Journal of Statistics Applications & Probability Letters
An International Journal

http://dx.doi.org/10.18576/jsapl/030102

General Families of Estimators for Estimating Population Mean in Stratified Random Sampling under Non-Response

Manoj K. Chaudhary and Saurabh*

Department of Statistics, Banaras Hindu University, Varanasi, India

Received: 4 Jul. 2015, Revised: 5 Aug. 2015, Accepted: 7 Aug. 2015
Published online: 1 Jan. 2016

Abstract: The present article concentrates on estimating the mean of a stratified population in the presence of non-response. In this article, we have suggested separate and combined-type families of estimators of population mean using the information of an auxiliary variable assuming that the non-response is observed on both study and auxiliary variables. The properties of the suggested families have been thrashed out. The suggested families have been discussed under the proportional, Neyman and some other allocation schemes proposed by Chaudhary et al. [6]. An empirical study has also been carried out in the support of theoretical results.

Keywords: Stratified random sampling, separate-type family, combined-type family, population mean and non-response.

1 Introduction

Non-response is a very serious issue in estimating the population parameters through a mail survey. Hansen and Hurwitz [1] were the first who coped up the problem of non-response while conducting mail surveys. They developed a technique of sub-sampling of non-respondents to deal with the problem of non-response and its adjustments. In point of fact, they suggested an unbiased estimator of population mean under non-response by dividing the population into two different groups, viz. group of respondents and group of non-respondents. In order to avoid the bias due to non-response, they suggested a technique of selecting a sub-sample from the non-respondents of the sample.

A lot of works have been done for estimating the population mean in stratified random sampling whenever the investigator suffers with the problem of non-response. Khare [2] has discussed the problem of optimum allocation in stratified random sampling in the presence of non-response. Khan et al. [3] described the method of optimum allocation in multivariate stratified random sampling under non-response. Chaudhary et al. [4] have proposed a general family of estimators in stratified random sampling in the presence of non-response by considering Khoshnevisan et. al. [5], Chaudhary et al. [6] have proposed some new allocation schemes based on response and non-response rates in stratified random sampling. Further, Chaudhary et al. [7] have suggested a class of factor-type estimators of population mean in stratified random sampling under non-response.

All of the works mentioned above have been carried out in the situations when non-response is observed on study variable and auxiliary variable(s) is free from non-response. But the situations in which both study and auxiliary variables are suffered from non-response, it would be inevitable to introduce the estimators of population parameters of study variable. In the light of above circumstances, we have suggested some families of factor-type estimators of population mean in stratified random sampling using an auxiliary variable under non-response. The optimum estimators of the proposed families have been discussed. We have compared the proportional and Neyman allocations with some other allocation schemes based on response and/or non-response rates through the suggested families of estimators. The theoretical study has also been supported with numerical analysis.

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

2 Sampling Strategy and Estimation Procedure

Let us suppose that a population consists of N units is divided into k strata. Let there be $N_i$units in the i^{th}stratum$(i = 1,2,...,k)$. A random sample of size $n$is selected from the entire population in such a way that n_i units are selected from the i^{th}stratum so that$\sum_{i=1}^{k} n_i = n$. It is noted that out of $n_i$units there are n_{i1}units who supply the information and n_{i2}units who do not respond. Using Hansen and Hurwitz [1] technique of sub-sampling of non-respondents, a sub-sample of $h_{i2}\,(= n_{i2}/L_i, L_i \geq 1)$ non-respondents is selected from the n_{i2} units by simple random sampling without replacement (SRSWOR) scheme and the information is obtained on all the h_{i2} units. Let X_0 and X_1 be the study and auxiliary variables respectively with their respective population means $\overline{X}_0$ and $\overline{X}_1$. Thus the Hansen and Hurwitz [1] estimators of $\overline{X}_0$ and$\overline{X}_1$ are respectively given by

$$T_{0st}^{*} = \sum_{i=1}^{k} p_i T_{0i}^{*} \tag{1}$$

$$T_{1st}^{*} = \sum_{i=1}^{k} p_i T_{1i}^{*} \tag{2}$$

where $T_{0i}^{*} = \frac{n_{i1}\overline{x}_{0i1} + n_{i2}\overline{x}_{0hi2}}{n_i}$, $T_{1i}^{*} = \frac{n_{i1}\overline{x}_{1i1} + n_{i2}\overline{x}_{1hi2}}{n_i}$, $p_i = \frac{N_i}{N}$,
$\overline{x}_{0i1}$ and $\overline{x}_{1i1}$ are the means based on n_{i1} respondent units for the study and auxiliary variables respectively. $\overline{x}_{0hi2}$and $\overline{x}_{1hi2}$ are the means based on h_{i2} non-respondent units for the study and auxiliary variables respectively. The variances of the unbiased estimators T_{0st}^{*} and T_{1st}^{*} are respectively given by

$$V\left(T_{0st}^{*}\right) = \sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right) p_i^2 S_{0i}^2 + \sum_{i=1}^{k}\frac{(L_i - 1)}{n_i} W_{i2} p_i^2 S_{0i2}^2 \tag{3}$$

and

$$V\left(T_{1st}^{*}\right) = \sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right) p_i^2 S_{1i}^2 + \sum_{i=1}^{k}\frac{(L_i - 1)}{n_i} W_{i2} p_i^2 S_{1i2}^2 \tag{4}$$

where $S_{0i}^2$and S_{0i2}^2 are the population mean squares of the entire group and non-response group respectively for study variable in i^{th} stratum. Similarly $S_{1i}^2$and S_{1i2}^2 are the population mean squares of the entire group and non-response group respectively for auxiliary variable in i^{th} stratum. W_{i2} is the non-response rate in the i^{th} stratum.

