Thailand Statistician
April 2023; 21(2): 305-313
http://statassoc.or.th
Contributed paper

On Efficient Estimation in Successive Sampling Over Two Occasions

Kuldeep Kumar Tiwari [a], Sandeep Bhougal*[a] and Sunil Kumar [b]
[a] School of Mathematics, Shri Mata Vaishno Devi University, Katra, Jammu and Kashmir, India
[b] Department of Statistics, University of Jammu, Jammu, Jammu and Kashmir, India
*Corresponding author; e-mail: sandeep.bhougal@smvdu.ac.in

Received: 21 July 2020
Revised: 12 June 2021
Accepted: 21 August 2021

Abstract

Surveys often get repeated on many occasions over years or seasons to study the change in the characteristics over the period. Using the data from the previous occasion also improves the estimation on the current occasion. We have worked on the problem of estimating the population mean in successive sampling over two occasions. To increase the efficiency in the estimation of the population mean on the current (second) occasion in two occasion successive sampling, we have proposed an estimator t by using the convex linear combination of the estimators t_u based on u units, which is drawn afresh at the current occasion and t_m based on m units, which are retained from the previous occasion. The expressions of bias and mean square error for the proposed estimator are calculated and optimal replacement policy is also discussed for the said case. To show the validity of the work, we made an empirical study followed by an application to a case study.

Keywords: Auxiliary variable, study variable, mean squared error, efficiency

1. Introduction

Surveying the same population at different intervals of time is called successive or rotation sampling. Successive sampling is useful when we wish to measure the characteristics of a parameter concerning time change. The sampling over successive occasion is appropriate when the aim is to know the change in the parameters of the population over different occasions, the average over all occasions, and the most important is the average for the most recent occasion. For example, in the survey of the production of wheat, one may be interested to estimate the average wheat production in the current season or the change in wheat production for two different seasons and the total production of wheat in the season. The theory of successive sampling was first introduced by Jessen (1942) and further improved by Patterson (1950), Tikkiwal (1951), Eckler (1955), and Kullduff (1963). Sen (1971) used two auxiliary variables to develop an estimator for population mean in successive sampling. Further, Sen (1972, 1973) generalized his work for p-auxiliary variables. Okafor (1987) used an auxiliary variable in two-stage successive sampling to compare some population total estimators. When the auxiliary information is available on both occasions, the authors Feng and Zou (1997), Singh (2005), Singh and Vishwakarma (2009), Singh and Karna (2009), Singh and Kumar (2010), Singh and Pal (2017), Sharma and Kumar (2018), Khalid and Singh (2020), and Jabeen et al. (2021), etc suggested estimators for estimating the population mean on the current occasion in rotation sampling.

In this paper, we have suggested an estimator to estimate the population mean of the study variable on current occasion in two occasion successive sampling. Section 2 discusses the sampling procedure and some existing estimators used in the paper. In Section 3, we have proposed an estimator followed by Section 4 in which we studied the properties of the estimators. In Section 5, we have studied the optimum replacement policy. Section 6 discusses the optimum conditions in which the proposed estimator is efficient than other estimators. Further, Sections 7 and 8 present the numerical illustration on some artificial data and real-life data, respectively. Section 9 comprises of conclusion followed by the references.

2. Sampling Procedure and Some Existing Estimators

Let us assume a finite population $U = (U_1, U_2, ..., U_N)$ of N units which remains unchanged over occasions. Let x and y be the study variable on the first and second occasion respectively, z is the auxiliary variable which is known on both occasions. Also, z is correlated with x and y on first and second occasions respectively. A sample of size n is drawn on the first occasion using simple random sampling without replacement. A random sub-sample of size $m = n\lambda$ unit retained (matched) for use on the second occasion.

A fresh sample of size $u = (n - m) = n\theta$ unit is taken using simple random sampling without replacement from the remaining $(N - n)$ units of population. The sample on the second occasion is chosen in such a way that the sample size on the second occasion is also n. Here λ and θ are the proportion of the matched and unmatched units in the sample such that $0 \leq \lambda \leq 1$, $0 \leq \theta \leq 1$ and $\lambda + \theta = 1$.

