Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/050210

Estimation of Mean of Finite Population Using Double Sampling Scheme under Non-Response

Manoj K. Chaudhary and Amit kumar*

Department of Statistics, Banaras Hindu University, Varanasi, India

Received: 1 Jul. 2015, Revised: 21 Mar. 2016, Accepted: 23 Mar. 2016
Published online: 1 Jul. 2016

Abstract: This paper presents a theoretical study on the estimation of finite population mean in simple random sampling using double sampling scheme under non-response. We have proposed the estimators of population mean utilizing the information on an auxiliary variable under the situation in which both study and auxiliary variables are suffered from non-response. The expressions for the biases and mean square errors of the proposed estimators up to the first order of approximation have been obtained. A theoretical study on the cost of the survey has been carried out. An empirical study has also been carried out to demonstrate the performances of the proposed estimators.

Keywords: Simple random sampling, double sampling scheme, population mean, auxiliary variable and non-response.

1 Introduction

The problem of non-response is a big issue in the mail surveys. In a selected sample of specified size, some of the units may not respond or these may not be contacted during the survey period. Due to this reason, the prescribed size of the sample may be reduced and hence the precision of the estimator may decrease in estimating the characteristics of the population. It is obvious that the error due to non-response is not so important if the non-responding units are similar in their characteristics to those of responding units. But, the non-similarity is inherent between the groups of responding and non-responding units in the population, it is imperative to consider the seriousness of non-response. Hansen and Hurwitz [1] were the first who tackled the seriousness of non-response in estimating the population mean. They introduced a technique of sub-sampling of non-respondents to deal with the problem of non-response and its adjustments.

The auxiliary information may be utilized in order to estimate the population characteristics if it is highly correlated with the study character. There is a lot of works which have been discussed in estimating the population parameters using auxiliary information in the presence of non-response. Khare [2] has suggested the estimation technique for estimating the population mean under optimum allocation in the presence of non-response. Khan *et al*. [3] have conferred the procedure of optimum allocation while conducting a mail survey in multivariate stratified random sampling under non-response. Chaudhary *et al*. [4] and Chaudhary and Singh [5] have suggested the families of estimators of population mean in various sampling schemes viz. stratified random sampling and two-stage sampling over non-response.

The auxiliary information may easily be utilized to improve the precision of the estimator for study character if the parametric values of auxiliary characters are known. The situations in which the parametric values of auxiliary characters are not known, one can utilize the procedure of double (two-phase) sampling scheme in estimating the parameters of study character. The two-phase sampling involves the method of selecting a larger sample for gathering the information on auxiliary character and then selecting a sub-sample from it for collecting the information on study character. Khare and Sinha [6] have proposed some estimators of population ratio in double sampling scheme under non-response. Singh and Kumar [7] have recommended a general class of estimators of population mean adopting two-phase sampling under non-response. Chaudhary *et al.* [8] have suggested a class of estimators for assessing the population mean using double sampling scheme in the presence of non-response. Recently, Chaudhary and Kumar [9] have proposed a class of estimators for estimating the mean of a stratified population in two-phase sampling scheme under non-response.

* Corresponding author e-mail: ritamanoj15@gmail.com

In most of the mail surveys, it is generally seen that the non-response occurs on both the study and auxiliary variables and leads to reduce the efficiency of the estimators. In such situations, it is very difficult to estimate the population parameters of study character. Thus, the present study aims to suggest the estimation procedures for estimating the finite population mean under the situation in which both study and auxiliary variables are suffered from non-response. The study about the cost of survey has also been carried out.

2 Proposed Estimators

Let us suppose that a population consists of N units $(U_1, U_2, ..., U_N)$. Let Y and X be the characteristics under study and auxiliary information respectively. Let their respective population means be $\overline{Y}$ and $\overline{X}$. Let $y_i\,(i = 1, 2, ...N)$ and $x_i\,(i = 1, 2, ...N)$ be the observation on the i^{th} unit in the population for study and auxiliary variables respectively. Let us assume that both study and auxiliary variables are suffered from non-response and the information on $\overline{X}$ is not available. In such situation, we use the two-phase sampling scheme to estimate the population mean$\overline{Y}$. Let a sample of $n^{'}$ units be selected from N units by simple random sampling without replacement (SRSWOR) scheme at the first phase and then a smaller sub-sample of n units be selected from $n^{'}$ $(n < n^{'})$ units by SRSWOR at the second phase. At the first phase, it is observed that there are $n_1^{'}$ units respond and $n_2^{'}$ units do not respond on auxiliary variable. Now, we select a sub-sample of $h_2^{'}$ units from the $n_2^{'}$ units by SRSWOR $(h_2^{'} = \frac{n_2^{'}}{L^{'}}, L^{'} > 1)$ where $L^{'}$ is the inverse sampling rate at the first phase and collect the information on all the $h_2^{'}$ units [see Hansen and Hurwitz [1]]. Thus, the estimate of $\overline{X}$ at the first phase is given by

$$\overline{x}^{'*} = \frac{n_1^{'}\overline{x}_{n1}^{'} + n_2^{'}\overline{x}_{h2}^{'}}{n^{'}} \tag{1}$$

where $\overline{x}_{n1}^{'}$ and $\overline{x}_{h2}^{'}$ are respectively the means based on $n_1^{'}$ responding units and $h_2^{'}$ non-responding units.

The estimator $\overline{x}^{'*}$ gives unbiased estimate of $\overline{X}$ and the variance of $\overline{x}^{'*}$ is given as

$$V\left(\overline{x}^{'*}\right) = \left(\frac{1}{n^{'}} - \frac{1}{N}\right)S_X^2 + \frac{\left(L^{'} - 1\right)}{n^{'}}W_2 S_{X2}^2 \tag{2}$$

where S_X^2 and S_{X2}^2 are the mean squares of entire group and non-response group respectively for the auxiliary variable. W_2 is the non-response rate in the population.