3 Proposed Families of Estimators

It is very difficult to estimate the population parameters using auxiliary variable(s) under the situations in which both study and auxiliary variables are suffered from non-response. In the sequence of estimating the population mean of study variable, we propose two different types of families of estimators for population mean in stratified random sampling using an auxiliary variable over the situation in which non-response is observed on both study and auxiliary variables.

3.1 Separate-type Family of Estimators

Following Singh and Shukla [8], we now propose a separate-type family of estimators of population mean $\overline{X}_0$ in stratified random sampling under non-response as

$$T_{FS}^{*}(\alpha) = \sum_{i=1}^{k} p_i T_{Fi}^{*}(\alpha) \tag{5}$$

where,

$$T_{Fi}^{*}(\alpha) = T_{0i}^{*}\left[\frac{(A+C)\overline{X}_{1i} + fBT_{1i}^{*}}{(A+fB)\overline{X}_{1i} + CT_{1i}^{*}}\right], \tag{6}$$

$A = (\alpha-1)(\alpha-2)$, $B = (\alpha-1)(\alpha-4)$, $C = (\alpha-2)(\alpha-3)(\alpha-4)$ for $\alpha > 0$and $f = n/N$. $\overline{X}_{1i}$ is the population mean of auxiliary variable for the i^{th} stratum.

The above family can generate a number of separate-type estimators of population mean $\overline{X}_0$ under non-response for the different choices ofα. Particularly, if we take $\alpha = 1,2,3$and4, we get

Case(i) If $\alpha = 1$ then $A = B = 0$ and $C = -6$,
so that $T_{Fi}^{*} = T_{0i}^{*}\frac{\overline{X}_{1i}}{T_{1i}^{*}}$,
hence $T_{FS}^{*}(1) = \sum_{i=1}^{k} p_i T_{0i}^{*}\frac{\overline{X}_{1i}}{T_{1i}^{*}}$
which is usual separate ratio estimator under non-response.

Case(ii) If $\alpha = 2$then $A = C = 0$and $B = -2$,
so that $T_{Fi}^{*} = T_{0i}^{*}\frac{T_{1i}^{*}}{\overline{X}_{1i}}$,
therefore $T_{FS}^{*}(2) = \sum_{i=1}^{k} p_i T_{0i}^{*}\frac{T_{1i}^{*}}{\overline{X}_{1i}}$ which is usual separate product estimator under non-response.

Case(iii) If $\alpha = 3$then $A = 2$, $B = -2$ and $C = 0$,
so that $T_{Fi}^{*} = T_{0i}^{*}\frac{\overline{X}_{1i} - fT_{1i}^{*}}{(1-f)\overline{X}_{1i}}$,
hence $T_{FS}^{*}(3) = \sum_{i=1}^{k} p_i T_{0i}^{*}\frac{\overline{X}_{1i} - fT_{1i}^{*}}{(1-f)\overline{X}_{1i}}$
which is separate dual to ratio-type estimator under non-response. The dual to ratio-type estimator was introduced by Srivenkataramana [9].

Case(iv) If $\alpha = 4$then $A = 6$, $B = 0$ and $C = 0$,
so that $T_{Fi}^{*} = T_{0i}^{*}$,
consequently $T_{FS}^{*}(4) = \sum_{i=1}^{k} p_i T_{0i}^{*} = T_{0st}^{*}$
which is usual mean estimator defined in equation (1)

In order to obtain the bias and mean square error (MSE) of the proposed family, we use large sample approximation. Let $T_{0i}^{*} = \overline{X}_{0i}(1+e_0)$ and $T_{1i}^{*} = \overline{X}_{1i}(1+e_1)$ such that $E(e_0) = E(e_1) = 0$,

$$E\left(e_0^2\right) = \frac{1}{\overline{X}_{0i}^2}\left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{0i}^2 + \frac{(L_i - 1)}{n_i}W_{i2}S_{0i2}^2\right],$$

$$E\left(e_1^2\right) = \frac{1}{\overline{X}_{1i}^2}\left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{1i}^2 + \frac{(L_i - 1)}{n_i}W_{i2}S_{1i2}^2\right]$$

and

$$E(e_0e_1) = \frac{1}{\overline{X}_{0i}\overline{X}_{1i}}\left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)\rho_{01i}S_{0i}S_{1i} + \frac{(L_i - 1)}{n_i}W_{i2}\rho_{01i2}S_{0i2}S_{1i2}\right]$$

where $\overline{X}_{0i}$ is the population mean of study variable for the i^{th} stratum. ρ_{01i} and ρ_{01i2} are the correlation coefficients between study and auxiliary variables of entire group and non-response group respectively for the i^{th} stratum.

Expressing equation (5) in terms of e_0 and e_1, we get

$$T_{Fi}^{*}(\alpha) = \overline{X}_{0i}\left[1 + e_0 - e_1\phi(\alpha) - e_0e_1\phi(\alpha) + \frac{C}{A+C+fB}e_1^2\phi(\alpha) + \ldots\right] \tag{7}$$

and

$$T_{FS}^{*}(\alpha) - \overline{X}_0 = \sum_{i=1}^{k} p_i\overline{X}_{0i}\left[e_0 - e_1\phi(\alpha) - e_0e_1\phi(\alpha) + \frac{C}{A+C+fB}e_1^2\phi(\alpha) + \ldots\right] \tag{8}$$

where $\phi(\alpha) = \dfrac{C - fB}{A+C+fB}$.

