ESTIMATION OF MEAN VECTOR IN PRESENCE OF NON-RESPONSE

T. P. Tripathi

Indian Statistical Institute,
203, B.T. Road, Calcutta, India

B.B. Khare

Department of Statistics
Banaras Hindu University
Varanasi-221005, India

Key words and Phrases : Mean vector; matrix mean squared error; non-response, cost function; subsampling.

ABSTRACT

Simultaneous estimation of means of several variables is considered for finite population in presence of non-response. Two types of nonresponses (partial and complete) are considered using the technique of sampling and subsampling with equal probabilities without replacement. The optimum sample size and the optimum value of subsampling fraction to be repeated from the nonresponding units of the sample have been obtained for fixed survey budget.

INTRODUCTION

In most of the socio-economic studies, several variables are considered simultaneously. For example, while conducting a household survey, the investigator may be interested in studying the characteristics such as number of wage earners, per capita income, land holding, number of illiterate persons, number of females etc. The problem of simultaneous estimation of means of several variables has been considered by Tripathi and Khattree (1989) and Chaubey and Tripathi (1993) based on SRSWOR, double sampling, general sampling designs and information on several auxiliary variables. The works related to the study of several variables assume that responses are available for each unit selected in the sample. In practice, however the problem of non-response often arises in sample surveys. In such situations while single survey variable is under investigation, the problem of estimating population mean using subsampling scheme has been first considered by Hansen and Hurwitz (1946). Further El-Badry (1956) has generalized the Hansen-Hurwitz method in another direction of several waves of follow up in case of mailed questionnaires.

In multivariate surveys, the non-response problem may assume complicated form. Broadly speaking, the non-response may be mainly of two types : (i) partial non-response and (ii) complete non-response.

If response from a unit i of the sample s is available only for some variables from the survey variables $y_1, y_2, \ldots, y_q$ while not available for other variables, such a situation may be termed as the situation of partial non-response.

The situation where for a unit selected in the sample, the response is either available for all variables or not at all available for any of the survey variables $y_1, y_2, \ldots, y_q$ may be characterised as the situation of complete non-response/response.

It may be noted that in case of complete non-response/ response situation, the number and the corresponding units belonging to respondents and non-respondents will be same for each of the variables $y_1, y_2, \ldots, y_q$ while in case of partial non-response situation it will vary from variable to variable. Further it may also be noted that there may be more complex situations where response is not available (or available) for any (all) of the survey variables in case of some units while it is not available (or available) only for some of the variables in case of some other units in the sample. However we have not considered these complicated situations in this paper.

In this paper we discuss the estimation of finite population mean vector in the presence of non-response using unrestricted and restricted simple random sampling without replacement (SRSWOR) considering the situation of partial non-response as well as that of complete non-response/response.

2. SAMPLING STRATEGY FOR MEAN VECTOR IN CASE OF COMPLETE NON-RESPONSE/RESPONSE

Let the finite population U = {1,2,...N} under c nsideration consist of N known units labelled 1 through N and $\underline{y}_i$ d notes the value of the vector variable $\underline{y} = (y_1, y_2, \ldots, y_q)$ for u it i.

We consider the problem of estimating the finite population m n vector

$$\underline{\mu} = \frac{1}{N} \sum_{i \in U} \underline{y}_i = (\mu_j) : q \times 1;\ \mu_j = \frac{1}{N} \sum_{i \in U} y_{ji}; \quad j = 1,2,\ldots q \tag{2.1}$$

i the presence of non-response.

