Journal of Modern Applied Statistical Methods

Volume 23 | Issue 1 Article 12

Enhanced Estimators and Effective Estimation Procedures for Population Variance under Missing at Random Data in Successive Sampling

Anup Kumar Sharma
Department of Mathematics, National Institute of Technology Raipur, India,
aksharma.ism@gmail.com

Alok Kumar Singh
Department of Statistics, Sri Venkateswara College, University of Delhi, India,
aksingh.ism@gmail.com

Recommended Citation

Anup Kumar Sharma, Alok Kumar Singh (2024). Enhanced Estimators and Effective Estimation Procedures for Population Variance under Missing at Random Data in Successive Sampling. Journal of Modern Applied Statistical Methods, 23(1), https://doi.org/10.56801/Jmasm.V23.i1.12

Journal of Modern Applied Statistical Methods
2024, Vol. 23, No. 1,
Doi: 10.56801/Jmasm.V23.i1.12

ISSN 1538 – 9472

Enhanced Estimators and Effective Estimation Procedures for Population Variance under Missing at Random Data in Successive Sampling

Anup Kumar Sharma
Department of Mathematics, National Institute of Technology Raipur, India

Alok Kumar Singh
Department of Statistics, Sri Venkateswara College, University of Delhi, India

This paper explores some potent estimation procedures of population variance under missing at random situations in two-occasion successive sampling. Information on auxiliary variables has been involved in supporting for effective estimation of some chain-type exponential and regression estimators under the assumption of sampling units on which information of study variables cannot be obtained due to missing at random. The non-respondents follow the Binomial type of distribution in the estimation. The proposed estimators are compared with the competent estimators under a complete response of sample units. The Empirical studies support theoretical results in the present probability of non-respondent. Results have been interpreted and suitable recommendations are made for data practitioners.

Keywords: Successive sampling, auxiliary variable, random non-response, variance estimator, bias, mean square error.

1. Introduction

When the population characteristics are changed quickly, census at long and infrequent intervals is not much used. It is desirable to take a sample at annual or even at shorter intervals regularly to capture these changes in such circumstances. Similarly, one may have to resort to the repeated sampling of a population where several kinds of data are to be collected and published at regular intervals. All this is possible while performing successive sampling or sampling on successive occasions. The sampling method from the same population on successive occasions is called multiple sampling, rotation sampling, or successive sampling. This kind of sampling involves the repetition of some units selected on previous occasions to improve the estimator's efficiency for subsequent occasions.

Jessen (1942) introduced successive sampling by utilizing the information obtained on a previous occasion with the partial replacement of sampling units. The theory of successive sampling was further extended by Eckler (1955), Patterson (1950), Tikkiwal (1951), Rao and Graham (1964), Gupta (1979), among others. Sen (1971) applied this theory with success in designing the population's estimator on the current occasion by using the information on two auxiliary variables available from the previous occasion. Further, Sen (1972, 1973) extended his work for several auxiliary variables. Singh *et al*. (1991) and Singh and Singh (2001) used the auxiliary information on the current occasion for estimating the current population mean in two occasions successive sampling.

The use of auxiliary information plays a vital role in estimating the population parameter to improve the survey estimates efficiency. In successive sampling, it is empirically verified that the utilization of full information gathered at the previous investigation is very advantageous. In many situations, information on auxiliary variables with known population parameters and have fair correlations with the study variable may be available on the first as well on the second occasion. For example, in survey sampling of transportation, each vehicle's seat capacity is known; the number of employees in different organizations may be known in employee satisfaction surveys.

Utilizing the auxiliary information on both occasions, Feng and Zou (1997), Birader and Singh (2001), Singh and Karna (2009), Singh and Priyanka (2010), Singh and Sharma (2014), Singh *et al.* (2015), Singh *et al.* (2018), Singh *et al.* (2020) have proposed varieties of estimators of population mean on current (second) occasions in two occasions successive sampling.

The approach used to estimate the current population mean was further extended by Sud *et al*. (2001) to estimate the current population variance in two successive sampling occasions. Azam *et al*. (2010) proposed an estimation strategy for current population variance by considering the linear combination of all available information from the current and previous occasions. Recently, Singh *et al*. (2011) suggested a class of estimators of finite population variance on current occasion in two occasions of successive sampling.

It is to be mentioned that all the above-referred works deal with the estimation of population variance on the current occasion in successive sampling based on the complete response available from the sample. In many practical situations encountered in sample surveys, observations are not available for all the selected units in the sample, i.e., some observations may be missing for various reasons such as the unwillingness of some selected units to supply the desired information.

Incompleteness (in the form of absence) is a troublesome feature of many data sets. Statisticians have long known that failure to account for incompleteness's stochastic nature can damage the actual conclusion. Rubin (1976) advocated three concepts: missing at random (MAR), observed at random (OAR), and parameter distribution (PD). Rubin defined: "The data are MAR if the probability of the observed missingness pattern, given the observed and unobserved data, does not depend on the value of the unobserved data". The response mechanism of missing at random (MAR) helps estimate the population variance in cost-effectiveness when the number of missing units is present in survey data. Tracy and Osahan (1994) studied the effect of random non-response on the conventional ratio estimator of the population mean in two situations: (i) non-response on the study as well as on the auxiliary variables and (ii) non-response on the study variable only. Singh and Joarder (1998), Singh *et al.* (2003), Ahmed *et al.* (2005) and Sharma (2017) have studied the effect of random non-response on different estimators of the population variance.