Some notations used throughout the article are given here. $\bar{X}$, $\bar{Y}$, $\bar{Z}$ are the population means of variables x, y and z respectively. $\bar{y}_u$, $\bar{y}_m$, $\bar{x}_m$, $\bar{x}_n$, $\bar{z}_u$, $\bar{z}_m$, $\bar{z}_n$ are the sample means of respective variables of size written in suffices. $\rho_{yx}, \rho_{yz}, \rho_{zx}$ are the correlation coefficients between the variables given in subscripts. C_x, C_y, C_z are the coefficients of variation for the variable x, y and z respectively.
$C_x = \frac{S_x}{\bar{X}}$, $C_z = \frac{S_z}{\bar{Z}}$, $C_y = \frac{S_y}{\bar{Y}}$, $C_{yx} = \rho_{yx} C_y C_x$, $C_{yz} = \rho_{yz} C_y C_z$, $C_{zx} = \rho_{zx} C_z C_x$,
$K_{yz} = \rho_{yz} \frac{C_y}{C_z}$, $K_{xz} = \rho_{zx} \frac{C_x}{C_z}$, $K_{zx} = \rho_{zx} \frac{C_z}{C_x}$,
$S_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2$, $S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{X})^2$, $S_z^2 = \frac{1}{N-1} \sum_{i=1}^{N} (z_i - \bar{Z})^2$,
$\lambda_u = \frac{1}{u} - \frac{1}{N}$, $\lambda_n = \frac{1}{n} - \frac{1}{N}$, $\lambda_m = \frac{1}{m} - \frac{1}{N}$, $\delta_0 = 1 - \rho_{yz}^2$.

When there is no auxiliary variable availabe, the usual unbiased estimator $\hat{\bar{y}}_n = \bar{y}_n$ used to estimate the population mean. The variance of $\hat{\bar{y}}_n$ is

$$Var(\hat{\bar{y}}_n) = \frac{S_y^2}{n}. \tag{1}$$

To estimate the population mean on current occasion in successive sampling, Shabbir, Azam and Gupta (2005) proposes the estimator

$$t_{SAG} = \phi t_{SAG_m} + (1 - \phi) t_{SAG_u}$$

where ϕ is unknown constant. t_{SAG_m} & t_{SAG_u} are extimators of $\bar{Y}$ based on matched and unmatched portion of the sample. t_{SAG_m} is defined as

$$t_{SAG_m} = w_1 \left\{ \bar{y}_m + b_1(\bar{x}_n - \bar{x}_m) \frac{\bar{x}}{\bar{X}} \right\} + w_2 \left\{ \bar{y}_m + b_2(\bar{z}_n - \bar{z}_m) \frac{\bar{z}}{\bar{Z}} \right\}$$

where w_1 and w_2 are constants such that $w_1 + w_2 = 1$, b_1 and b_2 are sample regression coefficient of y on x and y on z respectively for matched portion.

t_{SAG_u} is defined as

$$t_{SAG_u} = \bar{y}_u + b(\bar{Z} - \bar{z}_u)$$

where b is sample regression coefficient of y on z for unmatched portion. The minimum mean square error (MSE) of t_{SAG} is

$$MSE(t_{SAG}) = \frac{S_y^2}{n} \frac{\delta_0[1-\rho_{zx}^2 - \theta(\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{zx})]}{[(\delta_0 + \theta\rho_{yz}^2)(1-\rho_{zx}^2) - \theta^2(\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{zx})]}. \tag{2}$$

Shabbir, Azam and Gupta (2005) shows that t_{SAG} performs better than the estimator proposed by Tracy and Singh (1999).

Singh and Pal (2016) proposes an estimator to estimate the population mean $\bar{Y}$ on the second occasion as

$$t_{SP} = \phi t_{SP_m} + (1-\phi)t_{SP_u}$$

where ϕ is constant and t_{SP_m} and t_{SP_u} are defined as

$$t_{SP_m} = \bar{y}_m \exp\left\{\frac{\alpha_2(\bar{Z}-\bar{z}_m)}{(\bar{Z}+\bar{z}_m)}\right\} + b_{yx}\left[\bar{x}_n \exp\left\{\frac{\alpha_3(\bar{Z}-\bar{z}_n)}{(\bar{Z}+\bar{z}_n)}\right\} - \bar{x}_m \exp\left\{\frac{\alpha_3(\bar{Z}-\bar{z}_m)}{(\bar{Z}+\bar{z}_m)}\right\}\right]$$

where α_2 and α_3 are scalars and b_{yx} is sample regression coefficient based on matched portion.

$$t_{SP_u} = \bar{y}_u \exp\left\{\frac{\alpha_1(\bar{Z}-\bar{z}_u)}{(\bar{Z}+\bar{z}_u)}\right\}$$

where α_1 is constant.

