The Philippine Statistician
Vol. xxiv, Nos. 1 & 2, 1975

IN SEARCH OF OPTIMUM STRATEGIES FOR PHILIPPINE HOUSEHOLD SURVEYS — I. SAMPLING AND ITS LIMITATIONS

by I. P. David[1]

O. Foreword

Many of our current and continuing sample surveys, notably the National Census and Statistics Office's quarterly Survey of Households and the Bureau of Agricultural Economics' quarterly Integrated Agricultural Survey, have been designed to yield "good" statistics at the provincial level. However, only official statistics at the regional and national levels are released because of the "high variability" of provincial estimates, which means that the latter can fluctuate greatly from one survey round to another. To the statistical novice, and perhaps even to the not-so-novice, this state of affairs often seems inexplicable if not downright inexcusable. For what could go wrong in an apparently simple process of sampling at random of towns, barrios and households, sending out interviewers to fill out questionnaires, and averaging of results?

This article, which is the first of a series, aims to explain from a statistical standpoint why and how survey methodology and its attendant problems especially in the Philippines are not as simple as they may appear to many researchers. The entire series is geared towards searching for an optimum strategy or strategies for socio-economic surveys (with the exclusion initially of agricultural surveys because optimum sampling strategies for these can be very different). By sampling strategy we mean not just a sampling scheme but a combination of both sampling and estimation procedures. Our definition of an optimum sampling strategy is not the very formal one, but that which is limited to the context of the Philippine

[1] University of the Philippines at Los Banos. This article is a revised version of a Professorial Chair in Statistics inaugural lecture paper read on September 3, 1975, UPLB Campus. I wish to thank Miss T. A. Oliveros and Prof. S. M. Alviar of the UPLB Statistical Laboratory and Computing Center for their help in the computations and programming.

situation — namely, one that will yield at minimum cost provincial statistics with acceptable level of error. Just what is an acceptable level of error will be dealth with later. Also, the style of writing is slanted towards practicing survey statisticians and research workers who either run their own surveys or use secondary data from other surveys. This explains the absence of formal proofs of many statements and formulas.

1. Introduction

Simple random sampling (SRS) of n from N population units is an easy task of labelling the units from 1 to N and drawing n < N of these (without replacement) either by lottery or by a table of random numbers. This is also called equal probability sampling without replacement. If a unit drawn is replaced first before another is chosen, the method is referred to as equal probability sampling with replacement.

This very simplicity of SRS has made it the much used tool in social science surveys, often with very little regard to its statistical limitations and implications. This is quite contrary to the statistical opinion that in sampling from finite populations, equal probability sampling is not always a Good Thing. In fact, SRS alone probably should be used only when the amount of prior information about the target population will not allow the use of one of the more efficient sampling methods. We shall pursue this conjecture empirically with a minimum of theoretical regor. First, however, we lay down in the next section the basic ideas and language required in assessing goodness of results from surveys. In section three we use the 1970 Population Census figures to illustrate that SRS alone will seldom be satisfactory.[2] Section four presents some results when SRS is used along with some variance-reducing techniques.

2. When is a Sample Survey a Success?

There are two types of surveys, namely descriptive and analytical, or absolute and comparative. The first type is exemplified by the Survey of Households and the Integrated Agricultural Survey mentioned earlier, wherein the primary purpose is to produce estimates of parameters of the sampled population. The second type, which is usually but not necessarily smaller in scope, is used primarily for making analytical inferences about the sampled population, e.g. model building, contingency analysis and testing significance of differences of

2 1970 Census of Population and Housing, National Census and Statistics Office, Manila.

group means. This dichotomy of surveys however, is not mutually exclusive, for a survey can be both descriptive and analytical. Also, data and condensed results from mainly descriptive surveys often serve as secondary data for analytical studies, and vice versa.

The discussion here will focus on descriptive surveys although many of the points that come to light also apply to analytical surveys. Nevertheless, there exist important differences between the approaches to the designing, analysis and evaluation of these two surveys. Our time and space constraints here, however, would not allow an adequate treatment of both types.

In general, the success of a survey should be reckoned in terms of its objectives vis-a-vis output. In particular, a survey can be judged on the basis of (a) timelines of release of results, (b) cost and (c) accuracy of results. The factors that influence (a) are largely nonstatistical.[3] The total cost of a survey is a sum of overhead and actual survey (including analysis) costs. The latter component is affected by the choice of sampling strategy via sample size, sampling frame construction, sampling procedure and complexity of the analysis. For a given problem requiring survey data with a predetermined level of accuracy, there exist many alternative sampling strategies and one or a few of these would involve minimum cost. The fact remains, however, that after censuses, sample surveys are the next most expensive ways of collecting data.

Without belaboring (a) and (b) any further, we now devote the rest of the section on the elucidation of the concept and measurement of statistical accuracy.

The accuracy of an estimate is inversely related to its distance from the parameter being estimated. There are statistical formulas for measuring this accuracy, some of which we shall discuss presently. It should be noted beforehand, however, that it is not possible in practice to determine the exact degree of accuracy of an estimate since the true value of the parameter is not known; i.e. the measure of accuracy is also an estimate itself. In the end, there are no universal rules upon

3 The prerequisites for the timely release of statistics from a large survey include adequate preparation (training of supervisory, field and support personnel, construction and pretesting of questionnaries), a well-oiled machinery for the expedient collection and editing of just the right amount of data and (most important nowadays) availability of an efficient, operational computer-based data management system. These things, however, are easier said than done, for one need not search hard to find cases of surveys and censuses the results from which do not become available until after four or five years, if ever.

which we can all agree wheter an estimate is sufficiently accurate or not, this being relative to individual value judgment and the purpose for which the estimate will be used.

In statistics we prefer to think in terms of average accuracy. A classroom illustration for this is figure 1 which depicts the performance of our target shooters given ten shots apiece.

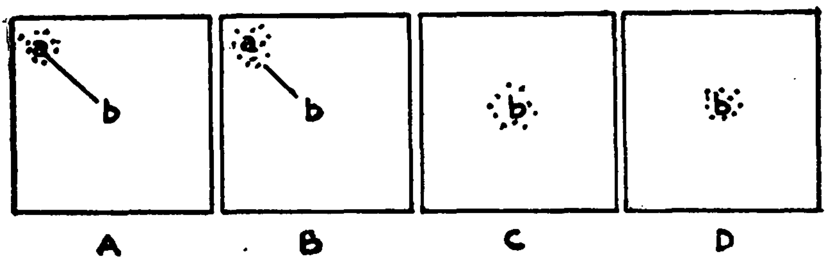

Figure 1

Shooter A is a marksman but unfortunately his rifle has faulty sights. The focal point, a, of his shots is his "average" or "expected hit" and the distance between a and the bull's-eye b is his "average error" or bias. It appears that A is very precise (his shots being close to each other) but inaccurate since he is off the mark by a considerable margin. Had he been forewarned of the rifle's defect, however, he could have been very accurate. Shooter B is definitely not handy with a gun as he scatters his shots; moreover, his gun is not good either for his expected hit is quite a distance from the bull's eye. B's shots therefore are imprecise and biased as well, hence he is inaccurate to a larger extent than A. Figure 1C depicts a case of a good gun in the wrong hands. Although C's expected hit is the bull's-eye, his shots, though unbiased, are too dispersed; thus he too is inaccurate. Figure 1D shows a marksman wielding a very good rifle. The expected hit of D is b and his chance of hitting it with one shot is highest; he is the most accurate of the four.

