ASSESSMENT AND CONTROL OF NON-SAMPLING ERRORS IN CENSUSES AND SURVEYS

By M. N. MURTHY

Indian Statistical Institute

SUMMARY. A comprehensive treatment of the theory of non-sampling errors is given in this paper by bringing together the work of a number of authors on this subject. Various aspects of the problem of non-sampling errors such as sources of non-sampling errors, non-sampling bias and variation, use of the technique of interpenetrating sub-samples and non-response have been discussed. It may be noted that the derivation of a number of results in this field has been considerably simplified by the use of the conditional approach.

1. SOURCES OF NON-SAMPLING ERRORS

Till recently the theory of sampling has been developed assuming that each unit in the population has a unique 'true' value and that it can be observed and tabulated without introducing any error. This would mean that a complete enumeration of all the units in the population would give us figures without any errors, which is not usually the case, since in practice there are bound to be some observational and tabulation errors in the final results. Of course, in some cases these errors may be negligible in the context of the use to which the results are to be put. Even the first part of the assumption that each unit has a unique true value is questionable. As these types of errors are different from the error due to sampling of units and are due to sources other than sampling of units they are termed 'non-sampling errors' or 'response errors'.

The broad sources of non-sampling errors, which are present in both complete enumeration and sample survey, though possibly to varying degrees, are incomplete coverage of the population or sample (including non-response), faulty definitions, defective methods of data collection and tabulation errors. In case of sample surveys, the errors may also arise from defective sampling frame and selection procedures. More specifically the non-sampling errors may arise due to omission or duplication of units, inaccurate and inappropriate methods of measurement, inappropriate arrangement or wording of questions, inadequate and ambiguous instructions, non-response, deliberate or unconscious misreporting of data by respondents, carelessness on the part of the investigators and clerks, lack of proper supervision, and defective methods of scrutiny and tabulation of data.

From what has been stated above, it is clear that the results of sample surveys are subject not only to sampling error but to non-sampling errors also. In many situations the non-sampling errors may even be larger and therefore more important than the sampling error. Though data obtained on the basis of a complete enumeration are free from sampling error, they are subject to non-sampling errors. To make the results of censuses and surveys useful, it is necessary to reduce the non-sampling

errors as much as possible. It may be noted that while, in general, sampling error decreases with increase in sample size, the non-sampling errors tend to increase with the sample size.

The question of assessment and control of non-sampling errors has been receiving considerable attention and suitable techniques are being developed for this purpose. Mahalanobis (1940, 1944, 1946), Mahalanobis and Lahiri (1961) and Lahiri (1957a,b) have given many important techniques for assessing and controlling errors in censuses and surveys. Hansen and others (1946, 1951, 1961) and Sukhatme and Seth (1952) have considered in detail the question of non-sampling errors and have developed a suitable mathematical model for it. Post-enumeration checks and re-interview surveys are being made part of some of the nation-wide censuses and surveys so as to enable assessment of non-sampling errors.

2. Conceptual set-up

The difference between the sample survey estimate and the parameter true value being estimated may be termed 'error'. If the units in the sample can be observed and tabulated accurately then this error consists of only the error due to sampling, namely, 'sampling error'. A measure of the sampling error is supplied by the mean square error which is the expected value of the square of the difference between the estimator and the true value. This mean square error is composed of two parts—'sampling bias' and 'sampling variance'. The former has been defined as the difference between the expected value of the estimator and the true value and the latter is a measure of the divergence of the estimator from its expected value. Of course, in some cases the sampling bias may be negligible or zero.

If the data are also subject to non-sampling errors, then the difference between a survey estimate and the parameter true value may be termed 'total error' and this consists of both sampling and non-sampling errors. In this section a conceptual set-up is developed, which would enable us to get a measure of the non-sampling errors in terms of 'non-sampling bias' and 'non-sampling variance'.

The 'true' value of a unit is to be conceived of as a characteristic of the unit independent of the survey conditions which may affect the value 'reported' for that unit. For instance, age of a person at a particular point of time, income of a person during a particular period of time and number of persons in a country at a point of time are examples of characteristics for which the true value exists and is clearly defined. There are many items of information, such as intelligence of a person, attitude to some social measures, consumer preference to certain articles, for which it is very difficult even to conceive of the true value. In such cases some suitable conceptually, defined value, which has to be to some extent arbitrary, may be taken as the true value. For the definition of a true value to be useful in practice, it should serve the purpose of the survey and it should be well defined and observable under 'reasonable conditions of survey' relating to subject coverage, method of enquiry, survey period and method of tabulation.

The non-sampling errors arise due to the fact that it may not be possible to collect and process data accurately even if the true value is well defined because of so many operational difficulties.

Suppose a sample has been chosen to be canvassed under reasonable conditions of survey and that there are two populations, one of investigators and the other of tabulators (clerks) qualified for doing the field and the processing work respectively of the same survey. If we repeatedly carry out the survey on the selected units with different samples of investigators and computors chosen with some suitable probability designs, we may get different results because of the various possible sources of error present under the usual operational conditions. Here there are three stages of randomization—selection of units, investigators and computors. The difference between the expected value of the estimator taken over all the three stages of randomization and the true value may be termed 'total bias'. This consists of both 'sampling bias' and 'non-sampling bias'. The variance of the estimator taken over all the three stages of randomization measures the divergence of the estimator from its expected value and consists of sampling variance, variance between investigators, variance between computors and some interactions between the three sources of error. For instance, the data collected by one investigator may be affected by his misunderstanding of the instructions, his preconceived notions about the survey, the earlier units canvassed by him etc. Thus we see that the total error consists of sampling bias and variance, non-sampling or response bias and variance and some interactions between the sample and the sources of non-sampling errors.

3. Non-sampling bias

For the sake of simplicity, let us assume only two stages of randomization ; one for selecting the sample of units and the other for selecting the survey personnel instead of the three stages of randomization considered earlier. Here we consider the survey personnel as a whole instead of as investigators and computors. Let $\hat{Y}_{rs}$ be the estimate for the s-th sample of units supplied by the r-th sample of the survey personnel. The conditional expected value of $\hat{Y}_{rs}$ taken over the second stage of randomization for a fixed sample of units is given by

$$E_r(\hat{Y}_{rs}) = \hat{Y}_{s}, \qquad \ldots \quad (3.1)$$

which may be different from the estimate $\hat{Y}_s$ based on the true values of the units in the sample. The expected value of this $\hat{Y}_s$ over the first stage of randomization gives

$$E_s(\hat{Y}_{s}) = Y', \qquad \ldots \quad (3.2)$$

which is the value that can be unbiasedly estimated by the specified survey process. This value Y' may be different from the true population total Y and the difference

$$B(t) = Y' - Y \qquad \ldots \quad (3.3)$$

may be termed the 'total bias'.