Motivated with the above works and utilizing the information on an auxiliary variable over both occasions, modified chain type exponential and regression estimators have been proposed to estimate the current population variance when random non-response occurs on the study and auxiliary variables both in two occasions successive sampling. Properties of the proposed estimators are studied and well supported with empirical studies. Results are interpreted and recommendations to the survey practitioners have been made.

2. Sample Structures and distribution of random non-response

Consider a finite population $U=(U_1, U_2, ..., U_N)$ of N units which has been sampled over two occasions. The character under study is denoted by x(y) on the first (second) occasion, respectively. It is assumed that the information on an auxiliary variable z (stable over occasions) whose population variance is known and readily available on both occasions. We assume that the missing at random occurs in study and auxiliary variables on both occasions. Let a simple random sample (without replacement) of size n be drawn on the first occasion. A random sub-sample of size m is retained (matched) from the sample on the first occasion for its use on the second occasion, while a fresh simple random sample (without replacement) of size u = (n-m) is drawn on the second occasion from the entire population so that the total sample size on this occasion is also n.

Let $r_1 \{r_1 = 0, 1, 2, ..., (n-2)\}$ be the number of sampling units on the first occasion on which information could not be obtained due to missing at random and then the remaining $(n-r_1)$ units in the sample can be treated as simple random sampling without replacements. It is assumed that r_1 is less than $(n-1)$ i.e., $0 \le r_1 \le (n-2)$. Similarly in fresh sample of size u on current occasion $r_2 \{r_2 = 0, 1, 2, ..., (u-2)\}$ denotes the number of non responding units and the remaining $(u-r_2)$ units in the sample can be treated as simple random sampling without replacements, such that $0 \le r_2 \le (u-2)$. Let p_1 and p_2 be the probabilities of non-response among $(n-2)$ and $(u-2)$ possible values of non-response respectively, hence r_1 and r_2 have the following Binomial type discrete probability distributions of:

$$P\left(r_1\right)=\frac{n-r_1}{nq_1+2p_1}\,{}^{n-2}C_{r_1}p_1^{r_1}q_1^{n-r_1-2},\ r_1=0,1,2,...,n-2\ .$$

and

$$P\left(r_2\right)=\frac{u-r_2}{uq_2+2p_2}\,{}^{u-2}C_{r_2}p_2^{r_2}q_2^{u-r_2-2},\ r_2=0,1,2,...,u-2$$

where $q_1=1-p_1$ and $q_2=1-p_2$.

${}^{n-2}C_{r_1}$ and ${}^{u-2}C_{r_2}$ are total number of ways to obtaining $r_i\,(i=1,2)$ non-response out of $(n-2)$ and $(u-2)$ non-responses respectively.

Further we use the following notations:

$\bar{x}_n^*=\frac{1}{n-r_1}\sum_{i=1}^{n-r_1}x_i,\ \bar{y}_u^*=\frac{1}{u-r_2}\sum_{i=1}^{u-r_2}y_i$: Sample means of the respective variables based on number of responding units of size $(n-r_1)$ and $(u-r_2)$ on first and second occasions respectively.

$\bar{z}_u^*=\frac{1}{u-r_2}\sum_{i=1}^{u-r_2}z_i$: Sample mean of the auxiliary variable z based on sample size ($u-r_2$).

$$s_{x_n}^{*2}=\frac{1}{n-r_1-1}\sum_{i=1}^{n-r_1}\left(x_i-\bar{x}_n^*\right)^2,\ s_{y_u}^{*2}=\frac{1}{u-r_2-1}\sum_{i=1}^{u-r_2}\left(y_i-\bar{y}_u^*\right)^2,\ s_{z_u}^{*2}=\frac{1}{u-r_2-1}\sum_{i=1}^{u-r_2}\left(z_i-\bar{z}_u^*\right)^2 :$$

Sample variances of study variables x and y and auxiliary variables z based on the responding part of sample of sizes $(n-r_1)$ and $(u-r_2)$ on first and second occasions, respectively.

$s_{y_m}^2, s_{x_m}^2, s_{z_m}^2$: Sample variances of the variables y, x and z respectively based on matched sample of size m.

