CONTRIBUTIONS OF STATISTICIANS TO SOCIAL AND ECONOMIC DEVELOPMENT

By

Dr. BURTON T. OÑATE *

Summary

Some contributions of statisticians to the development of the nation are presented and examples are given to illustrate the role of statistics in research and the partnership which could exist between the policy maker (or researcher) and the statistician. These contributions are discussed with special emphasis on their bearing to the Philippine development. The *extent of these contributions to the development of the nation* will depend by and large on the prestige which is enjoyed by our statistical system. This prestige is contingent on the objectivity, integrity, and independence of the statisticians and the general recognition of these basic traits by the public.

Statistics may be considered as a science and as an art. It is a science from the fact that its methods and approaches are basically objective, systematic, and applicable. The success in its application, however, will depend upon the skill, background and experience and special knowledge of the field in which it is applied. From this point of view, statistics is an art. It is, therefore, important that the statistician know something about the field in which statistics will be applied. This knowledge will determine the extent of his participation in the development and application of statistics to the various research fields. The contributions of statisticians to the development of the nation may, therefore, take the form of developing new statistical concepts and theories and of applying these concepts and theories to relevant subject matter fields. The ultimate end in view is the pursuit of truth.

* Acting Director, Office of Statistical Coordination and Standards, National Economic Council.

By and large, these contributions may be summarized as follows: I. Production of Official Statistics, II. Development of Statistical Methods as Scientific Tools, and III. Provision for Bases of Decision-Making. These contributions will be discussed with special emphasis on their bearing to the Philippine development.

I. Production of Official Statistics

Statisticians make available in the form of official statistics a major portion, if not all, of the objective quantitative description and information on the resources of the nation, its people and their economic and social activities. These activities which are prerequisites to socio-economic development plans are also necessary for the proper evaluation of these plans.

In the production of official statistics, the statistical organization, the statistical manpower and its training, and the overall statistical program of the country should be considered. The urgent need for objective statistical information on the various stages of economic development is not well recognized even by our government economists, administrators, and planners. This situation is one of the problems which has to be met by the statistical system. Approaches to the solution of this particular problem are given in Section II.

In the production of official statistics, the Philippines in 1960 allocated a per capita amount of about ₱0.15 or approximately less than one-fifth of that spent for statistical work in the United States. The amount of ₱0.15 spent per capita for statistical work may also be compared with that approximately spent on a per capita level for the other fields of governmental activity, namely education (₱8.95), national defense (₱6.80), health (₱1.95), public works (₱1.10) and agriculture (₱1.50). Since reliable and up-to-date statistics are prerequisites to sound policy and decision in each of these fields of governmental endeavor, it becomes imperative that financial support be given for the development of sound statistical framework.

Statistical Organization — In the Philippines, the statistical organization is decentralized with a central coordinating office, the Office of Statistical Coordination and Standards (OSCAS) in the National Economic Council (NEC). The details of the organization and functions of each of the Statistical agencies, and the extent and complexity of the statistical organization, may be studied from the "Statistical Services of the Philippine Government" published in July 1957 by the OSCAS, NEC. An evaluation survey of all statistical agencies of the Philippine Government has just been concluded. The results of this survey may appear late this year in a revised edition of this OSCAS publication.

It may be seen in this publication that an addition to the coordinating office, there are presently four major statistical agencies, namely: the Bureau of Census and Statistics of the Department of Commerce and Industry, the Department of Economic Research of the Central Bank, the Agricultural Economics Division of the Department of Agriculture and Natural Resources, and the Division of Labor Statistics of the Department of Labor. The Social Security System may soon be added as the fifth major statistical office in the Philippine Government. The other statistical offices are either research organizations, like the Statistical Center of the University of the Philippines, or agencies which collect, compile, summarize, and publish statistics as by-products of their administrative responsibilities. Examples of these administrative offices are the Bureau of Internal Revenue and Securities Exchange Commission.

Statistical Manpower and Training — The composition and number of statistical manpower responsible for the production, compilation, summarization, and publication of objective official statistics may be obtained from the recent evaluation survey conducted by the OSCAS, NEC, on the statistical services of the Philippine Government. The total number of statistical personnel employed in the different bureaus, offices, and corporations of the national government, the Central Bank of the Philippines, excluding the University of the Philippines shows the following:

Statistical Personnel	FY 1959-60	FY 1960-61	FY 1961-62
Professional	564	613	783
Intermediate	770	2,499*	865
Primary	1,516	24,153*	2,003
Total	2,850	27,265	3,651

* Includes the census enumerators, supervisors and other personnel of the 1960 Census of Housing, Population, and Agriculture totalling 24,364.

Note that primary statistical workers like the field interviewer and other auxiliary clerical personnel may be employed for short periods during the fiscal year in question. Examples of these are those field interviewers in the Philippine Statistical Survey of Households (PSSH) who are employed for two-three weeks during the months of October and May. This short-term employment may be observed in the other current statistical surveys and in the decennial censuses.

In the training of professional and intermediate statistical personnel, the U.P. Statistical Center counts with 13 professionals, 9 intermediates and 4 primaries. This group is engaged mostly in teaching and research. The academic course offered by the Center leads to the degrees of Master of Arts (M.A.S.) and Master of Science in Statistics (M.S.S.). It has an in-service training program for statistical personnel of government and non-government agencies. In addition, the Center conducts special training for statistical personnel from other countries. As of November 1961, the U.P. Statistical Center has graduated 30 students, 22 with M.A.S. and 8 with M.S.S. It has conducted 15 in-service training sessions with 470 successful participants. There are enrolled 131 graduate students in the first semester of academic year 1961-62.

In addition, no less than 35 trainees from various government entities left the country to pursue further studies abroad under training programs and fellowships. Under the ICA-NEC Technical Assistance Grant, 19 statisticians have been sent to the United States to study and/or observe on industrial and economic statistics, census methods and statistical operations, statistical standards and organization, national income measurement, agricultural statistics, methods and techniques on the construction of price indices, labor statistics, etc. The Acting Director of OSCAS, NEC, completed his doctoral degree in statistics from the Iowa State University under a Rockefeller Foundation scholarship. No less than five faculty members of the U.P. Statistical Center were sent abroad on U.P. and U. N. fellowships for doctoral studies in statistics either in the United States or England. Other trainees have participated in the statistical training given by the International Statistical Education Centre in India.

The trained statistical personnel at the three levels for staffing regular statistical services of government and private offices was, therefore, built and developed. The number of statistically trained personnel is still far from satisfactory. At this stage, the training of statisticians at the undergraduate level should be started. This training may be initiated by private institutions of learning with high standards of instruction and the curriculum of instruction must be geared to the training of professional and intermediate levels of statisticians.

In an effort to develop the statistical profession, the Bureau of Civil Service, acting on recommendation of the OSCAS, gave the first civil service examination for statisticians and statistical aides in 1959. This was a significant step in the development of statistical career workers with security of tenure.

Statistical Program — Pursuant to one of its functions and responsibilities, the OSCAS, NEC, prepares annually an overall statistical program of the Philippine Government for the purpose of improving the precision, uses and comparability of official statistics and of providing new statistical series where gaps exist. This program is presented in one of the issues of The Statistical Reporter, a quarterly publication of the OSCAS, NEC.

II. Statistical Methods as Scientific Tools

Statistical methods are the scientific tools used in creating innovation such as obtaining new facts and deriving relationship among other existing facts. These innovations usually result in technological progress prerequisite to socio-economic development.