It may be noted that the sampling bias is given by

$$B(s) = E_s(\hat{Y}_s) - Y, \qquad \ldots \quad (3.4)$$

which is the difference between the expected value of the estimator based on the true values and the true value of the population total. Since the total bias is the sum of sampling and non-sampling biases, the non-sampling or response bias is given by

$$B(r) = B(t) - B(s) = Y' - E_s(\hat{Y}_s) = E_s(\hat{Y}_{rs} - \hat{Y}_s) \qquad \ldots \quad (3.5)$$

which is the expected value of the non-sampling or response deviation for the s-th sample. If it is a complete enumeration, there is no sampling bias and the total bias consists of only the response bias. In case of many sample surveys also, the total bias consists of only the response bias, since usually unbiased estimators (from the point of view of sampling of units) are used.

To fix the ideas let us consider an example where a simple random sample of n units is drawn without replacement from a population of N units and surveyed by k persons chosen with equal probability from a large population of K persons qualified for this work, each person surveying m of the units assigned to him at random ($n = mk$). Let y_{rij} be the value reported by the j-th investigator for the i-th unit allotted to him in the s-th sample. Then an estimator of the population total is given by

$$\hat{Y} = \frac{N}{n} \sum_{j}^{k} \sum_{i}^{m} y_{rij}. \qquad \ldots \quad (3.6)$$

The conditional expected value over all possible samples of investigators where the s-th sample is fixed is

$$E(\hat{Y}/s) = \frac{N}{n} \sum_{i}^{n} y_{ri.}, \quad (n = mk),\ y_{ri.} = \frac{1}{K} \sum_{j}^{K} y_{rij}.$$

If there were no non-sampling errors in the survey, the estimator would be

$$\hat{Y} = \frac{N}{n} \sum_{i}^{n} y_i,$$

where y_i is the true value. The difference

$$d_{ri} = y_{ri.} - y_i \qquad \ldots \quad (3.7)$$

may be considered to be the 'response deviation'. This deviation may also depend on the particular sample being surveyed because of the possible influence of some units on those of the others in the sample. The response bias in this case is given by

$$B(r) = Y' - Y \qquad \ldots \quad (3.8)$$

where

$$Y' = \frac{1}{\binom{N}{n}} \frac{N}{n} \sum_{i}^{N} \sum_{s \supset i} Y_{ri.},$$

Σ standing for the summation over all samples containing the i-th unit. If the average response for a unit is not affected by that of another unit in the sample, i.e., if $y_{ri} = y'_i$, then

$$Y' = \sum_i^N Y'_i \Big(Y_i = \frac{1}{K} \sum_j^K Y_{ij} \Big). \qquad \ldots \quad (3.9)$$

There are a number of techniques available for the assessment of response bias (Lahiri, 1957a, 1957b). The survey figure may be compared with an external figure obtained by some other agency or by the same agency in some previous period after making the necessary adjustments for differences in coverage, definitions, survey period etc. This comparison may be taken as a broad check of the survey figure. This check is termed 'external aggregative check'. A better check would be to have unit by unit comparison of the survey data with the corresponding values in some other survey. This method is termed 'external unitary check'. It may be noted that there would be considerable difficulties in matching the units for this type of check. In these checks the assumption is that one source of data is more reliable than the other. If this assumption is not ture, it would be difficult to conclude which figure is subject to more bias in case of discrepancies. Another technique of assessing response bias is to draw the sample in the form of two or more interpenetrating subsamples and to get these surveyed by different groups of investigators. This procedure is known as the method of interpenetrating sub-samples and will be considered in detail in Section 10.

The response bias in a census can be estimated by surveying a sample of units in the population using better techniques of data collection and compilation than would be possible under census conditions. Such surveys which are usually conducted just after the census to study the quality of the census data are called 'post-enumeration surveys'. Even in case of a large scale sample survey, the response bias can be estimated by resurveying a sub-sample of the original sample using better survey techniques. Another method of checking survey data would be to compare the values of the units obtained in two surveys and to reconcile the figures by further investigation in case of discrepancies. This method of checking is termed 'reconciliation (check) surveys.

4. Non-sampling variance

The mean square error of the estimator $\hat{Y}_{sr}$, based on the s-th sample of units and supplied by the r-th sample of the survey personnel, is by definition

$$M(\hat{Y}_{sr}) = E_{sr}(\hat{Y}_{sr} - Y)^2 \qquad \ldots \quad (4.1)$$

where Y is the true value being estimated. This is a measure of the divergence of the estimator from the true value, taking into account both sampling and non-sampling errors. This measure consists of bias and variance, that is,

$$\begin{aligned} M(\hat{Y}_{sr}) &= V(\hat{Y}_{sr}) + B^2(\hat{Y}_{sr}) \\ &= E(\hat{Y}_{sr} - Y')^2 + (Y' - Y)^2 \qquad \ldots \quad (4.2) \end{aligned}$$

where Y' is the expected value of the estimator taken over both the stages of randomization. The variance of the estimator is a measure of the divergence of the estimator from its expected value and $Y'-Y$ is the bias. Taking the variance over the two stages of randomization, we get

$$V_{sr}(\hat{Y}_{sr}) = V_s E_r(\hat{Y}_{sr}) + E_s V_r(\hat{Y}_{sr})$$
$$= V_s(\hat{Y}_{s\cdot}) + E_s E_r(\hat{Y}_{sr} - \hat{Y}_{s\cdot})^2. \qquad \ldots \quad (4.3)$$

From (4.3) we see that the variance can be split up into two parts—sampling variance and response variance. The second term stands for the expected value of the square of the response deviations of the sample estimates from their expected value taken over both the stages of randomization. This term can be further split up by writing

$$\hat{Y}_{sr} - \hat{Y}_{s} = (\hat{Y}_{sr} - \hat{Y}_{s} - \hat{Y}_{\cdot r} + Y') + (\hat{Y}_{\cdot r} - Y')$$

where $\hat{Y}_{\cdot r} = E_s(\hat{Y}_{sr})$, and taking the variance we get

$$E_{sr}(\hat{Y}_{sr} - \hat{Y}_{s\cdot})^2 = E_{sr}(\hat{Y}_{sr} - \hat{Y}_{s\cdot} - \hat{Y}_{\cdot r} + Y')^2 + E_r(\hat{Y}_{\cdot r} - Y')^2. \qquad \ldots \quad (4.4)$$

The first term on the right in (4.4) is the interaction between the sampling and non-sampling errors and the second term is the variance between survey personnel. Thus we see that the mean square of the estimator consists of sampling variance, interaction between sampling and non-sampling errors, variance between survey personnel and square of sum of the sampling and non-sampling biases. In a complete census the mean square error is composed of only the non-sampling variance and square of the response bias.

