A NOTE ON A PRINCIPLE OF OBTAINING EXPANSION FACTOR IN MULTI-STAGE SAMPLE SURVEYS

By

DOMINGO C. ALONZO[1]

Multi-stage sample surveys admit several variants of sample size or fraction at each stage. With each variant there may be more than one choice of expansion factor to estimate a national total or mean. After sample fractions and mode of sampling (probability of selecting each sample unit) have been decided, the practical choice for mode of expansion usually lies between an unbiased linear estimator and a biased estimator which may have lower mean square error but whose bias may be uncertain. With a sample design such that one may be confident that bias will be less than an acceptable fraction of the sampling standard error the biased estimate may be preferred on account of its potentially greater precision. Usually however, such estimators require knowledge of the total number of sample elements per stratum as well as in each observed subsample unit. Such information is not at present known for the Philippines (with adequate precision). Consequently the Philippine Statistical Survey of Households (PSSH) has chosen to use the unbiased estimator.

A typical sample design used in the Philippines has three sample stages within strata. For instance, from each stratum sample towns are selected at random and from each selected sample towns sample barrios are also selected at random. Then from each selected sample barrio a proportion of households are selected for the observation of the characteristic under study. We use the following notation for any single stratum.

[1] Chief Statistician, Planning and Coordination Branch, Office of Statistical Coordination and Standards, National Economic Council.

T (t) = total (sample) number of towns in the stratum

B_i (b_i) = total (sample) number of barrios in the ith sample town

H_{ij} (h_{ij}) = total (sample) number of households in the ijth barrio

π_i = probability of selection in sampling (with replacement) for the ith town [1]

π_{ij} = probability of selecting the jth sample barrio within the ith sample town

So far, in the Philippine surveys the selection of households within sample barrios has always been such as to produce simple (equal probability) samples, usually systematic. The unbiased estimator for the stratum total is then

$$(1) \qquad \hat{X} = \frac{1}{t} \sum_i \frac{B_i}{\pi_i b_i} \sum_j \frac{1}{\pi_{ij}} \cdot \frac{H_{ij}}{h_{ij}} \sum_k X_{ijk}.$$

We will consider here specifically probabilities proportional to either area or population. Considering the case of the PSSH where the sample towns in the first stage were selected proportional to their population and sample barrios were selected with equal probability, the unbiased estimator for the stratum total (1) becomes

$$(2) \qquad \hat{X} = \frac{1}{t} \sum_i \frac{P_s}{P_i} \cdot \frac{B_i}{b_i} \sum_j \frac{H_{ij}}{h_{ij}} \sum_k X_{ijk}$$

$$= \sum_i \sum_j \sum_k \left(\frac{1}{t} \cdot \frac{P_s \, B_i \, H_{ij}}{P_i \, b_i \, h_{ij}} \right) X_{ijk}$$

The capital letters are fixed for the population being sampled, the lower case letters are at our disposal when choosing a sample design. If sample sizes be chosen arbitrarily, the $\hat{X}$ is laborious to compute. But the sample sizes can be chosen so

[1] Sampling with replacement greatly simplifies matters when π_i (π_{ij}) are variable. When simple random sampling (equal probabilities of selection) is used replacement gives no advantage and can only increase the sample variance. Sampling without replacement should then be used.

that the factor inside the parenthesis in (2) is a constant and can be taken outside the summation signs to give

$$\hat{X} = R \sum_i \sum_j \sum_k X_{ijk}$$

A sample so arranged is said to be self-weighing.

A suitable value of R must be determined to conform with available resources for conducting the survey. From the consideration of cost, administrative convenience and what is known about variability within and between each sample stage we may be able to fix tentatively the approximate sizes wanted for t, b and h. Substitution of the average values of these, and of the capital letters may be used to indicate approximately an R which is feasible, i.e.

$$R \sim \frac{1}{t} \cdot \frac{\bar{P}_s \bar{B} \bar{H}}{\bar{P} \bar{b} \bar{h}} \tag{3}$$

($\bar{H}$ will not be known exactly, but to the degree of approximation an estimate from an earlier survey will suffice).

In the PSSH it has been customary to fix b_i as constant for all sample towns and t as constant for all strata. The required sample fraction for households within any town to produce a self-weighing sample is then given by

$$\frac{h_{ij}}{H_{ij}} = \frac{1}{t} \cdot \frac{P_s B_i}{P_i b R} \tag{4}$$

(Note that barrio fraction is independent of j, i.e. is constant for all barrios in a given town but varies from all barrios in a given town but varies from town to town).

In the PSSH the number of sample towns per stratum, the number of sample barrios per sample town and the expected number of sample households were fixed at 5, 2 and 10, respectively, after considering allowable cost, administrative convenience, expected precision and the mode of computation. The average values of P_s, P_i, B_i and H_i as computed from available sources are as follows:

$$\bar{P}_s = 588{,}174$$
$$\bar{P} = 14{,}049$$
$$\bar{B} = 14$$
$$\bar{H} = 144$$

Then from (3), we have

$$R \sim \frac{1}{5} \cdot \frac{588{,}174.14.144}{14{,}049.\ 2\ .\ 10} = 846.72$$

For convenience the expansion factor R was decided to be 850 and from (4) $\frac{h_{ij}}{H_{ij}}$ were then fixed such that

$$\frac{h_{ij}}{H_{ij}} = \frac{1}{5} \cdot \frac{P_s\ B_i}{P_i\ .\ 2\ .\ 850}$$

I am grateful to Professor H. Fairfield Smith for the helpful suggestions and comments I received from him in the preparation of this note.

———o———

REFERENCES

1. Alonzo, Domingo C., "National Sample Survey of Households, Report on the Sampling Designs and Estimation Procedures", The Philippine Statistician, No. 2, Vol. V, pp. 82-97, June 1956.
2. Makanas, Elpidio D., "Recent Developments in the Design of the Philippine Statistical Survey of Households", The Statistical Reporter, No. 2, Vol. I, pp. 1-8, April 1957.
3. The Philippine Statistical Survey of Households Bulletin, No. 1, Vol. I, pp. 26-28, January 1957.