Statisticians and Applied Research Workers — Statisticians provide sound statistical methods which are necessary scientific tools for the optimum utilization of data by policy makers, government and private. These methods may be simple or elaborate. It is, therefore, important that there be familiarity with simple statistical methods. For example, most economists (themselves in short supply) do not posses highly developed skills in the use of statistical methods now available for social and economic investigation and analysis. Under this situation, statisticians, economists, engineers, other applied scientists, and administrators should contribute jointly and effectively to the greater and better use of statistics. This form of partnerships between the statistician and the applied research worker or administrator may provide the needed approach for the problem stated earlier in Section I. An example of this joint and effective effort on a government-wide basis is the organization by the OSCAS, NEC, of technical working committees composed of representatives from different government and private agencies. These are permanent standing committees which may be called upon for consultation on various problems relating to the production and use of statistics. These committees contribute vastly to the planning, coordination, standards and review of practically all phases of production and use of official statistics.

In the recent Food and Agriculture Organization (FAO) Seminar on the Analysis, Evaluation and Uses of Agricultural Census Results, held in Manila from December 5 to 20, in 1960 in which the Government of the Philippines acted as host, an attempt was made to provide a common platform for the exchange of views, ideas, experiences between the census takers and the economic planners so that the census takers may know what the planners need and that the planners may know what the census takers have to offer. From this point of view, this Seminar was considered unique, for it was the first attempt to bridge a partnership between producers and users of statistical information on an international level.

The present state of development of scientific machine computation was brought about through a strong partnership between statisticians, mathematicians, engineers and other applied research scientists. The electronic computer is presently the major tool in the efficient processing of data obtained from experiments, surveys, and other sources in many leading research centers all over the world. Programs are prepared, written and made available in the libraries of electronic computing centers. Some of us may be familiar with a few of these giant electronic computers such as the UNIVAC in the U.S. Bureau of the Census, the ILIAC in the University of Illinois, and the CYCLONE in the Iowa State University. The Philippines is now aware of the potentialities of these electric computers in the processing and analysis of data obtained from experimental and technological research. The potential and positive role of electronic computers in the development of a sound statistical system in the country is now under study.

Statisticians work out designs and analysis of scientific and technological experiments which are important phases of the role of statistics in research. Statistical procedure and experimental design are only two different aspects of the same whole and that the whole is the logical approach to the overall process of contributing to natural knowledge of experimentation.* The analysis of the result and the conclusions derived from these experiments are important to technological progress and, hence, to social and economic growth.

The fields of application and use of statistics are numerous and varied. Statistical methods are now used and applied in agronomic experimentation; genetics; plant breeding; animal breeding and animal nutrition, including research on vitamins; physics; many fields of engineering; industry; food technology research; marketing research; public health; epidemiology; biological essay; psychology; demography; economics education; home economics; hydrology; meteorology; and the arts. This is enough evidence of the contribution of statistics to the advancement of research.

* Fisher, R.A. The Design of Experiments. Edinburgh, Oliver & Boyd, 1937.

Experimental Designs — The objectives of the Commission in Agriculture of the first national agro-industrial conference sponsored by the National Science Development Board (NSDB) held on October 18-20, 1960 are directed toward higher yield of crops and increased production of livestocks. Very broadly, the variables which may be studied for increased crop production are varieties of hybrids; methods of cultivation; types and kinds of fertilizers; susceptibility to pests and diseases; physiological requirements; geographic conditions and other important factors such as institutions or agrarian structures.

Experimental designs and correlation studies with many ramifications have been applied to various types of agricultural research in which the variables under study were similar to those referred to earlier. It may be said that the statisticians have worked out the theory of experimental designs in such a pace that the applied research workers have lagged behind in applying the designs to their respective fields of study. These designs may be classified complete block and incomplete block designs. Under the complete block design are the completely randomized, randomized block, Latin square, crossover and Graeco-Latin square. There are two general types of incomplete block designs: the balanced and the partially balanced. The designs under the first category include balanced incomplete block, balanced lattice, incomplete Latin square, and lattice square; while under the latter category are the designs with high-order interactions confounded, quasi-Latin squares, and many other ramifications and modifications. Some related problems of experimental designs are those of size and shape of agricultural plot or its equivalent, the combining of experiments with perennial crops, testing large number of varieties or treatments, and general adjustment of data. An analytic survey was conducted a few years ago for the purpose of evaluating the existing conditions in PACD (Presidential Assistance on Community Development) areas and in non-PACD areas. This approach is really an extension of the concept of experimental design as applied to rather complex social situation.

Sampling Designs — Most of our observational data are obtained through properly designed sample surveys. Some of the more important current surveys in the Philippines are the Philippine Statistical Survey of Households (PSSH) and the Philippine Statistical Survey of Manufactures (PSSM) which are conducted by the Bureau of the Census and Statistics, Department of Commerce and Industry; the Crop and Livestock Survey (CLS) conducted by the Agricultural Economics Division, Department of Agriculture and Natural Resources; and the Nutrition Survey (NS) of the Food and Nutrition Research Center, National Science Development Board. These sample surveys provide current and reliable statistics on various components of the nation's economy. The development of these sample designs in the Philippines are discussed in detail in a paper by the author.*

It is very important to note that each of these experimental and sampling designs with their numerous ramifications, methods of analysis, and the conclusions which are derived therefrom, is based on sound theoretical and mathematical models. There are definite principles, conditions and assumptions which must be met in using any of these designs. If these requirements are met, then these designs may be used in any field of research, agriculture or otherwise. For example, the basic principles of experimental designs are randomization, replication and control of variability.

R. A. Fisher, father of the science of statistics, has defined the basic theoretical problems underlying modern applied statistics as those of specification, distribution, estimation and tests of hypothesis.** Each of these problems enters into the solution of the design of experiment. Thus, the sample data are specificied as having been drawn by some random scheme from a parent distribution (binomial, Poisson, normal, etc.) with unknown parameter or parameters. Functions of the sample data are used as estimators of these parameters and

* Oñate, B. T. Development of Multistage Designs for Statistical Surveys in the Philippines. Iowa State Statistical Laboratory Multi-lith Series No. 3, February 1960, Ames, Iowa, U.S.A. or The Statistical Reporter, OSCAS, NEC, October 1960.

**Anderson, R. L. and T. A. Bancroft. Statistical Theory in Research. McGraw Hill Book, 1952.

mathematical distributions of those estimators (statistics) are derived. These estimators must have some optimum properties which are unbiasedness, efficiency, consistency, and sufficiency. Appropriate test criteria are evolved in order that valid tests of hypothesis may be forwarded.

For example, in the process of steel-making, it is important for management to know the relationship between the percentage of iron in the form of pig iron (X) and the lime consumption in hundredweight (Y) per cast of steel. Let us illustrate with this setting the meaning of the problems of specification, distribution and estimation and tests of hypothesis.

Specification and distribution. — The sample pairs (X,Y) are specified as having been drawn by some random scheme from a parent bivariate normal distribution. These specification and distribution problems may be simplified if we specify that the Y's are drawn at random for a given X, and that the relationship between X and Y can be specified as $Y = A + BX$, where A and B are parameters or constants which are to be estimated by a and b, respectively.

Estimation and test of hypothesis — If we have a bivariate normal distribution, then we estimate independently the population mean of Y and X by

$\bar{y}$ and $\bar{x}$, respectively where

$$\bar{y} = \sum_{i}^{n} y_i/n, \text{ sample mean of the } Y_i\text{'s};$$

$$\bar{x} = \sum_{i}^{n} x_i/n, \text{ sample mean of the } X_i\text{'s};$$

and n is the number of sample pairs (X, Y).