5. Simple random sampling

To fix the ideas let us consider the case where a simple random sample of n units drawn with replacement from a population of N units is divided at random into k equal sub-samples of m units each and these sub-samples are surveyed by k investigators selected with equal probability from a large population of K investigators qualified for this work. Let Y_{ij} and Y_i be the value reported by the j-th investigator for the i-th unit in the population and its true value respectively. Suppose y_{ij} is the value reported for the i-th unit in the sample by the j-th selected investigator. Here it is assumed that the response for a unit is not affected by the response of other units in the sample. An estimator of the population mean is given by

$$\bar{y} = \frac{1}{n} \sum_{j}^{k} \sum_{i}^{m} y_{ij}, \qquad (n = km). \qquad \ldots \quad (5.1)$$

The expected value of the estimator taken over the two stages of randomization is

$$E(\bar{y}) = \frac{1}{N} \sum_{i}^{N} Y_i', \left(Y_i' = \frac{1}{K} \sum_{j}^{K} Y_{ij} \right) \qquad \ldots \quad (5.2)$$
$$= \bar{y}'$$

and the total bias, which in this case consists wholly of response bias, is

$$B(t) = B(\bar{y}) = \frac{1}{N} \sum_{i}^{N} (Y_i' - Y_i). \qquad \ldots \quad (5.3)$$

The variance of the estimator is given by

$$V_{sr}(\bar{y}) = V_s E_r(\bar{y}) + E_s V_r(\bar{y})$$

where the subscripts denote the stages of randomization. The conditional expected value of $\bar{y}$ over the second stage of randomization for a fixed sample of units is

$$E_r(\bar{y}) = \frac{1}{n} \sum_{i}^{n} y'_i, \left(y'_i = \frac{1}{K} \sum_{j}^{K} y_{ij} \right).$$

The unconditional variance of this over the first stage of randomization is given by

$$V_s E_r(\bar{y}) = \frac{\sigma_s^2}{n}, \qquad \sigma_s^2 = \frac{1}{N} \sum_{i}^{N} (Y'_i - \bar{Y}')^2. \qquad \ldots \quad (5.4)$$

The conditional variance of $\bar{y}$ over the second stage of randomization for a fixed sample of units is

$$V_r(\bar{y}) = \frac{1}{k} E_r \left(\frac{1}{m} \sum_{i}^{m} y_{ij} - \frac{1}{m} \sum_{i}^{n} y'_i \right)^2$$

$$= \frac{1}{knm^2} \frac{1}{K} \sum_{j}^{K} \left[\sum_{i}^{m} (y_{ij} - y'_i)^2 + \sum_{i}^{m} \sum_{i' \neq i} (y_{ij} - y'_i)(y_{i'j} - Y'_{i'}) \right]$$

for $V_r \left(\frac{1}{m} \sum_{i}^{m} y_{ij} \right)$ is the same for all j. Taking the unconditional expected value of this expression over the first stage of randomization, we get

$$E_s V_r(\bar{y}) = \frac{1}{knm^2} \frac{1}{K} \sum_{j}^{K} \left[\frac{m}{N} \sum_{i}^{N} (Y_{ij} - Y'_i)^2 + \frac{m(m-1)}{N(N-1)} \sum_{i}^{N} \sum_{i' \neq i} (Y_{ij} - Y'_i)(Y_{i'j} - Y'_{i'}) \right.$$

$$= \frac{1}{knm} \sigma_d^2 [1 + (m-1)\rho]. \qquad \ldots \quad (5.5)$$

where σ_d^2 is termed 'simple' or 'uncorrelated' response variance and is given by the variance of individual response deviations, that is,

$$\sigma_d^2 = \frac{1}{KN} \sum_{i}^{N} \sum_{j}^{K} (Y_{ij} - Y'_i)^2 \qquad \ldots \quad (5.6)$$

and ρ is the intra-class correlation among the response deviations in a sample canvassed by one investigator (intra-investigator correlation), and is given by

$$\rho \sigma_d^2 = \frac{1}{KN(N-1)} \sum_{j}^{K} \sum_{i}^{N} \sum_{i' \neq i} (Y_{ij} - Y'_i)(Y_{i'j} - Y'_{i'}). \qquad \ldots \quad (5.7)$$

Hence the variance and the mean square error of $\bar{y}$ are

$$V(\bar{y}) = \frac{\sigma_s^2}{n} + \frac{\sigma_d^2}{n} [1 + (m-1)\rho] \qquad \ldots \quad (5.8)$$

and

$$\text{m.s.e.}(\bar{y}) = V(\bar{y}) + (\bar{Y}' - \bar{Y})^2. \qquad \ldots \quad (5.9)$$

In case of a complete census, sampling variance would be zero and hence the variance and mean square error of the census figure $\bar{y}'$ are given by

$$V(\bar{y}') = \frac{\sigma_d^2}{N}[1+(m-1)\rho] \qquad \ldots (5.10)$$

$$\text{m.s.e.}(\bar{y}') = V(\bar{y}')+(\bar{Y}'- \quad)^2 \qquad \ldots (5.11)$$

The result in (5.8) shows the contribution to the total variance from the response variation and it also brings out the impact of the intra-class correlation among the responses in a sample canvassed by one investigator (intra-investigator correlation) on the response variance. The intra-class correlation will be positive if the response deviations for the different units have a consistent tendency to be in one direction for an investigator and in another direction for another investigator. Even when this correlation is small, the contribution to the response variation may be considerable if m, the number of units surveyed by each investigator is large. For instance, if $\rho = 0.01$ and $m = 1000$, the response variation becomes about ten times more than that in case of $\rho = 0$.