Accuracy is therefore a function of the degree of precision and the magnitude of the bias. In the absence of bias, accuracy is synonymous with precision; alternatively, the term efficiency is used in this case.

Let us now translate the ideas from this simple illustration into statistical, quantifiable terms. Again, we use a simple, albeit instructive example.

Consider drawing a SRS from the population {1, 2, 3, 4, 5}. The number of possible samples of size three from this population is 10. Likewise there are a number of formulas (estimators) available for estimating the population mean $\mu = 3$; e.g. y_{mean} or r where

y_{mean} = sample mean = (sum of sample values) /3
and r = midrange = (highest sample value - lowest sample value)/2

(The mean μ is to the bull's-eye in our previous illustration, the samples to bullets, and the estimators to the gun-man). Of course, in practice we normally choose a single sample only, along with one with one estimator. Hence, as shown in table 1, we obtain as an estimate of μ, any one of twenty possible values depending on the choice of sample and estimator.

Table 1. Samples and corresponding estimates.

Sample	y_{mean}	r
{1, 2, 3}	2	1
{1, 2, 4}	2 1/3	1 1/2
{1, 2, 5}	2 2/3	2
{1, 3, 4}	2 2/3	1 1/2
{1, 3, 5}	3	2
{1, 4, 5}	3 1/3	2
{2, 3, 4}	3	1
{2, 3, 5}	3 1/3	1 1/2
{2, 4, 5}	3 2/3	1 1/2
{3, 4, 5}	4	1
Total	30	15
Expected Value	3	1 1/2

The expected value of y_{mean} denoted by E (y_{mean}), is the average of its ten values. Since E (y_{mean}) = 3 = μ, y_{mean} is an unbiased estimator of μ. On the other hand, E(r) = 1.5 < μ; hence r is (negatively) biased and its use will (on the average) lead to understating μ. The bias in r is given by

$$\text{Bias (r)} = \text{E (r)} - \mu = -1.5. \quad (1)$$

The scatter plots of the values of y_{mean} and r are shown in figures 2. The unbiasedness property of y_{mean} and the biased nature of r are clear from this figure.

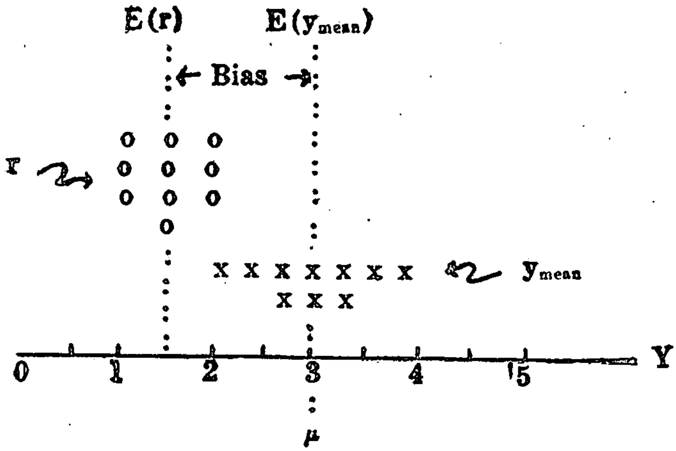

Figure 2. Scatter diagram of y_{mean} and r value.

Thus, on the criterion of unbiasedness, y_{mean} is preferred to r. Unbiasedness, however, is not the only basis for choosing between estimators. (This is the same as saying that biased estimators should not be rejected altogether). A second look at figure 2 will show another important distinction between y_{mean} and r that should not be ignored, namely that the estimates are dispersed about their respective expected values in varying degrees. The values of r are bunched more closely about $E(r) = 1.5$ than those of y_{mean} about $E(y_{mean}) = 3$. Hence r is more precise in the sense that a single SRS will tend to yield a r closer to E (r) than y_{mean} is to E (y_{mean}). We now see how this precision (and conversely, dispersion or variation) is measured quantitatively.

As is well known, one of the most useful measures of variation is the variance which is defined as the expected value of the squared deviations of observations about their mean; e.g.

$$V(y_{mean}) = (1/10) \{ (2\text{-}3)^2 + (2\ 1/2\text{-}3)^2 + \ldots + (4.3^2) \} = 1/3$$

$$V(r) = (1/10) \{ (1\text{-}1.5)^2 + (1.5\text{-}1.5)^2 + \ldots + (1\text{-}1.5)^2 \} = 3/20$$

As expected, $V(r) < V(y_{mean})$. Hence r is more precise than y_{mean}, but then the former is biased whereas the latter is not. A solution to this dilemma is the mean square error (MSE), which combines both variance and bias, as a measure of accuracy. The

MSE of an estimator, say T, is the expected value of the squared deviations of the values of T from the target parameter. That is, if T is intended to estimate μ, then

$$MSE(T) = E\{T - \mu\}^2 \quad (2)$$

If E(T) denotes the expected value of T, we can write

$$MSE(T) = E\{T-E(T) + E(T) - \mu\}^2$$
$$= E\{T-E(T)\}^2 + \{E(T) - \mu\}^2$$
$$= V(T) + \{Bias(T)\}^2 \quad (3)$$

so that MSE (T) = V(T) if and only if T is unbiased. For y_{mean} and r, we get

$$(MSE(y_{mean}) = V(y_{mean}) = 1/3$$

and

$$MSE(r) = (3/20) + \{-1.5\}^2 = 2.40$$

Hence y_{mean} is to be preferred to r.

There exist numerous practical situations where biased estimators have smaller MSE's than unbiased ones, hence the former are used especially if there is prior evidence to the effect that the bias is neglgible or if the bias can be estimated from the same sample.