Similarly, our estimate of A and B are

$$a = \bar{y} - b\bar{x} \text{ is the so-called y intercept,}$$

$$\text{and} \quad b = \frac{\Sigma x_i y_i}{\Sigma x_i^2} \quad \text{is the sample regression}$$

coefficient where x_i and y_i are deviations of the original observations from their sample means and Σ is the summation sign. The estimators $\bar{y}$, $\bar{x}$, a and b are least squares estimates and they possess the optimum properties mentioned earlier.

The test of the hypothesis may be as follows:
A = O or B = O. To test singly these hypothesis we use,

$$t = \frac{a - A}{s_a}$$

and $t = \frac{b - B}{s_b}$ where s_a and s_b are the sample

standards errors of a and b, respectively and A = O and B = O are the so-called "null" hypothesis. For example, in our test for B = O, we can't commit the so-called two types of errors, namely:

(1) error of the 1st kind, i.e. rejection of the hypothesis which in fact it is true (α level), and
(2) error of the 2nd kind, i.e. acceptance of the hypothesis when in fact it is false (β level).

The balance and consequences of these two types of errors serve as guides for both the statistician and policy maker. This simple linear regression technique is utilized in our projection of national income (Y) to 1965. The sample data assumed are for the years 1949 (X = O) to 1959 (X = 10). The predicting equation is Y = 5047.59 + 425.48 X where the values of a and b are in millions of pesos at 1955 prices. Estimates of national income and its rate of growth; estimates of investments (net or gross) and its rate of growth; capital-output ratios and other aggregate economic measures are used in economic development plans where lower and upper limits of desired targets are set. These limits are obtained through the help of these statistical tools.

Let us illustrate further these types of errors by some interesting examples. Pedro made a proposal of marriage to Maria. Maria can make two types of mistakes which are: (1) if she rejects the proposal of Pedro, when in fact Pedro will make the ideal husband, or (2) if she accepts Pedro, when in fact Pedro is not for her. The policy maker usually thinks along this line of reasoning which may also be compared to a judge who will decide whether a man is guilty or innocent. The judge will make the probability (α) of passing a guilty verdict when in fact the person is innocent as small as possible, as compared with the probability (β) of finding him innocent when in fact the person is not. An example involving administrative policy will be given in the next section.

In passing, we may also mention here the so-called third type of error which is committed by giving the right answer to the wrong problem.* This error usually arises because of inadequate communication between the statistician and the executive (policy maker or researcher). This error may be minimized by a closer partnership between these two groups.

III. Bases for Decision-Making

Statisticians provide the bases for decisions in many problems of administration, social and economic planning in government and management in industry. These include decisions in the control of quality of manufactured products, in management, and very broadly, in operational research under which fall the improvement of questionnaire, operation and survey designs of sample surveys and censuses. In some studies, the validity of results will depend on local social and cultural traditions and background.

The approaches to and benefits of quality control are found in various textbooks and articles. A recent article entitled "Quality Control of Hay** indicated the broad field into which this particular statistical method has been applied. A conference on quality control in industry was recently held under the auspices of the Philippine Statistical Association (PSA). The theme centered on how a well organized design on the collection, analysis, and interpretation of the data could lead to quality improvement of the product, maximum utilization of (scarce) material, the overall reduction in cost, and

* Kimball, A. W. Errors of the Third Kind in Statistical Consulting. Journal of American Statistical Association. 52: 133-142, 1957.

** Oliver, J.H. Quality Control of Hay. Read at the 1959 Annual Meeting of the American Society of Agricultural Engineers; New York. June 1959.

a general improvement in productivity. A number of papers dealing with the approaches to and benefits of quality control as they relate to local industry were read and discussed. The technical proceedings of this conference are now being prepared for publication.

One of the major contributions of statisticians is in the complex field of national decision-making wherein the statistician is increasingly being called upon to provide the basis of the decisions. The policy arrived at in each case will depend upon the official statistics produced.; the various statistical techniques used in the collection of these statistics and other statistical approaches and measurements relevant to other social and economic alternatives. An example will be given to illustrate an actual case.

Statistical Rules for Guides in Administrative Policy — Our Inter-agency Technical Working Committee on Rice and Corn Requirements meet periodically to assess existing and future indications on the rice and corn situation in the Philippines. Basically, the problem of the Committee consists in having estimate of available production, $\hat{P}$ (and its ramifications) with say, variance, Var $(\hat{P})$, and estimate of consumption $\hat{C}$ (and its components) with variance, Var $(\hat{C})$. These statistics will give an estimate of surplus or shortage $\hat{S} = \hat{P} - \hat{C}$ with variance, Var $(\hat{S})$ = Var $(\hat{P})$ + Var $(\hat{C})$, if there is no correlation between $\hat{P}$ and $\hat{C}$. The question which has to be aswered is: Is there a shortage or surplus of the cereal? This question may be paraphrased in terms of a hypothesis, namely: S is less than zero where S is the actual surplus, i.e., there is a shortage. The alternative hypothesis is that S is greater than or equal to zero, i.e., there is enough cereal. To test this hypothesis we may utilize the estimate S and its variance, Var $(\hat{S})$ or its estimate and use of one-tailed test. The decision are as follows:

Decisions			
Hypothesis S less than zero (Shortage)	Alternative S greater than or equal to zero (Surplus)	Types of Error	Probabilities
Reject	Accept	1st	α (smaller than β)
Accept	Reject	2nd	β

The magnitudes of α and β will depend, by and large, on the impact of committing either the first or second type of error on the socio-economic conditions of the general populace. The level of α and β may also depend upon the availability of other food crops which may be used as substitutes for the cereal in question. Another question will, therefore, arise: if and when a "real" shortage of cereal exists, are the people prepared to change their food habits for the duration? Other questions will, undoubtedly, arise such as the fine balance which must exist in order that both the consumer and the producer are equally protected. The author is strongly of the opinion that the rice problem should be viewed not the basis of the rice industry alone but in its relation to the overall Food Balance Sheet of the country.

This example again illustrates the form of partnership which could exist between the administrator and the statistician in arriving at important decisions.

IV. Conclusion

In all these undertakings, the statistician must (a) utilize **sound** and **objective** statistical methods; (b) maintain the **integrity** of his office; and (c) be **independent** of or from any pressure or influence. The prestige of our statistical system will depend upon the public's general recognition of these important requirements. The United Nations has recognized the importance of these three traits as necessary for the development of a sound statistical system. These traits should be the basic guide for the Philippine Statistical System.

ON THE SAMPLE DESIGN OF THE PHILIPPINE STATISTICAL SURVEY OF HOUSEHOLDS*

By Elpidio D. Makanas **

INTRODUCTION

The Philippine Statistical Survey of Households (PSSH) is a continuing survey aimed primarily to produce up-to-date statistical data on the size and composition of the population, manpower demand and supply, household income and expenditure and economic activities of the people. It was established as a NEC-ICA project in 1955 and the first round was conducted in May 1956 under direct supervision of PSSH staff and the Office of Statistical Coordination and Standards, NEC. In July 1958 the whole project was made into a permanent set-up of the government and was integrated with the Bureau of the Census and Statistics as the new Division of Surveys. At present the survey is conducted twice a year, one round taken around May and another around October, although plans are being considered to have it conducted at least three times a year.