Taking expectation both the sides of the equation(8) and neglecting the terms of e_0, e_1 having power greater than two, we get

$$E\left[T_{FS}^{*}(\alpha) - \overline{X}_0\right] = \sum_{i=1}^{k} p_i\overline{X}_{0i}\left[\frac{C}{A+C+fB}\phi(\alpha)E\left(e_1^2\right) - \phi(\alpha)E(e_0e_1)\right].$$

Thus the bias of $T_{FS}^{*}(\alpha)$ up to the first order of approximation is given by

$$B[T_{FS}^{*}(\alpha)] = \phi(\alpha)\left[\sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)p_i\overline{X}_{0i}\left(\frac{C}{A+C+fB}C_{1i}^2 - \rho_{01i}C_{0i}C_{1i}\right)\right.$$
$$\left. + \sum_{i=1}^{k}\frac{(L_i - 1)}{n_i}W_{i2}\frac{p_i}{\overline{X}_{1i}}\left(\frac{C}{A+C+fB}R_{01i}S_{1i2}^2 - \rho_{01i2}S_{0i2}S_{1i2}\right)\right] \tag{9}$$

where $C_{0i} = \dfrac{S_{0i}}{\overline{X}_{0i}}$, $C_{1i} = \dfrac{S_{1i}}{\overline{X}_{1i}}$ and $R_{01i} = \dfrac{\overline{X}_{0i}}{\overline{X}_{1i}}$.

Now the MSE of $T_{FS}^{*}(\alpha)$ can be obtained as

$$M\left[T_{FS}^{*}(\alpha)\right]=M\left[\sum_{i=1}^{k}p_iT_{Fi}^{*}(\alpha)\right]=\sum_{i=1}^{k}p_i^2M\left[T_{Fi}^{*}(\alpha)\right]$$
$$=\sum_{i=1}^{k}p_i^2E\left[T_{Fi}^{*}(\alpha)-\overline{X}_{0i}\right]^2. \tag{10}$$

Putting the value of $T_{Fi}^{*}(\alpha)$ from equation(7) into the above expression, we get

$$M\left[T_{FS}^{*}(\alpha)\right]=\sum_{i=1}^{k}p_i^2E\left[\overline{X}_{0i}\left(1+e_0-e_1\phi(\alpha)-e_0e_1\phi(\alpha)\right.\right.$$
$$\left.\left.+\frac{C}{A+C+fB}e_1^2\phi(\alpha)+..\right)-\overline{X}_{0i}\right]^2. \tag{11}$$

Expanding the above expression and neglecting the terms of e_0, e_1 having power greater than two, we get the MSE of $T_{FS}^{*}(\alpha)$ up to the first order of approximation as

$$M\left[T_{FS}^{*}(\alpha)\right]=\sum_{i=1}^{k}\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2\left[S_{0i}^2+\phi^2(\alpha)R_{01i}^2S_{1i}^2-2\phi(\alpha)R_{01i}\rho_{01i}S_{0i}S_{1i}\right]$$
$$+\sum_{i=1}^{k}\frac{(L_i-1)}{n_i}W_{i2}p_i^2\left[S_{0i2}^2+\phi^2(\alpha)R_{01i}^2S_{1i2}^2-2\phi(\alpha)R_{01i}\rho_{01i2}S_{0i2}S_{1i2}\right]. \tag{12}$$

3.1.1 Optimum Choice of α

In order to choose the optimum value ofα, we differentiate $M\left[T_{FS}^{*}(\alpha)\right]$ with respect to α and equate the derivative to zero.

$$\frac{\partial M\left[T_{FS}^{*}(\alpha)\right]}{\partial\alpha}=\sum_{i=1}^{k}\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2\left[2\phi(\alpha)\phi'(\alpha)R_{01i}^2S_{1i}^2-2\phi'(\alpha)R_{01i}\rho_{01i}S_{0i}S_{1i}\right]$$
$$+\sum_{i=1}^{k}\frac{(L_i-1)}{n_i}W_{i2}p_i^2\left[2\phi(\alpha)\phi'(\alpha)R_{01i}^2S_{1i2}^2-2\phi'(\alpha)R_{01i}\rho_{01i2}S_{0i2}S_{1i2}\right]=0 \tag{13}$$

where $\phi'(\alpha)$ is the first derivative of $\phi(\alpha)$ with respect to α. From the above equation , we get

$$\phi(\alpha)=\left\{\sum_{i=1}^{k}\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2R_{01i}\rho_{01i}S_{0i}S_{1i}+\sum_{i=1}^{k}\frac{(L_i-1)}{n_i}W_{i2}p_i^2R_{01i}\rho_{01i2}S_{0i2}S_{1i2}\right\}/$$
$$\left\{\sum_{i=1}^{k}\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2R_{01i}^2S_{1i}^2+\sum_{i=1}^{k}\frac{(L_i-1)}{n_i}W_{i2}p_i^2R_{01i}^2S_{1i2}^2\right\}=C_1(Constant). \tag{14}$$

The above equation is a cubic equation in α and for a given value of C_1 it provides three real roots of αat which the MSE of $T_{FS}^{*}(\alpha)$ would attain its minimum. In order to obtain the optimum value of α among the three real roots, the bias is taken into consideration. We compute the bias of the estimator at the three real roots separately and select the optimum value of α at which bias is the least.