The variance and MSE have useful derivates; e.g. the standard deviation (σ) and the root mean square error (RMSE) which have the same unit of measure as the original variable. Another is the coefficient of variation (CV) which is defined as the ratio of the standard deviation to the mean. In the preceding example,

$$CV(y_{mean}) = \frac{\sigma(y_{mean})}{E(y_{mean})} = \frac{\sqrt{1/3}}{3} = 0.19 \quad (4)$$

or 19 percent. Note however that the CV is independent of the bias, so that the CV of a biased estimator may not be a good indicator of its accuracy (unless the bias is negligible). Note further that the CV is free of any unit of measure.[4]

[4] Also, for the same characteristic or variable, the CV of the estimate of the mean is the same as that of the total. Hence we can talk interchangeably about the precision of the mean and the total. For biased estimators, an analogue of the CV is the ratio of the RMSE to the target parameter.

In a normal population with mean μ and standard deviation σ, the interval $(\mu - \sigma, \mu + \sigma)$ contains 68 percent of the population while $(\mu - 2\sigma, \mu + 2\sigma)$ covers a little more than 95 percent. Thus the CV has a simple, direct meaning; e.g. a $CV(y_{mean}) = 0.10$ implies a 0.68 chance that the mean from a randomly chosen sample will fall in the interval $(0.9\mu, 1.1\mu)$, and a little more than 0.95 likelihood that it would fall in the interval $(0.8\mu, 1.2\mu)$. This straightforward interpretation of the CV makes it a valuable measure of variation as well as a standard for setting the level of precision for surveys. It also enables one to find a rough but quick answer to the eternal question of sample size for surveys. In a large population with $CV(Y_{pop}) = \sigma/\mu$, for example, the mean y_{mean} from a SRS of size n is

$$CV(y_{mean}) = (\sigma/\sqrt{n}) / \mu = CV(Y_{pop}) / \sqrt{n}, \text{ so that}$$

$$\sqrt{n} = CV(Y_{pop}) / CV(y_{mean}). \tag{5}$$

If a reliable estimate of $CV(y_{mean})$, say $CV(y)$, is available from past data[5] and if the desired level of precision for y_{mean} is such that its CV should not execeed a predetermined limit, say d, then the sample size should satisfy

$$n \geq \{ CV(y) / d \}^2 \tag{6}$$

A point that is often overlooked is that, in surveys employing stratification wherein independent estimates are to be reported for each stratum, the specification of precision and hence of sample size should be at the stratum level. Thus, for a Philippine survey from which reliable statistics for provinces are desired, the specification of precision and sample size should be done at the provincial, not at the regional nor national level; for ultimately, we pass judgment upon the success of such survey on the basis of the accuracy of its provincial statistics, regardless of the fact that the CV of the national estimate is, say two percent.

3. SRS is not a Panacea

Far from being a panacea, SRS, on the contrary, should be used only in conjunction with other more efficient sampling schemes and/or estimation procedures. The basic reason behind

5 A desirable property of the CV is that, while the standard deviation of a population usually increases as the mean rises (with time), the CV often remains more or less the same. This statement is certainly true with town population counts (see table 10).

this is that in SRS of n from N poulation units the variance of the sample mean y_{mean} is [6]

$$V(y_{mean}) = \frac{N-n}{N-1} \frac{\sigma^2}{n} \qquad (7)$$

which reduces to the more familiar form

$$V(y_{mean}) = \frac{\sigma^2}{n} \qquad (8)$$

when N is much larger than n, where σ^2 is the variance of the individual units in the population. Thus we see that SRS leaves σ^2 untouched and reduces $V(y_{mean})$ only by an increase in n (which means increased costs at the same time). Unfortunately, most socio-economic (and also agro-economic) characteristics are inherently very variable that an unreasonably large SRS would be necessary to reduce the variance of estimates to an acceptable level. We illustrate this by considering the problem of estimating provincial population counts. We choose this particular example for three reasons. (a) The Population Censuses of 1960 and 1970 provide adequate data for an extensive empirical study. (b) The problem is not academic; in fact the quarterly Survey of Households of the NCSO mentioned previously is for this purpose. (c) Population count serves as an adequate design variable for general-purpôse socio-economic surveys because it correlates highly with many variables arising from human economic activities such as housing, income, unemployment, and other labor force characteristics.

Let us set as our goal the production of provincial estimates wth CV's not exceeding ten percent given a uniform sampling design for all provinces. This means that the population estimate of a province of 0.5 million will be approximately between 0.4 and 0.6 million with probability .95. (We say approximately for here we are dealing not with a normal but with a discrete skewed population). This goal may not appear ambitious but it represents a marked improvement over the present situation. To see how this translates to the regional estimates y_r and country estimate y_c, let y_{rp} be the estimate of the total population of the p-th province in the r-th region, $r = 2, 3, \ldots, 10$ (excluding region 1, Greater Manila), $p = 1, 2, \ldots P_r$.

6 See e.g., Cochron, W.G. (1963). Sampling Techniques. John Wiley.

Then $y_r = \Sigma y_{rp}$ and $y_c = \Sigma_r \Sigma_p y_{rp}$, so that (with independent sampling for the provinces)

$$V(y_r) = \Sigma_p V(y_{rp}) \tag{9}$$

and

$$V(y_c) = \Sigma_r \Sigma_p V(y_{rp}) \tag{10}$$

If we set $CV(y_{rp}) \ \{V(y_{rp})\}^{1/2}/y_{rp} = d$ for all provinces, where Y_{rp} is the total population of the p-th province in the r-th region, we obtain $V(y_r) = d^2 \underset{p}{\Sigma} y^2_{rp}$ and $V(y_c) = d^2 \underset{r}{\Sigma} \underset{p}{\Sigma} y^2_{rp}$

Hence

$$CV(y_r) = \sqrt{V(y_r)}/y_r = d\sqrt{\underset{p}{\Sigma} y^2_{rp}}\ /y_r \tag{11}$$

and

$$CV(y_c) = \sqrt{V(y_c)}\ \ y_c = d\ \sqrt{\underset{r}{\Sigma}\ \underset{p}{\Sigma} y^2_{rp}}\ /y_c \tag{12}$$

where Y_r and Y_c are regional and country populations, respectively. Using the 1970 Population Census results, the values of equations (11) and (12) are shown in table 2.

Table 2. CV (percent) of estimators of 1970 regional and country population.

Region	$d = 5\%$	$d = 10\%$	$d = 15\%$
2	2.0	3.9	5.8
3	3.1	6.2	9.3
4	2.1	4.2	6.3
5	2.4	4.8	7.2
6	2.3	4.6	6.9
7	2.6	5.8	8.7
8	2.1	4.2	6.3
9	1.7	3.4	5.1
10	2.0	4.0	6.0
Country (excluding Greater Manila)	0.8	1.6	2.4

With provincial CV's set at ten percent the regional CV's are within six percent and the country estimate has a CV below two percent.