1. PROPOSED SAMPLING AND ESTIMATION PROCEDURE

For the survey variable vector $\underline{y}$, the population U may be thought as consisting of two domains : $U_{(1)}$, the domain of respondents of $\underline{y}$ consisting of, (say) N_{01} (unknown) units and $U_{(2)}$, the domain of nonrespondents on $\underline{y}$ consisting of $N_{02} = N - N_{01}$ uni . We have

$$U = \{U_{(1)}, U_{(2)}\} \tag{2.2}$$

nd

$$\underline{\mu} = W_{01}\, \underline{\mu}_{01} + W_{02}\, \underline{\mu}_{02},\ W_{01} + W_{02} = 1 \tag{2.3}$$

where $\underline{\mu}_{oh} = (\mu_{1h}, \mu_{2h}, \ldots \mu_{qh})'$; $\mu_{jh} = \frac{1}{N_{oh}} \sum_{i \in U_{(h)}} y_{ji}$, $h = 1,2$;

$j = 1, 2, \ldots q$ and W_{01}, W_{02} be the population response rate and population non-response rate, which is same for all the variables.

Let a sample s of size n (known) be selected from U using SRSWOR method of sampling. Let $\underset{\sim}{s}$ and $\underset{\sim}{n}$ be represented by

$$\underset{\sim}{s} = (s_{(1)}, s_{(2)}) \ , \ \underset{\sim}{n} = (n_{(1)}, n_{(2)}) \qquad (2.4)$$

which is same for each survey variable y_j $(j = 1,2,\ldots q)$, where $s_{(1)}$ is the sample consisting of $n_{(1)}$ responding units for the survey variable vector $\underline{y}$ from the sample s of size n and $s_{(2)}$ is the sample consisting of $n_{(2)}$ nonresponding units obtained from the sample s for $\underline{y}$.

Let $w_{oh} = n_{(h)}/n$, $0 < w_{oh} < 1$; $h = 1,2$, where w_{01} and w_{02} are the response and non-response rate in the sample s, which is same for all $y_j (j = 1,2,\ldots q)$. Let a subsample $s^{*}_{(2)}$ of size $m = \nu n_{(2)}$, $(0 < \nu < 1)$ is drawn from $s_{(2)}$ and the values of $\underline{y}$ on these m units are obtained. It is to be noted here that these m units of $s^{*}_{(2)}$ are same for all y_j $(j = 1,2,\ldots q)$. It is assumed here that the necessary efforts have been made so that the responses are made available for each $i \in s^{*}_{(2)}$ on all y_j $(j = 1,2,\ldots q)$.

Illustration 2.1 : Situation of complete non-response/response ($q = 5$, $n = 10$)

Units in s

Variables	1	2	3	4	5	6	7	8	9	10
y_1	0	⊗	0	0	0	⊗	0	⊗	0	0
y_2	0	⊗	0	0	0	⊗	0	⊗	0	0
y_3	0	⊗	0	0	0	⊗	0	⊗	0	0
y_4	0	⊗	0	0	0	⊗	0	⊗	0	0
y_5	0	⊗	0	0	0	⊗	0	⊗	0	0

$n_{(1)} = 7$, $n_{(2)} = 3$, 0 : response, ⊗ : non-response.

An unbiased estimator of the mean vector $\underset{\sim}{\mu}$ is given by

$$\hat{\underset{\sim}{\mu}} = (\hat{\mu}_{j0}) = w_{01}\, \bar{\underline{y}}_{01} + w_{02}\, \bar{\underline{y}}_{02(m)}, \quad w_{01} + w_{02} = 1, \qquad (2.5)$$

where $\mu_{j0} = w_{01}\, \bar{y}_{j1} + w_{02}\, \bar{y}_{j(m)}$, $\bar{\underline{y}}_{01} = (\bar{y}_{11}, \bar{y}_{21}, \bar{y}_{31}, \ldots \bar{y}_{q1})'$,

$\bar{\underline{y}}_{02(m)} = (\bar{y}_{1(m)}, \bar{y}_{2(m)}, \bar{y}_{3(m)} \ldots \bar{y}_{q(m)})'$, $\bar{y}_{jh} = \dfrac{\sum_{i \in s_{(h)}} y_{ji}}{n_{(h)}}$, $h = 1,$ 2 and $\bar{y}_{j(m)} = \dfrac{\sum_{i \in s^*_{(2)}} y_{ji}}{m}$.