The Sample Design

Excluding Metropolitan Manila (Region I), each of the remaining 9 regions (II to X) was divided into three subdivisions, namely: (1) barrios, (2) poblaciones and (3) chartered cities and provincial capitals (heretofore referred to as cities and capitals for brevity). Stratified multi-stage sampling was used in drawing the sample for each of these subdivisions. The sample design was the same for barrios and poblaciones and another design was applied to Metropolitan Manila and the cities and capitals.

Sampling for barrios and poblaciones was carried out in three stages with the following units: first-stage, municipalities; second-stage, barrios/poblaciones; and third-stage, households. Prior to selection, municipalities were arranged in order according to density of population as of 1948. Strata

* Master's thesis submitted to the Graduate School, U.P., April, 1961.

** Supervising Statistician, Division of Surveys, Bureau of Census and and Statistics.

were then formed such that the populations within strata were approximately equal. A total of 30 strata was formed over all the 9 regions. In each stratum, 5 municipalities were drawn for the first-stage with probabilities proportional to population and with replacement. At the second-stage, 2 barrios were selected with equal probabilities without replacement in the selected municipality, while the poblacion was automatically made the sample poblacion. At the third-stage, households were selected systematically from a list of households in each sample barrio/poblacion. The number of sample households was such that a self-weighting sample with a uniform overall sampling fraction for either barrios or poblaciones was obtained.

For Metropolitan Manila and the cities and capitals, sampling was carried out in two stages with the following units: first-stage, precincts; and second-stage, households. Before selection, the precincts were grouped into strata in such a manner that the strata formed within each region had approximately the same number of precincts. For Metropolitan Manila stratification was carried out by first arranging the precincts according to geographical location, and the strata were then formed by counting off a specified number of precincts. For the remaining urban areas, the cities and capitals were first arranged within each region according to degree of urbanization (percentage of urban precincts to total precincts) in descending order and the strata were then formed by counting off a specified number of precincts. A total of 32 strata was formed for Metropolitan Manila and 30 for the cities and capitals. Five precincts were drawn with equal probabilities and with replacement in each of these strata and households were then selected systematically after listing all the households in each sample precinct. The number of sample households was again so obtained as to yield a self-weighting sample with a uniform sampling fraction for either Metropolitan Manila or the cities and capitals.

ESTIMATES OF TOTALS AND THEIR VARIANCES

Notations

For barrios and poblaciones, we adopt the following notations:

P_h is the population of the hth stratum.

P_{hi} is the population of the ith municipality in the hth stratum.

M_h is the number of municipalities in the hth stratum.

m_h is the number of municipalities in the sample for the hth stratum.

N_{hi} is the number of barrios/poblaciones in the ith municipality.

n_{hi} is the number of barrios/poblaciones in the sample for the ith municipality

Q_{hij} is the number of households listed in the jth barrio/poblacion of the ith municipality;

q_{hij} is the number of households in the sample for the jth barrio/poblacion of the ith municipality.

X_{hijk} is the value of the characteristic for hijkth household in the population, $k = 1$ to Q_{hij}, $j = 1$ to N_{hi}, $i = 1$ to M_h.

x_{hijk} is the value of the characteristic for the hijkth household in the sample, $k = 1$ to q_{hij}, $j = 1$ to n_{hi}, $i = 1$ to m_h.

For Metropolitan Manila and the cities and capitals:

M_h is the number precincts in the hth stratum.

m_h is the number of precincts in the sample for the hth stratum.

N_{hi} is the number of households listed in the ith precinct.

n_{hi} is the number of households in the sample for the ith precinct.

X_{hij} is the value of the characteristic for the hijth household in the population, $j=1$ to N_{hi}, $i=1$ to M_h.

x_{hij} is the value of the characteristic for the hijth household in the sample, $j=1$ to n_{hi}, $i=1$ to m_h.

Estimates of Totals

Estimates of totals for the characteristic X are obtained separately for Metropolitan Manila and for each of the three subdivisions of the 9 regions, II to X. The method is such that the estimate of total for each stratum is first obtained and then aggregated over the strata.

i. **Stratum estimate.** For barrios and poblaciones, the estimate for the hth stratum total obtained from the *ith* municipality is

$$x'_{hi} = \frac{P_h}{P_{hi}} \frac{N_{hi}}{n_{hi}} \sum_{j}^{n_{hi}} \frac{Q_{hij}}{q_{hij}} \sum_{k}^{q_{hij}} x_{hijk} \ . \qquad (1)$$

But since the sampling fraction was made uniform for either barrios or poblaciones, that is,

$$f = f_1 \ f_2 \ f_3 = \frac{m_h \ P_{hi}}{P_h} \frac{n_{hi}}{N_{hi}} \frac{q_{hij}}{Q_{hij}} = \text{constant} = \frac{1}{R_1} \ , \qquad (2)$$

and noting that 5 municipalities were selected in each stratum ($m_h = 5$), Eq. 1 reduces simply to

$$x'_{hi} = 5R_2 \ x_{hi} \ , \quad i = 1, 2, \ldots, 5 \ , \qquad (3)$$

where x_{hi} is the total for all the sample households in the ith municipality. There will be 5 independent estimates since each of the 5 sample municipalities will give an independent estimate for the stratum, and the unbiased estimate is the mean of these 5 estimates, that is,

$$x'_h = \frac{1}{5} \sum_{i=1}^{5} x'_{hi} = R_1 \ x_h \ . \qquad (4)$$

For Metropolitan Manila and the cities and capitals, the estimate for the hth stratum total from the ith sample precinct is,

$$x'_{hi} = M_h \frac{N_{hi}}{n_{hi}} \sum_{j}^{n_{hi}} x_{hij} \ . \qquad (5)$$

The sampling fraction is also uniform, that is,

$$f = f_1 \ f_2 = \frac{m_h}{M_h} \frac{n_{hi}}{N_{hi}} = \text{constant} = \frac{1}{R_2} \ , \qquad (6)$$

and since 5 precincts were selected in each stratum ($m_h = 5$), Eq. 5 becomes simply

$$x'_{hi} = 5R_1\, x_{hi}, \quad i = 1, 2, \ldots, 5 . \tag{7}$$

where x_{hi} is the total value for all the sample households in the ith precinct. Again there will be 5 independent estimates and the unbiased estimate is

$$x'_h = \frac{1}{5} \sum_{i=1}^{5} x'_{hi} = R_2\, x_h . \tag{8}$$

ii. **Region and Country Estimates.** The total characteristic for a given region is estimated by aggregating stratum estimates over all the strata in the region. There will also be five independent estimates. The first is obtained by aggregating stratum estimates from all first-stage units having order of selection 1; the second, all first-stage units having order of selection 2; etc. Hence, the ith independent estimate for region r is

$$x'_{ri} = \sum_{h}^{L_r} x'_{hi}, \quad i = 1, 2, \ldots, 5 , \tag{9}$$

where L_r is the number of strata in the rth region and x'_{hi} is by either Eq. 3 (for barrios and poblaciones) or Eq. 7 (for Metropolitan Manila and the cities and capitals). The unbiased estimate is given by

$$x'_r = \frac{1}{5} \sum_{i=1}^{5} x'_{ri} . \tag{10}$$

The expressions for the estimates of total for the whole country will have similar forms as those regional estimates, except that the summation will now be over all the strata in the country, L, that is,

$$x'_i = \sum_{h}^{L} x'_{hi}, \quad i = 1, 2, \ldots, 5 , \tag{11}$$

and the unbiased estimate is simply

$$x' = \frac{1}{5} \sum_{i=1}^{5} x'_i \; . \qquad (12)$$

Estimates of Variance

The estimated variance of total estimates is computed solely from the variation between primary units, as if no subsampling were involved in the design, in accordance with the principle of cluster sampling. Here, the simplified method employed by the PSSH in computing estimated variances is compared with the standard method of estimating variances in stratified cluster sampling.