3. The Proposed Estimator

To estimate the population variance S_y^2 of the study variable on the current (second) occasion, two different sets of estimators are considered. One set of estimators $\Delta_u=\left\{T_{1u},T_{2u}\right\}$ based on sample of size u drawn afresh on the second occasion and another set of estimators $\Delta_m=\left\{T_{1m},T_{2m}\right\}$ based on sample of size m which is common to both the occasions. Since the random non-response occurs on study and auxiliary variables both on two occasions and population variance S_z^2 of the auxiliary character is known then the estimators of sets Δ_u and Δ_m are formulated as

$$T_{1u} = s_{y(u)}^{*2} \exp\left[\frac{S_z^2 - s_{z(u)}^{*2}}{S_z^2 + s_{z(u)}^{*2}}\right] \tag{1}$$

$$T_{2u} = \left(\frac{s_{y(u)}^{*2}}{s_{z(u)}^{*2}} S_z^2\right) \exp\left[\frac{S_z^2 - s_{z(u)}^{*2}}{S_z^2 + s_{z(u)}^{*2}}\right] \tag{2}$$

$$T_{1m} = s_{y(m)}^{2} \left(\frac{s_{x(n)}^{*2}}{s_{x(m)}^{2}}\right)\left(\frac{S_z^2}{s_{z(m)}^{2}}\right) \tag{3}$$

and

$$T_{2m} = \left(s_{y(m)}^{2} + \beta\left(s_{x(n)}^{*2} - s_{x(m)}^{2}\right)\right) \exp\left[\frac{S_z^2 - s_{z(m)}^{2}}{S_z^2 + s_{z(m)}^{2}}\right] \tag{4}$$

where β is some suitably chosen constant and to be determined optimally.

Combining the estimators of sets Δ_u and Δ_m, we have the following estimators of the population variance S_y^2 as

$$T_{ij} = \varphi_{ij} T_{iu} + \left(1 - \varphi_{ij}\right) T_{jm};\ (i, j = 1, 2) \tag{5}$$

where $\varphi_{ij}\ (0 \le \varphi_{ij} \le 1)$ are the unknown constants and to be determined under certain criterions.

4. Properties of the Proposed Estimators

To obtain the Bias and Mean square errors (MSE) of the proposed estimators $T_{ij}\ (j = 1, 2)$ under large sample approximations, following transformations have been considered:

$s_{y(u)}^{*2} = S_y^2(1+e_0)$, $s_{z(u)}^{*2} = S_z^2(1+e_1)$, $s_{x(n)}^{*2} = S_x^2(1+e_2)$, $s_{x(m)}^{2} = S_x^2(1+e_3)$, $s_{z(m)}^{2} = S_z^2(1+e_4)$, $s_{y(m)}^{2} = S_y^2(1+e_5)$ such that $|e_i| < 1 \forall (i = 1, 2, ..., 5)$.

Thus the various expected values are obtained as;

$E(e_0^2) = f_u(\lambda_{040} - 1)$, $E(e_1^2) = f_u(\lambda_{004} - 1)$, $E(e_2^2) = f_n(\lambda_{400} - 1)$, $E(e_3^2) = f_m(\lambda_{040} - 1)$, $E(e_4^2) = f_m(\lambda_{004} - 1)$, $E(e_5^2) = f_m(\lambda_{040} - 1)$, $E(e_0 e_1) = f_u(\lambda_{022} - 1)$, $E(e_2 e_3) = f_n(\lambda_{400} - 1)$, $E(e_4 e_5) = f_m(\lambda_{022} - 1)$, $E(e_2 e_4) = f_n(\lambda_{202} - 1)$, $E(e_3 e_5) = f_m(\lambda_{220} - 1)$, $E(e_2 e_5) = f_n(\lambda_{220} - 1)$ $E(e_3 e_4) = f_m(\lambda_{202} - 1)$

where

$\lambda_{pqr}=\dfrac{\mu_{pqr}}{\mu_{200}^{p/2}\mu_{020}^{q/2}\mu_{002}^{r/2}}$; $\mu_{pqr}=\dfrac{1}{N}\sum_{i=1}^{N}\left(x_i-\bar{Y}\right)^{p}\left(y_i-\bar{X}\right)^{q}\left(z_i-\bar{Z}\right)^{r}$; p, q, r being the non negative integers.

$$f_u=\left(\frac{1}{uq_2+2p_2}-\frac{1}{N}\right),\ f_n=\left(\frac{1}{nq_1+2p_1}-\frac{1}{N}\right),\ f_m=\left(\frac{1}{m}-\frac{1}{N}\right).$$

Under the above transformations the estimators given in equations (1)-(4) takes the following forms;

$$T_{1u}=S_y^2(1+e_0)\exp\left[-\frac{e_1}{2}\left(1+\frac{e_1}{2}\right)^{-1}\right] \tag{6}$$

$$T_{2u}=\left(\frac{S_y^2(1+e_0)}{S_z^2(1+e_1)}S_z^2\right)\exp\left[-\frac{e_1}{2}\left(1+\frac{e_1}{2}\right)^{-1}\right] \tag{7}$$

$$T_{1m}=S_y^2(1+e_5)\left\{\frac{S_x^2(1+e_2)}{S_x^2(1+e_3)}\right\}\left\{\frac{S_z^2}{S_z^2(1+e_4)}\right\} \tag{8}$$

and

$$T_{2m}=\left[S_y^2(1+e_5)+\beta\left(S_x^2(1+e_2)-S_x^2(1+e_3)\right)\right]\exp\left[-\frac{e_4}{2}\left(1+\frac{e_4}{2}\right)^{-1}\right] \tag{9}$$