At the second phase, it is noted that there are $n_1$respondent units and n_2 non-respondent units out of n units for both study and auxiliary variables. Applying Hansen and Hurwitz [1] technique of sub-sampling of non-respondents, we select a sub-sample of $h_2$units from the $n_2$non-respondents by SRSWOR at the second phase $(h_2 = \frac{n_2}{L}, L > 1)$ and gather the information on all the h_2 units (L being inverse sampling rate at the second phase). Thus, the Hansen and Hurwitz [1] estimators of $\overline{Y}$ and $\overline{X}$ at the second phase are respectively given by

$$\overline{y}^{*} = \frac{n_1\overline{y}_{n1} + n_2\overline{y}_{h2}}{n} \tag{3}$$

and

$$\overline{x}^{*} = \frac{n_1\overline{x}_{n1} + n_2\overline{x}_{h2}}{n} \tag{4}$$

where $\overline{y}_{n1}$ and $\overline{x}_{n1}$ are the means based on $n_1$respondent units for study and auxiliary variables respectively. $\overline{y}_{h2}$ and $\overline{x}_{h2}$ are respectively the means based on h_2 non-respondent units for study and auxiliary variables.

The variances of $\overline{y}^{*}$and $\overline{x}^{*}$ are respectively represented as

$$V\left(\overline{y}^{*}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)S_Y^2 + \frac{(L-1)}{n}W_2 S_{Y2}^2 \tag{5}$$

$$V\left(\overline{x}^{*}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)S_X^2 + \frac{(L-1)}{n}W_2 S_{X2}^2 \tag{6}$$

where S_Y^2 and S_{Y2}^2 are respectively the mean squares of entire group and non-response group for the study variable.

We now propose some estimators of population mean $\overline{Y}$ using an auxiliary variable whenever the population mean $\overline{X}$ is not known under the condition that both study and auxiliary variables are suffered from non-response. Thus, the ratio, product and regression type estimators of population mean $\overline{Y}$ under the above circumstances are respectively given by

$$T_1^{'*} = \frac{\overline{y}^*}{\overline{x}^*}\overline{x}^{'*} \tag{7}$$

$$T_2^{'*} = \frac{\overline{y}^*}{\overline{x}^{'*}}\overline{x}^* \tag{8}$$

and

$$T_3^{'*} = \overline{y}^* + b^*\left(\overline{x}^{'*} - \overline{x}^*\right) \tag{9}$$

where $b^* = \frac{s_{xy}^*}{s_x^{2*}}$ is an estimator of population regression coefficient of Y on X i.e. $\beta = \frac{S_{XY}}{S_X^2}$, s_{xy}^* and s_x^{2*} are respectively the unbiased estimators of S_{XY} and S_X^2 based on $(n_1 + h_2)$ units.

In order to obtain the biases and mean square errors (MSE) of the estimators$T_1^{'*}$, $T_2^{'*}$and $T_3^{'*}$, we use large sample approximation. Let us consider

$\overline{y}^* = \overline{Y}(1+e_0)$, $\overline{x}^* = \overline{X}(1+e_1)$, $\overline{x}^{'*} = \overline{X}\left(1+e_1^{'}\right)$, $s_{xy}^* = S_{XY}(1+e_2)$ and$s_x^{2*} = S_X^2(1+e_3)$

such that $E(e_0) = E(e_1) = E\left(e_1^{'}\right) = E(e_2) = E(e_3) = 0$,

$$E\left(e_0^2\right) = \left(\frac{1}{n} - \frac{1}{N}\right)C_Y^2 + \frac{(L-1)}{n}W_2\frac{S_{Y2}^2}{\overline{Y}^2},\ \ E\left(e_1^2\right) = \left(\frac{1}{n} - \frac{1}{N}\right)C_X^2 + \frac{(L-1)}{n}W_2\frac{S_{X2}^2}{\overline{X}^2},$$

$$E\left(e_1^{'2}\right) = \left(\frac{1}{n^{'}} - \frac{1}{N}\right)C_X^2 + \frac{\left(L^{'}-1\right)}{n^{'}}W_2\frac{S_{X2}^2}{\overline{X}^2},\ \ E(e_0e_1) = \left(\frac{1}{n} - \frac{1}{N}\right)\rho C_X C_Y + \frac{(L-1)}{n}W_2\rho_2\frac{S_{X2}}{\overline{X}}\frac{S_{Y2}}{\overline{Y}}$$

$$E\left(e_0e_1^{'}\right) = \left(\frac{1}{n^{'}} - \frac{1}{N}\right)\rho C_X C_Y + \frac{\left(L^{'}-1\right)}{n^{'}}W_2\rho_2\frac{S_{X2}}{\overline{X}}\frac{S_{Y2}}{\overline{Y}},$$

$$E\left(e_1e_1^{'}\right) = \left(\frac{1}{n^{'}} - \frac{1}{N}\right)C_X^2 + \frac{\left(L^{'}-1\right)}{n^{'}}W_2\frac{S_{X2}^2}{\overline{X}^2}, C_X = \frac{S_X}{\overline{X}}, C_Y = \frac{S_Y}{\overline{Y}}$$

where ρ and ρ_2 are respectively the correlation coefficients between Y and $X$for entire group and non-response group.