It is to be noted here that $\hat{w}_{01}$ and w_{02} are unbiased for W_{01} and W_{02}. It is easy to see that $\hat{\underset{\sim}{\mu}} = (\hat{\mu}_{j0})$ is an unbiased estimator of $\underset{\sim}{\mu}$. Since the values of the variable y_j and y_k are same for the units selected in $s^*_{(2)}$, so the cov $(\bar{y}_{j(m)}, \bar{y}_{k(m)}) \neq 0$.

2.2. MATRIX MEAN SQUARED ERROR (MMSE) OF $\hat{\underset{\sim}{\mu}}$ AND ITS ESTIMATOR

Theorem 2.1. The matrix mean squared error of $\hat{\underset{\sim}{\mu}}$ is given by

$$\underline{M}(\hat{\underset{\sim}{\mu}}) = E(\hat{\underset{\sim}{\mu}} - \underset{\sim}{\mu})\,(\hat{\underset{\sim}{\mu}} - \underset{\sim}{\mu})'$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right)S + \frac{W_{02}\,(\frac{1}{\nu} - 1)}{n} S_{(2)}\,, \qquad (2.6)$$

where $S = (S_{jk})$, $S_{(2)} = (S_{jk(2)})$: $q \times q$, $j,k = 1,2,\ldots q$

$$S_{jk} = \begin{cases} S_j^2 = \frac{1}{N-1} \sum_{i \in U} (y_{ji} - \mu_j)^2 & \text{for } j = k \\ S_{jk} = \frac{1}{N-1} \sum_{i \in U} (y_{ji} - \mu_j)(y_{ki} - \mu_k) & \text{for } j \neq k \end{cases}$$

and

$$S_{jk(2)} = \begin{cases} S^2_{j(2)} = \frac{1}{N_{02}-1} \sum_{i \in U_{(2)}} (y_{ji} - \mu_{j2})^2 & \text{for } j=k \\ S_{jk(2)} = \frac{1}{N_{02}-1} \sum_{i \in U_{(2)}} (y_{ji} - \mu_{j2})(y_{ki} - \mu_{k2}) & \text{for } j \neq k. \end{cases}$$

Following Rao (1990), the unbiased estimators of S^2_j and S_{jk} are given as follows :

$$\hat{S}^2_j = \frac{1}{n-1} \left[(n_{(1)} - 1)\, s^2_{j(1)} + \left\{ n_{(2)}\left(1 - \frac{1}{m}\right) + \frac{w_{02}(\frac{1}{\nu} - 1)}{n} \right\} s^2_{jm} \right.$$

$$\left. + \, n\, w_{01}\, w_{02} (\bar{y}_{j1} - \bar{y}_{j(m)})^2 \right], \tag{2.7}$$

$$\hat{S}_{jk} = \frac{1}{n-1} \left[(n_{(1)} - 1)\, s_{jk(1)} + \left\{ n_{(2)}\left(1 - \frac{1}{m}\right) + \frac{w_{02}(\frac{1}{\nu} - 1)}{n} \right\} s_{jk(m)} \right.$$

$$\left. + \, n\, w_{01}\, w_{02} (\bar{y}_{j1} - \bar{y}_{j(m)})(\bar{y}_{k1} - \bar{y}_{k(m)}) \right], \tag{2.8}$$

where

$$s^2_{j(1)} = \frac{1}{(n_{(1)}-1)} \sum_{i \in s_{(1)}} (y_{ji} - \bar{y}_{j1})^2 ,$$

$$s^2_{jm} = \frac{1}{m-1} \sum_{i \in s^*_{(2)}} (y_{ji} - \bar{y}_{j(m)})^2 ,$$

$$s_{jk(1)} = \frac{1}{(n_{(1)}-1)} \sum_{i \in s_{(1)}} (y_{ji} - \bar{y}_{j1})(y_{ki} - \bar{y}_{k1}),$$

and

$$s_{jk(m)} = \frac{1}{m-1} \sum_{i \in s^*_{(2)}} (y_{ji} - \bar{y}_{j(m)})(y_{ki} - \bar{y}_{k(m)}).$$