3.1.2 Proposed Family under Different Allocations

It is a well known fact that the Neyman allocation is always preferable over the proportional allocation if the strata mean squares are known and there is no non-response in the population. But in the presence of non-response, Chaudhary et al. [6] have shown that the Neyman allocation is not always a better proposition even the strata mean squares are known. With this background, they proposed some new allocation schemes which utilize the knowledge of ‘response’ and/or ‘non-response’ rates in the case of non-response. Moreover, it is observed that in the presence of non-response, the knowledge of response and non-response rates for different strata can easily be obtained than the strata variabilities from the past records or experiences. Now, we consider the proposed family of estimators under 'proportional allocation' (PA), 'Neyman allocation' (NA) and some of the new allocation schemes proposed by Chaudhary et al. [6].

Under the proportional and Neyman allocations, we have the sample size for the i^{th} stratum $n_i = np_i$ for $i = 1,2,...,k$ and $n_i = np_iS_{0i}/\sum_{i=1}^{k} p_iS_{0i}$ for $i = 1,2,...,k$ respectively.

The $M\left[T_{FS}^{*}(\alpha)\right]$ under the proportional and Neyman allocations are respectively given by

$$M\left[T_{FS}^{*}(\alpha)\right]_{PA} = \left(\frac{1}{n_i} - \frac{1}{N_i}\right)\sum_{i=1}^{k} p_i\left[S_{0i}^2 + \phi^2(\alpha)R_{01i}^2S_{1i}^2 - 2\phi(\alpha)R_{01i}\rho_{01i}S_{0i}S_{1i}\right] +$$

$$\frac{1}{n}\sum_{i=1}^{k}(L_i - 1)W_{i2}p_i\left[S_{0i2}^2 + \phi^2(\alpha)R_{01i}^2S_{1i2}^2 - 2\phi(\alpha)R_{01i}\rho_{01i2}S_{0i2}S_{1i2}\right] \tag{15}$$

and

$$M\left[T_{FS}^{*}(\alpha)\right]_{NA} = \frac{1}{n}\left(\sum_{i=1}^{k} p_iS_{0i}\right)^2 - \frac{1}{N}\sum_{i=1}^{k} p_iS_{0i}^2 + \sum_{i=1}^{k}\left[\left(\frac{\sum_{i=1}^{k} p_iS_{0i}}{nS_{0i}} - \frac{1}{N}\right)p_i\left\{\phi^2(\alpha)R_{01i}^2S_{1i}^2 - 2\phi(\alpha)R_{01i}\rho_{01i}S_{0i}S_{1i}\right\}\right]$$

$$+\frac{1}{n}\left(\sum_{i=1}^{k} p_iS_{0i}\right)\sum_{i=1}^{k}\frac{(L_i-1)}{S_{0i}}W_{i2}p_i\left[S_{0i2}^2 + \phi^2(\alpha)R_{01i}^2S_{1i2}^2 - 2\phi(\alpha)R_{01i}\rho_{01i2}S_{0i2}S_{1i2}\right]. \tag{16}$$

Under the new allocation scheme 1 (NAS1), if $n_i = np_iW_{i1}/\sum_{i=1}^{k} p_iW_{i1}$ (W_{i1} is the response rate in the i^{th} stratum) then we have

$$M\left[T_{FS}^{*}(\alpha)\right]_{NAS1} = \sum_{i=1}^{k}\left(\frac{\sum_{i=1}^{k} p_iW_{i1}}{nW_{i1}} - \frac{1}{N}\right)p_i\left[S_{0i}^2 + \phi^2(\alpha)R_{01i}^2S_{1i}^2 - 2\phi(\alpha)R_{01i}\rho_{01i}S_{0i}S_{1i}\right] +$$

$$\frac{1}{n}\left(\sum_{i=1}^{k} p_iW_{i1}\right)\sum_{i=1}^{k}(L_i-1)\frac{W_{i2}}{W_{i1}}p_i\left[S_{0i2}^2 + \phi^2(\alpha)R_{01i}^2S_{1i2}^2 - 2\phi(\alpha)R_{01i}\rho_{01i2}S_{0i2}S_{1i2}\right]. \tag{17}$$

Under the new allocation scheme 2 (NAS2), if $n_i = \frac{np_i}{W_{i2}\sum_{i=1}^{k}\frac{p_i}{W_{i2}}}$ then we get

$$M\left[T_{FS}^{*}(\alpha)\right]_{NAS2} = \sum_{i=1}^{k}\left(\frac{W_{i2}\sum_{i=1}^{k}\frac{p_i}{W_{i2}}}{n} - \frac{1}{N}\right)p_i\left[S_{0i}^2 + \phi^2(\alpha)R_{01i}^2S_{1i}^2 - 2\phi(\alpha)R_{01i}\rho_{01i}S_{0i}S_{1i}\right] +$$

$$\frac{1}{n}\left(\sum_{i=1}^{k}\frac{p_i}{W_{i2}}\right)\sum_{i=1}^{k}(L_i-1)W_{i2}^2p_i\left[S_{0i2}^2 + \phi^2(\alpha)R_{01i}^2S_{1i2}^2 - 2\phi(\alpha)R_{01i}\rho_{01i2}S_{0i2}S_{1i2}\right]. \tag{18}$$