From the natural state of administrative and geographic affairs, it seems logical that we use as sampling units existing ones such as towns, barrios and households .(properly defined of course). Hence the sampling design most likely will be multistage with the households as ultimate sampling units and either the towns or barrios as first-stage units.

With multi-stage designs in general, the variance of an estimator is a sum of variance components from the different stages. For instance, let y be an estimator of provincial population. In a two-stage design with barrios and households as first-stage units (fsu) and second-stage units (ssu), respectively, and with SRS at both stages, the variance of y is given by [7]

$$H^{-2}V(y) = \left(\frac{1}{b} - \frac{1}{B}\right)S_B^2 + \frac{1}{bB}\sum_j^B u_j^2\left(\frac{1}{h_j} - \frac{1}{u_j}\right)S_j^2 \qquad (13)$$

where H = number of households in the province, B = number of barrios in the province, b = number of sample barrios, S^2b =weighted) between barrios variance of household population, H_j = number of household in the j-th barrio, h_j = number of sample households in the j-th sample barrio, S_j^2 = between household variance in the j-th barrio, $u_i = H_j/\overline{H}$, and $\overline{H}$ = average number of households per barrio. In a three-stage design with towns, barrios and households as fsus, ssus and tertiary sampling units, respectively, $V(y)$ will have three components; i.e.

$$H^{-2}\,V(y) = \left(\frac{1}{t} - \frac{1}{T}\right)S_T^2 + \frac{1}{tT}\sum_i^T u_i^2\left(\frac{1}{b_i} - \frac{1}{B_i}\right)S_i^2 + \frac{1}{tT}\sum_i^T \frac{u_i^2}{b_iB_i}\sum_j^{H_i} v_{ij}^2\left(\frac{1}{h_{ij}} - \frac{1}{H_{ij}}\right)S_{ij}^2 \qquad (14)$$

where T = number of towns, t = number of sample towns, $S^2{}_t$ = (weighted) between towns variance of household population, S_i^2 = weighted between barrios variance of household population

[7] See e.g. Sukhatme. P.V. and B.V. Sukhatme (1970). Sampling Theory of Surveys and its Applications. Iowa State Univ. Press.

in the i-th town, S_{ij}^2 = between households variance in the ij-th barrio, u_i = ratio of the number of households in town i to the average number of households per town, and v_{ij} = ratio of the number of households in the ij-th barrio to the average number of households in town i.

Intuitively one can guess correctly that for the same number of ultimate sampling units, a two-stage design will have a smaller variance than a three-stage design. The latter, however is cheaper to use since its sampling frame requirements are less and the sample barrios will be restricted within sample towns so that time and travel costs will probably be lower.

To investigate the precision of a three-stage design with SRS of towns at the first stage, we consider for convenience, cluster sampling of the same number of towns (i.e., assume complete enumeration of the population in all the sample towns). The variance and CV of y from this latter scheme may serve as lower bounds for (14) and its corresponding CV, respectively.

The coefficients of variation of town population for each province based on the 1970 Population Census are presented in table 3. The values range from 37 to 173 with a median of 67 percent. This enormous variability of population counts (and many related characteristics) would require, if at all possible, a very costly survey to reduce the CV of estimate to within 10 percent. On the other hand, for economy and expediency, most large-scale surveys usually have very few sample first-stage units (towns) per stratum (provinces) — two or three and very seldom more than five. To see how estimators of provincial population behave in SRS (cluster sampling) of towns, we give in table 4 the values of

$$CV(y) = CV(Y_{pop}) \times \sqrt{(T-t)/(T-1)t}$$

for t = 3, 4, 5. Note that the results are quite unsatisfactory including those for the regions. In fact, since CV (y) decreases to the order of $1/\sqrt{t}$ only, it simply is not possible for most of the provinces to reduce the CV to within 10 percent even if the sampling rate is raised to as high as one-half.

One might think of bypassing the towns and use the barrios instead as fsus. However, table 3 also shows that barrio population counts are even more variable than those of towns. We illustrate what happens by taking a (rather large) national sample of about 1500 barrios allocated to the provinces in three

Table 3. CV (percent) of 1970 town and barrio population count

Province	*Town*	*Barrio*	*Province*	*Town*	*Barrio*
Masbate	37	99	Eastern Samar	66	410
Cotabato	39	217	Bukidnon	67	88
Aklan	41	88	Negros Or.	68	105
Antique	41	109	Agusan S.	70	104
Bataan	41	96	Camarines S.	70	115
Davao N.	41	115	S. Leyte	70	114
Ifugao	44	69	Cavite	71	179
Camiguin	46	73	Nueva Ecija	72	138
Marinduque	46	85	Ilocos Sur	73	113
Sorsogon	47	150	Misamis Occ.	73	96
Mt. Prov.	49	55	Negros Occ.	74	187
Bohol	53	93	Leyte	75	240
Capiz	53	113	Northern Samar	75	129
Albay	54	146	Palawan	75	132
Cagayan,	56	124	Quezon	79	196
Mindoro Or.	57	100	Batangas	84	110
Surigao S.	57	157	Abra	85	130
Davao Or.	58	99	Ilocos Norte	86	210
Zamboanga N.	58	110	Occ. Mindoro	86	101
Batanes	59	116	Tarlac	94	110
Catand.	59	134	Laguna	100	172
Bulacan	60	108	Surigao N.	100	186
Lanao S.	61	137	Zambales	100	136
N. Viscaya	61	137	Western Samar	103	157
Pangasinan	61	95	Benguet	111	208
Camarines N.	62	160	Iloilo	116	116
La Union	62	79	Zamboanga S.	121	136
Romblon	63	61	Lanao N.	126	159
S. Cotabato	64	178	Misamis Or.	139	151
Isabela	65	125	Agusan N.	143	232
K-Apayao	65	75	Rizal	150	128
Pampanga	65	108	Cebu	152	118
Sulu	65	110	Davao S.	173	196

Table 4. CV (percent) of estimators of 1970 population in SRS of 3, 4, 5, towns per province

Area	towns	t=3	t=4	t=5
Region 2	151	24	20	17
Abra	27	45	38	34
Ilocos N.	23	46	38	33
Ilocos S.	34	40	34	30
La Union [a]	20	32	27	23
Mt. Prov.	47	56	48	43
Region 3	81	21	18	16
Cagayan	29	30	26	24
Isabela	34	35	30	27
N. Vizcaya	81	31	26	23
Region 4	168	15	13	11
Bataan	21	19	16	13
Bulacan	24	32	27	23
N. Ecija	32	39	33	29
Pampanga	22	34	29	25
Pangasinan	47	34	29	25
Tarlac	17	48	40	34
Zambales	14	49	41	34
Region 5	210	37	31	27
Batangas	34	45	39	34
Cavite	22	37	31	27
Laguna	30	54	45	40
Occ. Min.	11	40	33	27
Or. Min.	15	28	23	20
Palawan	20	39	33	28
Quezon	49	44	37	33
Rizal	29	80	63	60
Region 6	114	15	13	11
Albay	18	27	23	20
Cam. N.	11	29	23	19
Cam. S.	37	38	32	29
Catanduanes	11	27	22	18
Masbate	21	19	16	14
Sorsogon	16	24	20	17

