STATISTICS IN TRANSITION-new series, September 2008
Vol. 9, No. 2, pp. 233—258

EXISTENCE OF THE BLUE FOR FINITE POPULATION MEAN UNDER MULTIPLE IMPUTATION

Pulakesh Maiti

ABSTRACT

Missing values not only mean less efficient estimates because of reduction in sample size, but also mean that the standard complete data methods cannot be immediately used to analyse the data. Imputation, single or multiple, is a compensatory method for handling non-responses and takes care of the fact that once the values have been filled in, standard complete data methods of analysis can be used. Here, in this paper, using multiple imputation technique, an estimator for the finite population mean in the presence of unit non-response has been proposed and the estimator so proposed has been found to be the BLUE. A very general cost model has been discussed in the presence of non-response and an optimal solution of sample size for a given number of imputations or of number of imputations for a given sample size has been worked out.

1. Introduction

A survey may mean to include census which attempts to collect information from each member in the population, whereas a sample survey refers to a survey in which a scientific sample of the population is studies i.e., the same sort of information is sought only for some of the units, — those in the sample. The choice of the sample is carefully made in order to draw inferences about the parameters of the population under study, but in many censuses and sample surveys, some of the selected units may not be possible to be contacted, and even contacted do not respond to at least some of the items being asked. Such non-responses which are known as survey non-response, whether it arises from a census or a sample survey is common.

The problem created by non-responses is of course that of non-availability of the complete data i.e., some of the values intended by the sampling design to be observed are in fact missing, and these missing values not only mean less efficient estimates because of the reduction in size in the data base, but also mean that standard complete data methods can not be immediately used to analyse the data.

Non-response as a concept has been defined in a number of ways. Most definitions distinguish **unit non-response** from **item non-response**. In general, non-response has been attributed to failure to obtain a response to a particular unit and/or to particular item when the questionnaire has been canvassed and has been completed partially or not been responded at all. [Kendall and Buckland (1960), Kish (1965), Bureau of census (1957, 1976), Cochran (1979), Zarkovich (1966), Ford (1976), Sudman (1976), Suchman (1962), Wark and Lilinger (1975), Deghton et al. (1978), Deming (1953)]. An extended definition of non-response includes in which missing data arise from the processing of information provided by units rather than refusal of units.

1.1.Non-response of different types at different levels

In household surveys with multistage design, non-responses can occur at different hierarchical stages singly or jointly, say, at the PSU (village), at the household and at the individual data item singly or jointly.

The first stage sampling unit, say, the village/urban block may be temporarily inaccessible or might have altered in character as a village might have become urbanized since last census. In spite of all the efforts, a few casualties do occur in a large scale survey operation like NSS. The extent of non-response at this level is of the order of 0.5 to 1% [Sarma, Rao and Ambe (1980)].

The existence of non-responses at the household level has been well exhibited through many surveys from both developed and developing Countries [Thompson I.b. and Siring E., Scott and Singh (1980), Verma (1980), US Current Population Survey (CPS) (1959—1978)., etc. There, non-response rates have been found to differ by reason in all the surveys conducted both in developing and developed Countries. Non-responses due to refusal are more in developed Countries, whereas non-responses due to non-contact are more in developing Countries. It has also been observed in Verma (1980), that in fertility surveys conducted in each of twenty developing Countries, each of non-response rates due to "vacant-dwelling unit", "due to not at home" is greater than that due to refusal and in all the Countries except Malayasia, Jamaica, Costarica, Mexica, Panama, non-response rate due to "not possible to locate" is greater than non-response rate due to refusal. Also one may consult the article by Maiti (2007) for demonstration of existence of non-responses in personal interviews.

The third level at which non-response occurs is at individual data level, which is popularly known as **item non-response**. There may be many reasons behind **item non-response**. To mention a few socio-economic background as well as the sensitivity of the specific item of information may be responsible for such non-responses. In fact, item non-responses are very much sensitive to particular type of items. For a good account of the extent of item non-responses on different items in different surveys, one can refer to the work by Sarma et al. (1980). Also one may consult paper by Dhar, N.R. (1971) and Maiti P. (2007).

One can also find through many surveys conducted both in developed and developing Countries that non-response rates are gradually increasing over the years, and hence call for strong attention. The work of assessing non-responses, though dates back to the forties [Deming (1944), (1950); Mahalanobis (1940), (1944), (1964); Moser (1958), Zarkovich (1966), Dalenius (1977a), (1977b), (1977c); Kish (1965), Sarndal et al. (1992)], but handling them with mathematical rigour is a new addition [Rubin (1987), Lessler and Kalsbeek (1996)] to the literature.

1.2. Two views of non-response: Deterministic and Probabilistic

The mechanism which generates non-response/response may be deterministic or may be stochastic in nature. By preassuming that members of the population are other certain to respond $(p_i = 1)$ or $(p_i = 0)$, the deterministic view of non-response rules out any uncertainty on whether or not each member of the population would provide usable data for the survey, if selected. Thus decision on whether to response or non-response is pre-determined. Reviews by Ford (1976) and Kalton (1983) provide extensive analytical discussions of non-response and non-response compensation procedures developed from a deterministic procedure.

Under stochastic view, each R_i associated with i^{th} sample unit, is a random variable whose outcome is determined by an assumed chance element in the response process. Associated with each R_i is the response probability p_i which may differ among different members of the population. [Politz and Simmons (1949), Hartley (1946), Deming (1953), Platek et al. (1977), Lessler (1983) etc.]

1.3. Different methods of dealing with non-response: Compensatory as well as Preventive

Several methods have been tried to compensate for the effect of non-response on the survey result. Some of these methods are part of **data collection procedure**; for example, intensive follow up of a subsample of non-respondents [Hansen, Hurwitz (1946), Fellegi and Sunter (1974), Platek et al. (1977)] or the collection of limited data through proxy interview from neighbours (Roshwab (1982) or though Call-backs (Birbaum and Sirken (1950), Durbin (1954), Deming (1953), Kish (1965), Kendal and Buckland (1972), Moser and Kalton (1972), Cochran (1977), Deighton et al. (1978). The substitution of other units for non-responding units is a controversial practice.

Other procedures, generally less costly, are used during data processing. These come under the general headings of **imputation** and **estimation procedures** which attempt to compensate for missing data.

There are other types of measures which may be termed as **preventive measures**. The preventive measures are those that would be implemented for identification, solicitation and compilation of the questionnaires, so that after the sample member has agreed, at least, in principle to co-operate, relevant data can be made available smoothly. These methods include correcting the frame errors, if any; proper designing of the questionnaires / schedules, providing uniform training to the investigators etc. [Maiti, (2007)].