The unbiased estimators of $S^2_{j(2)}$ and $S_{jk(2)}$ are given by

$$\hat{S}^2_{j(2)} = s^2_{jm}, \qquad \hat{S}_{jk(2)} = s_{jk(m)}. \tag{2.9}$$

Now we have the following results.

Theorem 2.2. The unbiased estimator of $MMSE(\hat{\underset{\sim}{\mu}})$ is given by

$$\hat{\underline{M}}(\hat{\underset{\sim}{\mu}}) = \left(\frac{1}{n} - \frac{1}{N}\right)\hat{S} + \frac{W_{02}\left(\frac{1}{\nu} - 1\right)}{n}\hat{S}_{(2)}, \tag{2.10}$$

where $\hat{S} = (\hat{S}_{jk})$, $\hat{S}_{(2)} = (\hat{S}_{jk(2)})$: $q \times q$; $j,k = 1,2,\ldots q$

$$\hat{S}_{jk} = \begin{cases} \hat{s}_j^2 & \text{for } j = k \\ \hat{s}_{jk} & \text{for } j \neq k \end{cases}$$

and

$$\hat{S}_{jk(2)} = \begin{cases} \hat{s}_{j(2)}^2 & \text{for } j = k \\ \hat{s}_{jk(2)} & \text{for } j \neq k \end{cases}$$

2.3. OPTIMAL SAMPLE SIZE AND SUBSAMPLING FRACTION (FOR FIXED BUDGET)

The cost function is given by

$$C = nC_0 + n_{(1)} C_{01} + \nu\, n_{(2)} C_{02} \tag{2.11}$$

where

C_0 : Cost per unit for mailing the questionnaire or visiting units selected in the sample,

C_{01} : Cost per unit for collecting and processing the data on vector variable $\underline{y}$ for the responding units in the sample at the first attempt,

C_{02} : Cost per unit for collecting and processing the data on selected subsample units from $n_{(2)}$ non-responding units for the vector variable $\underline{y}$.

Since the values of $n_{(1)}$ and $n_{(2)}$ varies from sample to sample, so the expected total cost apart from overhead cost is expressed as

$$C' = E(C) = n\,[C_0 + W_{01} C_{01} + \nu W_{02} C_{02}] \tag{2.12}$$

For determining the optimum value of n and ν which minimizes the $\operatorname{tr} \underset{\sim}{M}(\hat{\mu})$ for the given fixed cost C_0', we define a function

$$\phi = \operatorname{tr} \underset{\sim}{M}(\hat{\mu}) + \lambda\,(C' - C_0'), \tag{2.13}$$

where λ is a Lagranges multiplier.

Now differentiating ϕ with respect to n and ν and equating to zero (noting that $\left(\frac{\partial^2\phi}{\partial n^2}, \frac{\partial^2\phi}{\partial \nu^2} > 0\right)$ and using $C' \le C_0'$, we have the following results.

Theorem 2.3. The optimum values of n and ν which minimizes the $MSE(\hat{\underset{\sim}{\mu}})$ for $C' \le C_0'$ is given by

$$n_{opt} = \frac{C_0' \sqrt{A}}{\sqrt{B}\left[\sqrt{AB} + \left\{\sqrt{C_{02} \sum_{j=1}^{q} S_{j(2)}^2}\right\} W_{02}\right]} \tag{2.14}$$

$$\nu_{opt} = \sqrt{\frac{B \sum_{j=1}^{q} S_{j(2)}^2}{C_{02}\, A}} \tag{2.15}$$

where $A = C_0 + W_{01} C_{01}$ and $B = \sum_{j=1}^{q} (S_j^2 - W_{02} S_{j(2)}^2)$.