i. **PSSH Method.** The PSSH method of estimating the variance is based on the fact that for the stratum, region or country total, there are always 5 independent estimates and the variability of these 5 estimates gives an estimate of the variance. The unbiased estimates of the variance for the stratum, region and conutry totals are given by Eqs. 13, 14, and 15, respectively.

$$v(x'_h) = \frac{1}{5(5-1)} \sum_{i=1}^{5} (x'_{hi} - x'_h)^2 \; ; \qquad (13)$$

$$v(x'_r) = \frac{1}{5(5-1)} \sum_{i=1}^{5} (x'_{ri} - x'_r)^2 \; ; \qquad (14)$$

$$v(x') = \frac{1}{5(5-1)} \sum_{i=1}^{5} (x'_i - x')^2 \; . \qquad (15)$$

The above equations suggest that there are 5 psu's from which the variance is computed in each case. The psu however varies in meaning from one equation to another. In Eq. 13, the psu is the first-stage sampling unit, the municipality in the case of barrios or poblaciones and the precinct in Metropolitan Manila or the cities and capitals; in Eq. 14, it is now the aggregate of all the first-stage units having the same order of selection in each of the Lr strata; and in Eq. 15, it is the aggregate of all first-stage units which have the same order of selection in each of the L strata.

It is important to note, however, that since the selection of the psu's was made indendently in each stratum, the order of selection is immaterial and the set of 5 psu's for the region and country can be formed in a multiplicity of ways. Combining the units with the same order of selection is just one of the many possible combinations and the resulting set of 5 estimates of the total for the region or country is therefore not unique. Consequently, the computed variance from these 5 cluster totals is just one of the many possible estimates of the variance. The whole process therefore suggests a certain arbitraries.

ii. **Standard Method.** Following the procedure for stratified cluster sampling the estimated variance for the stratum total estimate will be the same as that by the PSSH method given in Eq. 13, but for the region or country estimates, the variance is now the sum of the stratum variance, Hence, for the region, the unbiased estimate of the variance of the total is

$$v(x_r') = \sum_h^{L_r} v(x_h') = \frac{1}{5(5-1)} \sum_h^{L_r} \sum_{i=1}^{5} (x_{hi}' - x_h')^2 \quad (16)$$

and for the country

$$v(x') = \sum_h^{L} v(x_h') \quad (17)$$

Eqs. 16 and 17 will each give a single unique estimate of the variance of the total unlike the method employed by the PSSH.

Comparison Between PSSH Method and Standard Method of Estimating Variance

The two methods of estimating variance are likely to give different values in any particular case. Nevertheless, both of them are unbiased estimates of the true variance and their expectations will therefore be the same. It can be shown, however, that the estimated variance obtained by the PSSH method has a larger standard error, and is therefore less precise, than that obtained by the standard method.

Let

$$s_h^2 = \frac{1}{m-1}\sum_i^m (x_{hi} - \bar{x}_{h.})^2 \quad \text{and} \quad E(s_h^2) = \sigma_h^2 \;;$$

$$s^2 = \frac{1}{m-1}\sum_i^m (x_{.i} - \bar{x}_{..})^2 \quad \text{and} \quad E(s^2) = \sigma^2 \;.$$

Now by the PSSH estimated variance of $\bar{x}_{..}$ is

$$v_P = s^2/m \;,$$

and by the standard method

$$v_S = \sum_h^m s_h^2 \Big/ m \;.$$

Assuming normal distribution, we have the following expressions for the variances of the above estimates:

$$V(v_P) = \frac{V(s^2)}{m^2} = \frac{2\sigma^4}{m^2(m-1)} \;,$$

$$V(v_S) = \frac{1}{m^2}\sum_h V(s_h^2) = \frac{2}{m^2(m-1)}\sum_h \sigma_h^4 \;. \tag{18}$$

But

$$\sigma^2 = \sum_h \sigma_h^2 \;,$$

and

$$\sigma^4 = \sum_h \sigma_h^4 + 2\sum_{hj} \sigma_h^2\,\sigma_j^2 \;.$$

The second term on the right side is positive being the sum of products of variances of pairs of strata; this will be quite large when there are several strata. Therefore,

$$\sigma^4 > \sum_h \sigma_h^4 \text{ ,}$$

and from Eq. 18, it follows that

$$V(v_p) > V(v_S) \text{ .}$$

This shows that the standard method gives a more precise estimate of the variance than the PSSH method. This of course can also be seen from the fact that the standard method is actually based on 4H degrees of freedom, where H is the number of strata, while the PSSH method is based on only 4 degrees of freedom. (Please see Appendix Table 1)

ESTIMATE OF COMPONENTS OF VARIANCE

Information on the components of variance at the different stages of sampling is very helpful in designing future sample surveys on the same universe. In practice, the population variance and its components for a given multi-stage sample design are generally not known. But in a continuing survey like the the PSSH, realistic estimates of variances of some important characteristics can be obtained, which, together with cost considerations, can serve as guide in improving or redesigning the sample in order to attain optimum efficiency.

Barrios and Poblaciones

The variance of the estimate x'_h, given in Eq. 4, is

$$V(x'_h) = \frac{1}{m_h} S_h^2 + \frac{1}{m_h} \sum_i^{M_h} \frac{1}{Z_{hi}} N_{hi}^2 \frac{N_{hi} - n_{hi}}{N_{hi}} \frac{S_{hi}^2}{n_{hi}}$$

$$+ \frac{1}{m_h} \sum_i^{M_h} \frac{1}{Z_{hi}} \frac{N_{hi}}{n_{hi}} \sum_j^{N_{hi}} Q_{hij}^2 \frac{Q_{hij} - q_{hij}}{Q_{hij}} \frac{S_{hij}^2}{q_{hij}} \qquad (19)$$

where Z_{hi} is the probability of selecting the first-stage unit in a single draw, i.e., P_{hi}/P_h, and S_h^2, S_{hi}^2, and S_{hij}^2 are, respectively ,the population variances amongst all municipalities, all the barrios/poblaciones in the ith municipality and all the households in th ejth barrio/poblacion in the ith municipality, all within the hth stratum. Following Hansen, Hurwitz and Madow (1953), consistent estimate of the third, second, and first-stage contributions to the variance, that is, the third second and first terms of Eq. 19, are given below.