Thus, we have the following theorems:

Theorem 1: Bias of the estimators T_{ij} (i,j=1,2) to the first order of approximations are obtained as

$$B\left(T_{ij}\right)=\varphi_{ij}B\left(T_{iu}\right)+\left(1-\varphi_{ij}\right)B\left(T_{jm}\right)\ (i,j=1,2) \tag{10}$$

where

$$B\left(T_{1u}\right)=S_y^2f_u\left(\frac{3}{8}\lambda_{004}-\frac{1}{2}\lambda_{022}+\frac{1}{8}\right) \tag{11}$$

$$B\left(T_{2u}\right)=S_y^2f_u\left(\frac{15}{8}\lambda_{004}-\frac{3}{2}\lambda_{022}-\frac{3}{8}\right) \tag{12}$$

$$B\left(T_{1m}\right)=S_y^2\left(f_n\left(\lambda_{220}-\lambda_{400}-\lambda_{202}+1\right)-f_m\left(\lambda_{220}+\lambda_{022}-\lambda_{202}-\lambda_{400}-\lambda_{004}+1\right)\right) \tag{13}$$

and

$$B\left(T_{2m}\right)=S_y^2\left(\frac{3}{8}\lambda_{004}-\frac{1}{2}\lambda_{022}+\frac{1}{8}\right)+\frac{\beta S_x^2}{2}\left\{f_m-f_n\right\}\left(\lambda_{202}-1\right) \tag{14}$$

Proof: The bias of the estimators T_{ij} $(j=1,2)$ is given by

$$B\left(T_{ij}\right)=E\left[T_{ij}-S_y^2\right]=\varphi_{ij}E\left[T_{iu}-S_y^2\right]+\left(1-\varphi_{ij}\right)E\left[T_{jm}-S_y^2\right]$$
$$=\varphi_{ij}B\left(T_{iu}\right)+\left(1-\varphi_{ij}\right)B\left(T_{jm}\right);(j=1,2) \tag{15}$$

where

$$B(T_{iu})=E\left(T_{iu}-S_y^2\right) \text{ and } B(T_{jm})=E\left(T_{jm}-S_y^2\right)$$

Substituting the values of T_{1u}, T_{2u}, T_{1m} and T_{2m} from equations (6)-(9) in the equation (15), expanding, taking expectations and retaining the terms up-to the first order of approximations, we have the expression of the bias of the estimators T_{ij} $(j=1,2)$ as given in equation (10).

Theorem 2: Mean square errors of the estimators T_{ij} (i, j=1, 2) to the first degree of approximation are obtained as

$$M\left(T_{ij}\right)=\varphi_{ij}^2M\left(T_{iu}\right)+(1-\varphi_{ij})^2M\left(T_{jm}\right);(j=1,2) \tag{16}$$

where

$$M(T_{1u})=S_y^4f_u\left(\lambda_{040}+\frac{1}{4}\lambda_{004}-\lambda_{022}-\frac{1}{4}\right) \tag{17}$$

$$M\left(T_{2u}\right)=S_y^4f_u\left(\lambda_{040}+\frac{9}{4}\lambda_{004}-3\lambda_{022}-\frac{1}{4}\right) \tag{18}$$

$$M\left(T_{1m}\right)=S_y^4\begin{pmatrix} f_m\left(\lambda_{400}+\lambda_{004}+\lambda_{040}+2\lambda_{202}-2\lambda_{220}+2\lambda_{022}-1\right)+ \\ f_n\left(2\lambda_{220}-2\lambda_{202}-\lambda_{400}+1\right) \end{pmatrix} \tag{19}$$

and

$$\begin{aligned}M\left(T_{2m}\right)=&S_y^4f_m\left(\lambda_{040}+\frac{1}{4}\lambda_{004}-\lambda_{022}-\frac{1}{4}\right)+\beta^2S_x^4\left(f_m-f_n\right)(\lambda_{400}-1)+\\ &2\beta S_y^2S_x^2\left(\left(f_n-f_m\right)\left(\lambda_{220}-1\right)+\frac{1}{2}\left(f_m-f_n\right)\left(\lambda_{202}-1\right)\right)\end{aligned} \tag{20}$$

Proof. The mean square errors of the estimators T_{ij} (i,j=1,2) is given by

$$M\left(T_{ij}\right)=E\left[T_{ij}-S_y^2\right]^2=E\left[\varphi_{ij}\left(T_{iu}-S_y^2\right)+\left(1-\varphi_{ij}\right)\left(T_{jm}-S_y^2\right)\right]^2$$
$$=\varphi_{ij}^2M\left(T_{iu}\right)+\left(1-\varphi_{ij}\right)^2M\left(T_{jm}\right)+2\varphi_{ij}\left(1-\varphi_{ij}\right)C\left(T_{iu},T_{jm}\right) \tag{21}$$

Where $M(T_{iu}) = E\left[T_{iu} - S_y^2\right]^2$, $M(T_{jm}) = E\left[T_{jm} - S_y^2\right]^2$ and $C(T_{iu}, T_{jm}) = \left[\left(T_{iu} - S_y^2\right)\left(T_{jm} - S_y^2\right)\right]$.