The minimum MSE of t_{SP} is

$$MSE(t_{SP}) = \frac{S_y^2}{n} \frac{\delta_0[\delta_0 - \theta(\rho_{yx}^2 + \rho_{yx}^2\rho_{zx}^2 - 2\rho_{yx}\rho_{yz}\rho_{zx})]}{[\delta_0 - \theta^2(\rho_{yx}^2 + \rho_{yx}^2\rho_{zx}^2 - 2\rho_{yx}\rho_{yz}\rho_{zx})]}. \tag{3}$$

Singh and Pal (2016) concluded that the estimator t_{SP} works better than the estimator of Singh and Homa (2013).

3. Proposed Estimator

We propose the following estimator to estimate the population mean of the study variable y on current (second) occasion in two occasion successive sampling by using the information on auxiliary variable z,

$$t = (1-\phi)t_u + \phi t_m$$

where ϕ is constant to be determined for minimum mean square error.

For unmatched portion, the estimator t_u is defined as

$$t_u = \bar{y}_u \left(\frac{\bar{Z}}{\bar{z}_u}\right) \exp \alpha \left(\frac{\bar{Z}-\bar{z}_u}{\bar{Z}+\bar{z}_u}\right)$$

where α is real scalar.

The estimator t_m is defined for matched portion as

$$t_m = \left[\bar{y}_m + \alpha_1(\bar{x}_n - \bar{x}_m)\right] \exp \alpha_2 \left(\frac{\bar{Z}-\bar{z}_m}{\bar{Z}+\bar{z}_n}\right)$$

where α_1 and α_2 are suitably chosen constants.

4. Properties of the Proposed Estimator

To obtain the bias and mean square error (MSE) of the proposed estimator, the error terms are defined as

$$\begin{aligned}\bar{y}_u &= \bar{Y}(1+\epsilon_{yu}), \bar{y}_m = \bar{Y}(1+\epsilon_{ym}), \bar{x}_n = \bar{X}(1+\epsilon_{xn}), \bar{x}_m = \bar{X}(1+\epsilon_{xm}),\\ \bar{z}_u &= \bar{Z}(1+\epsilon_{zu}), \bar{z}_m = \bar{Z}(1+\epsilon_{zm}), \bar{z}_n = \bar{Z}(1+\epsilon_{zn})\end{aligned}$$

and the expected values are

$$E(\epsilon_{yu}) = E(\epsilon_{ym}) = E(\epsilon_{xn}) = E(\epsilon_{xm}) = E(\epsilon_{zu}) = E(\epsilon_{zm}) = E(\epsilon_{zn}) = 0$$

and

$$\begin{aligned}&E(\epsilon_{yu}^2) = \lambda_u C_y^2, E(\epsilon_{ym}^2) = \lambda_m C_y^2,\\ &E(\epsilon_{xn}^2) = \lambda_n C_x^2, E(\epsilon_{xm}^2) = \lambda_m C_x^2,\\ &E(\epsilon_{zu}^2) = \lambda_u C_z^2, E(\epsilon_{zm}^2) = \lambda_m C_z^2,\\ &E(\epsilon_{zn}^2) = \lambda_n C_z^2, E(\epsilon_{zn}\epsilon_{zm}) = \lambda_n C_z^2,\\ &E(\epsilon_{yu}\epsilon_{zu}) = \lambda_u C_{yz}, E(\epsilon_{ym}\epsilon_{xn}) = \lambda_n C_{yx}, E(\epsilon_{ym}\epsilon_{xm}) = \lambda_m C_{yx},\\ &E(\epsilon_{ym}\epsilon_{zm}) = \lambda_m C_{yz}, E(\epsilon_{xn}\epsilon_{xm}) = \lambda_n C_x^2, E(\epsilon_{xn}\epsilon_{zm}) = \lambda_n C_{xz}, E(\epsilon_{xm}\epsilon_{zm}) = \lambda_m C_{xz}.\end{aligned}$$

Now, to get bias and MSE, express t_u in terms of errors ϵ_{ij}, $i = x, y, z$; $j = u, m, n$. We have

$$t_u = \bar{Y}(1+\epsilon_{yu})\left[\frac{\bar{Z}}{\bar{Z}(1+\epsilon_{zu})}\right]\exp\alpha\left(\frac{\bar{Z}-\bar{Z}(1+\epsilon_{zu})}{\bar{Z}+\bar{Z}(1+\epsilon_{zu})}\right)$$

or

$$t_u = \bar{Y}(1+\epsilon_{yu})(1+\epsilon_{zu})^{-1}\exp\left[-\frac{\alpha\epsilon_{zu}}{2}\left(1+\frac{\epsilon_{zu}}{2}\right)^{-1}\right].$$

