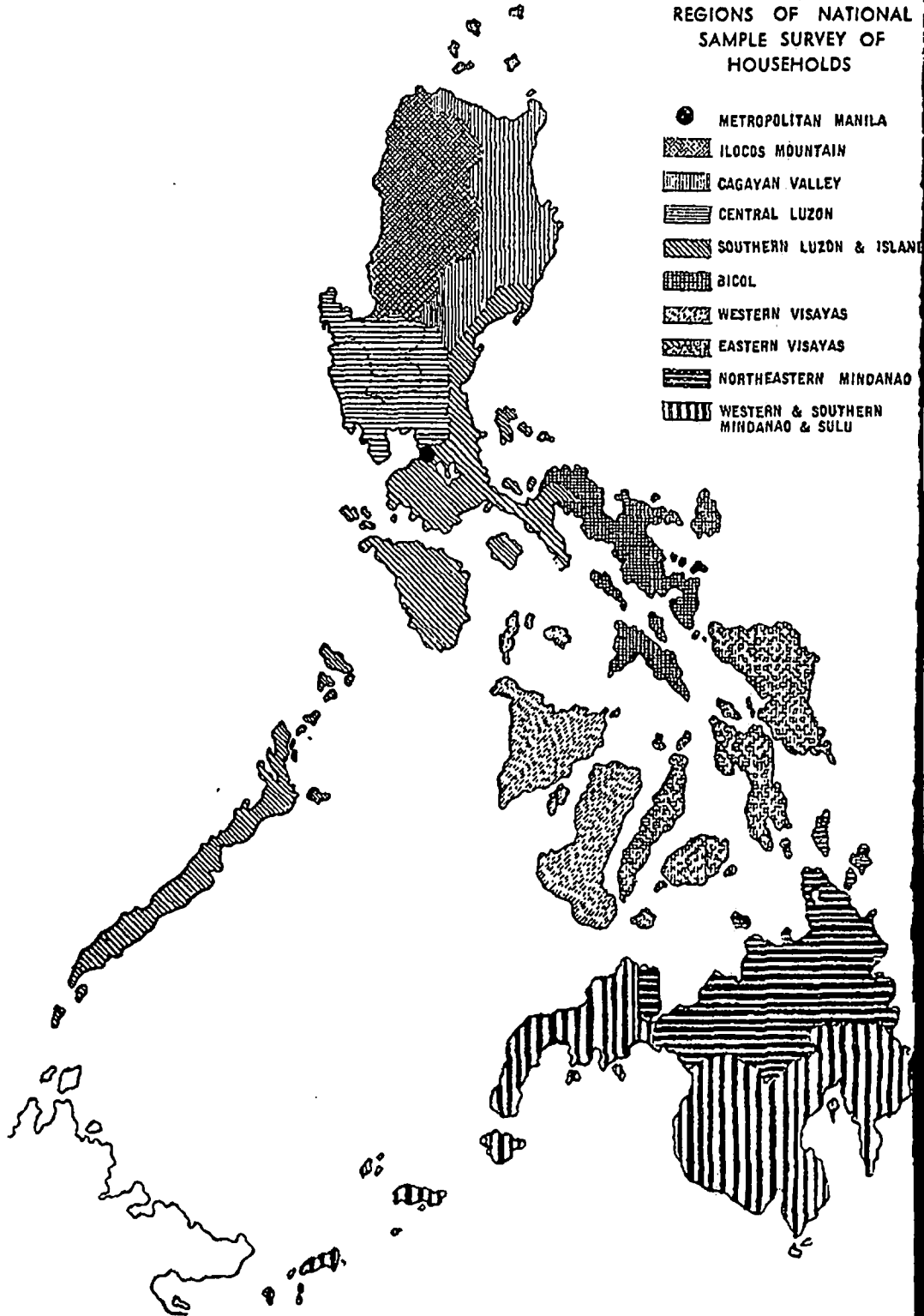
REGIONS OF NATIONAL SAMPLE SURVEY OF HOUSEHOLDS

● METROPOLITAN MANILA
ILOCOS MOUNTAIN
CAGAYAN VALLEY
CENTRAL LUZON
SOUTHERN LUZON & ISLAND
BICOL
WESTERN VISAYAS
EASTERN VISAYAS
NORTHEASTERN MINDANAO
WESTERN & SOUTHERN MINDANAO & SULU

Illustration I.