Now expressing the equation (7) in the terms ofe_0,e_1, $e_1^{'}$ and neglecting the terms involving powers ofe_0, e_1 and $e_1^{'}$ greater than two, we get

$$T_1^{'*} - \overline{Y} = \overline{Y}\left[e_0 + e_1^{'} - e_1 + e_0e_1^{'} - e_0e_1 - e_1e_1^{'} + e_1^2\right]. \tag{10}$$

Taking the expectation on both sides of the equation (10), we get
$E\left[T_1^{'*} - \overline{Y}\right] = \overline{Y}\left[E\left(e_0e_1^{'}\right) - E(e_0e_1) - E\left(e_1e_1^{'}\right) + E\left(e_1^2\right)\right] = \overline{Y}\left[\left(\frac{1}{n^{'}} - \frac{1}{N}\right)\rho C_X C_Y + \frac{\left(L^{'}-1\right)}{n^{'}}W_2\rho_2\frac{S_{X2}}{\overline{X}}\frac{S_{Y2}}{\overline{Y}} - \left(\frac{1}{n} - \frac{1}{N}\right)\rho C_X C_Y - \frac{(L-1)}{n}W_2\rho_2\frac{S_{X2}}{\overline{X}}\frac{S_{Y2}}{\overline{Y}}\right.$

$$\left. - \left(\frac{1}{n^{'}} - \frac{1}{N}\right)C_X^2 - \frac{\left(L^{'}-1\right)}{n^{'}}W_2\frac{S_{X2}^2}{\overline{X}^2} + \left(\frac{1}{n} - \frac{1}{N}\right)C_X^2 + \frac{(L-1)}{n}W_2\frac{S_{X2}^2}{\overline{X}^2}\right].$$

Thus the bias of $T_1^{'*}$ to the first order of approximation is given by

$$B\left(T_1^{'*}\right) = \overline{Y}\left[\left(\frac{1}{n} - \frac{1}{n^{'}}\right)\left(C_X^2 - \rho C_X C_Y\right) + W_2\left(\frac{S_{X2}^2}{\overline{X}^2} - \rho_2\frac{S_{X2}}{\overline{X}}\frac{S_{Y2}}{\overline{Y}}\right)\left(\frac{(L-1)}{n} - \frac{\left(L^{'}-1\right)}{n^{'}}\right)\right]. \tag{11}$$

Squaring both the sides of the equation (10) and then taking expectation on neglecting the terms involving powers in e_0, $e_1$and $e_1^{'}$ higher than two, we get

$$E\left[T_1^{'*} - \overline{Y}\right]^2 = \overline{Y}^2\left[E\left(e_0^2\right) + E\left(e_1^2\right) + E\left(e_1^{'2}\right) + 2E\left(e_0e_1^{'}\right) - 2E(e_0e_1) - 2E\left(e_1e_1^{'}\right)\right]$$

$$= \overline{Y}^2 \left[\left(\frac{1}{n}-\frac{1}{N}\right) C_Y^2 + \frac{(L-1)}{n} W_2 \frac{S_{Y2}^2}{\overline{Y}^2} + \left(\frac{1}{n}-\frac{1}{N}\right) C_X^2 + \frac{(L-1)}{n} W_2 \frac{S_{X2}^2}{\overline{X}^2} + \right.$$

$$\left(\frac{1}{n'}-\frac{1}{N}\right) C_X^2 + \frac{\left(L^{'}-1\right)}{n^{'}} W_2 \frac{S_{X2}^2}{\overline{X}^2} + 2\left(\frac{1}{n'}-\frac{1}{N}\right) \rho C_X C_Y + 2\frac{\left(L^{'}-1\right)}{n^{'}} W_2 \rho_2 \frac{S_{X2}}{\overline{X}} \frac{S_{Y2}}{\overline{Y}} -$$

$$\left. 2\left(\frac{1}{n}-\frac{1}{N}\right) \rho C_X C_Y - 2\frac{(L-1)}{n} W_2 \rho_2 \frac{S_{X2}}{\overline{X}} \frac{S_{Y2}}{\overline{Y}} - 2\left(\frac{1}{n'}-\frac{1}{N}\right) C_X^2 - 2\frac{\left(L^{'}-1\right)}{n^{'}} W_2 \frac{S_{X2}^2}{\overline{X}^2} \right].$$

Thus the MSE of $T_1^{'*}$ to the first degree of approximation is given by

$$MSE\left(T_1^{'*}\right) = \left(\frac{1}{n'}-\frac{1}{N}\right) S_Y^2 + \left(\frac{1}{n}-\frac{1}{n'}\right)\left(S_Y^2 + R^2 S_X^2 - 2\rho R S_X S_Y\right) + \frac{(L-1)}{n} W_2 S_{Y2}^2$$
$$+ W_2\left(R^2 S_{X2}^2 - 2\rho_2 R S_{X2} S_{Y2}\right)\left[\frac{(L-1)}{n} - \frac{\left(L^{'}-1\right)}{n^{'}}\right]. \qquad (12)$$

We now express the equation (8) in terms of e_0, e_1 and $e_1^{'}$ on neglecting the contribution of terms involving powers in e_0, e_1 and $e_1^{'}$ greater than two as

$$T_2^{'*} - \overline{Y} = \overline{Y}\left[e_0 + e_1 - e_1^{'} + e_0 e_1 - e_0 e_1^{'} - e_1 e_1^{'} + e_1^{'2}\right]. \qquad (13)$$

Thus the bias of $T_2^{'*}$ to the first order of approximation is given by

$$B\left(T_2^{'*}\right) = E\left[T_2^{'*} - \overline{Y}\right] = \overline{Y}\left[E\left(e_0 e_1\right) - E\left(e_0 e_1^{'}\right) - E\left(e_1 e_1^{'}\right) + E\left(e_1^{'2}\right)\right]$$

$$= \overline{Y}\left[\left(\frac{1}{n}-\frac{1}{n'}\right)\rho C_X C_Y + W_2 \rho_2 \frac{S_{X2}}{\overline{X}} \frac{S_{Y2}}{\overline{Y}} \left(\frac{(L-1)}{n} - \frac{\left(L^{'}-1\right)}{n^{'}}\right)\right]. \qquad (14)$$

Squaring both the sides of the equation (13) and taking expectation, we get

$$E\left[T_2^{'*} - \overline{Y}\right]^2 = \overline{Y}^2\left[E\left(e_0^2\right) + E\left(e_1^2\right) + E\left(e_1^{'2}\right) + 2E\left(e_0 e_1\right) - 2E\left(e_0 e_1^{'}\right) - 2E\left(e_1 e_1^{'}\right)\right].$$