Under the new allocation scheme 3 (NAS3), if $n_i = \frac{np_iW_{i1}}{W_{i2}\sum_{i=1}^{k}\frac{p_iW_{i1}}{W_{i2}}}$ then we have

$$M\left[T_{FS}^{*}(\alpha)\right]_{NAS3} = \sum_{i=1}^{k}\left[\frac{W_{i2}}{nW_{i1}}\left(\sum_{i=1}^{k}\frac{p_iW_{i1}}{W_{i2}}\right) - \frac{1}{N}\right]p_i\left[S_{0i}^2 + \phi^2(\alpha)R_{01i}^2S_{1i}^2 - 2\phi(\alpha)R_{01i}\rho_{01i}S_{0i}S_{1i}\right] +$$

$$\frac{1}{n}\left(\sum_{i=1}^{k}\frac{p_iW_{i1}}{W_{i2}}\right)\sum_{i=1}^{k}(L_i-1)\frac{W_{i2}^2}{W_{i2}}p_i\left[S_{0i2}^2 + \phi^2(\alpha)R_{01i}^2S_{1i2}^2 - 2\phi(\alpha)R_{01i}\rho_{01i2}S_{0i2}S_{1i2}\right]. \tag{19}$$

3.2 Combined-type Family of Estimators

In the similar manner, the combined-type family of estimators of population mean $\overline{X}_0$ in stratified random sampling while both study and auxiliary variables go through the non-response, is given as

$$T_{FC}^{*}(\alpha) = T_{0st}^{*}\left[\frac{(A+C)\overline{X}_1 + fBT_{1st}^{*}}{(A+fB)\overline{X}_1 + CT_{1st}^{*}}\right]. \tag{20}$$

The above proposed family may produce a number of estimators of population mean $\overline{X}_0$ under non-response. For example if$\alpha = 1$, we get $T^*_{FC} = T^*_{0st}\frac{\overline{X}_1}{T^*_{1st}}$ which is usual combined ratio estimator in stratified random sampling when both the variables are subjected to non-response. Similarly, for $\alpha = 2, 3$ and 4, we have $T^*_{FC} = T^*_{0st}\frac{T^*_{1st}}{\overline{X}_1}$(usual combined product estimator under non-response), $T^*_{FC}(3) = T^*_{0st}\frac{\overline{X}_1 - fT^*_{1st}}{(1-f)\overline{X}_1}$ (combined dual to ratio-type estimator under non-response) and $T^*_{FC}(4) = T^*_{0st} = T^*_{FS}(4)$(usual mean estimator in stratified random sampling under non-response) respectively.

To obtain bias and MSE of$T^*_{FC}(\alpha)$, we use large sample approximation. Let

$T^*_{0st} = \overline{X}_0(1+e_2)$ and $T^*_{1st} = \overline{X}_1(1+e_3)$

such that $E(e_2) = E(e_3) = 0$,

$$E\left(e_2^2\right) = \frac{1}{\overline{X}_0^2}\left[\sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)p_i^2 S_{0i}^2 + \sum_{i=1}^{k}\frac{(L_i - 1)}{n_i}W_{i2}p_i^2 S_{0i2}^2\right],$$

$$E\left(e_3^2\right) = \frac{1}{\overline{X}_1^2}\left[\sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)p_i^2 S_{1i}^2 + \sum_{i=1}^{k}\frac{(L_i - 1)}{n_i}W_{i2}p_i^2 S_{1i2}^2\right]$$

and $E(e_0e_1) = \frac{1}{\overline{X}_0\overline{X}_1}\left[\sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)p_i^2\rho_{01i}S_{0i}S_{1i} + \sum_{i=1}^{k}\frac{(L_i-1)}{n_i}W_{i2}p_i^2\rho_{01i2}S_{0i2}S_{1i2}\right]$.

Under the above assumptions, equation(20) can be expressed in the terms of e_2 and e_3 as

$$T^*_{FC}(\alpha) - \overline{X}_0 = \overline{X}_0\left[e_2 - e_2\phi(\alpha) - e_2e_3\phi(\alpha) + \frac{C}{A+C+fB}e_3^2\phi(\alpha) + ...\right]. \quad (21)$$

Taking expectation both the sides of equation (21) and neglecting the terms of $e_2$and e_3 having power greater than two, we get bias of $T^*_{FC}(\alpha)$ up to the first order of approximation

$$E\left[T^*_{FC}(\alpha) - \overline{X}_0\right] = \overline{X}_0\left[\frac{C}{A+C+fB}\phi(\alpha)E\left(e_3^2\right) - \phi(\alpha)E(e_2e_3)\right]$$

$$\Rightarrow B\left[T^*_{FC}(\alpha)\right] = \frac{\phi(\alpha)}{\overline{X}_1}\left[\sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)p_i^2\left(\frac{C}{A+C+fB}R_{01}S_{1i}^2 - \rho_{01i}S_{0i}S_{1i}\right) + \right.$$

$$\left.\sum_{i=1}^{k}\frac{(L_i - 1)}{n_i}W_{i2}p_i^2\left(\frac{C}{A+C+fB}R_{01}S_{1i2}^2 - \rho_{01i2}S_{0i2}S_{1i2}\right)\right] \quad (22)$$

where $R_{01} = \frac{\overline{X}_0}{\overline{X}_1}$.