1.4. Imputation versus Revising the weights under estimation procedures

Adjustment of estimates by revising the weights is for the fact that on measured information, non-respondents differ from respondents. It has been empirically observed through a number of surveys [Lundberg and Larsen (1949), Reuss (1943), Politz and Simon (1949), Birbaum and Sirken (1950), Pan (1951), King and Chen (1957), Buckland (1960), Suchman (1962), Lubin, Levit and Zuckerman (1962), Skelton (1963), Bennet and Hill (1964), Dunn and Hawks (1966), Ognibene (1970), Lessier (1974), U.S. Bureau of Census (1974), Warwick and Lininger (1975), Roy (1976—77, 1977—78, 1988—89), Sundman (1976), Deighton et al. (1978), Gower (1979), Kalton (1983), Madow et al. (1983), Maiti (1994—95, 1995—96) etc.] that who would form the set of respondents and who would be the non-respondents. Under this view of non-response, in revising the weights, the works, among others, due to Politz and Simon (1949), Hansen et al. (1953), Hartigan (1975), Platek et al. (1977), Kish and Anderson (1978), Bailar et al. (1978), Kohen and Kalsbeek (1981), Drew and Fuller (1980, 1981), Rizvi (1983), Madow (1983) may be mentioned.

In the broad sense, **imputation** means replacing missing or unusable information with usable data from other sources. These sources can include the same questionnaire, another questionnaire from the same survey or external sources, such as another survey or an administrative record.

In case of total/unit non-response, the choice of imputing the questionnaire has to be made from a large group of responding units. **It is at this level, one sees the similarity between weighting and imputation. Such imputation is equivalent to duplicating questionnaires. Duplicating a questionnaire to adjust for a missing unit is equivalent to giving that questionnaire an extra weighting factor of 2.**

From the view point of sampling error, adjusting for non-response by an estimation procedure is preferable to duplication of individual questionnaires. **The only reason for using the latter procedure would be to maintain a sample design that is self-weighting. This is accomplished by actually duplicating the computer record for the selected questionnaire rather than giving an extra weighty of 2.**

Operational simplicity and flexibility of procedures for automatic imputation make them attractive.

1.5. Single Imputation Versus Multiple Imputation

Imputation, whether single or multiple, takes care of the fact that once the values have been filled in, standard complete data methods of analysis can be used. The second advantage of imputation is that in many cases, imputation can be created by incorporating the knowledge of the data collector to reflect the uncertainty about which values to inputs.

Single imputation is useful when possibly substantial efforts are needed to create, but single value being imputed can reflect neither sampling variability about actual value when one model for non-response is being considered, nor additional uncertainty when more than one model is entertained. The obvious problem with single imputation is that the missing value is not known, but automatic application of complete data set treats missing values, as if they were known. Because of this, inferences based on the single imputed data set will be too sharp, since the extra variability due to unknown missing values is not being taken into account.

Multi imputation corrects the major flaws of single imputation. The idea behind multiple imputation is that for each missing value several values, say m, instead of just one, are computed. Thus m imputations for each missing datum create m-complete data sets. The practical difficulty with multiple imputations lies in the necessity of producing multiple data for each missing value. Where single set of imputed values is prohibited, repeating the process may be difficult. **Fortunately, there is some empirical evidence that the number of sets should not be large for multiple imputation to be effective (Rubin and Shanker 1986).**

Multiple imputation using modest m, say, $2 \leq m \leq 10$ is designed for situations with a modest fraction of missing information due to non-response.

The organization of this presentation is as follows.

Some results are presented in the next Section 2.1; under the assumption that responses/non-responses are purely random, followed in the next Section 2.2, assuming that each member of the population can be considered as having been labeled "respondent" or "non-respondent", or assigned to a respondent or non-respondent sub class prior to the survey. Finally, a non-linear cost-model has been considered to have an optimal solution of n or m in Section 3.

2. Existence of the BLUE for Population Mean

2.1. Non-responses occur randomly

Let us consider a finite population $U = \{1, 2, \ldots, N\}$ of N number of units labeled 1 through N. Let $y_i = y(i)$ be the value of the variable y for the i^{th} unit. Let the parameter to be estimated be $\overline{Y} = \Sigma y_i / N$.

Sampling Scheme under single imputation: Under SRSWOR (N, n), let a typical sample realized be $s(i_1, i_2, \ldots, i_n)$, which after the initial data collection, is partitioned into

$s_{(1)} = (i_1, i_2, \ldots, i_{n_1})$ and $s_{(0)} = (j_1, j_2, \ldots, j_{n_0})$, $n = n_1 + n_0$, where the suffixes i and j stand for the responding and non-responding units. Thus, after the initial field work, the following situation wises.

Data are available on n_1 number of units	Data are not available on n_0 number of units

To compensate for these unit non-responses, the following imputation method is adopted.

The incomplete data set is completed by imputing the missing values $y_{j_1}, y_{j_2}, \ldots, y_{j_{n_0}}$ through $Z_{j_1}, Z_{j_2}, \ldots, Z_{j_{n_0}}$, where $Z_k's\,(k = 1, 2, \ldots, n_0)$ are realized through a sample of SRSWR (n_1, n_0) and the incomplete data is completed as

$$\left(y_{i_1}, y_{i_2}, \ldots, y_{i_{n_1}}; Z_{j_1}, Z_{j_2}, \ldots, Z_{j_{n_0}}\right);$$

The incomplete data can be completed by adopting any other imputation method also.

Let $\bar{y}_1$ and $\bar{y}_0$ be the sample means based on $y'_{i_k}s\,(k = 1, 2, \ldots, n_1)$ and $Z'_{j_k}s\,(k = 1, 2, \cdots, n_0)$.

Let

$$\bar{y}_* = \frac{n_1}{n}.\bar{y}_1 + \frac{n_0}{n}\bar{y}_0 \tag{2.1}$$

then we have the following result.

Theorem 2.1: Under the above sampling schemes, $\bar{y}_*$ is unbiased for $\bar{Y}$ for any given (n_1, n_0) and we have,

$$V(\bar{y}_*) = \begin{cases} \left(S^2/n\right)\left(p_0 + \dfrac{1}{p_1} - f\right), & \text{under } SRSWR(n_1, n_0) \\ \left(S^2/n\right)\left\{\left(p_0 + \dfrac{1}{p_1} - f\right) - p_0^2/p_1\right\}, & \text{under } SRSWOR(n_1, n_0) \end{cases} \tag{2.2}$$

up to the order of $1/n$,

where, $(N-1)S^2 = \sum_{i=1}^{N} \left(y_i - \overline{Y}\right)^2, f = n/N,\ p_0 = \frac{n_0}{n}$ and $p_1 = \frac{n_1}{n}$.

Proof: The results follow immediately through the observations

$$E\ (\overline{y}_*) = E_{s_1} E_{z/s1}(\overline{y}_*) \text{and } V\ (\overline{y}_*) = V_{s_1} E_{z/s1}(\overline{y}_*) + E_{s_1} V_{z/s1}(\overline{y}_*)$$

and after routine calculations, where the variable z stands for imputation.

Remarks:

1. Clearly, $V(\overline{y}_*)\ |_{SRSWOR} \le V(\overline{y}_*)\ |_{SRSWR}$;
2. $\dfrac{V(\overline{y}_*)\ |\ SRSWR}{V(\overline{y}_*)\ |\ SRSWOR} = \dfrac{\left(p_0 + \dfrac{1}{n_1} - f\right)}{\left[\left(p_0 + \dfrac{1}{p_1} + f\right) - \left(p^2/p_1\right)\right]}$;
3. If each missing value $y_{jk}\left(k = 1, 2, \ldots, n_0\right)$ be imputed by the single value $\overline{y}_1$, then $\overline{y}_*$ becomes equal to $\overline{y}_1$.