Since, the estimators T_{iu} and T_{jm} are based on two independent samples of sizes u and m respectively, therefore, the covariance type term has been ignored i.e.

$C(T_{iu}, T_{jm}) = 0$

Substituting the values of T_{1u}, T_{2u}, T_{1m} and T_{2m} from equations (6)-(9) in the equation (21) and taking expectations retaining the terms up-to the first order of approximations, we have the expression of the mean square errors of the estimators T_{ij} $(j = 1, 2)$ as given in equation (16).

4.1 Optimum Choice of β

The mean square error of the estimator deriving in equation (18) consist an unknown constant β, hence to get the optimum choice of β, it is minimized with respect to β and subsequently the optimum choice of β say β^* is obtained below.

$$\beta^* = \frac{S_y^2}{2S_x^2}\frac{2(\lambda_{220}-1)-(\lambda_{202}-1)}{(\lambda_{400}-1)} \tag{22}$$

Now substitute the optimum value of β from equation (22) in equation (20) we get the minimum MSE of T_{2m} with respect to β as

$$M(T_{2m})_{\min} = S_y^4\left(\begin{array}{l} f_m\left(\lambda_{040}+\frac{1}{4}\lambda_{004}-\lambda_{022}-\frac{1}{4}\right)+(f_m-f_n)\frac{2(\lambda_{220}-1)-(\lambda_{202}-1)}{(\lambda_{400}-1)} \\ \left(\frac{1}{4}\lambda_{202}-\frac{1}{2}\lambda_{220}+\frac{1}{4}\right) \end{array}\right) \tag{23}$$

Remark: The optimum value of β derived in equation (22) consist some unknown population parameters. To make the estimator practicable these unknown population parameters may be estimated with respective sample estimates. Sometimes these population parameters may be guessed from the past surveys.

5. Minimum mean square errors of the estimators T_{ij}

Since the mean square errors of the estimator T_{ij} (j=1,2) in equation (16) are the function of the unknown constants (scalars) φ_{ij}, therefore, MSE's are minimized with respect to φ_{ij} and subsequently the optimum values of φ_{ij} are obtained as

$$\varphi_{ij_{opt}} = \frac{M\left(T_{jm}\right)}{M(T_{iu}) + M\left(T_{jm}\right)} \quad (i, j = 1, 2) \tag{24}$$

where

$$\varphi_{11_{opt}} = \frac{M\left(T_{1m}\right)}{M\left(T_{1u}\right) + M\left(T_{1m}\right)}, \quad \varphi_{12_{opt}} = \frac{M\left(T_{2m}\right)_{min}}{M\left(T_{1u}\right) + M\left(T_{2m}\right)_{min}}, \quad \varphi_{21_{opt}} = \frac{M\left(T_{1m}\right)}{M\left(T_{2u}\right) + M\left(T_{1m}\right)}$$

and $\varphi_{22_{opt}} = \dfrac{M\left(T_{2m}\right)_{min}}{M\left(T_{2u}\right) + M\left(T_{2m}\right)_{min}}$.

From equations (24) substituting the values of $\varphi_{ij_{opt}}$ in equation (16) we get the optimum mean square errors of the estimators T_{ij} (j=1, 2) as

$$M\left(T_{ij}\right)_{opt} = \frac{M\left(T_{iu}\right) M\left(T_{jm}\right)}{M\left(T_{iu}\right) + M\left(T_{jm}\right)} \tag{25}$$

where

$$M\left(T_{11}\right)_{opt} = \frac{M\left(T_{1u}\right) M\left(T_{1m}\right)}{M\left(T_{1u}\right) + M\left(T_{1m}\right)}, \qquad M\left(T_{12}\right)_{opt} = \frac{M\left(T_{1u}\right) M\left(T_{2m}\right)_{min}}{M\left(T_{1u}\right) + M\left(T_{2m}\right)_{min}},$$

$$M\left(T_{21}\right)_{opt} = \frac{M\left(T_{2u}\right) M\left(T_{1m}\right)}{M\left(T_{2u}\right) + M\left(T_{1m}\right)}, \text{ and } M\left(T_{22}\right)_{opt} = \frac{M\left(T_{2u}\right) M\left(T_{2m}\right)_{min}}{M\left(T_{2u}\right) + M\left(T_{2m}\right)_{min}}$$

6. Efficiency Comparison

An empirical study is carried out to illustrate the performances of the proposed estimators $T_{ij}\ (i, j = 1, 2)$ owing to non-response, the percent relative efficiencies of the estimators T_{ij} with respect to estimator ξ under the complete response situations have been examined. The estimator ξ is considered as

$$\xi = \psi \xi_u + \left(1 - \psi\right) \xi_m \tag{26}$$

where, $\xi_u = s_{y_u}^2$, $\xi_m = s_{y_m}^2 \left(\dfrac{s_{x_n}^2}{s_{x_m}^2}\right)\left(\dfrac{S_z^2}{s_{z_m}^2}\right)$ and $\psi\left(0 \le \psi \le 1\right)$ is an unknown constant to be determined by the minimization of mean square error of estimator ξ.