In the sequence of obtaining the MSE of $T^*_{FC}(\alpha)$ up to the first order of approximation, we square both the sides of equation (21) and take expectation ignoring the terms of e_2 and e_3 having power greater than two

$$E\left[T^*_{FC}(\alpha) - \overline{X}_0\right]^2 = \overline{X}_0^2\left[E\left(e_2^2\right) - \phi^2(\alpha)E\left(e_3^2\right) - 2\phi(\alpha)E(e_2e_3)\right]$$

$$\Rightarrow M\left[T^*_{FC}(\alpha)\right] = \sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)p_i^2\left[S_{0i}^2 + \phi^2(\alpha)R_{01}^2S_{1i}^2 - 2\phi(\alpha)R_{01}\rho_{01i}S_{0i}S_{1i}\right] +$$

$$\sum_{i=1}^{k}\frac{(L_i - 1)}{n_i}W_{i2}p_i^2\left[S_{0i2}^2 + \phi^2(\alpha)R_{01}^2S_{1i2}^2 - 2\phi(\alpha)R_{01}\rho_{01i2}S_{0i2}S_{1i2}\right]. \quad (23)$$

3.2.1 Optimum Choice of α

To obtain the minimum MSE of$T^*_{FC}(\alpha)$, we find the optimum value of α on differentiating the $M\left[T^*_{FC}(\alpha)\right]$ with respect to $\alpha$and equating the derivative to zero. Thus we have

$$\frac{\partial M\left[T^*_{FC}(\alpha)\right]}{\partial\alpha} = \sum_{i=1}^{k}\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2\left[2\phi(\alpha)\phi'(\alpha)R_{01}^2S_{1i}^2-2\phi'(\alpha)R_{01}\rho_{01i}S_{0i}S_{1i}\right]$$
$$+\sum_{i=1}^{k}\frac{(L_i-1)}{n_i}W_{i2}p_i^2\left[2\phi(\alpha)\phi'(\alpha)R_{01}^2S_{1i2}^2-2\phi'(\alpha)R_{01}\rho_{01i2}S_{0i2}S_{1i2}\right]=0$$
$$\Rightarrow\phi(\alpha)=\frac{\sum_{i=1}^{k}\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2\rho_{01i}S_{0i}S_{1i}+\sum_{i=1}^{k}\frac{(L_i-1)}{n_i}W_{i2}p_i^2\rho_{01i2}S_{0i2}S_{1i2}}{\sum_{i=1}^{k}\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2R_{01}S_{1i}^2+\sum_{i=1}^{k}\frac{(L_i-1)}{n_i}W_{i2}p_i^2R_{01}S_{1i2}^2}=C_2(Constant). \tag{24}$$

Since $\phi(\alpha)$ is a cubic function of the parameter α. Thus for a given value of C_2, equation (24) provides three real roots of αat which we get the minimum MSE of $T^*_{FC}(\alpha)$.

3.2.2 Proposed Family under Different Allocations

In this section, we discuss the proposed combined-type family of estimators under proportion allocation, Neyman allocation and some new allocation schemes considered in section 3.1.2.

The sample size of i^{th} stratum and MSE of the proposed family under proportional allocation are respectively given by

$n_i = np_i$ for $i = 1,2,...,k$ and

$$M\left[T^*_{FC}(\alpha)\right]_{PA}=\left(\frac{1}{n_i}-\frac{1}{N_i}\right)\sum_{i=1}^{k}p_i\left[S_{0i}^2+\phi^2(\alpha)R_{01}^2S_{1i}^2-2\phi(\alpha)R_{01}\rho_{01i}S_{0i}S_{1i}\right]+$$
$$\frac{1}{n}\sum_{i=1}^{k}(L_i-1)W_{i2}p_i\left[S_{0i2}^2+\phi^2(\alpha)R_{01}^2S_{1i2}^2-2\phi(\alpha)R_{01}\rho_{01i2}S_{0i2}S_{1i2}\right]. \tag{25}$$

Under Neyman allocation, the sample size of i^{th} stratum and MSE of $T^*_{FC}(\alpha)$ are respectively given by

$n_i=\frac{np_iS_{0i}}{\sum_{i=1}^{k}p_iS_{0i}}$ for $i = 1,2,...,k$ and

$$M\left[T^*_{FC}(\alpha)\right]_{NA}=\frac{1}{n}\left(\sum_{i=1}^{k}p_iS_{0i}\right)^2-\frac{1}{N}\sum_{i=1}^{k}p_iS_{0i}^2+$$
$$\sum_{i=1}^{k}\left[\left(\frac{\sum_{i=1}^{k}p_iS_{0i}}{nS_{0i}}-\frac{1}{N}\right)p_i\left\{\phi^2(\alpha)R_{01}^2S_{1i}^2-2\phi(\alpha)R_{01}\rho_{01i}S_{0i}S_{1i}\right\}\right]+$$
$$\frac{1}{n}\left(\sum_{i=1}^{k}p_iS_{0i}\right)\sum_{i=1}^{k}\frac{(L_i-1)}{S_{0i}}W_{i2}p_i\left[S_{0i2}^2+\phi^2(\alpha)R_{01}^2S_{1i2}^2-2\phi(\alpha)R_{01}\rho_{01i2}S_{0i2}S_{1i2}\right]. \tag{26}$$

The sample size of i^{th} stratum and MSE of $T^*_{FC}(\alpha)$ under the new allocation scheme 1 (NAS1) are respectively given by

$n_i=\frac{np_iW_{i1}}{\sum_{i=1}^{k}p_iW_{i1}}$ for $i = 1,2,...,k$ and

$$M\left[T^*_{FC}(\alpha)\right]_{NAS1}=\sum_{i=1}^{k}\left(\frac{\sum_{i=1}^{k}p_iW_{i1}}{nW_{i1}}-\frac{1}{N}\right)p_i\left[S_{0i}^2+\phi^2(\alpha)R_{01}^2S_{1i}^2-2\phi(\alpha)R_{01}\rho_{01i}S_{0i}S_{1i}\right]+$$
$$\frac{1}{n}\left(\sum_{i=1}^{k}p_iW_{i1}\right)\sum_{i=1}^{k}(L_i-1)\frac{W_{i2}}{W_{i1}}p_i\left[S_{0i2}^2+\phi^2(\alpha)R_{01}^2S_{1i2}^2-2\phi(\alpha)R_{01}\rho_{01i2}S_{0i2}S_{1i2}\right]. \tag{27}$$