Sampling Schemes under multiple imputation.

Under m-fold independent $SRS(n_1, n_0)$ from $SRSWOR$ (N, n) i.e., under m-tier imputation by SRS each time, tires being independent, let the estimator for population mean be defined by

$$\overline{\overline{y}}_* = \sum_{j=1}^{m} \overline{y}_*^{(j)} \Big/ m \tag{2.3}$$

where,

$$\overline{y}_*^{(j)} = \left(\frac{n_1}{n_0}\right)\overline{y}_1 + \left(\frac{n_0}{n}\right)\overline{y}_0^{(j)};$$

then we have the following result.

Theorem 2.2: Under the above sampling schemes and for a given sample (n_1, n_0), we have

(i) $\overline{\overline{y}}_*$ is BLUE for $\overline{Y}$

$$(ii)\quad V(\bar{y}_*) = \begin{cases} \left(S^2/n\right)\left(\dfrac{1}{p_1} - f\right) + \dfrac{S^2}{nm} p_0;\text{under} m - fold\text{independent } SRSWR(n_1, n_0) \\ \left(S^2/n\right)\left(\dfrac{1}{p_1} - f\right) + \dfrac{S^2}{nm} \alpha,\text{under} m - \text{foldindependent} SRSWOR(n_1, n_0) \end{cases} \qquad (2.4)$$

where S^2, p_0, p_1, f are defined as before and $\alpha = p_0 . p^*$, $p^* = \left(1 - \dfrac{p_0}{p_1}\right)$, normally $p_0 \le p_1$ and hence $\alpha \le 1$.

Lemma 2.1: Let each of $T_j = (j = 1,2,\ldots,m)$ be unbiased for θ and $T_j's$ are correlated i.e., $E(\underset{\sim}{T}) = \theta \underset{\sim}{1}$, $D(\underset{\sim}{T},) = \Sigma$; $T = (T_1, T_2, \ldots, T_m)'$ and $1 = (1,1,\ldots,1)'$, then the BLUE for θ i.e., $\hat{\theta}_{BLUE} = 1'\Sigma^{-1}T \big/ 1'\Sigma^{-1}1$. In particular, when $\Sigma = [(a-b)I + bJ]$,

then $\hat{\theta}_{BLUE} = \sum_{j=1}^{m} T_j \Big/ m$, where Σ is of $m \, x \, m$.

Proof of theorem 2.2: (i) follows, from Lemma 2.1 by setting $T = \left(\bar{y}_*^{(1)}, \bar{y}_*^{(2)}, \ldots, \bar{y}_*^{(m)}\right)'$ and observing a,b in Σ as $a = \sigma_*^2$ and $b = \rho^* \sigma_*^2$, where $\sigma_*^2 = V\left(\bar{y}_*^{(j)}\right)$ and $\rho^* \alpha_*^2 = Cov\left(\bar{y}_*^{(j)}, \bar{y}_*^{(k)}\right)$.

(ii) follows after routine calculations.

2.2. Occurrence of non-responses are non-random

One may consider the deterministic view to be one in which the outcome of the R-variable for each member of the population has been conditioned. The deterministic view then becomes a conditional form of the stochastic view. Here the population of N-units is assumed to be partitioned into two mutually exclusive and exhaustive subgroups, one with $N_0 = \sum_{i=1}^{N} R_i$ units, which with certainty would respond and $N_0 = N - N_1$ units, which with certainty, would not. The proportions in the respondent subgroup, $P_1 = N_1 / N$ and non-respondent subgroup, $P_0 = N_0 / N$ would depend on the characteristics of the study as well as on some specific features of the population members. Therefore, the population U can be thought of as consisting of two domains, $u_{(1)}$, the domain of respondents and $u_{(0)}$, the domain of non respondents.

$$U = \left(U_{(1)}, U_{(0)}\right), \; N = N_1 + N_0 .$$

Since, here the respondents are systematically different from the non-respondents, biases exist, unless further assumptions, on equality of two group means and/or group variances are assumed.

Let $\bar{y}_*^{(j)} = (j = 1, 2, \ldots, m)$, $\bar{\bar{y}}_*$ be defined as before, then we have the following result.

Theorem 2.3.: Under the sampling schemes of *SRSWOR (N, n)* and under m-tier $SRS(n_1, n_0)$ each time, we have,

(i) $E\left(\bar{y}_*^{(j)}\right) = \bar{Y}$ under the assumption of $\bar{Y}_1 = \bar{Y}_0$, for $j = 1, 2, \ldots, m$, $\bar{Y}_1$ and $\bar{Y}_0$ being to group means;

(ii) $$V\left(\bar{\bar{y}}_*\right) = \left(S_1^2/n\right)\left[P_0 + \frac{1}{P_1}(1-f)\left(1 + \frac{1}{C_1^2}\right)\right]$$

$$+ \left(S_1^2/n^2\right)\left[\left\{(1+f) + \frac{1}{C_1^2}\right\} + \frac{3P_0(1-f)}{P_1^2}\left(1 + \frac{(n-1)P_1}{C_1^2(1-f)}\right)\right]$$

$$\simeq \frac{S_1^2}{n}\left[\frac{P_0}{m} + \frac{1}{P_1}(1-f)\left(1 + \frac{1}{C_1^2}\right)\right], \text{ (up to the order of } 1/n) \qquad (2.5)$$

where, S_1^2, f are as before, C_1^2 is the square of coefficient of variation of y in the domain of respondents, and $P_1 = N_1 / N$, $P_0 = N_0 / N$. Here n_1, n_0 have been treated as random variables with $E(n_1) = np_1$ and $E(n_0) = n\,p_0$.

Proof: (i) we have

$$E\left(\bar{y}_*^{(j)}\right) = E_s E_{n_{1,s}} E_{z|n_{1,s}}\left(\bar{y}_*^{(j)}\right)$$

$$\simeq \bar{Y}_1$$

$$= \overline{Y} + \left(\frac{N_0}{N}\right)\left(\overline{Y}_1 - \overline{Y}_0\right).$$ (For calculation see the Appendix)

thus, under the assumption of $\overline{Y}_1 = \overline{Y}_0$, the result (i) follows and (ii) follows by observing the following fact and after routine calculations,

(ii) $V(\overline{\overline{y}}_*) = E_s E_{n_{1|s}} V_{z|n_1,s}(\overline{\overline{y}}_*)$

$$+ E_s V_{s_1|s} E_{z|s_1,s} (\overline{\overline{y}}_*)$$

$$+ V_s E_{s_1|s} E_{z|s_1,s} (\overline{\overline{y}}_*)$$

and

$$Cov\left(\overline{y}_*^{(j)}, \overline{y}_*^{(k)}\right) = E_s E_{s_1|s} Cov_{z|s_1,s} \left(\overline{y}_*^{(j)}, \overline{y}_k^{*}\right)$$

$$+ E_s Cov_{s_1|s} \left(E_{z|s_1,s}(\overline{y}_*^{(j)}), E_{z|s_1,s}(\overline{y}_*^{k}\right)$$

$$+ Cov\left(E_{s_1|s} E_{z|s_1,s}(\overline{y}_*^{(j)}, E_{s_1|s} E_{z|s_1,s}(\overline{y}_*^{(k)}\right)$$

where, suffix z stands over imputation. (For calculation see the Appendix)

3. General Cost Model

The optimal solution is always conditional on the appropriate cost-error model. Costs do vary across different activities in the total survey design and are very much dependent on the extent of efforts needed for their execution. Efforts at any stage can be translated into time of operations and finally, total cost can be evaluated with the available knowledge on the rate of cost per unit of time, rates being different for different survey operations. Thus specification of a cost model reduces to modeling of time allocation into different components under the total survey design. One of the major activities lies with the Survey Management Group who are mainly engaged in the survey operation leading to response or non response. It may be noted that time of response for complete or partial information from a respondent depends along with others on the efficiency of an investigator. The efficiency of an investigator need not necessarily be uniform in his whole course of action. In fact, efficiency of an investigator increases with the number of interviews completed (Mahalanobis, 1944).