Area	towns	t=3	t=4	t=5
Region 7	145	26	22	20
Aklan	17	21	18	15
Antique	18	21	17	15
Capiz	17	27	22	19
Iloilo	47	64	55	49
Neg. Occ.	31	40	34	30
Romblon	15	31	26	22
Region 8	268	30	25	22
Bohol	47	29	25	22
Cebu	53	84	72	64
Leyte	51	41	36	31
S. Leyte	17	36	30	25
Neg. Or.	31	37	31	28
Samar [b]	69	52	45	40
Region 9	185	20	18	16
Agusan N.	11	67	55	45
Agusan S.	13	34	28	24
Bukidnon	19	34	29	25
Lanao N.	21	66	55	48
Lanao S.	30	33	28	24
Mis. Or. [c]	31	74	63	55
Mis. Occ.	16	37	31	26
Surigao N.	26	53	45	39
Surigao S.	18	29	24	21
Region 10	169	22	19	17
Cotabato	35	21	18	16
S. Cot.	15	32	26	22
Davao N.	19	21	18	15
Davao S.	14	85	71	60
Davao Or.	11	27	22	18
Sulu	22	34	29	25
Zamb. N.	20	30	25	22
Zamb. S.	33	66	56	49

[a] Includes Kalinga-Apayao, Ifugao, Benguet
[b] Includes East, West and North Samar
[c] Includes Camiguin

Table 5. CV (percent) of estimators of 1970 population counts in SRS of about 1500 barrios.

Area	No. of barrios	Equal alloc. Size	Equal alloc. CV	Prop. alloc. Size	Prop. alloc. CV	Optimum alloc. Size	Optimum alloc. CV
Phils.[a]	32010	1536	4.9	1499	4.6	1500	3.6
Region 2	2450	192	12	116	16	73	15
Abra	267	24	26	13	36	6	53
Benguet	131	24	42	6	85	17	50
Ifugao	103	24	14	5	31	2	49
Ilocos N.	426	24	43	20	47	22	45
Ilocos S.	711	24	23	33	20	13	31
K-Apayao	189	24	15	9	25	3	43
La Union	440	24	16	21	17	8	28
Mt. Prov.	183	24	11	9	18	2	39
Region 3	1745	72	16	82	14	56	17
Cagayan	658	24	25	31	22	22	26
Isabela	877	24	26	41	20	25	25
Nueva Vis.	210	24	28	10	43	9	46
Region 4	3620	168	9	170	9	172	8
Bataan	148	24	20	7	36	6	39
Bulacan	515	24	22	24	22	27	21
Nueva Ecija	639	24	28	30	25	36	23
Pampanga	507	24	22	24	22	30	20
Pangasinan	1198	24	19	56	13	40	15
Tarlac	446	24	22	21	24	19	25
Zambales	167	24	28	8	48	14	36
Region 5	4301	216	14	202	14	287	8
Batangas	878	24	22	41	17	31	20
Cavite	315	24	37	15	46	28	34
Laguna	551	24	35	26	34	36	29
Marinduque	197	24	17	9	28	4	42
Occ. Mindoro	111	24	21	5	45	4	51
Or. Mindoro	337	24	20	16	25	10	32
Palawan	321	24	27	15	34	9	44
Quezon	1171	24	40	55	26	55	26
Rizal	420	24	26	20	29	110	12
Region 6	2884	144	12	135	11	117	12
Albay	593	24	30	28	28	30	27
Cam. N.	240	24	33	11	48	13	44
Cam. S.	942	24	24	44	17	33	20
Catanduanes	234	24	27	11	41	7	51
Masbate	464	24	20	22	21	15	26
Sorsogon	411	24	30	19	34	19	34

[a] Excludes Batanes and Camiguin

Area	No. of barrios	Equal alloc. Size	Equal alloc. CV	Prop. alloc. Size	Prop. alloc. CV	Optimum alloc. Size	Optimum alloc. CV
Region 7	3753	144	17	175	17	160	11
Aklàn	311	24	18	15	23	7	33
Antique	544	24	22	25	22	10	34
Capiz	425	24	23	20	25	14	30
Iloilo	1814	24	24	85	13	41	18
Neg. Occ.	480	24	38	22	40	85	20
Romblon	179	24	12	8	22	3	35
Region 8	5842	192	14	273	11	269	10
Bohol	1022	24	19	48	13	19	21
Cebu	1086	24	24	51	16	59	15
Leyte	1247	24	49	58	32	84	26
S. Leyte	382	24	23	18	27	9	38
Neg. Or.	629	24	21	29	19	23	22
E. Samar	378	24	84	18	97	42	63
N. Samar	416	24	26	19	30	12	37
W. Samar	682	24	33	32	28	21	34
Region 9	3631	216	10	169	13	127	13
Agusan N.	151	24	47	7	87	19	53
Agusan S.	153	24	21	7	39	5	46
Bukidnon	334	24	18	16	22	11	.27
Lanao N.	437	24	33	20	36	17	39
Lanao S.	1239	24	28	58	18	19	32
Mis. Occ.	393	24	20	18	23	9	32
Mis. Or.	368	24	31	17	37	22	32
Sur. N.	303	24	38	14	50	13	51
Sur. S.	253	24	32	12	45	12	45
Region 10	3784	192	14	177	13	239	10
Cotabato	1007	24	44	47	32	79	24
S. Cotabato	228	24	29	11	42	20	31
Davao N.	299	24	24	14	31	15	30
Davao N.	402	24	40	19	45	47	29
Davao Dr.	169	24	20	8	35	7	37
Sulu	415	24	22	19	25	14	29
Zamb. N.	372	24	22	17	27	14	29
Zamb. S.	892	24	28	42	21	43	21

different ways. One is equal allocation of 24 barrios per province which is done usually for no other reason except an equitable distribution of work load. A second more common method is proportional allocation; i.e. if b denotes the total sample size desired for the country, B_i the number of barrios in the i-th province and B the total number of barrios in the country, then $[b_i]$, the number of sample barrios for the i-th province, is the integer nearest to $b(B_i/B)$. However, proportional allocation does not take into account the differences in variability of the provinces. A still more efficient allocation scheme is one which strikes a balance between the size and variability of the provinces. One such scheme is prescribed by the Tschuprow-Neyman optimum allocation formula

$$b_i = b(B_i\sigma_i / \overset{p}{\Sigma} B_i\sigma_i)$$

where σ_i is the standard deviation of barrio population in the i-th province, $i = 1, \ldots, P$. The results are shown in table 5. It s clear that the strategy as a whole is again very inefficient. Note that while the CV of the country estimate is smallest with optimum allocation, those of the provincial estimates seem to be worst.