THE SURVEY DESIGN
Rural Areas

The rural areas in the Philippines were classified into two sectors, the barrios representing the true rural conditions and the poblaciones representing the intermediary conditions of the rural and urban areas. All barrios and poblaciones were grouped into their respective municipalities while the municipalities were arranged into regions. (See illustration 1.) According to the 1948 Census, there are 1,256 poblaciones and 17,603 barrios in the country. The rural areas have a total population of about 15.4 millions of which 12.2 relate to the barrios while the rest to the poblaciones.

Sampling Scheme. The sampling was carried out in three stages. Municipalities were selected at the first stage while barrios were selected from the selected municipalities at the second stage. (See illustration 2.) At the third, e.g., last stage, households will be selected from the selected barrios. Over all the barrios, 3,000 households will be systematically selected for enumeration while 1,500 households will be similarly taken from the poblaciones.

Illustration 2: Schematic Diagram.

Dulag

Calape

Pilar

Moalboal
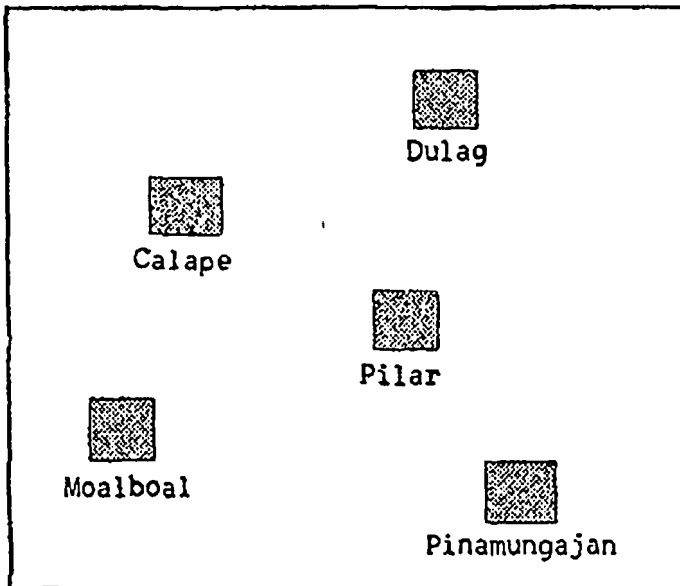
Pinamungajan

a) First stage: In stratum 25, five municipalities, i.e., Dulag, Calape, Pilar, Moalboal and Pinamungajan, were selected at random.
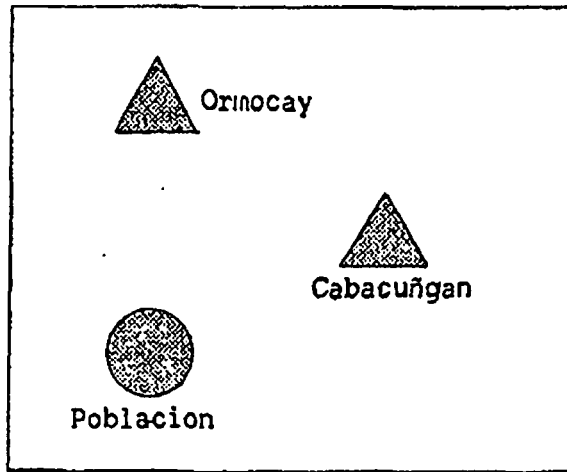
b) Second stage: In Dulag, a selected municipality, its poblacion and two barrios, i.e., Ormocay and Cabacuñgan, were selected at random.

Stratification. All the municipalities in the Philippines were first arranged into ten regions. (See illustration 1.) Then in each region the municipalities were stratified according to their densities of population. By this process some regions may have only one stratum while others may have more than one depending upon their populations.

In every region, the densities of the municipalities were arranged in ascending order preparatory to stratification. The strata were then formed by counting off their approximate specified population starting from the municipality with the lowest density. The strata were so adjusted that their populations were approximately equal and in no case was a municipality divided.

Illustration 3: Stratum 25 of Region VIII, Western Visayas.