A general cost model can be specified as follows:

$$C_T = C_0 + C(D) + C(n) + \underbrace{C(n+d)} \tag{3.1}$$

where, C_T = Total Cost:

C_0= Summary of fixed costs, which are independent of sampling design as well as sample size; It primarily includes costs of recruitment of human capital for administration as well as for technical work associated with programming as well as other computer job for data processing. It also includes machine capital. The cost of hardware/software and other equipment charges are also included here. Costs in specifying an association rule δ_{ik} between k^{th} frame units and i^{th} population unit and costs in preparing survey instruments for procuring and reconciliation of the survey results are also included here.

$C(D)$ = Summary of fixed costs mainly related to some sampling office work including computation of multipliers, estimators and their variances etc.

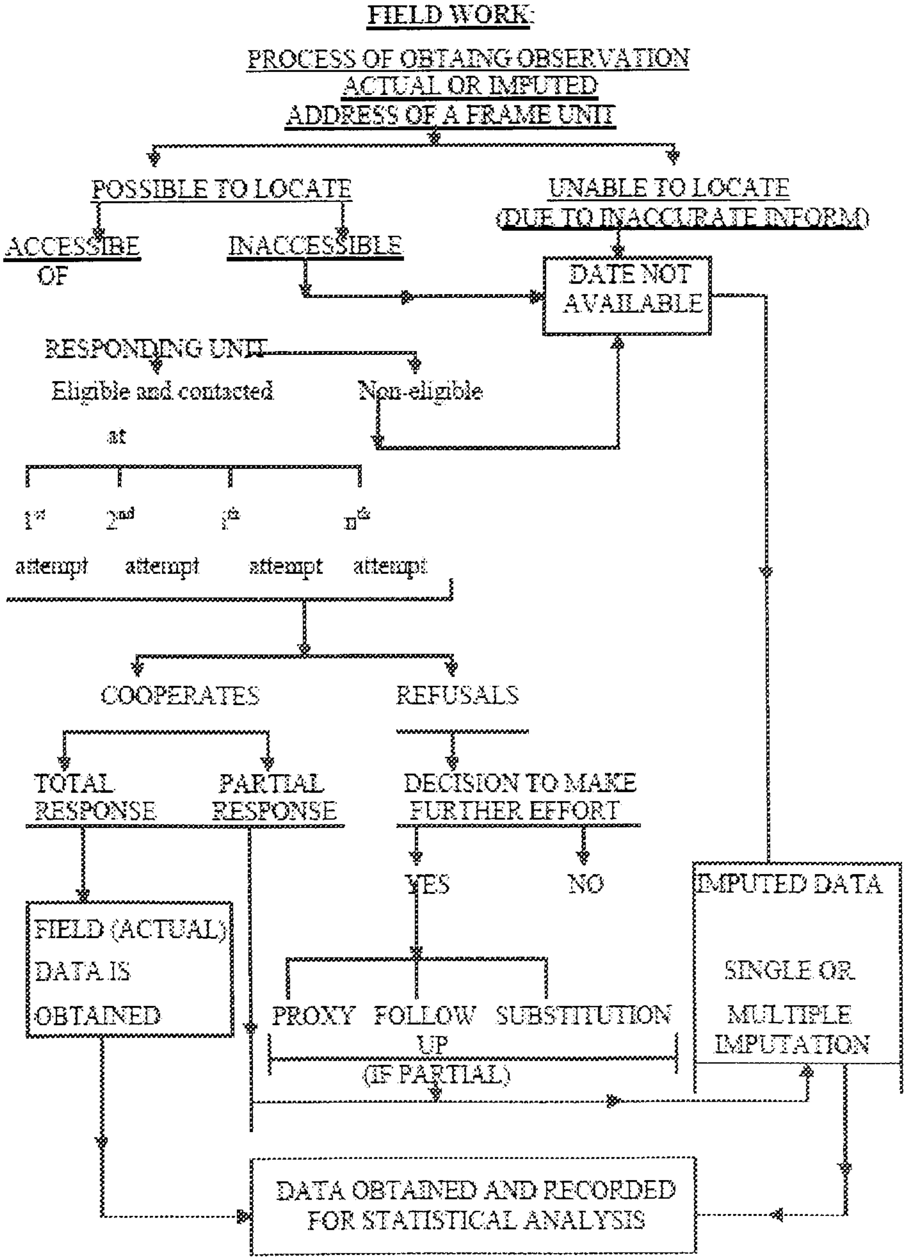
$C(n) = n\ C$ = All those costs dependent on the number of sampling units alone, but unaffected by the change of the sampling design; These fixed costs include printing/photocopying of the schedules/questionnaires, coding, editing etc.;

$C(n,d)$ = This is a **variable cost** dependent on the sample size and on the particular method of data collection adopted by the Survey Management Group;

It may be noted that though a response or a non-response is primarily the outcome of an interactive process between a respondent and an investigator under a given survey condition, such outcomes are dependent not only on interviewers and interviews, but also on all the instruments used in the whole system. The above general model tries to keep an account of all the costs.

The component C (n, d) which represents costs incurred during the data collection procedure will vary with different interviewers having different levels of efficiency. These efficiencies will further depend on different methods of data collation and collection. The component $C(n, d)$ may require further specification, as it will have a larger share in the total cost. We present the different field conditions through the following schematic diagram 1.

Modeling of C (*n*, *d*) in the presence of no non – response



Since the efficiency of interviewers increases with the number of interviews completed, the relationship between sample size and cost may not be linear, but curvilinear so that cost/unit is a decreasing function of sample size. A possible cost model in the presence of no non-response would be

$$C(n,d) = C(t_r) + (K_2 + K_3 n)\, n \qquad (3.2)$$

where,

$C(t_r)$ = the cost incurred in pertaining a training programme to the investigators;

K_2 = Base Cost

K_3 = a measure reflecting to decline in cost for interviewer with increasing efficiency.

Costs incurred in all efforts leading to having an effective interview would form a part of the base cost k_2. Such cost arises because of some or all of the following field conditions.