The minimum mean square error of the estimator ξ up to the first order of approximations is obtained as

$$M(\xi)_{min} = \frac{M(\xi_u).V(\xi_m)}{M(\xi_u)+V(\xi_m)} \tag{27}$$

where

$$M(\xi_u) = S_y^4 f_u^* (\lambda_{040} - 1)$$

and

$$V(\xi_m) = S_y^4 \left(f_m (\lambda_{400} + \lambda_{004} + \lambda_{040} + 2\lambda_{202} - 2\lambda_{220} + 2\lambda_{022} - 1) + f_n^* (2\lambda_{220} - 2\lambda_{202} - \lambda_{400} + 1)\right)$$

where $f_u^* = \left(\frac{1}{u} - \frac{1}{N}\right), f_n^* = \left(\frac{1}{n} - \frac{1}{N}\right)$

Thus, the expression of percent relative efficiencies E of the estimator T_{ij} with respect to estimator ξ under their optimality conditions are given by

$$E = \left[\frac{M(\xi)_{min}}{M(T_{ij})_{opt}}\right] \times 100. \ (i, j = 1, 2) \ .$$

7. Numerical Illustrations

To illustrate the performances of the proposed estimators we have considered two natural populations data sets. The source of the populations, the nature of the variables y, x, z and the values of the various parameters are given as :

Population I-Source: Murthy (1967, p-399)

y: Area under wheat in 1964.

x: Area under wheat in 1963.

z: Cultivated area in 1961.

$N = 34, \delta_{040} = 3.7256, \delta_{400} = 2.9122, \delta_{004} = 2.8082, \delta_{220} = 3.1045, \delta_{022} = 2.9789, \delta_{202} = 2.7389$

Population II-Source: Sukhatme & Sukhatme (1970, p-185)

y: Area under wheat in 1937.

x: Area under wheat in 1936.

z: Cultivated area in 1931.

$N = 34, \delta_{040} = 3.5469, \delta_{400} = 3.3815, \delta_{004} = 2.7425, \delta_{220} = 2.5068, \delta_{022} = 2.6868, \delta_{202} = 2.0652$

An empirical studies are carried out through two different natural populations data sets and the dominance of the proposed estimators are shown for different choices of sample size n on first occasion matched sample of size m, fresh sample size u on second occasion, non response probability (p_1) and (p_2) on first and second occasions respectively. Table 1-4 present the percent relative efficiencies E of the estimator T_{ij} $(i, j = 1, 2)$ with respect to the estimator ξ .

Table 1. Percent relative efficiencies of estimator T_{11} with respect to the estimator ξ

Population				I				II			
Non-response probability				p_1 ↓				p_1 ↓			
n	m	u	p_2 ↓	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
28	14	14	0.05	139.77	141.61	143.79	146.43	142.85	145.23	148.09	151.56
			0.10	133.71	135.55	137.73	140.37	137.28	139.67	142.52	145.99
			0.15	127.98	129.82	132.00	134.64	132.02	134.41	137.26	140.73
			0.20	122.56	124.39	126.58	129.21	127.04	129.43	132.28	135.75
23	11	12	0.05	141.81	143.32	145.09	147.20	145.06	147.04	149.39	152.20
			0.10	136.11	137.62	139.39	141.50	139.81	141.79	144.13	146.95
			0.15	130.66	132.17	133.94	136.05	134.78	136.76	139.11	141.92
			0.20	125.43	126.94	128.71	130.82	129.96	131.94	134.29	137.10
18	8	10	0.05	145.16	146.30	147.63	149.18	149.04	150.57	152.36	154.46
			0.10	139.69	140.83	142.16	143.71	143.94	145.47	147.25	149.35
			0.15	134.39	135.54	136.86	138.41	139.00	140.53	142.31	144.41
			0.20	129.26	130.41	131.73	133.28	134.21	135.75	137.53	139.63

Table 2. Percent relative efficiencies of estimator T_{12} with respect to the estimator ξ

Population				I				II			
Non-response probability				p_1 ↓				p_1 ↓			
n	m	u	p_2 ↓	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
28	14	14	0.05	205.37	203.15	200.79	198.26	258.15	247.01	236.17	225.62
			0.10	199.31	197.09	194.72	192.20	252.58	241.44	230.60	220.05
			0.15	193.58	191.36	188.99	186.47	247.32	236.18	225.34	214.79
			0.20	188.15	185.94	183.57	181.05	242.34	231.20	220.36	209.81
23	11	12	0.05	195.89	194.33	192.65	190.84	231.85	224.89	217.90	210.90
			0.10	190.19	188.63	186.95	185.15	226.60	219.63	212.65	205.65
			0.15	184.74	183.18	181.50	179.69	221.57	214.61	207.62	200.62
			0.20	179.51	177.95	176.27	174.47	216.75	209.79	202.80	195.80
18	8	10	0.05	191.16	190.06	188.88	187.61	220.00	215.34	210.58	205.71
			0.10	185.69	184.59	183.41	182.14	214.90	210.24	205.48	200.61
			0.15	180.39	179.30	178.12	176.84	209.96	205.30	200.54	195.66
			0.20	175.26	174.16	172.98	171.71	205.17	200.51	195.75	190.88

