11-1-2016

An Improved Generalized Estimation Procedure of Current Population Mean in Two-Occasion Successive Sampling

G. N. Singh
Indian School of Mines, Dhanbad, India, gnsingh_ism@yahoo.com

Alok Kumar Singh
Indian School of Mines, Dhanbad, India, aksingh.ism@gmail.com

Anup Kumar Sharma
Indian School of Mines, Dhanbad, India, aksharma.ism@gmail.com

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

An Improved Generalized Estimation Procedure of Current Population Mean in Two-Occasion Successive Sampling

An Improved Generalized Estimation Procedure of Current Population Mean in Two-Occasion Successive Sampling

G. N. Singh
Indian School of Mines
Dhanbad, India

Alok Kumar Singh
Indian School of Mines
Dhanbad, India

Anup Kumar Sharma
Indian School of Mines
Dhanbad, India

The present work is an attempt to make use of several auxiliary variables on both occasions for improving the precision of estimates for the current population mean in two-occasion successive sampling. A generalized exponential-cum-regression type estimator of the current population mean is proposed and its optimum replacement strategy has been discussed. Empirical studies are carried out to show the dominance of the proposed estimation procedure over the sample mean estimator and natural successive sampling estimator. Empirical results have been interpreted and suitable recommendations are put forward to survey practitioners.

Keywords: Successive sampling, auxiliary information, bias, mean square error, optimum replacement strategy

Introduction

There are many problems of practical interest in different fields of the applied and environmental sciences where the various characters of interest have tendencies to change over time. It is often required to monitor the behaviors of such characters at different points of time (occasions) and the patterns of variations occurring over the period of time. For example, an investigator or owner involved in the cold drinks industry may be interested (a) to know the average or total sale of cold drinks in the different seasons, (b) to know the pattern of change in average or total sale of cold drinks in two different seasons, or (c) they may be simultaneously interested to know both (a) and (b). These kinds of problems are well answered by the tools of successive (rotation) sampling.

A. K. Sharma is in the Department of Applied Mathematics. Email at: aksharma.ism@gmail.com. Alok Kumar Singh is a Junior Research Fellow in the Department of Applied Mathematics. Email at: aksingh.ism@gmail.com. G. N. Singh is in the Department of Applied Mathematics. Email at: gnsingh_ism@yahoo.com.

The theory of successive (rotation) sampling was initiated by Jessen (1942), where the idea of using the available information gathered on previous occasions during the past surveys was suggested. Jessen (1942) used past information in order to make current estimates more precise in agronomical surveys. This idea was further explored by Patterson (1950), Rao and Graham (1964), Gupta (1979), Das (1982), and Chaturvedi and Tripathi (1983), among others. Sen (1971) extended this theory by utilizing the information on two auxiliary variables, which was available on previous occasions, and suggested estimators of current the population mean in two-occasion successive sampling. Sen (1972; 1973) generalized his idea for several auxiliary variables. V. K. Singh, Singh, and Shukla (1991) and G. N. Singh and Singh (2001) used the auxiliary information from the current occasion for estimating the current population mean in two-occasion successive sampling. G. N. Singh (2003) extended this work for h-occasion successive sampling.

In many situations, information on an auxiliary variable may be readily available on the first as well as on the second occasion. For instance, to study the problems related to the public health and welfare of a state or a country, several factors that can be treated as auxiliary variables, such as the number of beds, doctors, and supporting staff in different hospitals, the amount of funds available for medicine, etc. may be known well in advance. Likewise, in other cases, there may be information available on several auxiliary variables and, if efficiently utilized, the estimates could be made more precise.

Utilizing the auxiliary information on both occasions, Feng and Zou (1997), Biradar and Singh (2001), G. N. Singh (2005), G. N. Singh and Priyanka (2006; 2007; 2008; 2010), G. N. Singh and Karna (2009), H. P. Singh and Vishwakarma (2009), G. N. Singh and Prasad (2010), G. N. Singh, Karna, and Prasad (2011), H. P. Singh, Tailor, Singh, and Kim (2011), G. N. Singh and Prasad (2013), and G. N. Singh and Homa (2013) proposed varieties of estimators of the population mean on the current (second) occasions in two-occasion successive sampling.

Motivated with these arguments, the objective of the present work is to propose a more precise estimator of the population mean on the current occasion using the information on p $(p \geq 2)$ stable auxiliary variables which are readily available on both occasions. Utilizing the information on p auxiliary variables, a generalized exponential-cum-regression type estimator of the current population mean in two-occasion successive sampling has been proposed. The dominance of the proposed estimator has been shown over the sample mean and natural successive sampling estimators. Empirical studies have been carried out to justify the proposition of estimator. Results are interpreted, and suitable recommendations have been made.