'ow putting the value of n_{opt} and ν_{opt} from (2.14) and (2.15) in 2.6), the minimum value of $\underset{\sim}{M}(\hat{\mu})$ can be obtained.

: SAMPLING STRATEGY FOR MEAN VECTOR IN CASE OF PARTIAL NON-RESPONSE

: 1. Proposed Sampling and Estimation Procedure

For each survey variable y_j ($j = 1,2,\ldots q$), the population may be thought as consisting of two domains : $U_{j(1)}$, the domain c respondents consisting of, say $N_{j(1)}$ (unknown) units and $U_{j(2)}$, the domain of non-respondents consisting of $N_{j(2)} = N - N_{j(1)}$ units with respect to y_j. This leads to q proportion of U.

$$U = \{U_{j(1)}, U_{j(2)}\}, \qquad j = 1,2,\ldots q \tag{3.1}$$

We may express

$$\mu = W_{(1)}\mu_{(1)} + W_{(2)}\mu_{(2)}, \quad W_{(1)} + W_{(2)} = I, \tag{3.2}$$

where $W_{(h)}$: diag $(W_{1(h)}, W_{2(h)}, \ldots, W_{q(h)})$: q × q, is the diagonal matrix, I is q × q identity matrix,

$$\mu_{(h)} = (\mu_{1(h)}, \mu_{2(h)}, \ldots, \mu_{q(h)})', \quad \mu_{j(h)} = \frac{1}{N_{j(h)}} \sum_{i \in U_{j(h)}} y_{ji}, \quad h=1,2,$$

$W_{j(1)} = N_{j(1)}/N$ is the population response rate for y_j and

$W_{j(2)} = 1 - W_{j(1)}$ is the population non-response rate for y_j.

We follow the sampling technique suggested by Hansen and Hurwitz (1946). Let a sample s of size n (known) be selected from U according to SRSWOR. Let s and n have the representation

$$\underset{\sim}{s} = (s_{j(1)}, s_{j(2)}), \quad \underset{\sim}{n} = (n_{j(1)}, n_{j(2)}) \tag{3.3}$$

for each survey variable y_j $(j = 1,2,\ldots,q)$, where $s_{j(h)}$ consisting of $n_{j(h)}$ units is the sample representation from $U_{j(h)}$, $h = 1,2$. Let $w_{j(h)} = n_{j(h)}/n$, where $w_{j(1)}$ and $w_{j(2)}$ are the response and non-response rates in the sample s respectively in the survey, with respect to the survey variable y_j.

From non-response part $s_{j(2)}$ of s for y_j; let a subsample $s^*_{j(2)}$ of $m_j = \nu_j\, n_{j(2)}$, $0 \le \nu_j \le 1$, units be selected for each j respectively using equal probability selection without replacement, where ν_j is fixed a priori. We assume that necessary efforts have been made so that the responses are available for each $i \in s^*_{j(2)}$. It may be noted that only y_j is obtained for units in $s^*_{j(2)}$.

Illustration 3.1 : Situation of partial non-response (q=5, n=10)

	Units in s									
Variables	1	2	3	4	5	6	7	8	9	10
y_1	⊗	0	0	0	⊗	0	⊗	⊗	0	0
y_2	0	⊗	⊗	0	0	⊗	0	⊗	0	0
y_3	0	⊗	⊗	⊗	⊗	0	0	⊗	⊗	0
y_4	⊗	0	⊗	0	0	⊗	0	0	0	0
y_5	0	0	0	0	0	0	⊗	0	0	⊗