Table 3. Percent relative efficiencies of estimator T_{21} with respect to the estimator ξ

Population				I				II			
Non-response probability				p_1 ↓				p_1 ↓			
n	m	u	p_2 ↓	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
28	14	14	0.05	133.09	134.92	137.11	139.75	173.72	176.11	178.96	182.43
			0.10	127.50	129.34	131.52	134.16	165.94	168.32	171.17	174.65
			0.15	122.23	124.06	126.25	128.89	158.58	160.97	163.82	167.29
			0.20	117.23	119.07	121.26	123.89	151.62	154.01	156.86	160.33
23	11	12	0.05	134.79	136.30	138.07	140.18	177.60	179.58	181.93	184.74
			0.10	129.54	131.05	132.83	134.94	170.25	172.24	174.58	177.39
			0.15	124.52	126.03	127.80	129.91	163.22	165.21	167.55	170.36
			0.20	119.71	121.22	122.99	125.10	156.48	158.47	160.81	163.63
18	8	10	0.05	137.58	138.72	140.05	141.60	184.61	186.14	187.92	190.02
			0.10	132.54	133.68	135.01	136.56	177.48	179.01	180.79	182.89
			0.15	127.66	128.81	130.13	131.68	170.57	172.10	173.88	175.98
			0.20	122.94	124.08	125.41	126.96	163.88	165.41	167.19	169.29

Table 4. Percent relative efficiencies of estimator T_{22} with respect to the estimator ξ

Population				I				II			
Non-response probability				p_1 ↓				p_1 ↓			
n	m	u	p_2 ↓	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
28	14	14	0.05	198.68	196.47	194.10	191.58	289.02	277.88	267.04	256.49
			0.10	193.10	190.88	188.52	186.00	281.23	270.09	259.26	248.70
			0.15	187.82	185.61	183.24	180.72	273.88	262.74	251.90	241.35
			0.20	182.83	180.61	178.25	175.73	266.92	255.78	244.94	234.39
23	11	12	0.05	188.87	187.31	185.63	183.83	264.39	257.43	250.44	243.44
			0.10	183.63	182.06	180.39	178.58	257.05	250.08	243.10	236.09
			0.15	178.61	177.04	175.36	173.56	250.01	243.05	236.07	229.06
			0.20	173.79	172.23	170.55	168.75	243.28	236.31	229.33	222.32
18	8	10	0.05	183.58	182.48	181.30	180.03	255.57	250.91	246.15	241.28
			0.10	178.54	177.44	176.26	174.99	248.43	243.78	239.02	234.14
			0.15	173.66	172.57	171.39	170.11	241.53	236.87	232.11	227.23
			0.20	168.94	167.84	166.66	165.39	234.83	230.18	225.42	220.54

8. Interpretations of empirical results

The following interpretation may be made from Table 1-4:

(1) From Table-1, it is clear that

(a) For fixed values of p_1 and p_2, percent relative efficiency E is increasing with the decreasing values of n, m and u which lead to reduction of the cost of the survey. This behaviour is useful in terms of increased precision of estimates with reduced cost of the survey.

(b) For fixed values of p_1, n, m and u, percent relative efficiency E is decreasing with the increasing values of p_2. This phenomenon is highly desirable to cope with the negative impact of non-response situations even if nonresponding elements are increasing on current occasion.

(c) For fixed values of p_2, n, m and u, percent relative efficiency E is increasing with the increasing values of p_1.

(2) From Table-2, it is observed that

(a) For fixed values of p_1 and p_2, percent relative efficiency E is decreasing with the decreasing values of n, m and u.

(b) For fixed values of p_1, n, m and u, percent relative efficiency E is decreasing with the increasing values of p_2. This phenomenon is highly desirable to reduce the negative impact of non-response situations even if nonresponding elements are increasing on current occasion.

(c) For fixed values of p_2, n, m and u, percent relative efficiency E is decreasing with increasing values of p_1. This behaviour is also cope with the negative impact of non-response even when nonresponding elements are increasing on first occasion.

(3) The trends of Tables 3-4 is similar to the trends seen in Table 1-2 respectively, hence their interpretations are also same.

9. Conclusions

From the above interpretations it is clear that the proposed estimators contribute significantly to cope with the different realistic situations of random non-response while estimating the population variance on current (second) occasion in two occasion successive sampling. It is also seen that the proposed estimator is more efficient than the estimator ξ under the complete response situations. Therefore, proposed estimators may be recommended to the survey statisticians for their practical applications.

References

Ahmed, M.S., Al-Titib, O., Al-Rawib, Z., Abu- Dayyeh, W. (2005): Estimation of finite population variance in presence of random non response using auxiliary variables. Information and Management science, 16(2), 73-82.