An unbiased estimator of the variance of the estimator $\bar{y}$ given in (5.8) is given by

$$\hat{V}(\bar{y}) = \frac{1}{k(k-1)} \sum_j^k (\bar{y}_{\cdot j}-\bar{y})^2, \qquad \left(\bar{y}_{\cdot j} = \frac{1}{m}\sum_i^m y_{ij}\right) \qquad \ldots (5.12)$$

for $$E\left(\sum_j^k \bar{y}_{\cdot j}^2 - k\bar{y}^2\right) = k[kV(\bar{y})+\bar{Y}'^2 - V(\bar{y})-\bar{Y}'^2] = k(k-1)\,V(\bar{y}).$$

This result shows that if k independent samples are surveyed by k investigators selected with equal probability from a large population of investigators, then it is possible to get an unbiased estimator of the total variance (and not the total mean square error). This procedure is known as the method of 'interpenetrating sub-samples' which is considered in detail in Section 10. The variance between investigators is given by

$$\sigma_r^2 = \frac{1}{K}\sum_j^K (\bar{Y}'_{\cdot j}-\bar{Y}')^2 \doteq \sigma_d^2\,\rho \qquad \ldots (5.13)$$

for $$\sigma_r^2 = \frac{1}{K}\sum_j^K \left[\frac{1}{N}\sum_i^N (Y_{ij}-Y'_i)\right]^2$$

$$= \frac{1}{KN^2}\sum_j^K \sum_i^N (Y_{ij}-Y'_i)^2 + \frac{1}{KN^2}\sum_j^K \sum_i^N \sum_{i'\neq i} (Y_{ij}-Y'_i)(Y_{i'j}-Y'_{i'})$$

$$= \frac{\sigma_d^2}{N} + \frac{N-1}{N}\,\sigma_d^2\rho \doteq \sigma_d^2\rho,$$

if N is assumed to be large. An unbiased estimator of σ_r^2 is given by

$$\hat{\sigma}_r^2 = k\hat{V}(\bar{y}) - \frac{1}{km(m-1)}\sum_j^k \sum_i^m (y_{ij}-\bar{y}_{\cdot j})^2, \qquad \ldots (5.14$$

for taking the conditional expected value of the second term in (5.14), we get

$$\frac{1}{mk}\sum_{j}^{k}\frac{1}{N}\sum_{i}^{N}(Y_{ij}-\bar{Y}'_{.j})^2$$

and the expected value of this expression over the sample of investigators, is given by

$$\frac{1}{m}\,\frac{1}{NK}\sum_{j}^{K}\sum_{i}^{N}(Y_{ij}-\bar{Y}'_{.j})^2=\frac{1}{m}\,(\sigma^2-\sigma_r^2),$$

where σ^2 is the total variance in the population and is given by

$$\sigma^2=\frac{1}{KN}\sum_{j}^{K}\sum_{i}^{N}(Y_{ij}-\bar{Y}')^2=\sigma_r^2+\sigma_d^2. \qquad \ldots \quad (5.15)$$

Hence
$$E(\delta_r^2)=k\left[\frac{\sigma^2}{mk}+\frac{(m-1)}{mk}\,\sigma_r^2\right]-\frac{1}{m}\,(\sigma^2-\sigma_r^2)=\sigma_r^2.$$

6. Estimation of population proportion

It is interesting to consider the question of response variance in estimating a population proportion. Let Y_{ij} be 1 or 0 according as the j-th investigator reports the i-th unit in the population as belonging to a particular class or not and let P'_i be the proportion of the investigators reporting the i-th unit in the population as belonging to that class. Suppose a simple random sample of n units is drawn with replacement from a population of N units to be surveyed by a sample of k persons selected with equal probability from a large population of K persons qualified for this work. An estimator of the population proportion P is given by

$$\hat{P}=\frac{1}{mk}\sum_{j}^{k}\sum_{i}^{m}y_{ij}\quad(n=mk) \qquad \ldots \quad (6.1)$$

where y_{ij} is 1 or 0 according as the j-th selected investigator reports the i-th unit in the sample as belonging to a given class or not. The expected value of this estimator over both the stages of randomization is

$$E(\hat{P})=\frac{1}{N}\sum_{i}^{N}P'_i=P' \qquad \ldots \quad (6.2)$$

and the bias, which in this case consists of only the response bias, is $P'-P$. In this case σ_r^2 and σ_d^2 defined in (5.4) and (5.6) respectively are given by

$$\sigma_r^2=\frac{1}{N}\sum_{i}^{N}(P'_i-P')^2 \qquad \ldots \quad (6.3)$$

and
$$\sigma_d^2=\frac{1}{N}\sum_{i}^{N}P'_iQ'_i,\quad(Q'_i=1-P'_i). \qquad \ldots \quad (6.4)$$

The variance of $\hat{P}$ is

$$V(\hat{P})=\frac{1}{Nn}\sum_{i}^{N}(P'_i-P')^2+\frac{1}{Nn}\sum_{i}^{N}P'_iQ'_i\,[1+(m-1)\rho]. \qquad \ldots \quad (6.5)$$

From (5.12) it can be seen that an unbiased estimator of the total variance given in (6.5) is

$$\hat{V}(\hat{P}) = \frac{1}{k(k-1)} \sum_{j}^{k} (p_{*j}-p)^2 \qquad (6.6)$$

where p_{*j} is the sample proportion reported by the j-th selected investigator in the sample assigned to him and p is the over-all sample proportion. From (5.14), an unbiased estimator of the variance between investigators σ_r^2 is given by

$$\hat{\sigma}_r^2 = \hat{V}(\hat{P}) - \frac{1}{m(m-1)k} \sum_{j}^{k} p_{*j}q_{*j} \quad (q_{*j} = 1-p_{*j}). \qquad \ldots \quad (6.7)$$

If the intra-clas correlation is assumed to be 0, then the variance given in (6.5) reduces to

$$V(\hat{P}) = \frac{P'Q'}{n}, \quad (Q' = 1-P'). \qquad \ldots \quad (6.8)$$

This result is interesting because it shows that the expression which is normally used as the sampling variance of a sample proportion includes not only the sampling variance but also the uncorrelated response variance (Hansen, Hurwitz and Bershad, 1961). An unbiased estimator of the variance is

$$\hat{V}(\hat{P}) = \frac{pq}{(n-1)}, \quad (q = 1-p) \qquad \ldots \quad (6.9)$$

since $E(pq) = E(p)-E(p^2) = P'-V(p)-P'^2 = (n-1)V(p)$. Here again we see that the variance estimator of a sample proportion usually used to estimate the sampling variance estimates unbiasedly the total variance including both the sampling variance and the uncorrelated response variance.