Under the new allocation scheme 2 (NAS2), the sample size of i^{th} stratum and MSE of $T^*_{FC}(\alpha)$ are respectively given by

$n_i = \frac{np_i}{W_{i2}\sum_{i=1}^{k}\frac{p_i}{W_{i2}}}$ for $i = 1,2,...,k$ and

$$M\left[T_{FC}^{*}(\alpha)\right]_{NAS2} = \sum_{i=1}^{k}\left(\frac{W_{i2}\sum_{i=1}^{k}\frac{p_i}{W_{i2}}}{n} - \frac{1}{N}\right)p_i\left[S_{0i}^2 + \phi^2(\alpha)R_{01}^2S_{1i}^2 - 2\phi(\alpha)R_{01}\rho_{01i}S_{0i}S_{1i}\right] + \frac{1}{n}\left(\sum_{i=1}^{k}\frac{p_i}{W_{i2}}\right)\sum_{i=1}^{k}(L_i - 1)W_{i2}^2p_i\left[S_{0i2}^2 + \phi^2(\alpha)R_{01}^2S_{1i2}^2 - 2\phi(\alpha)R_{01}\rho_{01i2}S_{0i2}S_{1i2}\right]. \tag{28}$$

The sample size of i^{th} stratum and MSE of $T_{FC}^{*}(\alpha)$ under the new allocation scheme 3 (NAS3) are respectively given by

$n_i = \frac{np_iW_{i1}}{W_{i2}\sum_{i=1}^{k}\frac{p_iW_{i1}}{W_{i2}}}$ for $i = 1,2,...,k$ and

$$M\left[T_{FC}^{*}(\alpha)\right]_{NAS3} = \sum_{i=1}^{k}\left[\frac{W_{i2}}{nW_{i1}}\left(\sum_{i=1}^{k}\frac{p_iW_{i1}}{W_{i2}}\right) - \frac{1}{N}\right]p_i\left[S_{0i}^2 + \phi^2(\alpha)R_{01}^2S_{1i}^2 - 2\phi(\alpha)R_{01}\rho_{01i}S_{0i}S_{1i}\right] + \frac{1}{n}\left(\sum_{i=1}^{k}\frac{p_iW_{i1}}{W_{i2}}\right)\sum_{i=1}^{k}(L_i - 1)\frac{W_{i2}^2}{W_{i1}}p_i\left[S_{0i2}^2 + \phi^2(\alpha)R_{01}^2S_{1i2}^2 - 2\phi(\alpha)R_{01}\rho_{01i2}S_{0i2}S_{1i2}\right]. \tag{29}$$

4 Empirical Study

In order to support the theoretical results, we have taken the data considered by Chaudhary et al. [7]. There are 284 municipalities divided into four strata consisting of 73, 70, 97 and 44 municipalities respectively. The population in the year 1985 has been considered as study variable whereas the population in the year 1975 is assumed to be auxiliary variable. The parameters of the population are given below:

Table 1: Particulars of Data

Stratum No.	N_i	$\bar{X}_{0i}$	$\bar{X}_{1i}$	S_{0i}^2	S_{1i}^2	S_{0i2}^2	S_{1i2}^2	ρ_{01i}	ρ_{01i2}
1	73	40.85	39.56	6369.0999	6624.4398	618.8844	495.1075	0.999	0.799
2	70	27.83	27.57	1051.0725	1147.0111	240.905	192.724	0.998	0.798
3	97	25.79	25.44	2014.9651	2205.4021	265.522	212.4176	0.999	0.799
4	44	20.64	20.36	538.4749	485.2655	83.6944	66.95552	0.997	0.797

Table 2 gives the values of $\phi(\alpha)$ for different sets of non-response rates W_{i2} on which $T^*_{FS}(\alpha)$ provides the optimum estimates under the different allocation schemes.

Table 2: $\phi(\alpha)$ for $T^*_{FS}(\alpha)$ under Different Allocation Schemes $(L_i = 2\ \forall\ i)$

Stratum No.	W_{i2}(%)	$\phi(\alpha)$ under				
		PA	NA	NAS1	NAS2	NAS3
1	5	0.9483	0.9507	0.9686	0.9508	0.9504
2	10					
3	15					
4	20					
1	5	0.9483	0.9506	0.948	0.947	0.9465
2	10					
3	20					
4	15					
1	10	0.9482	0.9507	0.9484	0.9501	0.9498
2	5					
3	15					
4	20					
1	20	0.9477	0.9503	0.9474	0.9466	0.9467
2	15					
3	10					
4	5					

Table 3 represents the MSE comparison of the proposed family $T^*_{FS}(\alpha)$ at $\alpha = 4$, $\alpha = 1$ and α_{opt} under different allocation schemes for the different sets of W_{i2}.