Thus the MSE of $T_2^{'*}$ to the first degree of approximation is represented as

$$MSE\left(T_2^{'*}\right) = \left(\frac{1}{n'}-\frac{1}{N}\right) S_Y^2 + \left(\frac{1}{n}-\frac{1}{n'}\right)\left(S_Y^2 + R^2 S_X^2 + 2\rho R S_X S_Y\right) + \frac{(L-1)}{n} W_2 S_{Y2}^2$$
$$+ W_2\left(R^2 S_{X2}^2 + 2\rho_2 R S_{X2} S_{Y2}\right)\left[\frac{(L-1)}{n} - \frac{\left(L^{'}-1\right)}{n^{'}}\right]. \qquad (15)$$

Expressing the equation (9) in terms of e_0,e_1, $e_1^{'}$, e_2 , e_3 and neglecting the contribution of terms involving powers in e_0,e_1, $e_1^{'}$, e_2, e_3 higher than two, we get

$$T_3^{'*} - \overline{Y} = \overline{Y} e_0 + \beta \overline{X}\left[e_1^{'} - e_1 + e_1^{'} e_2 - e_1^{'} e_3 + e_1 e_3 - e_1 e_2\right]. \qquad (16)$$

Taking expectation on both the sides of the equation (16), we get

$$E\left[T_3^{'*} - \overline{Y}\right] = \beta \overline{X}\left[E\left(e_1^{'} e_2\right) - E\left(e_1^{'} e_3\right) + E\left(e_1 e_3\right) - E\left(e_1 e_2\right)\right].$$

Thus the bias of $T_3^{'*}$ to the first order of approximation is given by

$$B\left(T_3^{'*}\right)=\beta\left[\frac{N^2}{(N-1)(N-2)}\left(\frac{1}{n}-\frac{1}{n'}\right)\left(\frac{\mu_{30}}{S_X^2}-\frac{\mu_{21}}{S_{XY}}\right)+W_2\left(\frac{\mu_{30(2)}}{S_X^2}-\frac{\mu_{21(2)}}{S_{XY}}\right)\left(\frac{(L-1)}{n}-\frac{\left(L'-1\right)}{n'}\right)\right] \tag{17}$$

where $\mu_{rs}=\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\overline{X}\right)^r\left(y_i-\overline{Y}\right)^s$, $\mu_{rs(2)}=\frac{1}{N_2}\sum_i^{N_2}\left(x_i-\overline{X}_2\right)^r\left(y_i-\overline{Y}_2\right)^s$,
$\overline{X}_2=\frac{1}{N_2}\sum_i^{N_2}x_i$ and $\overline{Y}_2=\frac{1}{N_2}\sum_i^{N_2}y_i$.

Squaring both the sides of the equation (16) and taking expectation on neglecting the terms involving powers of e_0, e_1, e_1', e_2, e_3 greater than two, we get

$$E\left[T_3^{'*}-\overline{Y}\right]^2=\overline{Y}E\left(e_0^2\right)+\beta^2\overline{X}^2\left[E\left(e_1^{'2}\right)+E\left(e_1^2\right)-2E\left(e_1e_1'\right)\right]+2\beta\overline{XY}\left[E\left(e_0e_1'\right)-E\left(e_0e_1\right)\right].$$

Thus the MSE of $T_3^{'*}$ to the first degree of approximation is expressed as

$$\begin{aligned}MSE\left(T_3^{'*}\right)=&\left[\left(\frac{1}{n'}-\frac{1}{N}\right)+\left(\frac{1}{n}-\frac{1}{n'}\right)\left(1-\rho^2\right)\right]S_Y^2+\frac{(L-1)}{n}W_2S_{Y2}^2\\&+W_2\left(\beta^2S_{X2}^2-2\beta\rho_2S_{X2}S_{Y2}\right)\left[\frac{(L-1)}{n}-\frac{\left(L'-1\right)}{n'}\right].\end{aligned} \tag{18}$$

3 Cost of the Survey and Optimum Values of n', n, L' and L

Let c' be the unit cost associated with the first phase sample of size n' on first attempt. Let c_1' and c_2' be the costs per unit of enumerating n_1' respondent units and h_2' non-respondent units respectively. Let c be the cost per unit on first attempt associated with the second phase sample of size n. Now, let c_1 and c_2 respectively be the costs per unit of enumerating n_1 respondent units and h_2 non-respondent units. Thus the total cost is given by

$$C=c'n'+cn+c_1'n_1'+c_1n_1+c_2'h_2'+c_2h_2.$$

Now, the total expected average cost is given as

$$C_0=E(C)=c'n'+cn+c_1'n_1'+c_1n_1+c_2'h_2'+c_2h_2$$

$$=n'\left(c'+c_1'W_1+c_2'\frac{W_2}{L'}\right)+n\left(c+c_1W_1+c_2\frac{W_2}{L}\right) \tag{19}$$

where W_1 is the response rate in the population.

Let us define the Lagrange functions

$$\phi_1=MSE\left(T_1^{'*}\right)+\mu C_0 \tag{20}$$

$$\phi_2=MSE\left(T_2^{'*}\right)+\mu C_0 \tag{21}$$

$$\phi_3=MSE\left(T_3^{'*}\right)+\mu C_0 \tag{22}$$

where μ is Lagrange's multiplier.