(a) Some of the **dwelling units** may not be possible to be identified due to faulty information in the frame population and/or because of in accessibility due to natural calamities and / or political disturbances;

(b) **Dwelling units**, though accessible, may be found to be vacant;

(c) The **dwelling unit**, though may not be vacant, but may not have an eligible respondent;

(d) The eligible respondent may not be temporarily at home;

(e) When contacted after a number of call backs, the respondent may be turned out to be an '**initial non-respondent**" and efforts would be needed to convert him into a respondent, failure to which he becomes a "**permanent non-respondent**";

(f) A respondent may be contact at the first time, but refusal may take place, and the interviewer may proceed further with an attempt to meet another eligible interviewee.

The above efforts may be viewed as the amount of time need until an interviewee is reached and all the costs should be attributed to the cost of a successful completion of the schedule. This cost may be termed as the **cost of exploration leading to the discovery of an interviewee**. In some cases, these costs may even be zero.

Let

$$\delta_j(i) = \begin{cases} 1, \text{when } j^{th} \text{ investigator finds } i^{th} \text{ respondent ready for cooperation} \\ 0, \text{ otherwise;} \end{cases}$$

When $\delta_j(i) = 1$, then the time passed through the process of investigation may be termed as operating time $\theta_j(i)$ or exploitation time and may further be split into the following components.

(a) **Rapport-time** $(R_j(i))$ taken by the j^{th} investigator in pursuing the i^{th} investigator for co-operation;

(b) **Enumeration time** $(E_j(i))$ taken by the j^{th} investigator in actually collecting data from the i^{th} investigator;

(c) **Editing time** $(ED_j(i))$ taken by the j^{th} investigator in assessing, if the information collected from i^{th} respondent needs to be monitored and re-interviewed;

(d) **Re-interview time** $(\mathrm{Re}_j(i))$ is the time of re-interview and reconciliation time;

(e) **Break time** $(Br_j(i))$ taken by the j^{th} investigator to depart from the i^{th} investigator;

(f) **Travel time** $(Tra._j(i))$ is the time taken by the j^{th} investigator moving after the completion of the i^{th} schedule in search of another interviewee allotted to him.

Thus, operating time:

$$0_j(i) = Ra._j(i) + E_j(i) + Ed_j(i) + \mathrm{Re}_j(i) + Br._j(i) + Tra._j(i).$$

Total operating time taken by the j^{th} investigator would be ,

$$0_j = \sum_{i \in U} \delta_j(i) 0_j(i)$$

and the associated cost would be,

$C_j = 0_j R_j$, R_j being the rate of the j^{th} investigator.

Cost of operation for all the schedules by all the investigators combined would be $C = \sum_{j \in V} C_j$, where V is the set of investigators. In fact the rate of cost combined with the base cost would be a decreasing function of the sample size and takes the form as mentioned in the model specified by (3.2).

The filed-in schedules are finally scrutinized by the supervision staff, and in this process, let $t_K(i,j)$ be the time needed by the k^{th} supervisor for scrutinizing the schedules completed by the j^{th} investigator from the i^{th} respondent $(i = 1,2,\ldots u; j = 1,2,\ldots v$ and $k = 1,2,\ldots,L)$. Therefore, the total time needed by the k^{th} supervisor to supervise all the filled–in schedules allotted to him would be,

$$S(k)=\sum_{i=1}^{I}\sum_{j=1}^{J}\lambda_k(i,j)t_K(i,j)\ ,\ k=1,2,\ldots,L\ ;$$

where,

$$\lambda_k(i,i)=\begin{cases}1, \text{ if } (i,j)^{th} \text{ schedule is supervised by the } k \text{ superviser;}\\ 0, \text{ otherwise.}\end{cases}$$

and the related cost for the k^{th} supervisor would be

$$C(k)=S(k)\times R(k)\,,$$

$R\ (k)$ being the rate of the k^{th} supervisor.

Therefore, the total amount needed in supervision work would be

$$C(S)=\sum_{k=1}^{k}S(k)R(k)\,.$$

Normally, supervisory staff forms the permanent staff and hence these costs are included in the fixed cost in the form of recruiting human capital.

Modeling of *C* (*n*, *d*) is the presence of unit non-response.

The previous model may be reformulated as

$$C=C_0+C_1n+\left(K_2+K_3\ (n-n_0)\right)(n-n_0)+\left(K_2+C^*m\right)n_0$$

where,

C_0 = fixed cost (including the training cost);

C_1 = cost/unit in preparing schedules;

n_0 = number of non-respondents;

C^* = imputation cost/unit non-response;

m = number of imputations;

Therefore, the cost at expected number of non-respondents would be

$$C=C_0+\left(C_1+K_2+C^*mP_0\right)n+K_3\,(1-P_0)^2\,n^2 \tag{3.3}$$

where, P_0 is the proportion of non-response in the population.

Let $\Phi = V(\bar{\bar{y}}_*) + \lambda \left\{C - C_0 - \left(C_1 + K_2 + C^* m P_0\right) n - K_3 \left(1 - P_0\right)^2 n^2\right\}$

$$= \frac{S_1^2}{n}\left[\frac{1}{P_1} + (1-f)\left(1 + \frac{1}{C_1^2}\right)\right] + \frac{S_1^2 P_0}{nm}$$

$$+ \lambda \left\{C - C_0 - \left(C_1 + K_2 + C^* m P_0\right) n - K_3 (1 - P_0)^2 n^2\right\}$$

$$\frac{\partial \Phi}{\partial n} = -\frac{S_1^2 A}{n^2} - \frac{S_1^2 P_0}{n^2 m} - \lambda \left[C_1 + K_2 + C^* m P_0 + 2n K_3 (1 - P_0)^2\right]$$

where, $A = \frac{1}{p_1} + (1-f)\left(1 + \frac{1}{C_1^2}\right) > 0$

$$\frac{\partial \Phi}{\partial m} = -\frac{S_1^2 P_0}{n m^2} - \lambda \; C^* n \; P_0$$

$$\frac{\partial \Phi}{\partial \lambda} = C - C_0 - \left(C_1 + K_2 + C^* m P_0\right) n - K_3 (1 - P_0)^2 n^2$$

$$= (C - C_0) - \left(C_1 + K_2 + C^* m P_0\right) n - K_3 P_1^2 n^2$$

$$\frac{\partial \Phi}{\partial m} = 0 \Rightarrow -\frac{S_1^2 P_0}{n\, m^2} = \lambda \; C^* n P_0$$

$$\Rightarrow \qquad \lambda = -\frac{S_1^2 P_0}{n^2 \, m^2 C^* P_0}$$

$$\frac{\partial \theta}{\partial n} = -\frac{S_1^2 A}{n^2} - \frac{S_1^2 P_0}{n^2 m} + \frac{S_1^2 P_0}{n^2 m^2 C^* P_0} \lambda \; \left[C_1 + K_2 + C^* m P_0 + 2n K_3 (1 - P_0)^2\right]$$

$$= -\frac{S_1^2 A}{n^2} - \frac{S_1^2 P_0}{n^2 m} + \frac{S_1^2 P_0}{n^2 m^2 C^* P_0}\left[\frac{C_1 - C_0 + K_3 + P_1^2 n^2}{n} + 2n K_3 P_1^2\right]$$

$$\left(\text{By } \frac{\partial \Phi}{\partial \lambda} = 0 \Rightarrow C_1 + K_2 + C^* m P_0 = \frac{C - C_0 - K_3 P_1^2 n^2}{n}\right)$$