Now, assuming $|\epsilon_{ij}| < 1$, $i = x, y, z$; $j = u, m, n$, expand the expression using Taylor's series approximation and terminate the terms having ϵ's degree greater than two, after simplification, we get

$$t_u - \bar{Y} \cong \bar{Y}\left[\epsilon_{yu} - \left(1+\frac{\alpha}{2}\right)\epsilon_{zu} + \left(1+\frac{3\alpha}{4}+\frac{\alpha^2}{8}\right)\epsilon_{zu}^2 - \left(1+\frac{\alpha}{2}\right)\epsilon_{yu}\epsilon_{zu}\right] \tag{4}$$

Taking expectation on both sides of Eqn. (4), we get the bias of t_u to the first degree of approximation as

$$B(t_u) \cong \bar{Y}\left[\left(1+\frac{3\alpha}{4}+\frac{\alpha^2}{8}\right)\lambda_u C_z^2 - \left(1+\frac{\alpha}{2}\right)\lambda_u C_{yz}\right]$$

Squaring Eqn. (4) on both sides and terminating the terms having ϵ's degree greater than two, we have

$$(t_u - \bar{Y})^2 \cong \bar{Y}^2\left[\epsilon_{yu}^2 + \left(1+\frac{\alpha}{2}\right)^2\epsilon_{zu}^2 - (2+\alpha)\epsilon_{yu}\epsilon_{zu}\right]. \tag{5}$$

Taking expectation on both sides of Eqn. (5), we get

$$MSE(t_u) \cong \bar{Y}^2\left[\lambda_u C_y^2 + \left(1+\frac{\alpha}{2}\right)^2\lambda_u C_z^2 - (2+\alpha)\lambda_u C_{yz}\right]. \tag{6}$$

To minimize $MSE(t_u)$, differentiate Eqn. (6) with respect to α and equate to zero,

$$\frac{d}{d\alpha}\big(MSE(t_u)\big) = 0$$

which gives optimum value of α

$$\alpha = 2K_{yz} - 2 = \alpha^o(say).$$

Put the value of α^o in Eqn. (6), we get the optimum MSE of t_u as

$$\begin{aligned} MSE_{min}(t_u) &\cong \bar{Y}^2\big(\lambda_u C_y^2 + \lambda_u K_{yz}^2 C_z^2 - 2\lambda_u K_{yz} C_{yz}\big) \\ \text{or} \quad MSE_{min}(t_u) &\cong \lambda_u \bar{Y}^2 C_y^2 (1 - \rho_{yz}^2). \end{aligned}$$

This equation is same as the MSE of the linear regression estimator $t_{lru} = \bar{y}_u + b_{yz}(\bar{Z} - \bar{z})$. Now, to obtain the bias and MSE of t_m, express t_m in terms of ϵ's, we have

$$t_m = \big[\bar{Y}(1+\epsilon_{ym}) + \alpha_1(\bar{X} + \bar{X}\epsilon_{xn} - \bar{X} - \bar{X}\epsilon_{xm})\big] \exp\ \alpha_2\left(\frac{\bar{Z} - \bar{Z}(1+\epsilon_{zm})}{\bar{Z} + \bar{Z}(1+\epsilon_{zn})}\right).$$

Expand using Taylor's series approximation and ignoring terms having ϵ's degree greater than two. After simplification, we have

$$\begin{aligned} t_m - \bar{Y} \cong \bar{Y}\Bigg(&\epsilon_{ym} + \alpha_1 p\epsilon_{xn} - \alpha_1 p\epsilon_{xm} - \frac{\alpha_2}{2}\epsilon_{zm} + \frac{\alpha_2^2}{8}\epsilon_{zm}^2 + \frac{\alpha_2}{4}\epsilon_{zn}\epsilon_{zm} \\ &-\frac{\alpha_2}{2}\epsilon_{ym}\epsilon_{zm} - \frac{\alpha_1\alpha_2 p}{2}\epsilon_{xn}\epsilon_{zm} + \frac{\alpha_1\alpha_2 p}{2}\epsilon_{xm}\epsilon_{zm}\Bigg) \end{aligned} \tag{7}$$

where $p = \frac{\bar{X}}{\bar{Y}}$

Taking expectation on both sides of Eqn. (7), one can obtain the bias of t_m to the first degree of approximation as

$$B(t_m) \cong \bar{Y}\left(\frac{\alpha_2^2}{8}\lambda_m C_z^2 + \frac{\alpha_2}{4}\lambda_n C_z^2 - \frac{\alpha_2}{2}\lambda_m C_{yz} - \frac{\alpha_1\alpha_2 p}{2}\lambda_n C_{xz} + \frac{\alpha_1\alpha_2 p}{2}\lambda_m C_{xz}\right).$$