Case (i): (Under Estimator $T_1^{'*}$)

To get the normal equations, we differentiate the equation (20) with respect to n', n, L' and L respectively, and equate the derivatives to zero. Thus, we have

$$\frac{\partial\phi_1}{\partial n'} = \frac{1}{n'^2}\left[R^2S_X^2 - 2\rho RS_XS_Y + \left(L' - 1\right)W_2\left(R^2S_{X2}^2 - 2\rho_2RS_{X2}S_{Y2}\right)\right] + \mu\left(c' + c_1'W_1 + c_2'\frac{W_2}{L'}\right) = 0 \quad (23)$$

$$\frac{\partial\phi_1}{\partial n} = -\frac{1}{n^2}\left[\left(S_Y^2 + R^2S_X^2 - 2\rho RS_XS_Y\right) + (L-1)W_2\left(S_{Y2}^2 + R^2S_{X2}^2 - 2\rho_2RS_{X2}S_{Y2}\right)\right] + \mu\left(c + c_1W_1 + c_2\frac{W_2}{L}\right) = 0 \quad (24)$$

$$\frac{\partial\phi_1}{\partial L'} = \frac{W_2}{n'}\left(R^2S_{X2}^2 - 2\rho_2RS_{X2}S_{Y2}\right) + \mu n'c_2'\frac{W_2}{L'^2} = 0 \quad (25)$$

and

$$\frac{\partial\phi_1}{\partial L} = \frac{W_2}{n}S_{Y2}^2 + \frac{W_2}{n}\left(R^2S_{X2}^2 - 2\rho_2RS_{X2}S_{Y2}\right) - \mu nc_2\frac{W_2}{L^2} = 0. \quad (26)$$

From equations (23), (24), (25) and (26), we respectively get

$$n' = \frac{\sqrt{2\rho RS_XS_Y - R^2S_X^2 + (L'-1)W_2\left(2\rho_2RS_{X2}S_{Y2} - R^2S_{X2}^2\right)}}{\sqrt{\mu\left(c' + c_1'W_1 + c_2'\frac{W_2}{L'}\right)}} \quad (27)$$

$$n = \frac{\sqrt{\left(S_Y^2 + R^2S_X^2 - 2\rho RS_XS_Y\right) + (L-1)W_2\left(S_{Y2}^2 + R^2S_{X2}^2 - 2\rho_2RS_{X2}S_{Y2}\right)}}{\sqrt{\mu\left(c + c_1W_1 + c_2\frac{W_2}{L}\right)}} \quad (28)$$

$$\sqrt{\mu} = \frac{L'\sqrt{\left(2\rho_2RS_{X2}S_{Y2} - R^2S_{X2}^2\right)}}{n'\sqrt{c_2'}} \quad (29)$$

and

$$\sqrt{\mu} = \frac{L\sqrt{\left(S_{Y2}^2 + R^2S_{X2}^2 - 2\rho_2RS_{X2}S_{Y2}\right)}}{n\sqrt{c_2}} \quad (30)$$

Substituting $\sqrt{\mu}$ from equation (29) into the equation (27), we get optimum value of L' as

$$L'_{opt} = \frac{D\sqrt{c_2'}}{AB} \quad (31)$$

where $A = \sqrt{c' + c_1'W_1}$, $B = \sqrt{\left(2\rho_2RS_{X2}S_{Y2} - R^2S_{X2}^2\right)}$ and

$$D = \sqrt{\left(2\rho RS_XS_Y - R^2S_X^2\right) + W_2\left(R^2S_{X2}^2 - 2\rho_2RS_{X2}S_{Y2}\right)}$$

Substituting the value of L'_{opt} from equation (31) into equation (27), we get

$$n' = \frac{\sqrt{D^2 + \frac{D\sqrt{c_2'}BW_2}{A}}}{\sqrt{\mu\sqrt{A^2 + \frac{\sqrt{c_2'}W_2AB}{D}}}}. \quad (32)$$

Putting $\sqrt{\mu}$ from equation (30) into the equation (28), we have

$$L_{opt} = \frac{\sqrt{c_2}D'}{A'B'} \quad (33)$$

where $A' = \sqrt{c + c_1W_1}$, $B' = \sqrt{\left(S_{Y2}^2 + R^2S_{X2}^2 - 2\rho_2RS_{X2}S_{Y2}\right)}$ and

$$D' = \sqrt{\left(S_Y^2 + R^2S_X^2 - 2\rho RS_XS_Y\right) - W_2\left(S_{Y2}^2 + R^2S_{X2}^2 - 2\rho_2RS_{X2}S_{Y2}\right)}.$$

On substituting the value of L_{opt} from equation (33) into the equation (28), we get

$$n = \frac{\sqrt{D'^2 + \frac{\sqrt{c_2} D' B' W_2}{A'}}}{\sqrt{\mu}\sqrt{A'^2 + \frac{\sqrt{c_2} W_2 A' B'}{D'}}}. \quad (34)$$

To obtain the value of $\sqrt{\mu}$ in terms of total costC_0, we put the values of$n^{'}$,$L^{'}$, n and L respectively from the equations (32), (31), (34) and (33) into equation (19), we get

$$\sqrt{\mu} = \frac{1}{C_0}\left[\sqrt{D^2 + \frac{D\sqrt{c_2^{'}}BW_2}{A}}\sqrt{A^2 + \frac{A\sqrt{c_2^{'}}BW_2}{D}} + \sqrt{D'^2 + \frac{\sqrt{c_2}D'B'W_2}{A'}}\sqrt{A'^2 + \frac{\sqrt{c_2}A'B'W_2}{D'}}\right] \quad (35)$$

On substituting the value of $\sqrt{\mu}$ from equation (35) into the equations (32) and (34), we respectively get the optimum values of $n^{'}$ and $n$as

$$n^{'}_{opt} = \frac{C_0\sqrt{D^2 + \frac{D\sqrt{c_2^{'}}BW_2}{A}}}{\sqrt{A^2 + \frac{\sqrt{c_2^{'}}W_2AB}{D}}\left[\sqrt{\left(D^2 + \frac{D\sqrt{c_2^{'}}BW_2}{A}\right)\left(A^2 + \frac{\sqrt{c_2^{'}}W_2AB}{D}\right)} + \sqrt{\left(D'^2 + \frac{\sqrt{c_2}D'B'W_2}{A'}\right)\left(A'^2 + \frac{\sqrt{c_2}A'B'W_2}{D'}\right)}\right]} \quad (36)$$

and

$$n_{opt} = \frac{C_0\sqrt{D'^2 + \frac{\sqrt{c_2}D'B'W_2}{A'}}}{\sqrt{A'^2 + \frac{\sqrt{c_2}W_2A'B'}{D'}}\left[\sqrt{\left(D^2 + \frac{D\sqrt{c_2^{'}}BW_2}{A}\right)\left(A^2 + \frac{\sqrt{c_2^{'}}W_2AB}{D}\right)} + \sqrt{\left(D'^2 + \frac{\sqrt{c_2}D'B'W_2}{A'}\right)\left(A'^2 + \frac{\sqrt{c_2}A'B'W_2}{D'}\right)}\right]} \quad (37)$$