7. Cost function

Let us consider the case of getting optimum values of k, the number of investigators, and m, the number of units assigned to each investigator, which would minimize the total variance for a given fixed cost. Suppose the cost function

$$C = kC_1 + nC_2 \qquad \ldots \quad (7.1)$$

where C_1 is the cost of recruiting and training one investigator, C_2 is the cost of surveying one unit and $n = km$. The total variance of the estimator $\bar{y}$ of the population mean $\bar{Y}$, given in (5.8), may be written as

$$V(\bar{y}) = \frac{\sigma^2-\sigma_r^2}{n} + \frac{\sigma_r^2}{k} \qquad \ldots \quad (7.2)$$

where [illegible] and [illegible]. Minimizing the variance in (7.2) with respect to n and k subject to the cost restriction (7.1), we get

$$[illegible] \quad (7.3)$$

$$[illegible] \quad (7.4)$$

8. [illegible] CORRELATIONS

A number of empirical studies have been conducted in recent years to assess the magnitude of the intra-interviewer correlation coefficient for different types of characteristics. The results obtained in some of these studies are presented in Table 1.

TABLE 1. [illegible]

Investigator	k*	type of items	range of ρ
(1)	(2)	(3)	(4)
[illegible]	[illegible]	[illegible]	[illegible]
		[illegible]	[illegible]
		[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
Hanson and Marks [illegible]	[illegible]	[illegible]	[illegible]
		[illegible]	[illegible]
		[illegible]	[illegible]
[illegible]	[illegible]	first study	[illegible]
	[illegible]	second study	[illegible]
[illegible]	[illegible]	[illegible] of major dental defects	[illegible]

* number of investigators

Source: Kish, L. [illegible]. Two studies of interviewer variance of socio-psychological variables. Presented at the Annual Meeting of the American Statistical Association.

9. Non-response errors

One of the sources of error in censuses and surveys, mentioned earlier, is the incomplete coverage of the population or sample. This incomplete coverage may arise due to respondents refusal to give information, respondents being not at home, inaccessible sample units etc. The error in this case arises because the population of non-respondents may have characteristics different from those who respond and the results based only on the surveyed units may be misleading. This type of error may be termed 'non-response error' since it arises from not surveying all the units in the population or sample. The non-response error may not be important if the units not responding in a survey have characteristics similar to those of the responding

units. But usually in practice this situation does not arise. For instance, if questionnaires are mailed to a number of farmers, the non-response rate may not be uniform among the farmers having land-holdings of different sizes and hence the results based only on the responses of the responding farmers may be misleading. It may be noted that in most cases of non-response, the response may be obtained by persuasion, repeated visits to the non-responding units etc.

One way of dealing with the problem of non-response is to make all efforts to collect information from a sub-sample of the units not responding in the first attempt (Hansen and Hurtwitz, 1946). Suppose, out of n units selected with equal probability without replacement from a population of N units, n_1 units respond and $n_2(=n-n_1)$ units do not respond in the first attempt. Let a sub-sample of n_2' units be selected from the n_2 non-responding units with equal probability without replacement for making sepcial efforts to collect the information. If $\bar{y}_1$ and $\bar{y}_2'$ are sample means based on the n_1 units responding in the first attempt and on the sub-sample of n_2' units respectively, then an unbiased estimator of the population total Y is given by

$$\hat{Y} = \frac{N}{n}(n_1\bar{y}_1 + n_2\bar{y}_2'). \qquad \ldots \quad (9.1)$$

It may be noted that there are three stages of randomization in this case; sampling of units, number of units in the sample not responding in the first attempt and sub-sampling of n_2' units from the n_2 units not responding in the first attempt. Taking the variance of the estimator given in (9.1) over these three stages of randomization, we have

$$V_{123}(\hat{Y}) = V_1E_2E_3(\hat{Y}) + E_1V_2E_3(\hat{Y}) + E_1E_2V_3(\hat{Y})$$

where E and V stand for conditional expected value and variance and the subscripts denote the stages of randomization. The conditional expected value and variance over the third stage of randomization are given by

$$E_3(\hat{Y}) = \frac{N}{n}(n_1\bar{y}_1 + n_2\bar{y}_2) = N\bar{y}, \left(\bar{y}_2 = \frac{1}{n_2}\sum_{i}^{n_2} y_{2i}\right)$$

and

$$V_3(\hat{Y}) = \frac{N^2}{n^2} n_2(n_2-n_2')\frac{s_2^2}{n_2'}, \quad s_2^2 = \frac{1}{(n_2-1)}\sum_{i}^{n_2}(y_{2i}-\bar{y}_2)^2.$$

where $\bar{y}$ is the sample mean based on all the n units in the sample and y_{2i} is the value of the i-th non-responding unit in the sample. Further it can be seen that

$$E_2(N\bar{y}) = N\bar{y} \text{ and } V_2(N\bar{y}) = 0.$$

Hence the variance of the estimator is given by

$$V(\hat{Y}) = V_1(N\bar{y}) + E_1E_2\left[\frac{N^2}{n}\left(\frac{n_2}{n}\right)(k-1)s_2^2\right],$$

where $k = n_2/n_1$. Since

$$E_1(\sigma_2^2) = \frac{N_2}{N_2-1}\ \sigma_b^2 \text{ and } E_2\left(\frac{n_2}{n}\right) = \frac{N_2}{N},$$

where σ_b^2 is the variance between the units in the population not responding in the first attempt and N_2 is the number of such non-responding units in the population, the variance becomes

$$V(\bar{y}) = \frac{N^2(N-n)}{n(N-1)}\ \sigma^2 + \frac{N}{n}(k-1)\ \frac{N_2^2}{(N_2-1)}\ \sigma_b^2 \qquad \ldots\ (9.2)$$

where σ^2 is the variance between the N units in the population.