Squaring both sides of Eqn. (7) and simplify, we get

$$\begin{aligned} (t_m - \bar{Y})^2 \cong \bar{Y}^2\Bigg[&\epsilon_{ym}^2 + \alpha_1^2 p^2\epsilon_{xn}^2 + \alpha_1^2 p^2 \epsilon_{xm}^2 + \frac{\alpha_2^2}{4}\epsilon_{zm}^2 + 2\alpha_1 p\epsilon_{ym}\epsilon_{xn} - 2\alpha_1 p\epsilon_{ym}\epsilon_{xm} \\ &-\alpha_2\epsilon_{ym}\epsilon_{zm} - 2\alpha_1^2 p^2\epsilon_{xn}\epsilon_{xm} - \alpha_1\alpha_2 p\epsilon_{xn}\epsilon_{zm} + \alpha_1\alpha_2 p\epsilon_{xm}\epsilon_{zm}\Bigg] \end{aligned} \tag{8}$$

Taking expectation on both sides of Eqn. (8), we get

$$\begin{aligned} MSE(t_m) \cong \bar{Y}^2\Bigg[&\lambda_m C_y^2 + (\lambda_m - \lambda_n)\alpha_1^2 p^2 C_x^2 + \frac{\alpha_2^2}{4}\lambda_m C_z^2 - (\lambda_m - \lambda_n)2\alpha_1 p C_{yx} \\ &-\alpha_2\lambda_m C_{yz} + (\lambda_m - \lambda_n)\alpha_1\alpha_2 p C_{xz}\Bigg]. \end{aligned} \tag{9}$$

To get the minimum MSE of t_m, differentiate Eqn. (9) partially with respect to α_1 & α_2 and equating to zero, we get

$$2pC_x\alpha_1 + \rho_{zx}C_z\alpha_2 = 2\rho_{yx}C_y \tag{10}$$

$$(\lambda_m - \lambda_n)2pC_{xz}\alpha_1 + \lambda_m C_z^2\alpha_2 = 2\lambda_m C_{yz}. \tag{11}$$

On solving Eqns. (10) and (11) simultaneously, we get the optimum values of α_1 and α_2 as

$$\begin{aligned} \alpha_1 &= \frac{\lambda_m C_y(\rho_{yx} - \rho_{zx}\rho_{yz})}{pC_x[\lambda_m - (\lambda_m - \lambda_n)\rho_{zx}^2]} = \alpha_1^o(say) \\ \alpha_2 &= \frac{2C_y[\lambda_m\rho_{yz} - (\lambda_m - \lambda_n)\rho_{yx}\rho_{zx}]}{C_z[\lambda_m - (\lambda_m - \lambda_n)\rho_{zx}^2]} = \alpha_2^o(say) \end{aligned}$$

Substituting α_1^o & α_2^o in Eqn. (9), we get

$$MSE_{min}(t_m) \;\cong\; \lambda_m \bar{Y}^2 C_y^2 \left[1-\rho_{yz}^2-\frac{(\lambda_m-\lambda_n)(\rho_{yx}-\rho_{zx}\rho_{yz})^2}{\lambda_m-(\lambda_m-\lambda_n)\rho_{zx}^2}\right]$$

From Eqn. (3), we have

$$MSE(t)=(1-\phi)^2 MSE(t_u)+\phi^2 MSE(t_m) \tag{12}$$

Since estimators t_u and t_m are independent, the covariance term is not contributing in Eqn. (12). Substitute the values of $MSE(t_u)$ and $MSE(t_m)$ from Eqns. (6) and (9) in Eqn. (12), we get

$$\begin{aligned} MSE(t) \cong \bar{Y}^2 \Bigg[(1-\phi)^2 \Bigg\{\lambda_u C_y^2+\left(1+\frac{\alpha}{2}\right)^2 \lambda_u C_z^2-(2+\alpha)\lambda_u C_{yz}\Bigg\}+\phi^2\Bigg\{\lambda_m C_y^2 \\ +(\lambda_m-\lambda_n)\alpha_1^2 p^2 C_x^2+\frac{\alpha_2^2}{4}\lambda_m C_z^2-(\lambda_m-\lambda_n)2\alpha_1 p C_{yx} \\ -\alpha_2\lambda_m C_{yz}+(\lambda_m-\lambda_n)\alpha_1\alpha_2 p C_{xz}\Bigg\}\Bigg]. \end{aligned} \tag{13}$$