Azam, M., Zaman, Q., Salahuddin and Shabbir, J. (2010). Estimation of current population variance in two successive occasions. J. Statist. 17:54-65.

Biradar, R. S. and Singh, H. P. (2001): Successive sampling using auxiliary information on both occasions. Cal. Statist. Assoc. Bull. 51, 243-251.

Das, A. K. (1982): Estimation of population ratio on two occasions, Jour. Ind. Soc. Agr. Statist. 34, 1-9.

Eckler, A. R. (1955): Rotation sampling. Annals of Mathematical Statistics, 26, 664-685.

Feng, S. and Zou, G. (1997): Sample rotation method with auxiliary variable. Communications in Statistics-Theory and Methods, 26, 6, 1497-1509.

Gupta, P. C. (1979): Sampling on two successive occasions. Jour. Statist. Res. 13, 7-16.

Jessen, R.J. (1942): Statistical Investigation of a Sample Survey for obtaining farm facts, Iowa Agricultural Experiment Station Research Bulletin No. 304, Ames, Iowa, U. S. A., 1-104.

Murthy, M. N. (1967): Sampling theory and Methods. Statistical Publishing society. Calcutta, India.

Patterson, H. D. (1950): Sampling on successive occasions with partial replacement of units, Journal of the Royal Statistical Society, 12, 241-255.

Rao, J. N. K. and Graham, J. E. (1964): Rotation design for sampling on repeated occasions. Jour. Amer. Statist. Assoc. 59, 492-509.

Rubin, D.B. (1976): Inference and missing data. Biometrica, 63(3), 581-592.

Sen, A. R. (1971): Successive sampling with two auxiliary variables, Sankhya, 33, Series B, 371-378.

Sen, A. R. (1972): Successive sampling with p ($p \geq 1$) auxiliary variables, Ann. Math. Statist., 43, 2031-2034.

Sen, A. R. (1973): Theory and application of sampling on repeated occasions with several auxiliary variables, Biometrics 29, 381-385.

Sharma, A. K. (2017): An Effective Imputation Method to Minimize the Effect of Random Non-Response in Estimation of Population Variance on Successive Occasions, Kuwait Journal of Science, 44(4) 91-97.

Singh, G. N. and Karna, J. P. (2009): Estimation of population mean on current occasions in two-occasion successive sampling, Metron, 67(1), 69-85.

Singh, G. N. and Sharma, A. K. (2014): An Alternative Rotation Patterns in Two Occasion Successive Sampling. International Journal of Mathematics and Statistics, 15(3), 9-22.

Singh, G. N. and Singh, V. K. (2001): On the use of auxiliary information in successive sampling. Jour. Ind. Soc. Agri. Statist., 54(1), 1-12.

Singh, G. N., Majhi, D., Maurya, S & Sharma, A. K. (2015). Some Effective Rotation Patterns in Estimation of Population Mean in Two-Occasion Successive Sampling, Communications in Statistics-Theory & Methods, 44(12) 2571-2585.

Singh, G. N., Pandey, A. K. and Sharma, A. K., (2020): Some improved and alternative imputation methods for finite population mean in presence of missing information, Communications in Statistics-Theory & Methods. 50(19) 4401-4427.

Singh, G. N., Sharma, A. K., Bandyapadyay, A. and Paul, C. (2018). Effective Estimation of Population Mean, Ratio and Product in Two-Phase Sampling in Presence of Random Non-Response, Hacettepe Journal of Mathematics and Statistics, 47(1) 279-295.

Singh, G.N. and Priyanka, K. (2010): Estimation of population mean at current occasion in presence of several varying auxiliary variates in two – occasion successive sampling. Statistics in Transition- new series, 11(1), 105 – 126.

Singh, H. P., Tailor, R., Singh, S., Kim, J. M., (2011): Estimation of population variance in successive sampling, Quality and Quantity, 45, 477-494.

Singh, H.P. Chandra, P. and Singh, S. (2003): Variance estimation using multiauxiliary information for random non-response in survey sampling. Statistica. LXIII, 23-40.

Singh, S. and Joarder, A.H. (1998): Estimation of finite population variance using random non-response in survey sampling, Metrika, 47, 241-249.

Singh, V. K., Singh, G. N. and Shukla, D. (1991): An efficient family of ratio-cum-difference type estimators in successive sampling over two occasions, J. Sci. Res. 41 C,149-159.

Sud, U. C., Srivastava, A. K. and Sharma, D. P. (2001). On the estimation of population variance in repeat surveys. Journal of Indian Society of Agricultural Statistics, 54(2), 354-369.

Sukhatame, P. V. and Sukhatame, B. V. (1970): Sampling Theory of Survey with Applications. 2nd Edition, New Delhi, India.

Tikkiwal, B. D. (1951). Theory of Successive Sampling, Unpublished Diploma Dissertation, Institute of Agricultural Research Statistics, New Delhi, India

Tracy, D.S, and Osahan, S.S. (1994): Random non-response on study variable versus on study as well as auxiliary variables. Statistica, 54, 163-168.