The cost function in this case may be of the form

$$C = C_1 n + C_2 nP + C_3 \frac{n}{k}\ Q \qquad \ldots\ (9.3)$$

where C_1 is the cost per unit for the first attempt at data collection, C_2 is the cost per unit for tabulation, C_3 is the cost per unit sampled from the non-responding units (for obtaining data by additional efforts and for tabulation), P is the proportion of units in the population that would have responded in the first attempt, and $Q = 1-P$. The optimum values of n and k which would minimize the cost, ensuring at the same time a given value V^2 for the variance of the estimator, are given by

$$n = \hat{n}\left[\ 1+(k-1)Q^2\ \frac{N-1}{N_2-1}\ \frac{\sigma_b^2}{\sigma^2}\ \right] \qquad \ldots\ (9.4)$$

$$k = \sqrt{\left[\ \frac{N^2(N_2-1)\sigma^2}{N_2^2(N-1)\sigma_b^2} - 1\ \right]\ \frac{C_3 Q}{C_1+C_2 P}}, \qquad \ldots\ (9.5)$$

where

$$\hat{n} = \frac{N\sigma^2}{\sigma^2 + \frac{N-1}{N^2}\ v^2},$$

the sample size required for ensuring the value v^2 for the variance if there were complete response. If it is assumed that $\sigma^2 = \sigma_b^2$ and $\frac{N}{N-1} = \frac{N_2}{N_2-1} \doteq 1$, the optimum values of n and k reduce to

$$n = \hat{n}[1+(k-1)\ Q], \qquad \ldots\ (9.6)$$

$$k = \sqrt{\frac{C_3 P}{C_1+C_2 P}}.$$

An interesting device in dealing with 'not at home' cases has been considered by Politz and Simmons (1949). This procedure consists in ascertaining from the responding households the chance of their being at home at a particular point of time and weighting the results with the inverse of this chance. For instance, the households may be asked whether they were at home at some specified time during the previous

5 days. Then the households may be classified as being at home once in 6 visits, twice in 6 visits etc, and the data obtained for the different classes may be weighted by the inverse of the respective probabilities of being at home. In practice some bias would still persist because of persons not at home during the entire investigation period, who cannot be contacted.

10. INTERPENETRATING SUB-SAMPLES

The technique of interpenetrating sub-samples, which is due to Mahalanobis (1940, 1944, 1946), in its most general sense consists of drawing the sample in the form of k sub-samples according to any probability sampling design which would enable in getting valid estimates of the population parameter under consideration and subjecting these sub-samples to k different operations to study the differential effects of these operations. This technique has many possibilities in the field of censuses and survey sin assessing non-sampling error (Lahiri, 1953, 1957a, 1957b; Mahalanobis, 1956; Mahalanobis and Lahiri, 1961). One of the advantages has been mentioned in Section 5. There it was shown that if k independent interpenetrating sub-samples are drawn from a population of investigators, it would be possible to estimate the total variance of the estimator including both sampling and response variations.

Linked sub-samples. Originally, Mahalanobis (1940) made use of this technique in crop surveys to find out the differential investigator bias. For this purpose, linked pairs of grids (square parcel of land) were located at random on the maps in the form of dumb-bell shaped figures, one end of each figure representing the grid belonging to sub-sample 1 and the other end representing the grid belonging to sub-sample 2. One sub-sample was investigated by one set of investigators and the other sub-sample by an entirely different set of investigators independently. Under certain well-known assumptions Student's t-test may be applied to the difference between the estimates based on the two sub-samples to test the hypothesis that there is no differential investigator bias at any specified level of singificance. If the difference turns out to be significant, it means that the direction and magnitude of investigator bias are not of the same nature for all the investigators. It may be noted that if the difference turns out to be statistically insignificant, it does not mean that the investigator bias is zero. For, this result may be due to the fact that the biases are all of the same order and in the same direction.

The above method can well be applied to bring out the differential effect of different tabulation procedures, methods of data collection, etc., and to bring out the variation over time. Suppose one is interested in finding out whether intensive training of the investigators for a given survey is essential or not. For this purpose, one sub-sample may be assigned to intensively trained investigators and the other sub-sample to investigators who have got only superficial training. If the difference in the results obtained from these two sub-samples turns out to be significant, there is a strong case for adopting the method of intensive training in future surveys of a similar nature. On the other hand, if the difference were not significant, it would mean that for this type of survey intensive training is perhaps not essential.

The technique of interpenetrating sub-samples may be used as a check on the different operations involved in large scale surveys. Suppose one wishes to have a check on the calculations at the time of tabulation. For this purpose, the sample may be divided into k linked sub-samples assigned to k different groups of computors at random and the estimates may be obtained from each of these sub-samples. If there is good agreement between these estimates, for all practical purposes it may be assumed with certain amount of confidence that the calculations have been done correctly. If one of these estimates differs from the others (assuming k is more than 2) and if there is good agreement between the remaining $k-1$ estimates, one naturally suspects the calculations done on that sub-sample and gets that estimate recalculated. Thus it is seen that suitable action can be taken on the basis of the sub-sample estimates thereby increasing the accuracy and utility of the final results.

It is to be noted that detailed interpenetration of the sub-samples would require additional preparatory time and would increase the complexity of work at the field and the tabulation stages to some extent (Mokashi, 1950). It is also found that the powor of the interpenetrating sub-sample check is generally low due to the fact that the estimate of variance usually used in the test is based only on a few degrees of freedom (Sukhatme and Seth, 1952). It may be noted that the larger the positive correlation between the sub-samples, the greater will be the sensitivity of the test and lower will be the efficiency of the joint estimate based on the sub-samples. So if the main object of the survey is to test the differential investigator bias, then it would be desirable to have the same sample investigated by both the sets of investigators independently under the same conditions. If this is not possible due to the presence of conditioning effect between successive investigations of the same unit, it would be desirable to use sub-samples which are linked in such a way that the estimates from these samples are highly correlated.

Independent sub-samples. As has already been pointed out, linked samples are to be used only if the main objective is to find out the differential effect of two operations. But if the main object is to get a reliable estimate of the population parameter and the study of differential effects is only a subsidiary objective, then it is preferable to have independent interpenetrating sub-samples. The difference between the estimates based on two independent interpenetrating sub-samples provide a measure of the sampling as well as non-sampling errors present in the results.