Differentiating Eqn. (13) with respect to α, α_1, α_2 and ϕ, and equating to zero, we get

$$\left(1+\frac{\alpha}{2}\right)\lambda_u C_z^2-\lambda_u C_{yz} \;=\; 0 \tag{14}$$

$$2pC_x\alpha_1+\rho_{zx}C_z\alpha_2 \;=\; 2\rho_{yx}C_y \tag{15}$$

$$(\lambda_m-\lambda_n)2pC_{xz}\alpha_1+\lambda_m C_z^2\alpha_2 \;=\; 2\lambda_m C_{yz} \tag{16}$$

$$\begin{aligned} (1-\phi)\big\{\lambda_u C_y^2+\big(1+\frac{\alpha}{2}\big)^2\lambda_u C_z^2-(2+\alpha)\lambda_u C_{yz}\big\}+\phi\big\{\lambda_m C_y^2+(\lambda_m-\lambda_n)\alpha_1^2 p^2 C_x^2 \\ +\frac{\alpha_2^2}{4}\lambda_m C_z^2-(\lambda_m-\lambda_n)2\alpha_1 p C_{yx}-\alpha_2\lambda_m C_{yz}+(\lambda_m-\lambda_n)\alpha_1\alpha_2 p C_{xz}\big\}=0 \end{aligned} \tag{17}$$

On solving Eqns. (14), (15), (16) and (17), we get the optimum values of α, α_1, α_2 and ϕ as

$$\alpha^o=2K_{yz}-2,\ \alpha_1^o=\frac{\lambda_m C_y V}{pC_x[\lambda_m-(\lambda_m-\lambda_n)\rho_{zx}^2]}$$

$$\alpha_2^o=\frac{2C_y[\lambda_m\rho_{yz}-(\lambda_m-\lambda_n)\rho_{yx}\rho_{zx}]}{C_z[\lambda_m-(\lambda_m-\lambda_n)\rho_{zx}^2]},\ \phi^o=\frac{\lambda_u\delta_0}{\lambda_u\delta_0+\lambda_m\delta_0 V_1}$$

where $\delta_0=1-\rho_{yz}^2,\;\; V=(\rho_{yx}-\rho_{zx}\rho_{yz}),\;\; V_1=\frac{(\lambda_m-\lambda_n)V^2}{\lambda_m-(\lambda_m-\lambda_n)\rho_{zx}^2}$.

Put the optimum values of α^o, α_1^o, α_2^o and ϕ^o in Eqn. (13), we get

$$MSE_{min}(t)\cong\frac{\lambda_u\lambda_m S_y^2\delta_0(\delta_0-V_1)}{\lambda_u\delta_0+\lambda_m(\delta_0-V_1)}$$

Now, ignoring finite population correction terms, i.e. using $\frac{u}{N}\cong\frac{m}{N}\cong\frac{n}{N}\cong 0$, we get

$$MSE_{min}(t)\cong\frac{S_y^2}{n}\frac{\delta_0[\delta_0(1-\theta\rho_{zx}^2)-\theta V^2]}{[\delta_0(1-\theta\rho_{zx}^2)-\theta^2 V^2]}. \tag{18}$$

5. Optimal Replacement Policy

To get the estimated value of population mean $\bar{Y}$ with maximum precision, we have to find the optimal value of θ which is the fraction of sample taken fresh at second occasion. For optimal value of θ, we minimize the $MSE_{min}(t)$ with respect to θ in the Eqn. (18) which gives

$$\theta_0 = \frac{\delta_0 \pm \sqrt{\delta_0^2(1-\rho_{zx}^2) - \delta_0 V^2}}{V^2 + \delta_0 \rho_{zx}^2}.$$

Real values of θ_0 exists, whenever

$$\delta_0^2(1-\rho_{zx}^2) - \delta_0 V^2 \geq 0.$$

Here two real values of θ_0 are possible, we choose the one which lies in $0 \leq \theta_0 \leq 1$. If both of them are lying in said interval then we chose smallest one, because smallest value of θ_0 minimize the cost of the survey. So put value of θ_0 in Eqn. (18), we get the least value of minimum variance of t.

$$MSE_{min}(t) \cong \frac{S_y^2}{n} \frac{\delta_0[\delta_0(1-\theta_0\rho_{zx}^2) - \theta_0 V^2]}{[\delta_0(1-\theta_0\rho_{zx}^2) - \theta_0^2 V^2]}.$$