Serial Number	Municipality	Density	Population
1.	Toledo	199	39,225
2.	Minglanilla	204	15,774
3.	Cordoba	205	7,116
4.	Corrella	206	5,194
5.	Daram	208	21,848
6.	Zumarraga	208	8,745
7.	San Remigio	209	20,645
8.	Moalboal	213	15,019
9.	Anda	214	11,688
10.	Catigbian	215	16,233

Serial Number	Municipality	Density	Population
11.	Poro	215	12,523
12.	San Francisco	218	19,286
13.	Alburquerque	219	5,966
14.	Pinamungajan	223	20,364
15.	Tolosa	229	8,569
16.	Cabalian	230	7,447
17.	Mayorga	232	8,411
18.	Dulag	232	31,185
19.	Maripipi	232	6,550
20.	San Antonio	234	6,781
21.	Panglao	235	11,274
22.	Hinundayan	241	7,199
23.	Calape	246	21,383
24.	Sogod	248	15,687
25.	Liloan	249	12,292
26.	Daanbantayan	250	29,484
27.	Pilar	251	8,663
28.	Carigara	256	26,803
29.	Barili	258	27,267
30.	Dalaguete	261	29,333
31.	Maribojoc	263	13,906
32.	Mercedes	267	5,265
33.	Naga	272	24,911
	Total		513,625

Selection of Municipalities. In each stratum, five municipalities were randomly selected with probability proportional to size of population. This would enable big and small municipalities to have proportionate chances of being included in the sample. Each selected municipality in each stratum was replaced before a succeeding selection was made and this was done until five independent municipalities were selected. There were 150 municipalities selected from all the 30 strata.

Illustration 4: Selection of sample municipalities with probability proportional to size using Lahiri's Method.[2]

Each stratum consists of N sampling units, i.e., municipalities; and P is the largest population. Choose a pair of random numbers A and B, the first in the range $1 — N$ and the second between 1 and P. Let the population of the Ath municipality be P_A, then if B exceeds the population of the Ath municipality reject, otherwise accept the municipality. Repeat the operation until selection is made if rejected.

From stratum 25, illustration 3,

$$N = 33 \text{ and } P = 39225,$$

suppose we selected at random pairs of numbers from a table of random numbers,

    a) $A = 18$ $B = 19682$

    b) $A = 17$ $B = 31780$

In a), $A = 18$ denotes the serial number of the municipality in the stratum. The eighteenth municipality is Dulag with a population of 31,185. Since $B = 19,682$ does not exceed 31,185, Dulag is selected. In b), $A = 17$ and the seventeenth municipality is Mayorga with a population of 8,411. In this case, $B = 31,780$ exceeds 8,411, so Mayorga is not selected. The process is repeated until the desired number of samples is selected.

Selection of Poblaciones and Barrios. Two barrios and the poblacion of a selected municipality were selected for complete listing of their households. Thus, there were 150 sample poblaciones and 300 sample barrios selected from the 150 sample municipalities. Sample barrios were selected with equal probability and without replacement while the poblaciones of the selected municipalities were automatically made the sample poblaciones.

Illustration 3: Selection of sample barrios from the municipality of Dulag.

Sample barrios were selected at random using two digits random numbers drawn from a table of random numbers. Two random numbers between 01 to 30, the number of barrios in the municipality, will be drawn from the table. Discard any random number that is greater than the number of barrios in the list.

Barrio list of Dulag

1. Alegre	5. Burgos
2. Andres Bonifacio	6. Cabacuñgan
3. Arado	7. Cabarasan
4. Bolongtohan	8. Cabatoan

9.	Calbasag	20.	Rizal
10.	Calipayan	21.	San Agustin
11.	Camote	22.	San Isidro
12.	Cogon-Bincay	23.	San Jose
13.	Dacay	24.	San Pedro
14.	Dachay-Cabacuñgan	25.	San Roque
15.	Liberty	26.	Tabo
16.	Mati	27.	Talisay
17.	Mavorga	28.	Tigbao
18.	Ormocay	29.	Union
19.	Rawis	30.	Wilson

The random numbers drawn are 06 and 18. The barrios having the numbers corresponding to the random numbers drawn are Cabacuñgan and Ormocay. These are the sample barrios.

Selection of Households. All the households in each selected barrio or poblacion will be completely listed. The number of sample households to be enumerated will be determined by the product of the specified proportion and the total number of households listed in a selected barrio or poblacion. A systematic selection of households will be made in all the selected barrios or poblaciones.

Illustration 6: Systematic selection of households.