The technique of interpenetrating sub-samples is of help in calculating the total variation especially in large scale sample surveys where a number of characteristics are under consideration. If there are k independent interpenetrating sub-samples subjected to k different operations each providing a valid estimate of the population parameter under consideration, then an unbiased estimator of the variance of the combined estimator (mean of the sub-sample estimates) is given by

$$\hat{V}(\bar{y}) = \frac{1}{k(k-1)} \sum_{i}^{k} (y_i - \bar{y})^2, \quad (\bar{y} = \frac{1}{k} \sum_{i}^{k} y_i), \qquad \ldots \quad (10.1)$$

where y_i is the estimate based on the i-th sub-sample. It may be noted that this procedure gives a simple method of getting an estimator of the variance of a ratio estimator. If $r_i \left(=\frac{y_i}{x_i}\right), (i = 1, 2, \ldots, k)$ is an estimate of the population ratio $R \left(=\frac{Y}{X}\right)$ based on the i-th sub-sample, then an unbiased estimator of the variance of

$$R' = \frac{1}{k} \sum_{i}^{k} r_i \qquad \ldots \quad (10.2)$$

is given by

$$\hat{V}(R') = \frac{1}{k(k-1)} \sum_{i}^{k} (r_i - R')^2. \qquad \ldots \quad (10.3)$$

Since the variance of R' and that of the combined ratio estimator

$$R'' = \frac{\sum_{i}^{k} y_i}{\sum_{i}^{k} x_i} \qquad \ldots \quad (10.4)$$

are approximately the same (Murthy and Nanjamma, 1959), (10.3) can be taken as an estimator of the variance of R''.

It is to be noted that the variance estimator given in (10.1) holds even if the variances of sub-sample estimates are different, provided the combined estimator is taken as the arithmetic mean of the sub-sample estimates. An unbiased estimator of variance can be obtained on the basis of independent interpenetrating sub-sample estimates even if the sub-sample estimates are weighted to obtain the combined estimator. If y_i and w_i are respectively the estimate and the weight for the i-th sub-sample $(i = 1, 2, \ldots, k)$, then an unbiased estimator of the variance of the combined estimator

$$\hat{Y} = \sum_{i}^{k} w_i y_i, \qquad \left(\sum_{i}^{k} w_i = 1\right) \qquad \ldots \quad (10.5)$$

is given by

$$\hat{V}(\hat{Y}) = \hat{Y}^2 - \frac{2}{1-\sum w_i^2} \sum_{i}^{k} \sum_{j>i} w_i w_j y_i y_j$$

since $E(\hat{Y}^2) = V(\hat{Y}) + Y^2$ and the second term in the above expression estimates unbiasedly Y^2. This expression after simplification becomes

$$\hat{V}(\hat{Y}) = \frac{(\Sigma w_i^2 y_i^2) - (\Sigma w_i^2)(\Sigma w_i y_i)^2}{(1-\Sigma w_i^2)}. \qquad \ldots \quad (10.6)$$

Since this estimator, may be difficult to compute in practice, the following unbiased variance estimator is suggested.

$$\hat{V}'(\hat{Y}) = \hat{Y}^2 - \frac{1}{\binom{k}{2}} \sum_{i=1}^{k} \sum_{j>i} y_i y_j = \hat{Y}^2 - \frac{1}{k(k-1)} \left[\left(\sum_{i=1}^{k} y_i \right)^2 - \sum_{i=1}^{k} y_i^2 \right].$$

In case of $k = 2$ this becomes simply $\hat{V}(\hat{Y}) = \hat{Y}^2 - y_1 y_2$ which is quite simple to calculate since $\hat{Y}_1 \cdot y_1$ and y_2 would be readily available.

Suppose in a stratified sample design, there are k independent interpenetrating sub-samples in each stratum. Let y_{si} denote the estimate of the s-th stratum total based on the i-th sub-sample ($s = 1, 2, \ldots, L$; $i = 1, 2, \ldots, k$). The variance estimator based on sub-sample estimates may be obtained either using strata sub-sample estimates or just the sub-sample estimates pooled over the strata. That is

$$\hat{V}_1(\hat{Y}) = \frac{1}{k(k-1)} \sum_{s}^{L} \sum_{i}^{k} (y_{si} - \bar{y}_{s\cdot})^2 \qquad \ldots (10.7)$$

$$\hat{V}_2(\hat{Y}) = \frac{1}{k(k-1)} \sum_{i}^{k} (y_{\cdot i} - \bar{y})^2 \qquad \ldots (10.8)$$

where $\bar{y}_{s\cdot} = \frac{1}{k} \sum_{i}^{k} y_{si}$, $y_{\cdot i} = \sum_{s}^{L} y_{si}$ and $\bar{y} = \frac{1}{k} \sum_{i}^{k} y_{\cdot i}$.

Of these two estimators (10.7) is more efficient than (10.8) (Murthy, 1962), though the calculation of the latter will be less time consuming than that of the former. In a stratified sample design with k independent interpenetrating sub-samples, if y_{si} and x_{si} denote the estimates of the s-th stratum total for the characteristics y and x respectively based on the i-th sub-sample, then an estimator of the variance of the ratio estimator $\hat{R}(=\bar{y}/\bar{x})$ is given by

$$\hat{V}(\hat{R}) = \frac{1}{\bar{x}^2} \frac{1}{k(k-1)} \sum_{s}^{L} \left[\sum_{i}^{k} (y_{si} - \bar{y}_{s\cdot})^2 - 2\hat{R} \sum_{i}^{k} (y_{si} - \bar{y}_{s\cdot})(x_{si} - \bar{x}_{s\cdot}) + \hat{R}^2 \sum_{i}^{k} (x_{si} - \bar{x}_{s\cdot})^2 \right] \qquad \ldots (10.9)$$

and an estimator of the bias in $\hat{R}$ is given by

$$\hat{B}(\hat{R}) = \frac{k\hat{R} - R'}{(k-1)} \qquad \ldots (10.10)$$

where $R' = \frac{1}{k} \sum_{i}^{k} \frac{y_{\cdot i}}{x_{\cdot i}}$, (Murthy and Nanjamma, 1959).

Operationally this technique is convenient because it simplifies the computation of variance in case of complicated designs and at the same time helps having a broad internal check on the results. The efficiency of the variance estimator is, however, impaired due to the reduction in the number of degrees of freedom on which such estimates are based. However, the range of the sub-sample estimates provides a confidence interval for the median of the estimator (which is the same as the mean

if the distribution is symmetric) with a confidence coefficient of $1-\left(\frac{1}{2}\right)^{k-1}$ irrespective of the distribution of the estimator. It may be noted that the interpenetrating sub-samples are of value if the survey has to be carried out in successive stages due to the necessity of providing preliminary results. The agreement of the sub-sample estimates is likely to be more convincing to the layman than any statement of sampling and non-sampling errors.