Table 3: MSE of $T^*_{FS}(\alpha)$ under Different Allocation Schemes $(L_i = 2\forall i)$

St. No.	W_{i2}(%)	$M[T^*_{FS}(\alpha)]$ under														
		PA			NA			NAS1			NAS2			NAS3		
		$\alpha = 4$	$\alpha = 1$	α_{opt}	$\alpha = 4$	$\alpha = 1$	α_{opt}	$\alpha = 4$	$\alpha = 1$	α_{opt}	$\alpha = 4$	$\alpha = 1$	α_{opt}	$\alpha = 4$	$\alpha = 1$	α_{opt}
1	5	36.046	0.38	0.274	28.683	0.37	0.294	33.486	0.368	0.271	31.624	0.418	0.334	32.674	0.434	0.346
2	10															
3	15															
4	20															
1	5	36.109	0.403	0.297	28.745	0.393	0.317	34.002	0.399	0.298	35.697	0.517	0.406	37.923	0.555	0.435
2	10															
3	20															
4	15															
1	10	36.134	0.413	0.306	28.689	0.373	0.296	34.85	0.406	0.304	43.01	0.515	0.397	44.467	0.534	0.411
2	5															
3	15															
4	20															
1	20	36.397	0.512	0.402	28.85	0.434	0.356	38.295	0.535	0.418	62.311	0.842	0.646	69.941	0.934	0.714
2	15															
3	10															
4	5															

Table 4 reveals the values of $\phi(\alpha)$ for the different sets of W_{i2} on which the proposed family $T^{*}_{FC}(\alpha)$ gives the optimum estimates under the different allocation schemes.

Table 4: $\phi(\alpha)$ for $T^{*}_{FC}(\alpha)$ under Different Allocation schemes $(L_i = 2)$

Stratum No.	W_{i2}(%)	$\phi(\alpha)$ under				
		PA	NA	NAS1	NAS2	NAS3
1 2 3 4	5 10 15 20	0.9537	0.9512	0.953	0.9501	0.9492
1 2 3 4	5 10 20 15	0.9536	0.9511	0.9522	0.9455	0.9446
1 2 3 4	10 5 15 20	0.9535	0.9512	0.9533	0.9541	0.9534
1 2 3 4	20 15 10 5	0.9531	0.9509	0.9532	0.9545	0.955

Table 5 depicts the MSE of $T^{*}_{FC}(\alpha)$ at $\alpha = 4$, $\alpha = 1$ and α_{opt} under different allocation schemes for the choices of different sets of W_{i2}.

Table 5: MSE of $T^{*}_{FC}(\alpha)$ under Different Allocation Schemes $(L_i = 2)$

St. No	W_{i2}(%)	$M[T^{*}_{FC}(\alpha)]$ under														
		PA			NA			NAS1			NAS2			NAS3		
		$\alpha=4$	$\alpha=1$	α_{opt}	$\alpha=4$	$\alpha=1$	α_{opt}	$\alpha=4$	$\alpha=1$	α_{opt}	$\alpha=4$	$\alpha=1$	α_{opt}	$\alpha=4$	$\alpha=1$	α_{opt}
1 2 3 4	5 10 15 20	36.046	0.361	0.277	28.683	0.371	0.299	33.486	0.355	0.274	31.624	0.422	0.336	32.674	0.44	0.348
1 2 3 4	5 10 20 15	36.109	0.385	0.3	28.745	0.394	0.318	34.002	0.386	0.301	35.697	0.526	0.409	37.923	0.567	0.438
1 2 3 4	10 5 15 20	36.134	0.394	0.309	28.689	0.373	0.298	34.85	0.389	0.306	43.01	0.498	0.399	44.467	0.518	0.413
1 2 3 4	20 15 10 5	36.397	0.492	0.405	28.85	0.433	0.357	38.295	0.513	0.422	62.311	0.792	0.652	69.941	0.874	0.72

5 Conclusion

In the present article, we have proposed some families of factor-type estimators of population mean in stratified random sampling using an auxiliary variable under non-response. We have suggested separate and combined-type families of estimators of population mean whenever non-response is observed on both study and auxiliary variables. The optimum properties of the suggested families have been conferred. The theoretical study of the suggested families has been carried out under the proportional, Neyman and some new allocation schemes based on response or/and non-response rates. In order to sustain the theoretical results, an empirical study has also been done.

The tables 3 and 5 present a salient feature of comparison of proportional and Neyman allocations with some new allocation schemes based on response or/and non-response rates through the suggested families $T_{FS}^{*}(\alpha)$ and $T_{FC}^{*}(\alpha)$ respectively. In both the tables, the optimum estimators provide better estimates than the usual separate (combined) ratio and mean estimators under non-response. It is also revealed that in most of the situations (for different choices of W_{i2}), the allocation schemes NAS1, NAS2 and NAS3, depending upon the knowledge of p_i and W_{i1} (or/andW_{i2}), provide more prcised estimates as compared to proportional allocation as well as Neyman allocation.

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, 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. Khoshnevisan, R. Singh, P. Chauhan, N. Sawan and F.Smarandache, A general family of estimators for estimating population mean using known value of some population parameter(s), Far East Journal of Theoretical Statistics, 22, 181–191(2007).
[6] M. K. Chaudhary, V. K. Singh, R. Singh and F. Smarandache, On some new allocation schemes in stratified random sampling under non-response, Studies in Sampling Techniques and Time series, Zip Publishing 1313 Chesapeake Avenue Columbus, Ohio 43212, USA, 40-54(2011).
[7] M. K. Chaudhary, V. K. Singh and R. K. Shukla, Combined-type family of estimators of population mean in stratified random sampling under non-response, Journal of Reliability and Statistical Studies (JRSS), Vol. 5, No. 2, 133-142(2012).
[8] V. K. Singh and D. Shukla, One parameter family of factor-type ratio estimators, Metron, 45 (1-2), 273-283(1987).
[9] T. Srivenkataramana, A dual to ratio estimator in sample surveys, Biometrika, 67 , 199-204(1980).

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

Saurabh is research scholar in the Department of Statistics, Banaras Hindu University, Varanasi, INDIA. His area of research is sampling theory. He has published more than 5 research papers in national and international journals.