If the number of households listed in barrio Cabacuñgan is 330 then the number to be enumerated is 26. This number is obtained by multiplying 330 by .077, the proportion of households to be enumerated to total households listed in barrio Cabacuñgan, Dulag. (See illustration 8.) These households will be selected systematically with a random start. The sampling interval for barrio Cabacuñgan is $1/f = 13$, where $f = .077$. A random number will be selected in the range 1 to 13 in the central office to be given to the supervisor. This will be given to the enumerator after listing. The random number selected will indicate the starting point for systematic selection. The household with the given starting number and every 13th in the list will compose the number of sample households to be enumerated. Suppose the random number start is 8, then the households in the list with serial numbers 8, 21, 24, 37, 50, 63,

76, 89, 102, 115, 128, 141, 154, 167, 180, 193, 206, 219, 232, 245, 258, 271, 284, 297, 310, and 323 will be enumerated.

Urban Areas

The urban areas in the Philippines were also classified into two sectors: the chartered cities including provincial capitals, and Metropolitan Manila. Metropolitan Manila includes the four congressional district of Manila and its suburbs, namely, Quezon City, Pasay City, Caloocan, Makati, San Juan, Parañaque, and Mandaluyong. The chartered cities and provincial capitals have a combined populations of 2.5 millions while Metropolitan Manila has about 1.5 millions. The city of Manila alone has a population of about one million (1948 Census).

Sampling Scheme. The sampling was carried out in two stages. At the first stage precincts were randomly selected while at the second, e.g., last stage, households were selected from each selected precinct. A total of 1,200 households will be selected systematically for enumeration from all the chartered cities and provincial capitals while 800 households will be selected systematically from Metropolitan Manila.

Stratification. The chartered cities and provincial capitals were grouped into ten regional divisions after the three big cities, e.g., Cebu City, Iloilo City, and Davao City were separated. All other cities and provincial capitals in each region were stratified according to degree of urbanization, i.e., percentage of urban precincts. After arranging the cities and provincial capitals in each region in the descending order of urbanization, the number of precincts was counted, starting from the highly urbanized city or provincial capital, to form a stratum. In each region, the strata were so adjusted that they have approximately equal number of precincts and in no case was a city or provincial capital divided. The three big cities were also stratified so that the over-all total was 30 strata.

All precincts in Metropolitan Manila were grouped into the following areas: the four congressional districts of Manila City, Quezon City, Pasay City, Caloocan, Makati, San Juan, Parañaque and Mandaluyong. The precincts within each area were

stratified according to geographical location because information regarding their characteristics was not available. In each area, the strata were adjusted so that they have approximately equal number of precincts. In such a case they were considered a single stratum. There were 32 formed for Metropolitan Manila.

Selection of Precincts. In both the chartered cities including provincial capitals and Metropolitan Manila, sample precincts were selected at random and with replacement from each stratum. Five precincts from each stratum were selected with equal probability. In each stratum the selected precinct was replaced before a second one was selected and selection continued until five precincts were selected. Thus, there was 150 precincts selected from all the chartered cities including provincial capitals while 150 precincts were selected from Metropolitan Manila.

Selection of Households. All the households in each selected precinct will be completely listed. The product of the specified proportion and the total number of households listed will determine the number of sample households to be selected for enumeration. A systematic selection of sample households will be used in all the sample precincts.

Miscellaneous Problems

With replacement and repetition of first-stage units. The first-stage units, i.e., municipalities and precincts, were selected with replacement so that each selected first-stage unit would provide an independent estimate of the stratum characteristics and an unbiased estimate of the sampling variance. In this scheme there would be greater flexibility in the succeeding stages leading to an increase in the precision of the estimates and to simple operational facilities, e.g., systematic selection with a random start in the selection of households to be enumerated.

In case of repetition of first-stage units, an independent group of second-stage units were randomly selected for each occurrence. The selection of second-stage units was done without replacement to provide a wider scatter of selected units.

In this scheme, the loss of efficiency caused by repetition of first-stage units would be decreased.

Sampling fractions. The sampling fractions within each of the four sectors, Metropolitan Manila, chartered cities including provincial capitals, poblaciones, and barrios, were kept uniform for simplicity of operation. The table below shows the sampling fractions for each sector of the country based on the estimated population for 1955 and an average households size of 5.5.