6. Efficiency Comparison

An estimator i is said to be more efficient than estimator j if $MSE(i) < MSE(j)$. Here, we have to find the condition under which the proposed estimator t works better than $\hat{\bar{y}}$, t_{SAG} and t_{SP}. Using Eqns. (1), (2), (3) and (18), it is concluded that

- $MSE_{min}(t) < Var(\hat{\bar{y}})$ if $\delta_0\rho_{yz}^2 + \delta_0\theta(\rho_{yx}^2 - 2\rho_{yx}\rho_{yz}\rho_{zx}) > \theta^2 V^2$
- $MSE_{min}(t) < MSE(t_{SAG})$ if $\frac{\delta_0(1-\theta\rho_{zx}^2)-\theta V^2}{\delta_0(1-\theta\rho_{zx}^2)-\theta^2 V^2} < \frac{1-\rho_{zx}^2-\theta(\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{zx})}{(\delta_0+\theta\rho_{yz}^2)(1-\rho_{zx}^2)-\theta^2(\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{zx})}$.
- $MSE_{min}(t) < MSE(t_{SP})$ if $\rho_{zx}^2 + \theta(\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{zx}) < 1$.

7. Empirical Study

We have generated some artificial populations using R software to compare the performance of the proposed estimator with existing estimators. The data statistics are given in Table 1.

Table 1 Data statistics

Population	N	n	m	u	θ	S_y^2	ρ_{yx}	ρ_{yz}	ρ_{zx}
1	500	150	37	113	0.7533	963.12	0.8283	0.9662	0.8573
2	1000	300	120	180	0.6000	72.24	-0.8284	0.8823	-0.9388
3	220	105	60	45	0.4286	192.02	0.5312	0.6919	0.7676

To compare the estimators, we have computed the percent relative efficiency (PRE) of estimators with respect to $\hat{\bar{y}}_n$. The PRE of estimator i with respect to $\hat{\bar{y}}_n$ can be calculated by

$$PRE(i, \hat{\bar{y}}_n) = \frac{Var(\hat{\bar{y}}_n) \times 100}{MSE(i)}. \quad (19)$$

From Table 2, it can be seen that the PRE values of the proposed estimator t are higher than the other considered estimators for all three populations. This concludes that the proposed estimator t is more efficient than the estimators of Shabbir, Azam and Gupta (2005) and Singh and Pal (2016).

Table 2 Percent relative efficiency of the estimators with respect to $\hat{\bar{y}}_n$

Estimator	Percent Relative Efficiency		
	Population-1	Population-2	Population-3
$\hat{\bar{y}}_n$	100	100	100
t_{SAG}	1216.64	345.88	154.11
t_{SP}	1254.76	418.80	182.31
t	1504.72	451.37	191.84

8. Application to a Case Study

In this section, we use the data from a case study of HBAT industries to show the performance of the proposed estimator over the usual estimator. HBAT is a premium manufacturer of paper in the U.S. It sold paper products to print media and magazines. To lead the market against the competitors, the HBAT marketing unit has hired Crimson Consulting Co. This is an established marketing research company. They surveys over different variables and take observations on 100 units. The observations are taken on a scale of 0 to 10 with 0 being poor and 10 as excellent. We use their observations as a population. Let $N = 100$ be the population size among which $n = 30$ units taken as a sample using simple random sampling without replacement (SRSWOR) on the first occasion. Let $m = 10$ units taken as matched from the first occasion. So let a fresh sample of size $u = 20$ from the remaining 70 population units. Using some variables and their correlation from the case study, we consider two cases.

Case-1

The variables are defined as $Z =$ Complaint resolution, $X =$ Ordering & bill and $Y =$ Customer satisfaction. The correlation between variables are $\rho_{yx} = 0.521732$, $\rho_{yz} = 0.603263$ and $\rho_{zx} = 0.756869$.

Using Eqn. (19), the percent relative efficiency of the estimators with respect to $\hat{\bar{y}}_n$ are calculated as $PRE(t_{SAG},\ \hat{\bar{y}}_n) = 135.43$, $PRE(t_{SP},\ \hat{\bar{y}}_n) = 154.34$ and $PRE(t,\ \hat{\bar{y}}_n) = 157.56$.

Case-2

Let $Z =$ Complaint resolution, $X =$ Product line and $Y =$ Customer satisfaction. The correlation between variables are $\rho_{yx} = 0.550546$, $\rho_{yz} = 0.603263$ and $\rho_{zx} = 0.561417$.