Case (ii): (Under Estimator$T_2^{'*}$)

Now, we differentiate the equation (21) with respect to$n^{'}$,n, $L^{'}$ and L respectively, and equate the derivatives to zero. Thus, we get

$$\frac{\partial \phi_2}{\partial n^{'}} = \frac{1}{n'^2}\left[R^2S_X^2 + 2\rho RS_XS_Y + \left(L^{'} - 1\right)W_2\left(R^2S_{X2}^2 + 2\rho_2RS_{X2}S_{Y2}\right)\right] + \mu\left(c^{'} + c_1^{'}W_1 + c_2^{'}\frac{W_2}{L^{'}}\right) = 0 \quad (38)$$

$$\frac{\partial \phi_2}{\partial n} = -\frac{1}{n^2}\left[\left(S_Y^2 + R^2S_X^2 + 2\rho RS_XS_Y\right) + (L-1)W_2\left(S_{Y2}^2 + R^2S_{X2}^2 + 2\rho_2RS_{X2}S_{Y2}\right)\right] + \mu\left(c + c_1W_1 + c_2\frac{W_2}{L}\right) = 0 \quad (39)$$

$$\frac{\partial \phi_2}{\partial L^{'}} = \frac{W_2}{n^{'}}\left(R^2S_{X2}^2 + 2\rho_2RS_{X2}S_{Y2}\right) + \mu n^{'}c_2^{'}\frac{W_2}{L'^2} = 0 \quad (40)$$

and

$$\frac{\partial \phi_2}{\partial L} = \frac{W_2}{n}S_{Y2}^2 + \frac{W_2}{n}\left(R^2S_{X2}^2 + 2\rho_2RS_{X2}S_{Y2}\right) - \mu nc_2\frac{W_2}{L^2} = 0 \quad (41)$$

Solving the above equations in the similar manner as in case (i), we respectively get the optimum values of $L^{'}$, L, $n^{'}$and n

$$L^{'}_{opt} = \frac{D^*\sqrt{c_2^{'}}}{AB^*}, \quad (42)$$

$$L_{opt} = \frac{\sqrt{c_2}D'^*}{A'B'^*}, \quad (43)$$

$$n_{opt}^{'} = \frac{C_0\sqrt{-D^{*2}-\frac{D^*\sqrt{c_2^{'}}B^*W_2}{A}}}{\sqrt{A^2+\frac{\sqrt{c_2^{'}}W_2AB^*}{D^*}}\left[\sqrt{\left(-D^{*2}-\frac{D^*\sqrt{c_2^{'}}B^*W_2}{A}\right)\left(A^2+\frac{\sqrt{c_2^{'}}W_2AB^*}{D^*}\right)}+\sqrt{\left(D^{'*2}+\frac{\sqrt{c_2}D^{'*}B^{'*}W_2}{A^{'}}\right)\left(A^{'2}+\frac{\sqrt{c_2}A^{'}B^{'*}W_2}{D^{*'}}\right)}\right]} \tag{44}$$

and

$$n_{opt} = \frac{C_0\sqrt{D^{'*2}+\frac{\sqrt{c_2}D^{'*}B^{'*}W_2}{A^{'}}}}{\sqrt{A^{'2}+\frac{\sqrt{c_2}W_2A^{'}B^{'*}}{D^{'*}}}\left[\sqrt{\left(-D^{*2}-\frac{D^*\sqrt{c_2^{'}}B^*W_2}{A}\right)\left(A^2+\frac{\sqrt{c_2^{'}}W_2AB^*}{D^*}\right)}+\sqrt{\left(D^{'*2}+\frac{\sqrt{c_2}D^{'*}B^{'*}W_2}{A^{'}}\right)\left(A^{'2}+\frac{\sqrt{c_2}A^{'}B^{'*}W_2}{D^{'*}}\right)}\right]} \tag{45}$$

where

$$B^* = \sqrt{(R^2S_{X2}^2 + 2\rho_2RS_{X2}S_{Y2})}$$

$$D^* = \sqrt{(R^2S_X^2 + 2\rho RS_XS_Y) - W_2\,(R^2S_{X2}^2 + 2\rho_2RS_{X2}S_{Y2})},$$

$$B^{'*} = \sqrt{(S_{Y2}^2 + R^2S_{X2}^2 + 2\rho_2RS_{X2}S_{Y2})}$$

and

$$D^{'*} = \sqrt{(S_Y^2 + R^2S_X^2 + 2\rho RS_XS_Y) - W_2\,(S_{Y2}^2 + R^2S_{X2}^2 + 2\rho_2RS_{X2}S_{Y2})}.$$

Case (iii): (Under Estimator$T_3^{'*}$)

