Journal of Applied Business Research - Spring 1987

ECONOMIC RESULTS OF SAMPLING PROCESSES AS A CONSEQUENCE OF SAMPLING AND TESTING (or MEASURING) ERRORS IN THE CASE OF STRATIFIED RANDOM SAMPLING

Vassilios K. Benos, Piraeus Grad. School of Industrial Studies, Greece

1. General

Sampling is one of the methods of selecting and analyzing specific qualities which refer to statistical populations. The basic problem of sampling technique is that the estimates of population parameters from one part of it (sample) have the errors of sampling and testing (or measuring).

Sampling errors are distinguished into random and systematic ones. A characteristic quality of random sampling errors is their variability and their tendency of shelf-neutralization, while systematic sampling errors have as a characteristic quality their constant size and the way of their direction. Random sampling errors are actually due to chance variation in the combination of items of a sample, while systematic sampling errors are probable to appear during the selection of a sample. During the process of testing (or measuring) may also occur errors which are called testing (or measuring errors). Sampling and testing (or measuring) errors definitely cause a certain cost, the behavior of which will be examined in this paper.

2. Process to be Followed

Let us assume that the total number N of the population has been divided into sub-populations consisted of N1, N2 . . .Nh numbers, where N1 + N2 + . . .Nh = N. These "h" sub-populations are called stratifications.

Further on:

a) From each stratification we take some simple random samples (e.g.k).

b) We carry out replicated tests in every sample. Let it be "m" replicated tests in every sample. Finally, < hxkxm > tests (or measurements) are carried out with random process over a period of time. We introduce the following symbols and definitions:

h = number of stratifications (hEN*), N* = 1,2,3. . .

k = " " samples per stratification (kEN*)

m = " " replicated tests sample (mEN*)

During the measuring (or testing) process we find out:

S2h = Variance appeared by the surveyed characteristic among the samples from the same stratifications.

S2k = Variance appeared by the surveyed characteristic among the samples from the same stratification.

S2m = Variance appeared by the surveyed characteristic among the tests (measures) in the same sample.

<u>1.</u> The variance tests differ among them only to testing (or measuring) error.

The variance among repeated tests

on the same sample is an estimation of variance S2m with hxkx(m-1) degrees of freedom. The "variance of Mean" of "m" replicated tests in every sample is S2m.

2. The "k" sample (selection) means in the same stratification, differ from each other, partly because of the measurement error, expressed by the "variance of Mean" S2m/m and partly because of the sampling error (during the selection of samples) expressed by the variance S2k.

Measurement errors are independent from the errors of selection (sampling errors) and therefore the variance of sampling means will be:

$$S2k + S2m/m \qquad (1)$$

The variance of (1) is estimated by the variance of sampling means in the stratifications with hx(k-1) degrees of freedom. Therefore, the "variance of the Mean" of k samples per stratification, will be due to sampling and measuring errors and

$$(S2k+(S2m)/m)/k = S2k/k+S2m/k\ m \qquad (2)$$

The variance of (2) will apply to the following cases: (i) When "k" selections (samples) are independent and "k" selections are tested "m" times independently, and (ii) when "k" selections (samples) are unified into one sample and then kxm tests are carried out.

3. The mens of "h" stratifications are also subject to a variance S2h which is independent of S2k and S2m. Therefore, the total appearance (T.var.) among stratifications mean is :

$$T.var.=(2)+S2h=S2h+(S2k/k)+S2m/k.m \qquad (3)$$

The variance of (3) is estimated by the variance of the means of stratifications with (h - 1) degrees of freedom.

III. Results

We define sampling and testing cost assuming that these costs vary linearly.

If C1 = selection Cost of a sample and
C2 = test cost (each test costs the same)

The best scheme is the one mentioned, "variance (V) of the Mean" of "k" samples and "m" trials per sample, is:

$$V=(S2k/k)+(Sm/k.m)=(S2k+(S2m/m)/k)$$

$$\text{i.e. } k=(S2k+((S2m)/m)/V) \qquad (4)$$

The total cost of "k" selections and "kxm" trials (tests), will be:

$$C=k.C1+k.C2.m=k(C1+C2.m) \qquad (5)$$

From (4) and (5), we shall have

$$C=(S2k+(S2m/m)/V)\ (C1+C2.m) \qquad (6)$$

$$C=[(C1)(S2k)+(C2)(S2m)]/V) + [(m)(C2)(S2k)]/V) + [(C1)(S2m)]/Vm) \qquad (7)$$

Differentiate C with respect to m and minimize:

$$(dc/dm) = [(C2)(S2k)/V]-[(C1)(S2m)]/(V)(m2)= 0 = (C2)(S2k)-[(C1)(S2m)/(m2) = 0$$

where m2 = (S2m)/(S2k)*(C1)/(C2) and

$$m = (Sm/Sk)\sqrt{(C1)/(C2)} \qquad (8)$$

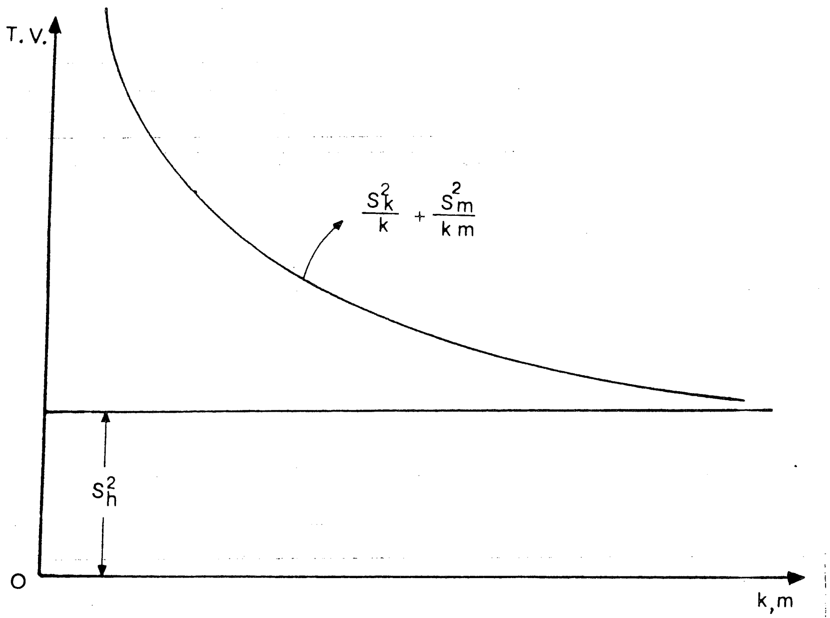
The second derivative

$$(d2)(C)/(d)(m2) = [(2C1)(S2m)]/(m3)$$

for $m = (Sm/Sk)\sqrt{\ }\ (C1/C2)$, is $(d2)(C)/(d)(m2) > 0$

Therefore, the best scheme (which gives the minimum cost) is the one resulting form (8) and (4). The influence on the whole accuracy of results is given in the following diagram.

In this diagram, the axis of the ordinate (T.V) refers to the total vari-



ance, while the axis of abscissae refers to the number of samples (k) by stratification and the number of repeated tests by sample (m). The curve refers to the velocity change in T.V when k and m are increased.

IV. Example

1. Let S2k = 900 , S2m = 400 and the permissible variance of the mean:V=200

Let C1/C2 = 1/10. Then, m = (30/20) (0,10) = 0,474 ≐ 1 (to nearest integer) and from (4):

k = (S2k)+(S2m)/(m)/(V) = (900 + ((400-/1)/200) = 6,5 ≐ 7

Thus, k = 7 and m = 1

2. If k = 10 > 7 and m = 1

10 = 900 +(400/1)/V) = V = 130 < 200

i.e. an increase in the number of samples per stratification (k) from 7 to 10, results a decrease by 35 % in the variance of the mean, thus an increase in the efficiency of stratified random sampling.

3. If k =7 and m = 2

7 = (900 + (400/2)/V) = V = 157 < 200

i.e., an increase in the number of replicated tests per sample (m) from 1 to 2, results a decrease by 21.5 % in the variance of the mean, thus an increase in the efficiency of stratified random sampling.

REFERENCES

1. Athanaspoulos, Dim., Statistical Inference. Piraeus, (1975).

2 Cochran, W.G., Sampling Techniques. John Wiley and Sons, New York, (1966).

3. Deming, W.E., Sample Design in Business Research. John Wiley and Sons, New York, (1960).

4. Hansen, M.H., "Response Errors in Surveys," Jour Amer. Stat. Assoc., 46, 147-1-90, (1951).

5. Nordin, J.A., "Determining Sample Size". Jour. Amer. Stat. Assoc., 39, 497-506, (1944).

6. Stephan, F.F. and McCarthy, P.J., Sampling Opinions. John Wiley and Sons, New York, (1958).

7. Yates, F., Sampling Methods for Censuses and Surveys. Hafner, New York, (1953).

8. Zaprokovic, S.S., "Computation of Errors for Sampling Estimates". Monthly Bull. Agr. Econ. and Stat. F.A.O. Rome, 9, No.4, (1960).