Here

$$n_{1(1)} = 6,\ n_{2(1)} = 6,\ n_{3(1)} = 4,\ n_{4(1)} = 7,\ n_{5(1)} = 8$$

$$n_{1(2)} = 4,\ n_{2(2)} = 4,\ n_{3(2)} = 6,\ n_{4(2)} = 3,\ n_{5(2)} = 2.$$

An unbiased estimator of the mean vector $\underline{\mu} = (\mu_1, \mu_2, \ldots \mu_q)'$ is given by

$$\hat{\underline{\mu}} = (\hat{\mu}_j) = w_{(1)}\bar{y}_{(1)} + w_{(2)}\bar{y}^*_{(2)},\ w_{(1)} + w_{(2)} = I \qquad (3.4)$$

where

$$\hat{\mu}_j = w_{j(1)}\ \bar{y}_{j(1)} + w_{j(2)}\ \bar{y}^*_{j(2)},\quad \bar{y}_{(1)} = (\bar{y}_{1(1)}, \bar{y}_{2(1)}, \ldots \bar{y}_{q(1)})',$$

$$\bar{y}^*_{(2)} = (\bar{y}^*_{1(2)}, \bar{y}^*_{2(2)}, \ldots \bar{y}^*_{q(2)})',\quad \bar{y}_{j(h)} = \frac{1}{n_{j(h)}} \sum_{i \in s_{j(h)}} y_{ji},$$

$$\bar{y}^*_{j(2)} = \frac{1}{m_j} \sum_{i \in s^*_{j(2)}} y_{ji}$$

and $w_{(h)} = \mathrm{diag}(w_{1(h)}, w_{2(h)}, \ldots w_{q(h)}) : q \times q,\ (h = 1,2)$ is the diagonal matrix. It may be noted that $w_{(h)}$ is unbiased for $W_{(h)}$.

Since $s^*_{j(2)}$ are selected separately from $s_{j(2)}$ for each $j = 1,2,\ldots q$ and y_j is observed on $s^*_{j(2)}$ and the probability of selecting a unit for $s^*_{j(2)}$ is $1/n_{j(2)}$ which varies for $j = 1, 2, \ldots q$, so $\mathrm{cov}(\bar{y}^*_{j(2)}, \bar{y}^*_{k(2)}/n, m_j, m_k) = 0$ and we have $\mathrm{cov}\ (\hat{\mu}_j, \hat{\mu}_k) = \left(\frac{1}{n} - \frac{1}{N}\right) S_{jk}$.

3.2. MMSE($\hat{\underline{\mu}}$) AND ITS ESTIMATOR

Theorem 3.1. The matrix mean squared error (MMSE) of $\hat{\underline{\mu}}$ is given by

$$\underline{M}(\hat{\underline{\mu}}) = E(\hat{\underline{\mu}} - \underline{\mu})\ (\hat{\underline{\mu}} - \underline{\mu})'$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) S + S_2, \qquad (3.5)$$

where $S = (S_{jk}),\ S_2 = (S^*_{jk(2)}) : q \times q, \qquad j,k = 1,2,\ldots q;$

$$S_{jk} = \begin{cases} S_j^2 & \text{for } j = k \\ S_{jk} & \text{for } j \neq k \end{cases},$$

$$S^{*}_{jk(2)} = \begin{cases} \dfrac{W_{j(2)}\left(\dfrac{1}{\nu_j} - 1\right) S^{*2}_{j(2)}}{n} & \text{for } j = k \\ 0 & \text{for } j \neq k \end{cases}$$

and $S^{*2}_{j(2)} = \dfrac{1}{(N_{j(2)} - 1)} \sum_{i \in U_{j(2)}} (y_{ji} - \mu_{j(2)})^2.$