TABLE 1

Sector	Sample Households	Sampling Fraction
1. Metropolitan Manila	800	1/400
2. Chartered cities and provincial capitals	1200	1/450
3. Poblaciones	1500	1/450
4. Barrios	3000	1/850

These sampling fractions were obtained using the formulae:

(1) For rural areas, $R_1 = \dfrac{1}{5} \cdot \dfrac{\bar{P}}{\bar{p}} \cdot \dfrac{\bar{B}}{\bar{b}} \cdot \dfrac{\bar{N}}{\bar{n}}$,

where $\bar{R}_1$ is the raising factor or barrios or poblaciones

$\bar{P}$ is the average population per stratum

$\bar{p}$ is the average population per municipality

$\bar{B}$ is the average number of barrios (poblacion) per municipality

$\bar{b}$ is the average number of sample barrios (poblacion) per municipality

$\bar{N}$ is the average number of households per barrio (poblacion)

$\bar{n}$ is the average number of sample households to be enumerated per barrio (poblacion).

Illustration 7: Computation of the raising factor for barrios.

$$\bar{P} = 516{,}660$$

$$\bar{p} = 12{,}340$$

$$\bar{B} = 14$$

$$\bar{b} = 2$$

$$\bar{N} = 670$$

$$\bar{n} = 10$$

$$R_1 = \frac{1}{5} \cdot \frac{516{,}660}{12{,}340} \cdot \frac{14}{2} \cdot \frac{670}{10} = 742$$

We take $R_1 = 850$ for the raising factor for barrios to take care of an increase of population.

(2) For urban areas, $R_2 = \dfrac{1}{5} \cdot \bar{P} \cdot \dfrac{\bar{N}}{\bar{n}}$,

where R_2 is the raising factor for chartered cities including provincial capitals or Metropolitan Manila

$\bar{P}$ is the average number of precincts per stratum

$\bar{N}$ in the average number of households per selected precinct

$\bar{n}$ is the average number of sample households to be enumerated per selected precinct.

Proportion of enumerated to listed households. All the households will be listed completely within each selected barrio, poblacion, and precinct. A specified proportion of the households listed in each case will be selected systematically for enumeration. These specified proportions were obtained from the formulae:

93

(3) For rural areas, $\dfrac{1}{F_1} = \dfrac{1}{5} \cdot \dfrac{P_s}{p_m} \cdot \dfrac{B_i}{b_i} \cdot \dfrac{N_{ij}}{n_{ij}}$,

where F_1 is the sampling fraction for barrios or poblaciones $= 1/R_1$

P_s is the population of the stratum

p_m is the population of the selected municipality

B_i is the number of barrios (poblacion) in the ith selected municipality

b_i is the number of sample barrios (poblacion) in the ith selected municipality

N_{ij} is the number of households listed in the ith selected municipality in the jth selected barrio (poblacion)

n_{ij} is the number of sample households to be enumerated in the ith selected municipality in the jth selected barrio (poblacion).

All are known in this formula except n_{ij} and N_{ij}. The ratios of n_{ij} to N_{ij} were computed for each barrio or poblacion.

Illustration 8: Computation of proportion of enumerated to listed households.

Stratum 25, population 513,625

Municipality of Dulag, population .. 23,460

Number of barrios 30

$$850 = \dfrac{1}{5} \cdot \dfrac{513,625}{23,460} \cdot \dfrac{30}{2} \cdot \dfrac{N}{n},$$

$$\therefore f = \dfrac{n}{N} = \dfrac{513,625 \times 30}{8,500 \times 23,460} = \dfrac{15,408,750}{199,410,000} = .077.$$

(4) For urban areas, $\dfrac{1}{F_2} = \dfrac{1}{5} \cdot P_s \cdot \dfrac{N_i}{n_i}$,

where F_2 is the sampling fraction for chartered cities including provincial capitals or Metropolitan Manila $= 1/R_2$

P_i is the number of precincts in the stratum

N_i is the number of households listed in the ith selected precinct

n_i is the number of sample households to be enumerated in the ith selected precinct.

All are known in this formula except n_i and N_i. The ratios of n_i to N_i were computed for each precinct.