$$v_3 = \frac{1}{m_h^2} \sum_{i}^{M_h} \frac{1}{Z_{hi}^2} \frac{N_{hi}^2}{n_{hi}^2} \sum_{j}^{r_{hi}} Q_{hij}^2 \frac{Q_{hij} - q_{hij}}{Q_{hij}} \frac{s_{hij}^2}{q_{hij}} , \qquad (20)$$

where s_{hij}^2 is the unbiased estimate of S_{hij}^2 and is given by

$$s_{hij}^2 = \frac{1}{q_{hij} - 1} \sum_{k}^{q_{hij}} (x_{hijk} - \bar{x}_{hij})^2 . \qquad (21)$$

$$v_2 = \frac{1}{m^2} \sum_{i}^{m_h} \frac{N_{hi}^2}{Z_{hi}^2} \frac{N_{hi} - n_{hi}}{N_{hi}} \frac{s_{hi}^2}{n_{hi}} ,$$

where

$$s_{hi}^2 = \hat{s}_{hi}^2 - \bar{s}_{hi}^2 ,$$

with

$$\hat{s}_{hi}^2 = \frac{1}{n_{hi} - 1} \sum_{j}^{n_{hi}} (x_{hij} - \bar{x}_{hi})^2$$

and

$$\bar{s}^2_{hi} = \frac{1}{n_{hi}} \sum_j^{n_{hi}} Q^2_{hij} \frac{Q_{hij} - q_{hij}}{Q_{hij}} \frac{s^2_{hij}}{q_{hij}} .$$

$$v_1 = s^2_h / m_h , \qquad (22)$$

where

$$s^2_h = \hat{s}^2_h - \bar{s}^2_h$$

with

$$\hat{s}^2_h = \frac{1}{m_h - 1} \sum_i^{m_h} \left[\frac{x'_{hi}}{Z_{hi}} - \frac{1}{m_h} \sum_i^{m_h} \frac{x'_{hi}}{Z_{hi}} \right]^2$$

and

$$\bar{s}^2_h = \frac{1}{m_h} \sum_i^{m_h} \frac{1}{Z^2_{hi}} \frac{N^2_{hi}}{n_{hi}} \left[\frac{N_{hi} - n_{hi}}{N_{hi}} s^2_{hi} + \frac{n_{hi}}{N_{hi}} \bar{s}^2_{hi} \right] .$$

Metropolitan Manila and the Cities and Capitals

The variance of the estimate x'_h (Eq. 8) for Manila and the cities and capitals in terms of the population variances S^2_h (amongst all precincts in the hth stratum) and S^2_{hi} (amongst all households in the ith precinct) is given by

$$V(x'_h) = \frac{M^2_h}{m_h} S^2_h + \frac{M_h}{m_h} \sum_i^{M_h} \frac{N^2_{hi}}{n_{hi}} \frac{N_{hi} - n_{hi}}{N_{hi}} S^2_{hi} . \qquad (23)$$

Consistent estimates of the within-precinct (second term of Eq. 23) and the between-precinct (first term) components of the variance are given in Eqs. 24 and 25 below.

$$v_2 = \frac{1}{m_h^2} \sum_i^{m_h} \frac{N_{hi}^2}{n_{hi}} \frac{N_{hi} - n_{hi}}{N_{hi}} s_{hi}^2 , \qquad (24)$$

where s_{hi}^2 is the unbiased estimate of S_{hi}^2.

$$v_1 = \frac{N_h^2}{m_h} \Big/ \frac{\sum_i^{m_h} (x'_{hi} - \bar{x}'_h)^2}{m_h - 1} = \frac{1}{m_h} \sum_i^{m_h} \frac{N_{hi}^2}{n_{hi}} \frac{N_{hi} - n_{hi}}{N_{hi}} s_{hi}^2 \Big/ . \qquad (25)$$

Using Eqs. 20 to 25, numerical computations for the estimates of the variance of total population estimates are made for Regions IV and IX and also for Metropolitan Manila (Region I). Please see Appendix Tables 2 and 3.

From said tables we note the following observations:

Region IV: i. The between-household variance contributions are smallest in the totals for each of the three subdivisions.

ii. For barrios, the largest contributions to the variance is from second-stage units (barrios) and represents about 70 percent of the total variance.

iii. For poblaciones and the cities and capitals, the contribution from the first-stage units (poblaciones or precincts) constitutes about 90 percent of the total variance.

Region IX: i. The last stage (between households) components are the smallest.

ii. The largest contributions in each of the three subdivisions are this time between the first-stage components.

iii. For barrios, the between-barrio components are smaller than the between-municipality components, but still are relatively much larger than the between-household components.

Metropolitan Manila: i. The total between-household components are smaller than the between-precinct components, which is true for almost all the strata except for a few cases in which the reverse is true.

ii. The between-precinct component represents more than 80 percent of the total variance.

It may be noted also that there are a few isolated cases of negative values of estimated first-stage components. This means that the sum of squares between units turns out to be smaller than the sum of squares within units. Cochran (1953) explains that this happens when the elements within the same unit are negatively correlated.

OPTIMUM ALLOCATION

The main purpose of studying the various components of the variance in multi-stage sampling is to have a basis for determining the number of units at the various stages that should be included in the sample in order to achieve maximum precision per unit cost. The estimated variance components together with cost considerations can now be used to achieve efficiency in the design.

Optimum Allocation for the Barrio Sample.

We shall consider the allocation of municipalities, barrios and households within a given stratum. We assume a constant number of sample barrios in all selected municipalities, i.e., $n_{hi} = n_h$, and let q_h denote the expected number of sample households per barrio. If the total cost for the hth stratum is

$$C_h = c_1 m_h + c_2 m_h n_h + c_3 m_h n_h q_h , \qquad (26)$$

then the optimum values of m_h, n_h and q_h subject to a fived total cost are:

$$q_h = \sqrt{\frac{c_2}{c_3} \, \frac{\sum_i \frac{N_{hi}}{Z_{hi}} \sum_j Q_{hij}^2 S_{hij}^2}{\sum_i \frac{N_{hi}^2 S_{hi}^2}{Z_{hi}} - \sum_i \frac{N_{hi}}{Z_{hi}} \sum_j Q_{hij} S_{hij}^2}} \; ; \qquad (27)$$

$$n_h = \sqrt{\frac{C_1}{C_2}\frac{\sum_i \frac{N_{hi}^2 S_{hi}^2}{Z_{hi}} - \sum_i \frac{N_{hi}}{Z_{hi}} \sum_j Q_{hij} S_{hij}^2}{S_h^2 - \sum_i \frac{N_{hi} S_{hi}^2}{Z_{hi}}}} \; ; \qquad (28)$$

$$m_h = \frac{C_h}{C_1 + C_2 n_h + C_3 n_h q_h} \, , \qquad (29)$$

where C_1 and C_2 are the unit costs per municipality, barrio and household, respectively. Note, however, that the optimal values for n_h and q_h (Eqs. 27 and 28) are given in terms of population values of the variance components. From Eqs. 20, 21 and 23, we obtain the optimal values of n_h and q_h in terms of estimated variances as follows:

where m' and n' denote the number of municipalities and number of barrios included in the original PSSH sample to distinguish them from the optimal values

$$q_h = \sqrt{\frac{C_2}{C_3}\frac{\frac{1}{n'}\sum_i^{m'} \frac{N_{hi}^2}{Z_{hi}^2} \sum_j^{n'} Q_{hij}^2 s_{hij}^2}{\sum_i^{m'} \frac{N_{hi}^2 s_{hi}^2}{Z_{hi}^2} - \frac{1}{n'}\sum_i^{m'} \frac{N_{hi}^2}{Z_{hi}^2} \sum_j^{n'} Q_{hij} s_{hij}^2}} \qquad (30)$$

$$n_h = \sqrt{\frac{C_1}{C_2}\frac{\sum_i^{m'} \frac{N_{hi}^2 s_{hi}^2}{Z_{hi}^2} - \frac{1}{n'}\sum_i^{m'} \frac{N_{hi}^2}{Z_{hi}^2} \sum_j^{n'} Q_{hij} s_{hij}^2}{\sum_i^{m'} \frac{N_{hi} s_{hi}^2}{Z_{hi}^2}}} \qquad (31)$$