4. Some Variance-Reducing Techniques

By now what we hope to have conveyed is this: In the attempt to reduce the variance of the mean, σ^2/n, the "obvious" way — increase n - is not always the best way. Since there is a limit to the value of n, it may not even be possible at times to reduce σ^2/n to a desired level.

The other alternative is to dampen the contribution of the population variance σ^2 by changing either the sampling procedure, the estimation procedure, the structure of the population, or a combinaton of these. This alternative actually consists of a host of alternative strategies (the study of which is what sample survey theory is all about). With SRS still as the sampling procedure, we present here two such strategies.

4.1 *Restructuring and stratification of sampling units.*

The high variability of town and barrio population counts is due mostly to the presence of a few units with extremely high densities. One approach towards variance reduction is to group these large units into a separate stratum which can be treated independently; e.g. these units may be automatically part of the sample with a lower internal sampling rate so that costs can be kept in check. Another possibility is to divide these large units into smaller ones with sizes which are similar to the majority of the population units.[8] Conversely the small units may be grouped to form bigger ones, although for economy of time and travel, the grouping should be limited to contiguous units only.

Stratification for variance reduction involves partitioning of the population into k nonoverlapping, internally homogeneous strata. This can be done very effectively by arranging the sampling units according to decreasing population count so that bigger units are in the first stratum and the last stratum contains the smallest units. This technique, sometimes called paper stratification, has been employed on barrios earlier by Oñate (1965).[9] In determining stratum boundaries, Mahalanobis (1952) has shown that stratification is efficient when the stratum contributions to the population total are the same and the stratum CV's are nearly equal.[10]

To approximate the results of this technique, we truncated the 1970 barrio population counts by excluding those with zero or greater than 20,000 population; then built paper strata of sizes 70-80 thousand population (without dissecting any barrio in the process). The stratification for region 2 is shown in table 6. Note that the CV's of the individual strata are generally lower than those of the provinces. However, with this technique we have not been able to obtain more or less the same CV's for the strata. We consider SRS of around 1500 barrios distributed equally(proportionately and optimally (using Tschuprow-Neyman's formula) to the provinces. The CV's of the simple expansion estimators from these schemes are given in table 7. It is clear that the technique has not been successful. This is partly due to the fact that SRS is really a relatively inefficient sampling procedure and partly because of the following reasons: (a) Although the common practice of setting the total national sample size and then allocating this either proportionately or optimally to the provinces reduces the CV of the country estimates, this is achieved usually at the expense of the provincial

[8] This is the idea behind the creation of enumeration districts (ed) by partitioning large poblacions, provincial capitals and chartered cities, which were first used by the NCSO during the 1970 Population Census. Aside from the initial costs in their creation, one serious drawback of the ed's is the difficulty in locating ed boundaries, which may lead to bias due to multiple coverage and/or omission of units in the population. Lately, the barangays have been considered as replacements for the ed's and barrios.

[9] Onate, B.T. (1965). Estimation of population count by province with the 1960 Population Census as sampling frame, IRRI Library.

[10] Mahalanobis, P. C. (1952). Some aspects of the design of sample surveys, Sankhya. 12, 1-7.

Table 6. Paper stratification of barrios according to decreasing order of 1970 population counts, Region 2.[a]

Province/ Strata	*Size* (B_i)	*Average pop'n.*	*σ*	*CV*
Abra (0,0)[b]	267	545	696	128
1	67	1088	1217	112
2	200	363	138	38
Benguet (1,0)	130	1675	1691	101
1	13	5596	2863	51
2	37	1964	475	24
3	80	905	285	31
Ifugao (0,0)	103	898	617	69
I. Sur (0,1)	710	542	613	113
1	34	2278	1859	82
2	81	947	116	12
3	118	650	61	9
4	166	465	56	12
5	311	247	80	32
K-Apayao (0,7)	182	749	529	71
1	53	1287	687	53
2	129	527	183	35
La Union (0,1)	439	736	618	84
1	35	2309	1108	47
2	77	1047	145	14
3	118	683	76	11
4	209	387	112	29
Mt. Prov. (0,5)	178	808	415	51
1	58	1248	432	35
2	120	596	166	28
I. Norte (1,2)	423	741	919	124
1	23	3423	2603	76
2	80	980	157	16
3	118	661	71	11
4	202	388	114	30

[a] The stratifications for the other regions are not presented for brevity's sake. The total number of strata for the country (excluding Manila) is 458; each stratum has 70-80 thousand population.

[b] The first number denotes excluded barrios with populations exceeding 20,000, the second denotes excluded barrios with zero population.

Table 7. CV (percent) of estimators of (truncated) 1970 population with SRS of barrios.

Area/ Allocation	*Equal* [b]	*Proportional* [c]	*Optimum* [d]
Philippines [a]	3.4	3.1	2.8
Region 2	8	10	12
Abra	26	35	45
Benguet	21	41	34
Ilocos N.	14	19	26
Ilocos S.	23	28	31
K-Apayao	14	24	35
La Union	17	18	25
Mt. Prov.	10	18	30
Ifugao	14	31	39
Region 3	12	10	11
Cagayan	11	10	15
Isabela	25	19	21
N. Viscaya	6	10	18
Region 4	8	8	7
Bataan	14	25	27
Bulacan	17	18	16
N. Ecija	23	21	18
Pampanga	20	20	16
Pangasinan	20	13	13
Tarlac	23	24	22
Zambales	24	41	30
Region 5	10	9	7
Batangas	22	17	17
Cavite	22	28	24
Laguna	30	29	23
Marinduque	17	28	38
Occ. Mindoro	20	45	41
Or. Mindoro	20	25	26
Palawan	26	34	36
Quezon	31	20	19
Rizal	18	21	10
Region 6	10	9	9
Albay	17	16	17
Cam. Norte	19	28	31
Cam. Sur	23	17	17
Catanduanes	27	40	45
Masbate	20	21	22
Sorsogon	30	34	29

Area/ Allocation	*Equal*	*Proportional*	*Optimum*
Region 7	11	9	8
Aklan	17	21	28
Antique	22	21	30
Capiz	23	25	26
Iloilo	22	12	15
Neg. Occ.	19	20	13
Romblon	12	21	30
Region 8	10	8	7
Bohol	19	13	18
Cebu	24	16	13
Leyte	26	16	17
S. Leyte	23	27	34
Neg. Or.	21	19	19
E. Samar	23	27	32
N. Samar	26	29	32
W. Samar	32	27	29
Region 9	8	9	10
Agusan N.	23	42	34
Agusan S.	21	39	39
Bukidnon	18	22	22
Camiguin	14	50	50
Lanao N.	32	35	33
Lanao S.	26	17	26
Mis. Occ.	19	22	26
Mis. Or.	23	28	25
Surigao N.	19	24	32
Surigao S.	19	27	32
Region 10	9	8	7
Cotabato	22	15	15
S. Cotabato	21	31	24
Davao N.	22	29	25
Davao S.	24	27	21
Davao Or.	20	34	31
Sulu	22	24	24
Zamb. N.	23	27	26
Zamb. S.	24	17	17

[a] Excluding Greater Manila and Batanes

[b] 24 sample barrios per province, or a total of 1560.