Following Rao (1990), in this situation the unbiased estimator of S_j^2 will be given by

$$\hat{S}_j^2 = \frac{1}{n-1}\left[(n_{j(1)} - 1)\, s^{*2}_{j(1)} + \left\{\frac{n_{j(2)}(m_j - 1)}{m_j} + \frac{w_{j(2)}(\frac{1}{\nu_j} - 1)}{n}\right\} s^2_{(m_j)} + n(w_{j1}\, w_{j2}(\bar{y}_{j(1)} - \bar{y}^{*}_{j(2)})^2\right] \qquad (3.6)$$

where

$$s^{*2}_{j(1)} = \frac{1}{(n_{j(1)} - 1)} \sum_{i \in s_{j(1)}} (y_{ji} - \bar{y}_{j(1)})^2,$$

$$s^2_{(m_j)} = \frac{1}{m_j - 1} \sum_{i \in s^{*}_{j(2)}} (y_{ji} - \bar{y}^{*}_{j(2)})^2.$$

Since $n_{j(1)} \neq n_{k(1)}$, $m_j \neq m_k$ and the units belonging to $s_{j12} = \{s_{j(1)} \cup s^{*}_{j(2)}\}$ are also not same as the units of $s_{k12} = \{s_{k(1)} \cup s^{*}_{k(2)}\}$. So in this situation it is a difficult problem to estimate S_{jk}. Further let us assume that $m_{jk} (\neq 0)$ units are common in s_{j12} and s_{k12} on which the data on y_j and y_k variables are available. So we can provide an estimate of S_{jk} on the basis of available data on y_j and y_k. We know $S_{jk} = \frac{1}{N-1}\,[T - N\mu_j\mu_k]$, where $T = \sum_{i\in U} y_{ji}\ y_{ki}$. Now ignoring the term $\frac{1}{N}$ for large N and using $\hat{T} = N\left[\sum_{i\in s_{j12}\cap s_{k12}} y_{ji}\ y_{ki}\right]/m_{jk}$, $\hat{\mu}_j = \frac{n_{j(1)}\ \bar{y}_{j(1)}}{n} + \frac{n_{j(2)}\bar{y}^*_{j(2)}}{n}$ and $\hat{\mu}_k = \frac{n_{k(1)}}{n}\bar{y}_{k(1)} + \frac{n_{k(2)}}{n}\bar{y}^*_{k(2)}$, we have

$$\hat{S}_{jk} = \frac{1}{m_{ij}}\sum_{i\in s_{j12}\cap s_{k12}} y_{ji}\ y_{ki} - \frac{1}{n^2}\,(n_{j(1)}\ \bar{y}_{j(1)} + n_{j(2)}\ \bar{y}^*_{j(2)})$$

$$(n_{k(1)}\ \bar{y}_{k(1)} + n_{k(2)}\ \bar{y}^*_{k(2)})\ . \tag{3.7}$$

Now we have the following results.

Theorem 3.2. The estimator of $\underline{M}(\hat{\underline{\mu}})$ is given by

$$\hat{\underline{M}}(\hat{\underline{\mu}}) = \left(\frac{1}{n} - \frac{1}{N}\right)\hat{S} + \hat{S}_2, \tag{3.8}$$

where $\hat{S} = (\hat{S}_{jk})$, $\hat{S}_2 = (\hat{S}^*_{jk(2)}) : q \times q$; $\quad j,k = 1,2,\ldots q$

$$\hat{S}_{jk} = \begin{cases} \hat{s}^2_j & \text{for } j = k \\ \hat{S}_{jk} & \text{for } j \neq k \end{cases},$$

and

$$\hat{S}^*_{jk(2)} = \begin{cases} \dfrac{W_{j(2)}\left(\frac{1}{\upsilon_j} - 1\right)\sum_{i\in s^*_{j(2)}}(y_{ji} - \bar{y}^*_{j(2)})^2}{n(m_j - 1)} & \text{for } j = k \\ 0 & \text{for } j \neq k. \end{cases}$$

3.3. OPTIMAL SAMPLE SIZE AND SUBSAMPLING FRACTIONS (FIXED BUDGET)

Let C be total cost of the survey apart from overhead cost and the cost function is given by

$$C = n\, C_0 + \sum_{j=1}^{q} n_{j(1)}\, C_{j(1)} + \sum_{j=1}^{q} n_{j(2)}\, C_{j(2)}\, \nu_j \qquad (3.9)$$

where

- C_0 : Cost per unit for mailing questionnaire or visiting the units selected in the sample,
- $C_{j(1)}$: Cost per unit for collecting and processing the data on responding units of y_j variable,
- $C_{j(2)}$: Cost per unit for collecting and processing the data on the subsample units from $n_{j(2)}$ non-responding units of y_j variable.