Optimum Allocation for Metropolitan Manila

We take the number of sample households per precinct in the hth stratum to be equal to the average, i.e., $n_{hi} = n_h$' and assume the simple cost function

$$C_h = C_1 m_h + C_2 m_h n_h \tag{32}$$

where C_h is the total field cost, C_1 is the cost per precinct and C_2 is the cost per household. Then the optimum values of m_h and n_h subject to a fixed total cost are

$$n_h = \sqrt{\frac{C_1}{C_2} \frac{\sum_i N_{hi}^2 S_{hi}^2}{M_h S_h^2 - \sum_i N_{hi} S_{hi}^2}}\,, \tag{33}$$

$$m_h = \frac{C_h}{C_1 + C_2 n_h} \tag{34}$$

Denoting m' as the constant number of sample precincts per stratum in the original PSSH sample, the estimated optimum n_h is given by

$$n_h = \sqrt{\frac{C_1}{C_2} \frac{\sum_i^{m'} N_{hi}^2 s_{hi}^2}{m' s_h^2 - \sum_i^{m'} N_{hi}^2 s_{hi}^2}}\,. \tag{35}$$

Cost Estimates

Since the survey has a permanent staff which includes regional supervisors, only costs for wages and travel expenses of temporary field personnel were considered. Based on actual payments made during previous survey rounds, unit costs for Regions IV and IX and Region I (Metropolitan Manila) are given in the table below.

Table

Field Cost Estimates for the Various Phases of Field Operations, for Regions IV, IX and I.

Region and Unit	Total Unit Cost	Phase of Field Operations: Training	Listing	Enumeration
Region IV				
c_1 Municipality	₱42.50	₱35.00	——	₱ 7.50 a/
c_2 Barrio	12.50	——	₱11.00	1.50
c_3 Household	1.85	——	——	1.85
Region IX				
c_1 Municipality	87.50	60.00	——	27.50 a/
c_2 Barrio	15.25	——	11.50	3.75
c_3 Household	1.95	——	——	1.95
Region I				
Manila Proper				
c_1 Precinct	8.95	3.15	5.50	0.30 a/
c_2 Household	1.35	——	——	1.35
Suburbs				
c_1 Precinct	9.90	3.40	5.90	0.60 a/
c_2 Household	1.40	——	——	1.40

Optimal Sizes and Precision of Resulting Sample

The optimal sizes of the various units can now be obtained by substituting estimates of both cost and variances in the

a/ Cost of supervision by Chief Interviewer.

equations for optimum values. Wherever negative values of the components of variance occur, they are taken to be zero and the overall variance is correspondingly adjusted to the sum of the other components.

The precision of the new sample may also be estimated from the expressions of variances; ratios of the new variances to those of the original sample in terms of the corresponding sizes of the two samples can be derived. Let k_1 , k_2 and k_3 denote the ratios of the first-, second- and third-stage variance components, respectively, of the new sample to those of the original sample for barrios. We denote m', n' and q'_h as the units in the original sample, and m_h, n_h and q_h the corresponding units in the new sample. Then the estimated variance of the new sample is

$$v(x_h') = k_1 v_1' + k_2 v_2' + k_3 v_3' , \qquad (36)$$

where v'_1 , v'_2 and v'_3 are the first-, second- and third-stage components for the original sample and where, approximately,

$$k_1 = m' / m_h \qquad (37)$$

$$k_2 = m' n'(\bar{N}_h - n_h) / m_h n_h(\bar{N}_h - n') \qquad (38)$$

$$k_3 \quad m' n' q_h'(\bar{\bar{Q}}_h - q_h) / m_h n_h q_h(\bar{\bar{Q}}_h - q_h') . \qquad (39)$$

For Metropolitan Manila we have similar considerations except that there are only two stages in the sampling and therefore only k_1 and k_2 are involved for the two components of the variance.

Optimum samples together with their total costs and cv's for the three regions (I, IV and IX) are shown in Appendix Tables 4, 5 and 6.

Alternative Allocations: Self-weighting Samples

In a large operation such as the PSSH, it is desirable to have a self-weighting sample with a uniform overall sampling fraction in order to simplify processing and computation procedures. Further, it is desirable to have a uniform number of first-stage units for all strata and, in the case of barrios, a uniform number of second-stage units for all psu's. The problem therefore is to find alternative combinations of the units which will yield self-weighting samples with maximum precision.

By the use of cut-and-try methods, but guided by the optimal values already obtained, some alternative samples are obtained and together with their total costs and cv's are shown in Appendix Tables 7, 8 and 9. Note that for Region IX, alternative allocations 3 and 4 represent attempts to bring down the cv's in the neighborhood of 5% which however require considerably higher costs. Alternative 3 involves a cost almost 4 times that of the original and yields a cv of 5.03%, and alternative 4 which costs about 3.4 times that of the original yields a cv of 5.48%.

EVALUATION OF THE SAMPLE FOR POBLACIONES

Method of Selection of Sample Poblaciones

As earlier stated, the design for poblaciones is such that the poblacion was drawn automatically for each selected municipality, while the municipality was selected with probability proportional to its population size. This means that the poblacion was selected with probability proportional to the population of the municipality containing it. While it is true that, in general, there is positive correlation between the populations of municipalities and poblaciones, the degree of correlation may not justify the use of populations of municipalities as "sizes" in the PPS sampling of poblaciones.

Based on the population census of 1948, the degrees of correlation between the populations of poblaciones and municipalities for Regions IV and IX are 0.598 and 0.631, respectively. It is expected that for the other regions the correlation coefficient will be of the same order of magnitude, that is, of order 0.6.

Sampling with PPS Versus Simple Random Sampling

The expressions for the estimates of totals and their variances are based on the assumption that there is a one to one correlation between the populations of poblaciones and municipalities. The correlation being of the order 0.6, is there any gain (or loss) in sampling poblaciones with probabilities proportional to the size of municipalities over that of sampling with equal probabilities?

Two methods of approach are employed here to answer the question. In the first method, a relationship between the variance of a PPS sample and that of a comparable simple random sample is derived. The estimated variance of a simple random sample, $s^2_{(R)}$, in terms of qualities that can be obtained from a PPS sample is given by

$$s^2_{(R)} = \frac{1}{M_h}\left[\frac{1}{m_h}\sum_{i}^{m_h}\frac{x^2_{hi}}{Z_{hi}} - \frac{x'^2_h}{M_h} + \frac{s^2_{(P)x'}}{M_h}\right] \qquad (40)$$

where $s^2_{(P)x'}$ is the estimated variance for a PPS sample given by Eq. 13 with $m_h = 5$ as in the original PSSH sample.

In the other method, we take the original PPS sample and consider the psu's (poblaciones) as having been selected with equal probabilities, that is, $Z_{hi} = 1/M_h$. With this assumption we proceed to compute the estimates of totals and their variances by putting $1/M_h$ in place of Z_{hi} in the expressions for said estimates.

The results of computations by the two methods described above show convincing proof of the fact that a simple random sample design for poblaciones is superior to that in which the poblaciones are selected with probabilities proportional to the population of corresponding municipalities, as may be seen from the table below.

Table —. Coefficients of Variation of Population Estimates for Comparable PPS and Simple Random Samples, by Stratum, for Poblaciones of Regions IV and IX, March 1957.