[c] Total sample = 1502.

[d] Total sample size = 1501.

and regional estimates; i.e. CV (optimum allocation) < CV (proportional allocation) < CV (equal allocation) for the country estimates but this hierarchy is not true for the provincial and regional estimates. (b) Paper stratification, when applied uniformly across provinces with the same population count per stratum, is not fully effective as it leaves some provinces intact and the strata consisting of the bigger barrios still have much higher CV's.

Since the primary goal is to produce precise statistics at the provincial level, sample allocation and paper stratification should be applied directly to each province on a case-to-case basis. For instance, if we set the CV of the estimate at 10 percent and assume optimum allocation of sample barrios in each province, the resulting allocations, with the paper stratification indicated in table 6, are shown in table 8. The allocations for Abra and Ifugao appear to be unrealistically large, which is an indication that the stratification is not sufficiently efficient; i.e. the allocations for these provinces can be reduced further by increasing the number of paper strata.

Table 8. Optimum allocations with 10 percent CV in each province for the paper stratification in Table 5.

Stratum		*Allocation*	*Stratum*		*Allocation*
Abra	1	29	K-Apayao	1	12
	2	10		2	8
Benguet	1	7	La Union	1	3
	2	3		2	1
Ifugao	1	46	Mt. Prov.	1	6
Ilocos S.	1	4		2	4
	2	1	Ilocos N.	1	7
	3	1		2	1
	4	1		3	1
	5	2		4	3

A more realistic and efficient approach is to consider stratification and sample allocation simultaneously. One formulation of the problem is minimization of the total sample size $n = n_1 + \ldots + n_k$ where the number of strata k is allowed to vary subject to $n_i \geq n_i^o$, $i = 1, \ldots, k$, given a specified level of precision of the estimator, say $CV \leq CV^o$. In particular, the minmum possible allocations n_i^o may be set depending upon whether a straightforward variance estimate is desired ($n_1^o = \ldots = n_k^o = 2$), every stratum is represented ($n_1^o = \ldots = n_k^o =$

2), every stratum is represented ($n_1^o = \ldots n_k^o = 1$), one or more stratum is enumerated completely ($n_i^o = N_i$ for some i), etc. A case which allows zero allocation for some strata is not entirely unthinkable.[11]

We illustrate the scheme with the 1970 barrio population of Abra, assuming SRS, paper stratification (of equal population counts), Tschuprow-Neyman allocation and $CV(Y) = \sqrt{(Y)}/$ (mean population count) ≤ 0.10 or $V(Y) \leq 2959$ since the 1970 mean population per barrio of Abra is 544.
Starting with a conveniently low k and n, we construct paper strata and compute

$$n_i = n(w_i\sigma_j/\overset{k}{\Sigma} w_i\sigma_i) \quad , \quad i = 1, \ldots, k$$

$$V(Y|n) = \sum_{i}^{k} w_i^2\sigma_i^2 \; (N_i - n_i)/\{n_i \; (N_i - 1)\}.$$

n is increased and ($n_1, \ldots, n_k$) recomputed progressively until such time that $V(Y|n) = 2959$. k is incremented (by 1) successively and the whole process is repeated until such time that $n_i \geq n_i^o$, $i = 1, \ldots, k$ are no longer satisfied. The whole procedure lends itself easily to programming even on a desk calculator. The results for Abra are given in table 9. Note the considerable decline in n as k increases. I we require $n_i \geq 2$ for all i, then we need to construct 3 paper strata, with sizes (34, 76, 157) and minimum allocations (15, 2, 4). If instead, $n_i \geq 1$ for all i, then k = 5 with sizes (13, 33, 45, 64, 112) and minimum allocation (6, 1, 1, 1, 1), or a reduction of the sample by more than one-half.

Table 9. Optimum stratification — allocation of barrios given a 10 percent CV of estimator, Abra, 1970

	Stratum size and allocation				
k =	1	2	3	4	5
1	267(164)	67(29)	34(15)	21(9)	13(6)
2		200(10)	76(2)	46(1)	33(1)
3			157(4)	69(1)	45(1)
4				131(2)	64(1)
5					112(1)
Total	267(164)	267(39)	267(21)	267(13)	267(10)

[11] See Eriscon. W.A. (1965). Optimum stratified sampling using prior information. Journal of the American Statistical Association 60 (311), 750-771.

4.2 *Ratio method of estimation.*

When a SRS is drawn, the only required information are labels of the units in the population. If the simple mean $Y = \Sigma Y_i/n$ or simple expansion NY is used to estimate the true mean or total, respectively, no further prior information about the population is utilized. The irony of this strategy is that we expend so much to gain more insight about the population, at the same time that we ignore whatever old knowledge we have about the same population. The amount of information about the 1970 Philippine population (Y) that is in the 1960 Population Census (X), for example, is tremendous as indicated by the correlations between municipal population counts given in table 10.[12]

Table 10. Range of provincial values of town correlations and CV's of 1960(X) and 1970(Y) population counts.

Region	*Range of correlations*	*Range of CV's*	
		X	*Y*
2	0.967 - 0.996	55-87	55-101
3	0.954 - 0.984	40-75	53-86
4	0.918 - 0.998	39-87	38-96
5	0.947 - 0.997	44-151	42-147
6	0.804 - 0.996	38-74	36-68
7	0.882 - 0.996	38-100	40-115
8	0.882 - 0.996	52-132	52-150
9	0.728 - 0.998	36-130	49-136
10	0.814 - 0.999	55-149	38-167

In fact, some correlations are almost perfect, in which case one can predict almost exactly the value of y from X and vice versa.