The values of $n_{j(1)}$ and $n_{j(2)}$ varies from sample to sample, so the expected total cost apart from overhead cost will be given by

$$C^{*} = E(C'') = n\left[C_0 + \sum_{j=1}^{q} W_{j(1)}\, C_{j(1)} + \sum_{j=1}^{q} W_{j(2)}\, C_{j(2)}\, \nu_j\right] \qquad (3.10)$$

Since the values of $\underline{M}(\hat{\underline{\mu}})$ depends on the values of n and ν_j (j = 1,2,...q) and also on the total cost C_0^{*}. So for determining the optimum values of n and ν_j which minimizes the $\underline{M}(\hat{\underline{\mu}})$ for the given cost C_0^{*}, we define a function ϕ which is given as

$$\phi = \mathrm{tr}\ \underline{M}(\hat{\underline{\mu}}) + \lambda\ (C^{*} - C_0^{*}), \qquad (3.11)$$

where λ is a Lagranges multiplier.

Now minimizing ϕ with respect to n and ν_j for $C^{*} \le C_0^{*}$, we have the following result.

Theorem 3.3. The optimum value of n and ν_j which minimizes the $\underline{M}(\hat{\underline{\mu}})$ for the given cost $C^{*} \le C_0^{*}$ is given by

$$n_{opt} = \frac{C_0^{*} \sqrt{A_1}}{\sqrt{B_1} \left\{ \sqrt{A_1 B_1} + \sum_{j=1}^{q} \left(\sqrt{C_{j(2)}\, S^2_{j(2)}} \right) W_{j(2)} \right\}} \tag{3.12}$$

and

$$\nu_{j\ opt} = \sqrt{\frac{B_1\, S^2_{j(2)}}{C_{j(2)}\, A_1}} \tag{3.13}$$

where $A_1 = \sum_{j=1}^{q} (S_j^2 - W_{j(2)}\, S^2_{j(2)})$ and $B_1 = C_0 + \sum_{j=1}^{q} W_{j(1)}\, C_{j(1)}$.

It is easy to see that for $q = 1$, we get the values of n_{opt} and ν_{opt} which is same as in case of Hansen and Hurwitz (1946). The minimum value of $\underline{M}(\hat{\underline{\mu}})$ can be obtained by putting the value of n_{opt} an $\nu_{j\ opt}$ from (3.12) and (3.13) in (3.5).

ACKNOWLEDGEMENT

The authors are thankful to the Associate Editor for his va uable suggestions.

BIBLIOGRAPHY

C ibey, Y.P. and Tripathi, T.P. (1993). "Estimation of finite population mean vector using multivariate auxiliary in formation". Technical Report # 8/93, Stat Math Division, ISI, Calcutta, India.

l adry, M.A. (1956). "A sampling procedure for mailed questionnaires". Journal of the American Statistical Association, 51, 209-227.

a sen, M.H. and Hurwitz, W.N. (1946). "The problem of non-response in sample surveys". Journal of the American Statistical Association, 41, 517-529.

o, P.S.R.S. (1990). "Regression estimators with subsampling of nonrespondents. Data Quality Control, Theory and

Pragmatics (Gunar. e. Liepins and V.R.R. Uppuluri, Eds) Marcel Dekker, New York, 191-208.

Tripathi, T.P. and Khattree, R. (1989). "Simultaneous estimation of several means using multivariate auxiliary information". Technical Report # 13/89, Stat - Math Division, ISI, Calcutta, India.

Received November, 1995; Revised March, 1997.