$$= -\frac{S_1^2 A}{n^2} - \frac{S_1^2 P_0}{n^2 m} + \frac{S_1^2}{n^2 m^2 C^*}\left(\frac{a}{n} + bn\right), \quad a = C - C_0 > 0, b = K_3 P_1^2 < 0,$$

as $k_3 < 0$

Thus, by $\frac{\partial \Phi}{\partial n} = 0$, we have,

$$\frac{S_1^2}{n^2 m^2 C^*}\left(\frac{a + bn^2}{n}\right) = \frac{\left(m S_1^2 A + S_1^2 P_0\right)}{n^2 m}$$

or, $\quad S_1^2\left(a + bn^2\right) = \left(m^2 S_1^2 A + m S_1^2 P_0\right) n C^*$

$$n^2 b - C^* m \,(P_0 + m\ A)\,.n + a = 0$$

or, $\quad n^2 b - nD + a = 0$, where $D = m C^*\,(P_0 + m)$

or, $\quad n = \dfrac{D \pm \sqrt{D^2 - 4ab}}{2b} = \dfrac{D \pm \sqrt{D^2 + 4ab}}{2b}$,

Thus, for given value of *m*, $(2 \le m \le 10)$, we shall have different pairs (*n*, *m*) $(2 \le m \le 10)$ and from these choice of (n,m), the optimum pair, say, $(n,m)_0$ can be obtained by comparing the cost at expected number of non-respondents as specified in (3.3) for different pairs (*n*, *m*)

REFERENCES

BABBIE, EARL R. (1973): Survey Research Methods. Belmont, CA, Wadsworth.

BACKSTORM, CHARLES H. and GERALD HURSH-CESAR (1981). Survey Research, 2nd edition, New York, Wiley.

BOWLEY, A.L. (1906): Address to the Economic and Statistics Section of the British Association for the Advancement of Science, York, 1906, J. Roy Statist. SOC. 69, 540—558

----------(1926): Measurement of the Precision attained in Sampling. Bulletin of the International Statistical Institute, 22, 6—62.

BENNET, C.M. and HILL, R.E (1964): A comparison of selected personality characteristics of respondents and non-respondents to a mailed Questionnaire. Journal of Educational Research, 58, No. 4 178—180.

BIRBAUM, Z.W. and MONROE G. SIRKEN (1950): Bias due to non-availability in Sampling Survey. JASA, 45, 98—111.

BAILER, BARBARA A. (1979): Rotation Sampling Biases and their effects on estimates of changes 43rd session of the International Statistical Institute, Manita.

BERGMAN, L.R. HONVE, R. and RAPPA, J. (1978): Why do some people refuse to participate interview surveys? Statistik Tidskrift.

BROOKS, CAMILLA and BARBARA BAILAR (1978): An Error Profile Employment as Measured by current Population Survey Statistical Policy Working Paper 3, office Federal Statistical Policy and Standards. U.S. Department of Commerce.

BANDYOPADHYAY, S. CHAUDHURY, A., GHOSH, J.K. and MAITI, P. (1999): A Draft Proposal for an Enterprise Survey Scheme as a substitute for Economic Census. Indian Statistical Institute, Calcutta.

COCHRAN, W.G. (1979): Sampling Techniques, Wiley Eastern Limited, New Delhi, III edition.

COLE, D (1956): Field Work in Sample Surveys of Household Income and Expenditure, Applied Statistics, Volume 5, 49—61.

COBB, J.M., KING S., and CHEN, E. (1957): Differences between respondents and non-respondents in a morbidity survey involving clinical examination, Journal of Chronic Diseases, 6.

CHEVRY GABRIEL (1949): Control of General Census by means of an area sampling method, JASA, 44, 373—379.

CHAPMAN, DAVID D. and ROGERS CHARLES, E.(1978): Census of Agriculture- Area Sample design and methodology. Proceedings of the American Statistical Association Section on Survey Research Methods, 141—147.

DEMING, W (1960): Sampling Design and Business Research, New York, Wiley.

--------------(1944): On Errors in Surveys" American Sociological Review, 9, 359—369.

-------------(1950): Some Theory of Sampling, John Wiley and Sons, New York.

-------------(1953): On a Probability mechanism to attain an Economic Balance between in resultant error and the bias of non-response . JASA 48, 743—772.

DALENIUS, TORE (1974): The Ends and Means of Total Survey Design; Stockholm, The University of Stockholm.

---------------(1957): Sampling in sweden Contribution to the Methods and Theories of Sample Survey Practice, Stockholm, Almquist and Wicksell.

----------------(1962): Recent Advances in Sample Survey Theory and Methods, AMS, 33, 325—349.

----------------(1977a): Bibliography of non-sampling errors in Surveys. I(A—G), International Statistical Review, 3, 71—89.

----------------(1977b): Bibliography of non-sampling errors in Surveys II(A—Q), International Statistical Review, 45, 181—197.

-----------------(1977c): Bibliography of non-sampling errors in Surveys, III(R—Z), International Statistical Review, 45, 313—317.

DASGUPTA, A and MITRA, S.N. (1958): A Technical Note on Age Grouping. The National Sample Survey No.12, New Delhi.

DHAR, N.R. (1971): A note on non-sampling errors in NSS data. The National Sample Survey Working Paper No. 57/71/1.

DUNN, J.P. and HAWKES, R (1966): Comparison of non-respondents and respondents in a Periodic Health Examination Program to a mailed questionnaire, American Journal of Public Health, 56, 230—236.

DEMAIO, T.Y.(1980): Refusals, who where and why? Public Opinion Quarterly 44.

ERICKSON, W.A. (1967): "Optimal Sample Design with non-response", JASA, 62, 63—78.

EMRICH, LAWRENCE (1983): "Randomised Response Technique" In William G. Madow and Ingram olkin eds. Incomplete data in Sample Surveys; Volume 2, Theory and Bibliographies, New York, Academic, 73—80.

FELLEGI, IVAN P. (1963): The Evaluation of the Accuracy of Survey Results Some Canadian Experiences. International Statistical Review, 41, 1—14.

---------------(1964): Response Variance and its Estimation, JASA, 59, 1016—1041.

FELLEGI, IVAN and SUNTER, A.B. (1974): Balance between Different Sources of Survey Errors, Some Canadian Experiences, Sankhya, 36, Series C), 119—142.

FERVER, REBERT (1966): Items non-response in a consumer survey, Public Opinion Quarterly, 12, 669—676.

FORD, BARRY L. (1976): Missing Data Procedures, A Comparative Study, American Statistical Association, Proceedings of the Social Statistics Section 1976, Pt. 1, 326—329.

GHOSH, J.K. and MAITI, PULAKESH (2003): The Indian Statistical System at cross roads an appraisal of Past, Present and Future, To be presented at the IMS meet during 2—3 January - 2004.

GHOSH, A (1953): Accuracy of Family Budget Data with reference to period of re-call, Calcutta Statistical Association Bulletin, 5, 16—23.

GOWER, A.R. (1979): Characteristics of non-respondents in the Labour Force Survey, Statistics Canada.

GROVES, ROBERT, M. and KAHN ROBERT LOUIS (1979): Surveys by Telephone, A national comparison with personal interview, New York; Academic.