Using equation (22) and following the procedure adopted in the cases (i) and (ii), we can respectively obtain the optimum values of $L^{'}$, L, $n^{'}$ and $n$for the estimator $T_3^{'*}$ as

$$L_{opt}^{'} = \frac{E\sqrt{c_2^{'}}}{AF}, \tag{46}$$

$$L_{opt} = \frac{\sqrt{c_2}E^*}{A^{'}F^*}, \tag{47}$$

$$n_{opt}^{'} = \frac{C_0\sqrt{E^2+\frac{E\sqrt{c_2^{'}}FW_2}{A}}}{\sqrt{A^2+\frac{\sqrt{c_2^{'}}W_2AF}{E}}\left[\sqrt{\left(E^2+\frac{E\sqrt{c_2^{'}}FW_2}{A}\right)\left(A^2+\frac{\sqrt{c_2^{'}}W_2AF}{E}\right)}+\sqrt{\left(E^{*2}+\frac{\sqrt{c_2}E^*F^*W_2}{A^{'}}\right)\left(A^{'2}+\frac{\sqrt{c_2}A^{'}F^*W_2}{E^*}\right)}\right]} \tag{48}$$

and

$$n_{opt} = \frac{C_0\sqrt{E^{*2}+\frac{\sqrt{c_2}E^*F^*W_2}{A^{'}}}}{\sqrt{A^{'2}+\frac{\sqrt{c_2}W_2A^{'}F^*}{E^*}}\left[\sqrt{\left(E^2+\frac{E\sqrt{c_2^{'}}FW_2}{A}\right)\left(A^2+\frac{\sqrt{c_2^{'}}W_2AF}{E}\right)}+\sqrt{\left(E^{*2}+\frac{\sqrt{c_2}E^*F^*W_2}{A^{'}}\right)\left(A^{'2}+\frac{\sqrt{c_2}A^{'}F^*W_2}{E^*}\right)}\right]} \tag{49}$$

where $E = \sqrt{\rho^2S_Y^2 + W_2\,(\beta^2S_{X2}^2 - 2\beta\rho_2S_{X2}S_{Y2})}$, $F = \sqrt{-\beta^2S_{X2}^2 + 2\beta\rho_2S_{X2}S_{Y2}}$, $E^* = \sqrt{(1-\rho^2)\,S_Y^2 - W_2\,(S_{Y2}^2 + \beta^2S_{X2}^2 - 2\beta\rho_2S_{X2}S_{Y2})}$ and $F^* = \sqrt{(S_{Y2}^2 + \beta^2S_{X2}^2 - 2\beta\rho_2S_{X2}S_{Y2})}$.

4 Numerical Study

To demonstrate the theoretical study, we have considered the two different data sets. One is used for showing the performances of the estimators $T_1^{'*}$ and $T_3^{'*}$ where the study and auxiliary variables are positively correlated. To show the performances of the estimators $T_2^{'*}$ and$T_3^{'*}$, another data set is considered where the study and auxiliary variables are negatively correlated.

Data Set 1:

Here we have considered the data used by Srivastava [10]. The data consist of the population of seventy villages in a Tehsil of India along with their cultivated area (in acres) in 1981. The cultivated area (in acres) and the population are respectively considered as study and auxiliary variables. The particulars of the population are given below:
$N = 70$, $n^{'} = 40$, $n = 25$, $\overline{Y} = 981.29$, $\overline{X} = 1755.53$, $S_Y = 613.66$, $S_X = 1406.13$, $S_{Y2} = 244.11$, $S_{X2} = \frac{4}{5}S_X = 1124.905$, $\rho = 0.778$, and $\rho_2 = \frac{4}{5}\rho = 0.622$.

The table 1 represents the variance of$\overline{y}^*$, MSE of$T_1^{'*}$, $T_3^{'*}$ and percentage relative efficiency (PRE) of $T_1^{'*}$ and $T_3^{'*}$ with respect to the sample mean estimator $\overline{y}^*$for different choices of non-response rate $(W_2 = 0.1, 0.2, 0.3, 0.4)$ and inverse sampling rates$\left(L^{'} = L = 1.5, 2.0, 2.5, 3.0\right)$.

Table 1: Variance of$\overline{y}^*$, MSE of$T_1^{'*}$, $T_3^{'*}$ and PRE of$T_1^{'*}$and $T_3^{'*}$with respect to $\overline{y}^*$

W_2	$L^{'} = L$	$Var./MSE$			PRE	
		$\overline{y}^*$	$T_1^{'*}$	$T_3^{'*}$	$T_1^{'*}$	$T_3^{'*}$
0.1	1.5	9802.63	7964.93	6405.94	123.07	153.02
	2.0	9921.81	8084.10	6525.12	122.73	152.06
	2.5	10040.99	8203.28	6644.30	122.40	151.12
	3.0	10160.17	8322.46	6763.48	122.08	150.22
0.2	1.5	9921.81	8237.33	6547.49	120.45	151.54
	2.0	10160.17	8475.69	6785.84	119.87	149.73
	2.5	10398.53	8714.05	7024.20	119.33	148.04
	3.0	10636.88	8952.41	7262.55	118.82	146.46
0.3	1.5	10040.99	8509.74	6689.02	117.99	150.11
	2.0	10398.53	8867.28	7046.56	117.27	147.57
	2.5	10756.06	9224.82	7404.10	116.60	145.27
	3.0	11113.60	9582.36	7761.64	115.98	143.19
0.4	1.5	10160.17	8782.15	6830.57	115.69	148.75
	2.0	10636.88	9258.87	7307.29	114.88	145.57
	2.5	11113.60	9735.59	7784.00	114.15	142.77
	3.0	11590.32	10212.31	8260.72	113.49	140.31

Data Set 2:

In this data set, we have used the data considered by Chaudhary and Shukla [11]. The data relate to the scores elicited by a diagnostic technique, called AMDN and the forced expiratory volume (FEV_1) scores for twenty two patients. Here, AMDN and FEV_1 are respectively considered as study and auxiliary variables. The population parameters are given below:
$N = 22$, $n^{'} = 15$, $n = 9$, $\overline{Y} = 1.76$, $\overline{X} = 78.18$, $S_Y^2 = 0.1212$, $S_X^2 = 631.83$, $S_{Y2}^2 = 0.1041$, $S_{X2}^2 = \frac{4}{5}S_X^2 = 505.46$, $\rho = -0.877$, and $\rho_2 = \frac{4}{5}\rho = -0.705$.