Suppose there are m agencies and n parties of investigators within each agency to conduct the survey. Then 8 or a multiple of mn (say kmn) independent interpenetrating sub-samples may be selected and each party of investigators in each agency may be assigned k sub-samples at random for being surveyed. With this arrangement the total variation of the estimator may be analysed as given below.

source of variation	degrees of freedom
between agencies	$m-1$
between parties	$m(n-1)$
within error	$mn(k-1)$
total	$mnk-1$

This analysis will help in locating the stages of operation where there is much of discrepancy. For instance if the between agency difference turned out to be statistically significant, if would mean that the survey has not been carried out according to the same specifications by one or more of the agencies. Similarly a significant result for the parties would mean that some parties have not functioned according to the specifications.

11. USE OF QUALITY CONTROL TECHNIQUES

The technique of statistical quality control (SQC) may be applied to census and survey work to assess the quality of the work and to improve the out-going quality with suitable corrective action. For this purpose it is desirable to use those SQC techniques which have built-in devices for initiating corrective action. More attention is to be paid to control of errors through SQC techniques than to acceptance plans for finished work. For a particular situation, the best plan is defined as that which ensures the highest out-going quality for a given cost or alternatively the lowest cost for a specified out-going quality. There is considerable scope to apply SQC techniques for control of errors in censuses and surveys because of the large amount of routine repetitive operations involved such as coding, punching etc.

No attempt will be made here to describe all the SQC techniques which may be applied to control errors in surveys. Instead, one procedure is described which is indicative of such applications. Suppose k operators are doing a particular routine operation where the out-put can be checked and the permissible error-rate in the finished work is specified. The work of each operator is first completely checked for

a suitable length of time. If the error-rate is less than the specified rate, only a sample of this work is verified in the subsequent periods of time. The decision regarding whether to continue verification on a sample basis or to have complete verification is taken separately for each operator on the basis of his cumulated error-rate over the past period. It may be noted that this procedure will help considerably in reducing the cost of verification and at the same time will ensure a specified quality level for the finished work. It may be mentioned that this type of procedure is being used in the United States Bureau of the Census and that this has been found to be helpful in controlling errors in census and survey work.

References

Bienbaum, Z. W. and Stoker, M. G. (1950): Bias due to non-availability in sampling surveys. *J. Amer. Stat. Ass.*, **45**, 98-111.

Durbin, J. (1956): Non-response and call-backs in surveys. *Bull. Inter. Stat. Inst.*, **34**(2), 72-86.

Deming, W. E. (1944): On errors in surveys. *Amer. Soc. Rev.*, **9**, 359-369.

——— (1960): *Sampling Design in Business Research*, John Wiley and Sons.

Elwadat, M. A. (1956): A sampling procedure for mailed questionnaires, *J. Amer. Stat. Ass.*, **51**, 209-227.

Gales, K. and Kendall, M. G. (1957): An enquiry concerning interviewer variability. *J. Roy. Stat. Soc.*, **120A**, 121-147.

Ghosh, B. (1949): Interpenetrating (net-work of) samples. *Bull. Cal. Stat. Ass.*, **2**, 108-119.

Gray, P. G. (1956): Examples of interviewer variability taken from two sample surveys. *Appl. Stat.*, **5**, 73-86.

Hansen, M. H. and Hurwitz, W. N. (1946): The problem of non-response error in sample surveys. *J. Amer. Stat. Ass.*, **41**, 517-529.

Hansen, M. H., Hurwitz, W. N. and Bershad, M. A. (1961): Measurement of errors in censuses and surveys. *Bull. Inter. Stat. Inst.*, **38**, (2), 359-374.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953): *Sample Survey Methods and Theory*, John Wiley and Sons.

Hansen, M. H., Hurwitz, W. N., Marks, E. S. and Mauldins, W. P. (1951): Response errors in surveys. *J. Amer. Stat. Ass.*, **46**, 146-190.

Hanson, R. H. and Marks, E. S. (1958): Influence of the interviewer on the accuracy of survey results. *J. Amer. Stat. Ass.*, **53**, 635-655.

Kish, L. (1962): Two studies of interview variance of socio-psychological variables. *J. Amer. Stat. Ass.*, **57**, 92-115.

Lahiri, D. B. (1957a): Recent developments in the use of techniques for assessment of errors in national surveys in India. *Bull. Inter. Stat. Inst.*, **36**(2), 71-93.

——— (1957b): Observations on the use of interpenetrating samples in India. *Bull. Inter. Stat. Inst.*, **36**(3), 144-152.

Macura, H. and Balaban, V. (1961): Yugoslavian experience in evaluation of population censuses and sampling. *Bull. Inter. Stat. Inst.*, **38**, (2), 375-399

Mahalanobis, P. C. (1940): A sample survey of the acreage under jute in Bengal. *Sankhyā*, **4**, 511-530.

——— (1944): On large scale sample surveys. *Phil. Trans. Roy. Soc.*, **231 B**, 329-451.

——— (1946): Recent experiments in statistical sampling in the Indian Statistical Institute. *J. Roy. Stat. Soc.*, **109**, 325-370.

——— (1950): Cost and accuracy of results in sampling and complete enumeration. *Bull. Inter. Stat. Inst.*, **33**(2), 210-219.

——— (1956) : Statistics must have purpose. Presidential Address at Pakistan Statistical Conference.

——— and LAHIRI, D. B. (1961) : Analysis of errors in censuses and surveys with special reference to experience in India. *Bull. Inter. Stat. Inst.*, **38**. (2), 409–433. Reprinted in *Sankhyā*, **23**, A, 1961, 325-358.

MURTHY, M. N. (1962) : Variance and confidence interval estimation. *Sankhyā*, **24**, B, 1-12.

MURTHY, M. N. and NANJAMMA, N. S. (1959) : Almost unbiased ratio estimates based on interpenetrating sub-sample estimates. *Sankhyā*, **21**, 381-392.

POLITZ, A. N. and SIMMONS, W. R. (1940) : An attempt to get the not-at-homes into the sample without call-backs. *J. Amer. Stat. Ass.*, **44**, 9-31.

SUKHATME, P. V. (1953) : *Sampling Theory of Surveys with Applications*, Iowa State College Press, 444-485.

——— and SETH, G. R. (1952) : Non-sampling errors in surveys. *J. Ind. Soc. Agr. Stat.*, **4**, 5-41.

Paper received : May, 1961.