GRAY, P. and GEE, F.E.N. (1972): A Quality check on the 1966 ten percent sample census of England Wales, office of the population census and surveys, London.

GHOSH, J.K., MAITI, P. MUKHOPADHYAY, A.C., PAL, M.P (1977): Stochastic Modeling and Forecasting of Discovery, Reserve and Production of Hydrocarbon-with an application, Sankhya, Series B, 59, pt. 3, 288—312.

HANSEN, M.H., MADOW WILLIAM G., and TEPPING B.J. (1983): An Evaluation of Model dependent and Probability Sampling inference in Sample Surveys, JASA, 78, 776—807.

HANSE, M.H., HURWITZ WILLIAM N. (1946): The Problem of non-response in Sample Survey, JASA, 41, 516—529.

---------and NISSELSON, H., STEINBERG, J. (1955): The redesign of the current population survey, JASA, 50, 701—719.

----------JUBINE, TOMAS B. (1963): The use of imperfect lists for Probability Sampling at U.S. Bureau of Census, Bulletin of the International Statistical Institute, 40(1), 497—517.

--------and PRITZKER, LENON (1964): The Estimation and interpretation of Gross differences and the simple response variance. In C.R. Rao with D.B. Lahiri, K—P.

NAIR, P. PANT and S.S. SHRIKHANDE eds. Contributions to Statistics Presented to Professor P.C. Mahalanobis on the occasion of his 70th birth day

Oxford, England, Pergaman, Calcutta Statistical Publishing Society, 111—136.

----------and BERSHAD, MAX A. (1961): Measurement errors in censuses and surveys, Bulletin of the International Statistical Institute, 38, 359—374.

----------MARKS, ELIS MAULDIN, PARKER W. (1951): Response Errors in Surveys, JASA, 46, 147—190.

------------(1976): Some Important Events in the Historical Development of Sample Surveys in Donald Bruce Owen ed., on the History of Statistics and Probability, Statistics Text Books and Monographs, Volume 17, New York Dekker, 73—102.

HURSCGBERG, DAVID FREDERICK, J. SCHEUREN and YUSKAVAGE ROBERT (1977): the impact on Personal and Family income of adjusting the current population survey for under coverage, Proceedings of the Social Statistics Section, American Statistical Association, 70—80.

HUBBACK, J.A. (1927): Sampling for rice yields in Bihar and Orissa, Imp. Agr. Res. Inst. Bulletin, Pusha (reprinted in Sankhya (1946), 7, 282—294)

HALDEN, J.B.S. (1957): The Syadvada System of Prediction, Sankhay 18, 195—2000.

HACKING, J. (1965): Logic of Statistical Inference Cambridge University Press.

HOINVILLE, GERALD and ROBERT JOELL (1978): Survey Research Practic , London, Heinemann.

JESSEN, RAYMUND J. (1978): Statistical Survey Techniques, New York, Wiley.

KIAWER, A. (1895): Observations et experiences concernant des denombrements representatives, Bull. Int. Statist. Inst. 9, 176—183.

KRUSKEL, WILLIAM and FREDERICK MOSTELLER (1980): Representative Sampling, IV, the History of the concept in Statistics, 1895—1939, International Statistical Review, 48, 169—195.

KISH, L. (1965): Survey Sampling Wiley and Sons, New York.

------- and HESS I. (1958): on non-coverage of sampling dwellings, JASA, 54, 509—524.

KALTON, GRAHAM and DANIEL KASPRZYK (1982): Imputing of missing survey Response, American Statistical Association 1982, Proceedings of the Section on Survey Research Methods, 22—31.

KOOP, J.C. (1974): Notes for a unified theory of estimation for sample surveys taking into account response errors, Metrika, 21, 19—39.

KALTON, GRAHAM (1983): Compensating for missing survey data Research Report Series, Ann. Arbor MI, Institute for Social Research, University of Michigan.

KENDAL, MAURICE GEORGE and WILLIAM R. BUCKLAND (1960): A Dictionary of Statistical Terms, 2nd edition, London, Oliver and Boyd.

LUBIN, B. LEVITT, E. and ZUCKERMAN, M. (1962): Some personality differences between respondents and non-respondents in a survey questionnaire, Journal of Consulting Psychology, 26—192.

LUNDBERG, G.A. and LARSEN, O.A. (1949): Characteristics of Hard-to-reach individuals in field surveys, Public Opinion Quarterly, 13, 487—494.

LYBERG, L. and RAPAPORT, E. (1979): Unpublished non-response problems at the national central Bureau of Statistics, Sweden.

LITTLE, RODERICK J.A. (1982): Models for non-response in Sample Surveys, JASA, 77, 237—250.

------------------------(1983): Super Population models for non-response, Part IV. In William G. Madow and Ingram Olkin eds. Incomplete data in Sample Surveys, Volume 2, Theory and Bibliographies, New York, Academic, 337—413.

LESSLER, J.T. (1974): A double sampling scheme model for eliminating measurement process bias and estimating measurement errors in surveys, Institute of Statistics Mimeo Series No. 949, University of North Carolina, New Chapel.

-------------------(1980): Errors associated with the frames, Proceedings of the American Statistical Association Section on Survey Research Methods, 125—130.

MADOW, W.G. NISSELSON, HAROLD and OLKIN, INGRAM (1983): Incomplete data on Sample Survey, Volume 1, Report and Case studies; New York, Academic.

MC. NEIL, JOHN M. (1981): Factors affecting the 1980 census content and the effort to develop a post census disability survey. Presented at the annual meeting of the American Public Health Association.

MAHALANOBIS, P.C. and LAHIRI, D.B. (1961): Analysis of errors in censuses and surveys, Bulletin of the International Statistical Institute, 38(2), 359—374.

----------------and SEN, S.B. (1954): On some aspects of the Indian National Sample Survey, Bulletin of the International Statistical Institute, 34, pt. 2.

MAHALANOBIS, P.C.(1944): On Large Scale Sample Surveys, Philosophical Transactions of Royal Society, 231—(B), 329—451.

-------------------(1946): Recent Experiments in Statistical Sampling in the Indian Statistical Institute.

---------------------(1941): A Sample Survey of the Acre-age under jute in Bengal, 4, 511—30.

---------------------(1954): The Foundations of Statistics, Dialectica, 8, 95—111 (reprinted in Sankhya 18, 183—194)

MAITI, P (1983): unpublished Ph.D. Thesis entitled Some Contributions to the Sampling Theory using auxiliary information" submitted to the Indian Statistical Institute, Calcutta.

----------, PAL, M. and SINHA B.K. (1992): Estimating unknown Dimensions of a Binary matrix with application to the estimation of the size of a mobile population. Statistics and probability, 220—233.

MOSER, CLAYS ADOLF and GRAHAM KALTON (1972): Survey methods in Social investigation , 2nd edition, New York, Basic Books.

MOONEY, H. (1962): On Mahalanobis' contributions to the development of sample survey theory and method in C.R. Rao et al. (eds) contributions of statistics, Pergamon Press.

------------------- (1967): Sampling Theory and Methods, Statistical Publishing Society, Calcutta.

NEYMAN JERZY (1934): On the two different aspects of the representative method, the method of stratified sampling and the method of purposive selection, J. Roy, Statist. SOC. 97, 5589625.