The percent relative efficiency of the estimators with respect to $\hat{\bar{y}}_n$ are calculated using Eqn. (19) as $PRE(t_{SAG},\ \hat{\bar{y}}_n) = 138.29$, $PRE(t_{SP},\ \hat{\bar{y}}_n) = 160.67$ and $PRE(t,\ \hat{\bar{y}}_n) = 162.09$.

In both cases the PRE of proposed estimator are greater than other considered estimators. Hence, using the proposed estimator in this case study results in smaller MSE.

9. Conclusion

In the present study, we have proposed an estimator for estimating the population mean in two occasion successive sampling techniques. The bias and MSE expressions for the proposed estimator derived. Also, the theoretical conditions have been obtained where the proposed estimator is more efficient than existing estimators. We have carried out an empirical study by using some artificial populations. The performance of the proposed estimator is compared with the usual unbiased estimator, Shabbir, Azam and Gupta (2005) and Singh and Pal (2016). It has been observed that the proposed estimator always performs better among the estimators in terms of having higher percent relative efficiency. For the application of the proposed estimator, we have used the case study data of HBAT industries and found that the proposed estimator performs better compared to the estimators under question. So, the proposed estimator justifies its worth theoretically and empirically. Hence, we recommend our estimator for future study and use in practice for real-life problems.

Acknowledgments

The authors are thankful to the two anonymous referees for their valuable comments which improved the quality of the paper.

References

Eckler AR. Rotation sampling. Ann Math Stat. 1955; 26(4): 664-685.

Feng S, Zou G. Sampling rotation method with auxiliary variable. Commun Stat Theory. 1997; 26(6): 1497-1509.

Jessen RJ. Statistical investigation of a sample survey for obtaining form facts. Res Bull. 1942; 26(304). https://lib.dr.iastate.edu/researchbulletin/vol26/iss304/1/.

Jabeen R, Sanaullah A, Hanif M, Zaka A. Two-exponential estimators for estimating population mean. AIMS Math. 2021; 6(1): 737-753.

Khalid M, Singh GN. General Estimation Procedures of Population Mean in Two-Occasion Successive Sampling under Random Non-response. Proc Natl Acad Sci India Sect A Phys Sci. 2020; https://doi.org/10.1007/s40010-020-00713-w.

Kullduff G. Some problems of optimum allocation for sampling on two occasions. Rev Int Stat Inst. 1963; 31(1): 24-57.

Okafor FC, Arnab R. Some strategies of two-stage sampling for estimating population ratio over two occasions. Aust J Stat. 1987; 29(2): 128-142.

Patterson HD. Sampling on successive occasions with partial replacement of units. J Roy Stat Soc B. 1950; 12(2): 241- 255.

Sen AR. Successive sampling with two auxiliary variables. Sankhya B. 1971; 33(3): 371-378.

Sen AR. Successive sampling with $p(p \geq 1)$ auxiliary variables. Ann Math Stat. 1972; 43(6): 2031-2034.

Sen AR. Theory and application of sampling on repeated occasions with several auxiliary variables. Biometrics. 1973; 29(2): 381-385.

Singh GN. On the use of chain type ratio estimator in successive Sampling. Stat in Trans - new ser. 2005; 7(1): 21-26.

Singh GN, Karna JP. Search of effective rotation patterns in presence of auxiliary information in successive sample over two-occasions. Stat in Trans - new ser. 2009; 10(1): 59-73.

Singh HP, Vishwakarma GK. A general procedure for estimating population mean in successive sampling. Commun Stat Theory. 2009; 38(2): 293-308.

Singh HP, Kumar S. Estimation of population product in the presence of non-response in successive sampling. Stat Pap. 2010; 51: 975-996.

Singh HP, Pal SK. An efficient effective rotation pattern in successive sampling over two occasions. Commun Stat Theory. 2016; 45(17): 5017-5027.

Singh HP, Pal SK. An exponential method for estimating the population mean in successive sampling. J Probab Stat Sci. 2017; 15(1): 79-93.

Sharma V, Kumar S. Generalized Estimator for Estimating the Population Mean in Successive Sampling. J Prob Stat Sci. 2018; 16(2): 209-228.

Shabbir J, Azam M, Gupta S. Estimation of current population mean in successive sampling. Proc ICCS-VIII. Lahore Pakistan. 2005; 13: 77-85.

Tracy DS, Singh HP. Efficient use of two auxiliary variables in two phase sampling as well as successive sampling. Pak J Stat. 1999; 15(1): 27-39.

Tikkiwal BD. Theory of successive sampling. Thesis for Diploma. ICAR, New Delhi. 1951.