The table 2 depicts the variance of$\overline{y}^*$, MSE of$T_2^{'*}$, $T_3^{'*}$ and PRE of the estimators $T_2^{'*}$ and $T_3^{'*}$ with respect to the sample mean estimator $\overline{y}^*$for different choices of non-response rate $(W_2 = 0.1, 0.2, 0.3, 0.4)$ and inverse sampling rates$\left(L^{'} = L = 1.5, 2.0, 2.5, 3.0\right)$.

Table 2: Variance of $\overline{y}^*$, MSE of $T_2^{'*}$, $T_3^{'*}$ and PRE of $T_2^{'*}$ and $T_3^{'*}$ with respect to $\overline{y}^*$

W_2	$L^{'} = L$	$Var./MSE$			PRE	
		$\overline{y}^*$	$T_2^{'*}$	$T_3^{'*}$	$T_2^{'*}$	$T_3^{'*}$
0.1	1.5	0.0085	0.0075	0.0043	114.16	199.26
	2.0	0.0091	0.0081	0.0049	113.14	187.45
	2.5	0.0097	0.0086	0.0054	112.26	178.16
	3.0	0.0103	0.0092	0.0060	111.49	170.65
0.2	1.5	0.0091	0.0081	0.0048	112.30	191.75
	2.0	0.0103	0.0093	0.0059	110.77	173.79
	2.5	0.0114	0.0104	0.0071	109.57	161.72
	3.0	0.0126	0.0116	0.0082	108.62	153.03
0.3	1.5	0.0097	0.0088	0.0052	110.71	185.59
	2.0	0.0114	0.0105	0.0070	108.94	164.25
	2.5	0.0132	0.0122	0.0087	107.67	151.43
	3.0	0.0149	0.0140	0.0104	106.72	142.87
0.4	1.5	0.0103	0.0094	0.0057	109.34	180.45
	2.0	0.0126	0.0117	0.0080	107.50	157.20
	2.5	0.0149	0.0140	0.0103	106.26	144.38
	3.0	0.0172	0.0163	0.0126	105.37	136.25

5 Conclusion

In the present paper, we have proposed the ratio, product and regression type estimators utilizing the information on an auxiliary variable under the situations in which both study and auxiliary variables are suffered from non-response. The biases and mean square errors of the proposed estimators have been obtained. The theoretical study on the cost of the survey has been presented. The optimum values of the first phase sample size $n^{'}$, second phase sample size n and inverse sampling rates at first phase and second phase i. e. $L^{'}$ and L have been obtained under each estimator. The table 1 represents PRE of the proposed estimators $T_1^{'*}$ and $T_3^{'*}$ with respect to sample mean estimator $\bar{y}^*$ for the different choices of non-response rate W_2 and inverse sampling rates $L^{'}$ and L where study and auxiliary variables are positively correlated. Similarly, table 2 depicts PRE of the proposed estimators $T_2^{'*}$ and $T_3^{'*}$ with respect to sample mean estimator $\bar{y}^*$ for the different choices of non-response rate W_2 and inverse sampling rates $L^{'}$ and L where study and auxiliary variables are negatively correlated. In both the tables, it is seen that the MSE of the suggested estimators increases with the increase in the non-response rate as well as to increase in the inverse sampling rates. The results are also intuitively expected.

References

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

[2] B. B. Khare (1987), Allocation In stratified sampling in presence of non-response, Metron, 45 (I/II), 213-221 (1987).

[3] M. G. M. Khan, E. A. Khan and M. J. Ahsan, Optimum allocation in multivariate stratified sampling in presence of non-response, J. Ind. Soc. Agril. Statist., 62, 42-48 (2008).

[4] M. K. Chaudhary, R. Singh, R. K. Shukla and M. Kumar, A family of estimators for estimating population mean in stratified sampling under non-response, Pakistan Journal of Statistics and Operation Research, Vol. 5-No. 1, 47-54 (2009).

[5] M. K. Chaudhary and V. K. Singh, Estimating the population mean in two-stage sampling with equal size clusters under non-response using auxiliary characteristic, Mathematical Journal of Interdisciplinary Sciences, 2, 43-55 (2013).

[6] B. B. Khare and R. R. Sinha, Estimation of finite population ratio using two-phase sampling scheme in the presence of non-response, Aligarh J. Stat., 24, 43-56 (2004).

[7] H. P. Singh and S. Kumar (2009), A general procedure of estimating the population mean in the presence of non-response under double sampling using auxiliary information, SORT, 33, 71-84 (2009).

[8] M. K. Chaudhary, R. Singh, M. Kumar, A. K. Singh and A. A. Adewara, A class of estimators for assessing the population mean under non-respone using auxiliary information, ASM Science Journal, 6, 23-30 (2012).

[9] M. K. Chaudhary and A. Kumar, Estimating the population mean in stratified random sampling using two-phase sampling scheme in the presence of non-response, World Applied Sciences Journal (WASJ), Vol. 33, No. 6, 874-882 (2015).

[10] S. Srivastava, Some problems on the estimation of population mean using auxiliary character in presence of non-response in sample surveys, Ph. D. Thesis, Banaras Hindu University, Varanasi (1993).

[11] M. K. Chaudhary and R. K. Shukla, Estimation of population mean with predictive approach under non-response, Journal of Progressive Science (JPS), Vol. 3, No. 1, 89-94 (2012).

Manoj K. Chaudhary is Professor (Assistant) in the Department of Statistics, Banaras Hindu University, Varanasi, INDIA. He has made significant contributions in the field of sampling theory and methods of estimation. He has published more than 25 research papers in national and international journals.

Amit Kumar is research scholar in the Department of Statistics, Banaras Hindu University, Varanasi, INDIA. His research interests are in the areas of sample surveys and estimation procedures. He has published more than 3 research papers in national and international journals.