NETER, J. and WAKSBERG, J. (1965): Response errors in collection of Expenditures data by household interview. An Experimental Study Technical Report No. 11 U.S. Bureau of the Census.

NEWMAN, S. (1962): Difference between early and late respondents in a mailed survey, Journal of Advertising Research, volume 2, 37—39.

OGNIBENE, P. Traits affecting questionnaire response, Journal of Advertising Research Volume 10, 18—20.

PAN, J.S. (1951): Social Characteristics of respondents and non-respondents in a questionnaire study of later maturity, Journal of Applied Psychology, 35, 780—781.

POLITZ, A.N. and SIMMONS, W.R. (1949): An attempt to get Not-at Homes into the sample without call-backs, JASA, 44, 9—31.

PLATEK, R. (1977): Some factors affecting non-response, Survey Methodology, 3.

PLAN, V.T. (1978): A Critical appraisal of household surveys in Malaysia Multipurpose household survey in developing Countries, Development Centre, OECD, Paris.

REUSS, C.F. (1943): Differences between persons responding and not responding to mail questionnaires, American Sociological Review, 8, 433—438.

RAO, V.R. and SASTRY, N.S. (1975): Evolution of a total survey design, The Indian Experience, Invited paper presented to the International Association of Survey Statisticians Warsaw.

Report of Research Projects

(1975—76): Cost Benefit Analysis of Rural Electrification, Project Leader Professor J. Roy, Computer Science Unit, ISI, Calcutta.

(1977—78): Calcutta Urban Poverty Survey, Project Leader Professor J. Roy, Computer Science Unit, ISI, Calcutta.

(1988—89): A Survey on Domestic Tourists in Orissa, Project Co-ordinator, P. Maiti, ISI, Calcutta.

(1994—95): An Enquiry into the Quality of Life in five communities in selected districts of Rural West Bengal, Project Co-ordinator, P. Maiti, ISI, Calcutta.

(1995): Community attitudes and Preferences pertaining to cemetery and cremated related issues in the East Rand in the Republic of South Africa, CENSIAT, HSRC, Pretoria, South Africa, Principal Statistician – P. Maiti.

(1995): Survey of family and Community life in the Selected Communities of the cape Peninsula of the Republic of South Africa, CENSTAT< HSRC, Principal Statistician – P. Maiti.

(1995): the Socio-economic demographic and cultural pattern of the female labour force participation in the North West and the Cape; CENSTAT, HSRC South Africa, Principal Statistician – P. Maiti.

(1996): Stanza-Bopape Project; CENSTAT, HSRC, South Africa, Principal Statistician – P. Maiti.

(1998): Mid. Term Review of IPP—VIII in Calcutta Metropolitan Area, ISI, Calcutta, Survey Statistician – P. Maiti.

(2001, August): National Statistics Commission , Government of India.

ROSHWALB, ALAN (1982): Respondent Selection Procedures within Households, American Statistical Association 1982 Proceedings of the section on Survey Research Methods, 93—98.

RUBIN DONALD B. (1983): Conceptual issues in the presence of non-responses, In William G. Madow and Ingram Olkin eds. Incomplete data in Sample Surveys 2, Theory and Bibliographies, New York, Academic, 123—142.

-------------------(1977): formalizing Subjective notions about the effect of non-respondents in Sample Surveys", JASA, 72, 538—543.

---------------(1978): Multiple imputations in Sample Surveys – A Phenomenological Bayesian Approach to non-response, American Statistical Association 1978 proceedings of the Section on Survey Research Methods, 20—28.

---------------(1987): Multi imputation for non-response in Surveys, New York, Wiley.

RIZVI, M. HASEEB (1983): Hot-Deck Procedures Imputation in William G. Madow and Ingram Olkin eds., Incomplete Data in Sample surveys, 3, Proceedings of the symposium, New York, Academic, 351—352.

SARMA, V.R.R., RAO, G. D., AMBE, V.N. (1980): Non-responses in household surveys of National Sample Survey.

STEPHEN, FREDERICK F. (1948): History of the uses of Modern Sampling Procedures, JASA, 43, 12—39.

SMITH, T.M.F. (1976): The Foundations of Survey Sampling, A Review, JRSS, 139A, 183—195.

SARNDAL, C.E., SWENSSON, B. and WRETMAN, J. (1992): Model Associated Survey Sampling , Springer Verlag, New York, Inc.

SCOTT CHRISTOPHER (1973): Experiments on recall error in African Budget Surveys, paper presented to the International Association of Survey Statisticians, Viena.

SHAH, NASRA M. (1981): Data from Tables used in the paper presented at Weekly Seminar of East West Population Institute, October 28, Honolulu.

SUDMAN, SEYMOUR (1976): Applied Sampling, New York, Academic.

SUCHMAN, EDWARD A. (1962): An analysis of bias in Survey Research, Public Opinion Quarterly, 26, 102—111.

SCHEAFFER, RICHARD, L. MENDENHALL, WILLIAM and OTT LYAN (1979): Elementary Survey Sampling, 2nd edition, North Scituate, MA: Duxbury Press.

SZAMEITAT, KLEUS and SCHAFFER, KARL AUGUST (1963): Imperfect Frames in Statistics and the consequences for their use in sampling, Bulletin of the International Statistical Institute, 40, 517—544.

SINGH, BAHADUR, SEDRANSK, JOSEPH (1978): A two phase sampling design for estimating the finite population mean when there is non-response, In N. Krishnan Namboodiri ed. Survey Sampling and measurement, New York, Academic, 143—155.

THOMPSON, IB and SIRING, E. (unpublished): On the causes and effect of non-response, Norwegian Experiences, Central Bureau of Statistics, Norway.

TUYGAN, KUTHAN and CAVADOR, TEVFIK (1975): Comparison of self and presale responses related to children of ever married women. In laboratories for population Statistics Scientific Report Series No. 17, 22—28.

TURNER, ANTHONY G., WALTMEN, HENRY. F, FAY ROBERT and CARLSON BEVERLY (1977): Sample Survey Design in developing Countries – three illustrations of methodology, Bulletin of the International Statistical Institute.

U.S. Bureau of the Census (1974): Standards for the discussion and presentation of errors in data, Technical Report No. 32.

-------------------(1976): An overview of population and housing census evaluation programmes conducted at the U.S. Bureau of Census, Census Advisory Committee of the American Statistical Association.

VERMA, VIJAY (1980): Sampling for national fertility surveys, World fertility survey conference, London.

WARWICK, DONALD P. and CHARTES A. LININGER (1975): The Sample Survey, Theory and Practice, New York, Mc. Graw Hill.

WOLTMAN, HENRY and BUSHERY, JOHN (1975): A panel bias study to the national crime survey. Proceedings of the Social Statistics Section. American Statistical Association.

WILLIAM, W.H. and MALLOWS, C.L. (1970): Systematic biases in panel surveys JASA, 65, 1338—1349.

WARNER, STANLEY L. (1965): Randomised Response, A Survey Technique for eliminating Evasive answer bias, JASA, 60, 63—69.

ZARKOVICH. S.S. (1966): Quality of Statistical data, Rome; FAO of the United Nations.