Region and Stratum No.	Population Estimate	Coefficent of Variation (%)		
		PPS Sample	Simple Random Sample	
			Method 1	Method 2
Region IV	592,650	12.5	3.6	10.0
4	266,650	25.0	a	17.0
5	111,150	14.4	14.6	21.9
6	48,150	33.3	27.9	32.0
7	99,900	20.3	a	19.1
8	66,600	16.5	a	29.1
Region IX	161,550	24.3	17.9	16.8
27	90,900	24.8	13.3	23.1
28	70,650	45.4	37.1	30.9

a — Negative values obtained for estimated variance

PROBLEM OF A FIXED SAMPLE IN SUCCESSIVE SURVEYS

The Sample in Different Rounds of the PSSH

In the PSSH the same sample design has been employed since the first round in May 1956, except for the revision of the sample for barrios made in March 1957 in which barrio boundaries within sample municipalities were redefined but still following the same sampling scheme. Furthermore, until October 1960, practically the same sample of households had been used since the initial round although partial replacement of households was employed. The procedure for replacement was such that one-fourth of the sample was dropped and replaced by households dropped in the preceding round. This means that the sample had been confined to 5 groups of households, 4 of which were included in the sample in each round. A household was therefore included in the sample for 4 successive rounds, dropped in the 5th round, and drawn again in the 6th round for another 4 successive rounds, etc.

Sampling Theory In Succesive Surveys

In sample surveys conducted successively with the same design, there are three alternative methods of dealing with the sample:

i. A fixed sample, used throughout all surveys.

ii. A partial replacement of units from survey to survey.

iii. A new and independent sample for each survey.

Theory shows that a fixed sample (alternative i) is best for estimating the change in the characteristics from one point of time to another. If on the other hand, the need is to have an estimate of average (or total) based on all surveys over a period of time, then it is best to draw a new sample for each survey (alternative iii). If the main interest is to have current estimates, i.e., estimates only for the most recent occasion, then keeping the same sample or drawing a new one on each occasion will give the same degree of precision. In this last case, partial replacement of the sample (alternative ii) may prove to be more useful than the other alternatives.

The method of partial replacement of the sample has been studied by Yates (1949) and Patterson (1950) and the results of their work provide the basis for later studies and applications. Cochran (1953) suggests a practical working rule on partial replacement based on the optimum proportion of units to be replaced when the correlation between values for the same units in two successive rounds is sufficiently large, to wit: Retain a fraction one-third or one-fourth of the first sample for the second occasion; thereafter, retain one-half of the sample on each occasion and draw one-half anew. This is in agreement with the results of studies made by Patterson, and in general this method will yield an adequate sample for the three types of estimates (change, current and overall mean or total).

Advantages and Disadvantages of the Fixed Sample in the PSSH

Although there had been rotation of households in the PSSH, the sample may be considered as fixed since the same five sample groups of households had been used round after round, four at a time. As earlier stated, this type of sample is suitable for estimating the change of a characteristics from one round to another, but will not be as efficient in estimating the mean or total for a period comprising several rounds. Being a continuous survey, the need for estimates of averages over all completed rounds will inevitable arise, even if this need is not immediately felt.

There are also many administrative problems in using the same sample from one round to another. Respondents can be expected to give the same type of information for a few successive occasions, but there is always the danger of reaching that point beyond which the respondent will begin to show resistance, if not complete refusal. Some respondents react unfavorably to an investigator in the first round, thus creating a strained interviewer-respondent relationship, and continue to reach the same way through subsequent rounds. There is also the possibility that the respondent, after a few rounds becomes familiar with the questions and tries to give the same ready made answers round after round. On the part of the interviewer, there is the danger of supplying the answers himself on the basis of data obtained from the same respondent in the past rounds, thus making the data collected less representative.

Nevertheless, there are some advantages in keeping the same sample throughout. First, there is the question of cost in drawing a new sample in each round. Second, the field cost of travel of enumerators and supervisors will be less as they could be stationed permanently in the same areas. Third, their knowledge of the sample units will enable speedier operations than if they have to operate in new areas each round.

Some of the administrative advantages can be retained and the interviewer-respondent interaction effects can be minimized if a proper partial replacement of the sample is effected. At any rate, the rotation of households should be such that all the households in the sample areas have a chance of being included in the sample.

Adjustment Made on the Barrio Sample

A change in the barrio sample was made in March 1957 in an effort to reduce sampling variability of regional estimates. This adjustment indeed resulted in large gains in the precision of estimates, but at the same time drew criticism on the manner it was applied.

The adjustment was made by redefining the barrio boundaries within each sample municipality such that the resulting units were approximately equal in size of population. This process involved combining of small barrios to adjoining ones and splitting extremely large barrios into two or more units each. The new unit thus formed was either a part of a barrio, a whole barrio, or a group of barrios. The units in which the original sample fell were retained, and in case the two original sample barrios were combined into one unit, another unit was selected at random from among the newly created units.

Based only on the number of redefined units new probabilities were then attached to these units entirely disregarding the initial probabilities with which the original units were drawn. It is here that the question is raised. It is pointed out that disregarding the old probabilities would be valid if a fresh sample of barrios were drawn after redefining the units. But since the original sample barrios were retained, their initial probabilities should be taken into account and reflected somehow in the new probabilities.

It is important to note, however, that the redefined units are approximately equal in size and any set of units drawn would therefore yield essentially the same precision of population estimates. Bringing back the old probabilities into the new setup would only tend to destroy the equality of size already achieved and thus perpetuate the major source of variability. Perhaps the point that really in question is the bias that must have been introduced in the automatic selection of the units containing or contained in the old sample barrios. Efforts however were made to minimize, if not altogether eliminate, this bias by assigning the work of redefining the units to persons who did not have any knowledge about the identities of the original sample barrios. This in a way can be considered to give some degree of randomness in the selection of the retained portion of the sample.

CONCLUSION

This paper was prepared in an attempt to analyze and evaluate as many of the sampling techniques currently used by the PSSH as can possibly be made from available information. The study of variances has revealed the relative variability amongst municipalities, barrios and rural households and also amongst precincts and urban households, a knowledge so vital in improving the sample design or planning a new survey. With the present sampling scheme of the PSSH, a near optimum result for estimating population can be attained by reducing the number of sample municipalities from 5 to 3 or 4 and increasing the sample barrios from 2 to 4 or 5 in Region IV, while in Region IX the number of municipalities can remain at 5 while the barrios are increased to 3. For Metropolitan Manila, the optimum number of precincts remains at 5 for Manila Proper but 6 for the Suburbs. Although only these three regions (1, IV and IX) are studied for this purpose, the results can very well serve as guide in the sample design for the other regions by comparing conditions observed from previous rounds. Furthermore, since the other socio-economic characteristics are well related to population, the allocations arrived at on the basis of population estimates will also yield efficient result for the other characteristics measured in the survey. A similar study was made based on household income and expenditure estimates and the results (not presented here) showed similar qualities as those for population.

The limitations of the method employed by the PSSH in the rotation of households have been partly diminished by dropping one-fourth of the households and replacing with fresh ones, a procedure which was started in October 1960 and followed subsequently thereafter. Nevertheless, the possibility of drawing a fresh sample of households each time and that of selecting a fresh sample of barrios while keeping the municipalities fixed need to be explored.

The analysis made on the sample for poblaciones definitely established the fact that a simple random sample of poblaciones will yield more efficient results than a pps sample based on the population of municipalities. While it is true that selecting the poblaciones of sample muncipalities provides operational convenience and costs savings, these gains should be weighed against the significant loss in the precision of results.

In addition to the problems of design considered in this paper, there is the problem of stratification which needs examination. Questions like whether there is any gain in stratifying municipalities by density of population and whether the provincial capitals by degree of urbanization of precincts is better than that by population require further detailed study.