ESTIMATION PROCEDURE

Stratum estimates. The hth stratum total X_h has five independent estimates, e.g., X'_{h_1}, X'_{h_2}, $\cdots$, X'_{h_5} where

(1) For rural areas, $X'_{hi} = \dfrac{P_h}{p_m} \cdot \dfrac{B_i}{b_i} \cdot \displaystyle\sum_{j=1}^{b_i} \cdot \dfrac{N_{ij}}{n_{ij}} \cdot \displaystyle\sum_{k=1}^{n_{ij}} X_{ijk}$

$$X_{ijk}, i = 1, 2, \cdots, 5$$

X_{ijk} is the characteristic of the kth household in the jth selected barrio (poblacion) in the ith selected municipality

N_{ij} is the number of households listed in the jth selected barrio (poblacion) in the ith selected municipality

n_{ij} is the number of households enumerated in the jth selected barrio (poblacion) in the ith selected municipality

B_i is the number of barrios in the ith selected municipality

b_i is the number of selected barrios in the ith selected municipality

P_h is the population of the hth stratum

p_m is the population of the ith selected municipality in the stratum.

where F_2 is the sampling fraction for chartered cities including provincial capitals or Metropolitan Manila $= 1/R_2$

P_i is the number of precincts in the stratum

N_i is the number of households listed in the ith selected precinct

n_i is the number of sample households to be enumerated in the ith selected precinct.

All are known in this formula except n_i and N_i. The ratios of n_i to N_i were computed for each precinct.

ESTIMATION PROCEDURE

Stratum estimates. The hth stratum total X_h has five independent estimates, e.g., $X'_{h_1}, X'_{h_2}, \cdots, X'_{h_5}$ where

(1) For rural areas, $X'_{hi} = \dfrac{P_h}{p_m} \cdot \dfrac{B_i}{b_i} \cdot \dfrac{b_i}{\sum\limits_{j=1}} \cdot \dfrac{N_{ij}}{n_{ij}} \cdot \sum\limits_{k=j}^{n_{ij}} X_{ijk}$

$$X_{ijk}, \; i = 1, 2, \cdots, 5$$

X_{ijk} is the characteristic of the kth household in the jth selected barrio (poblacion) in the ith selected municipality

N_{ij} is the number of households listed in the jth selected barrio (poblacion) in the ith selected municipality

n_{ij} is the number of households enumerated in the jth selected barrio (poblacion) in the ith selected municipality

B_i is the number of barrios in the ith selected municipality

b_i is the number of selected barrios in the ith selected municipality

P_h is the population of the hth stratum

p_m is the population of the ith selected municipality in the stratum.

(2) For urban areas, $X'_{hi} = P_h \cdot \dfrac{N_i}{n_i} \cdot \sum\limits_{j=i}^{n_i} X_{ij}; \; i = 1, 2, \cdots, 5$

X_{ij} is the characteristic of the jth household

N_i is the number of households listed in the ith selected precinct

n_i is the number of households enumerated in the ith selected precinct

$P_h B$ is the number of precincts in the hth stratum

The unbiased estimate of X_h is

(3) $\qquad\qquad X_h = \dfrac{1}{5} \sum\limits_{i=1}^{5} X'_{hi}$ Substituting the

values of X'_{hi} from (1) & (2) in (3) respectively, we obtain

(3') $\quad \bar{X}'_h = \dfrac{1}{5} \cdot \dfrac{P_h}{p_m} \cdot \sum\limits_{i=1}^{5} \dfrac{B_i}{b_i} \cdot \sum\limits_{j=1}^{b_i} \dfrac{N_{ij}}{n_{ij}} \sum\limits_{k=1}^{n_{ij}} X_{ijk}$ (for rural areas)

$\qquad = R_1 \sum\limits_{i=1}^{5} \sum\limits_{j=1}^{b_i} \sum\limits_{k=1}^{n_{ij}} X_{ijk} \; ; \; R_1 = $ raising factor

(3'') $\quad X'_h = \dfrac{1}{5} \cdot P_h \cdot \sum\limits_{i=1}^{5} \dfrac{N_i}{n_i} \sum\limits_{j=1}^{n_i} X_{ij}$ (for urban areas)

$\qquad = R_2 \sum\limits_{i=1}^{5} \sum\limits_{j=1}^{n_i} X_{ij}; \qquad R_2 = $ raising factor

The variance of the stratum estimate X'_h can be obtained simply as

$$\text{Var } X'_h = \dfrac{1}{5} \left\{ \sum\limits_{i=1}^{5} (X'_{hi} - X'_{hi})^2 \, / \, (5-1) \right\}$$

$$= \dfrac{\text{Var } X'_h}{5}$$

Country estimates. The total characteristic of the country like the stratum has also five independent estimates. The first independent estimate is obtained by taking from each stratum all first stage units having number one order of selection; the