Also, whereas the 1970 town populations $y_1, \ldots, y_t$ exhibit high variation, the ratios $y_1/X_1, \ldots y_t/X_t$, where the X_i's are the corresponding 1960 town populations, will be very stable. In a SRS of paired values $(X_1, Y_1), \ldots. (X_t, Y_t)$, the ratio of means

$$R = Y/x$$

12 The data used here were furnished by the Central Research Division, NCSO which did the difficult task of matching 1960 and 1970 town populations.

will be even more stable. Now R is the sample analogue of the true population ratio $R = \mu_y / \mu_x$; since μ_x is known (from the 1960 Census), we can use

$$\mathbf{Y}_r = R\,\mu_x \qquad (15)$$

as an estimator of the mean. The corresponding ratio estimator of the provincial total is

$$Y_r = T\,\mathbf{Y}_r \qquad (16)$$

This then is one method of using prior information after a sample is drawn.

It is known that R is a biased estimator of R, and so $\mathbf{Y}_r$ and Y_r are also biased for the mean and total, respectively. The bias however, is usually negligible in many situations where ratio estimators are applicable, vanishing in fact when the relationship between Y and X is a straight line through the origin.[13] Also, the MSE of $\mathbf{Y}_r$ (for large samples is smaller than V(Y) if the correlation, p, between X and Y satisfies

$$p > \frac{1}{2}\,\frac{CV\ (X)}{CV\ (y)}.$$

When X is some past value of y, then CV (X) — CV (y) (see table 10), and the last inequality simplifies to p > 1/2 which is certainly true in our example.

The bias and MSE of $\mathbf{Y}_r$ can be expressed as Taylor series expansions of population moments in powers of 1/t. General expressions for these series are given in David and Sukhatme (1974).[14]

Consider again the estimation of 1970 provincial population in SRS of t = 3, 4, 5 sample towns, using this time the ratio estimator (16) with 1960 town populations as concomitant infomation. The CV's of y_r, based on the MSE (y^r) expansion up to and including terms of order $1/t^2$, are given in table 11. Note that many of the CV's are below 10 percent and those of the regions are all within 8 percent even when t = 3. The CV's

13 Cochran, W. G. ibid.

14 David, I. P. and B. V. Sukhatme. (1974). On the bias and mean square error of the ratio estimator. Journal of the American Statistical Association. 69(346). 464-466.

of provinces in regions 9 and 10 are generally higher because of the relatively lower correlations of 1960 and 1970 municipal populations in these areas. In general, however, this strategy is still not very satisfactory since the CV's in some provinces are above the 10 percent limit. The inclusion of some other variance reducing techniques here, e.g. optimum stratification — allocation, is an interesting possibility which could lead to more interesting results.[15]

Finally, if we are to compare the results in table 3 and table 11, we see clearly the amount of precision gained with the use of the classical ratio estimator in place of the simple expansion estimator. Moreover, this extra precision is gained with very little cost — more complicated computations — an inconvenience that has become less important in this age of computers.

[15] Unfortunately, we cannot try ratio estimators with barrios as sampling units because of significant mismatching of data in the 1960 and 1970 Censuses arising from the birth and death processes of barrios in a span of 10 years.

Table 11. $\hat{CV}$ (percent) of ratio estimators of 1970 population in SRS of 3, 4, 5 towns per province.

Area	*No. of towns*	*t—3*	*t—4*	*t—5*
Region 2	151	5.2	4.5	3.9
Abra	27	4.4	3.7	3.2
Ilocos N.	34	7.4	6.2	5.4
Ilocos S.	34	4.1	3.5	3.1
La Union	20	3.6	3.0	2.6
Mt. Prov.[a]	47	15.6	13.4	11.8
Region 3	81	4.4	3.7	3.3
Cagayan	29	6.5	5.5	4.8
Isabela	34	6.4	5.4	4.8
N. Vizcaya	18	13.8	11.6	10.0
Region 4	168	3.4	2.8	2.5
Bataan	12	4.4	3.6	3.0
Bulacan	24	12.9	10.9	9.5
N. Ecija	32	7.9	6.7	5.9
Pampanga	22	7.6	6.4	5.6
Pangasinan	47	3.7	3.0	2.7
Tarlac	17	3.5	3.0	2.5
Zambales	14	21.4	17.7	15.0
Region 5	210	4.9	4.2	3.7
Batangas	34	3.4	2.9	2.6
Cavite	22	9.5	8.0	7.0
Laguna	30	7.4	6.3	5.5
Occ. Min.	11	11.2	9.1	7.5
Or. Min	15	4.6	3.8	3.2
Palawan	20	12.9	10.8	9.4
Quezon	49	11.1	9.5	8.4
Rizal	29	10.5	8.9	7.8
Region 6	114	5.2	4.4	3.8
Albay	18	2.8	2.4	2.0
Cam. N.	11	4.4	3.6	3.0
Cam. S.	37	14.2	12.1	11.1
Catand.	11	9.5	7.7	6.3
Masbate	21	12.8	10.8	9.4
Sorsogon	16	5.1	4.2	3.6
Region 7	145	8.2	7.0	6.2
Aklan	17	5.3	4.5	3.9
Antique	18	4.3	3.6	3.1
Capiz	17	3.9	3.3	2.8
Iloilo	47	9.8	8.4	7.4
Neg. Occ.	31	19.2	16.3	14.3
Romblon	15	3.5	2.9	2.4
Region 8	268	5.8	5.0	4.4
Bohol	47	8.5	7.3	6.4
Cebu	53	13.8	11.8	10.5
Leyte	51	10.0	8.5	7.6
S. Leyte	17	5.9	4.9	4.2
Neg. Or.	31	21.6	18.4	16.2
Samar[b]	69	11.7	10.1	9.0
Region 9	185	7.0	6.0	5.2
Agusan N.	11	5.1	4.1	3.4
Agusan S.	13	9.1	7.5	6.3
Bukidnon	19	22.5	18.9	16.3
Lanao N.	21	25.5	21.4	18.6
Lanao S.	30	22.9	19.5	17.1
Mis. Or.[c]	31	22.5	19.1	16.8
Mis. Occ.	16	6.9	5.8	4.9
Surigao N.	26	10.2	8.6	7.5
Surigao S.	18	17.5	14.6	12.6
Region 10	169	7.5	6.4	5.6
Cotabato	35	17.8	15.1	13.3
S. Cot.	15	19.5	16.2	13.8
Davao N.	19	11.2	9.4	8.1
Davao S.	14	10.1	8.3	7.1
Davao Or.	11	16.5	13.4	11.1
Sulu	22	11.1	9.4	8.2
Zamb. N.	20	12.6	10.6	9.2
Zamb. S.	33	26.3	22.4	19.7

[a] Includes Kalinga-Apayao, Ifugao, Benguet
[b] Includes East, West and North Samar
[c] Includes Camiguin