Formulation of Estimator

Let $U = (U_1, U_2,\ldots, U_N)$ be a finite population of N units which has been sampled over two occasions, and let the character under study be denoted by x (y) on the first (second) occasion. It is assumed that the information on p stable (non-negative integer constant) auxiliary variables z_j ($j = 1, 2,\ldots, p$), whose population means are known and closely related to x and y, are available on the first (second) occasion. Let a simple random sample (without replacement) of size n be drawn on the first occasion. A random subsample of size $m = n\lambda$ is retained (matched) for its use on the second occasion, while a fresh simple random sample (without replacement) of size $u = (n - m) = n\mu$ is drawn on the second occasion from the entire population so that the sample size on the second occasion is also n. Here λ and μ ($\lambda + \mu = 1$) are the fractions of the matched and fresh samples, respectively, on the current (second) occasion. The values of λ or μ would be chosen optimally.

The following notations have been considered for use below:

$\bar{X}$ ($\bar{Y}$): The population mean of the study variable x (y) on the first (second) occasion.

$\bar{Z}_j$: Population mean of the jth ($j = 1, 2,\ldots, p$) auxiliary variable.

$\bar{x}_n, \bar{x}_m, \bar{y}_u, \bar{y}_m, \bar{z}_{jn}, \bar{z}_{ju}, \bar{z}_{jm}$, ($j = 1, 2,\ldots, p$): The sample means of the respective variables based on the sample sizes shown in the subscript.

$\rho_{yx}, \rho_{yz_j}, \rho_{xz_j}, \rho_{z_j z_k}$: Population correlation coefficients between the variables shown in the subscript.

$S_x^2 = (N-1)^{-1} \sum_{i=1}^{N} (x_i - \bar{X})^2$: Population variance of the variable x.

$S_y^2, S_{z_j}^2$: Population variances of the variables y and z_j ($j = 1, 2,\ldots, p$), respectively.

To estimate the population mean $\bar{Y}$ on the current (second) occasion, two independent estimators are suggested. One is a generalized exponential type estimator based on a sample of size u ($= n\mu$) drawn afresh on the second occasion and given by

$$T_u = \bar{y}_u \exp \sum_{j=1}^{p} \left\{ \frac{\bar{Z}_j - \bar{z}_{ju}}{\bar{Z}_j + \bar{z}_{ju}} \right\} \tag{1}$$

The second estimator is a generalized exponential-cum-regression type estimator based on the sample of size $m\ (= n\lambda)$ common to both the occasions and is defined as

$$T_m = \bar{y}_m^* + \sum_{j=1}^{p} b_{yz_j}^{(m)} \left(\bar{Z}_j - \bar{z}_{jm} \right) \tag{2}$$

where

$$\bar{y}_m^* = \bar{y}_m + b_{yx}^{(m)} \left(\bar{x}_n^* - \bar{x}_m^* \right), \quad \bar{x}_n^* = \bar{x}_n \exp \sum_{j=1}^{p} \left\{ \frac{\bar{Z}_j - \bar{z}_{jn}}{\bar{Z}_j + \bar{z}_{jn}} \right\}$$

$$\bar{x}_m^* = \bar{x}_m \exp \sum_{j=1}^{p} \left\{ \frac{\bar{Z}_j - \bar{z}_{jm}}{\bar{Z}_j + \bar{z}_{jm}} \right\}$$

Combining the estimators T_u and T_m, we have the final estimator T of $\bar{Y}$ given as

$$T = \varphi T_u + (1 - \varphi) T_m \tag{3}$$

where $\varphi\ (0 \leq \varphi \leq 1)$ is an unknown constant (scalar) to be determined under certain criterion.

Remark 1:   The estimator T_u is suitable for estimating the population mean on the current occasion, while the estimator T_m is more appropriate for estimating change over two occasions. These two estimators may be derived from the estimator T by choosing φ as 1 or 0, respectively. To handle both problems simultaneously, an optimum choice of φ is required.

Properties of the Proposed Estimator

Bias and Mean Square Error

Because the estimators T_u and T_m are generalized exponential and generalized exponential-cum-regression type estimators, they are biased estimators of the population mean $\bar{Y}$. Therefore, the resulting estimator T is also a biased estimator of $\bar{Y}$. The bias B(.) and mean square error M(.) of the estimator T is derived under

large sample assumption and up to the first order of approximations using the following transformations:

$$\bar{y}_u = (1+e_1)\bar{Y}, \quad \bar{y}_m = (1+e_2)\bar{Y}, \quad \bar{x}_m = (1+e_3)\bar{X}, \quad \bar{x}_n = (1+e_4)\bar{X},$$

$$\bar{z}_{ju} = (1+e_{5j})\bar{Z}_j, \quad \bar{z}_{jm} = (1+e_{6j})\bar{Z}_j, \quad \bar{z}_{jn} = (1+e_{7j})\bar{Z}_j, \quad s_{yz_j} = (1+e_{8j})S_{yz_j},$$

$$s_{z_j}^2(m) = (1+e_{9j})S_{z_j}^2, \quad s_{xz_j}(m) = (1+e_{10j})S_{xz_j}, \quad s_{yx}(m) = (1+e_{11})S_{yx},$$

$$s_x^2(m) = (1+e_{12})S_x^2$$

such that $E(e_i) = 0$ and $E(e_{hj}) = 0$, $|e_i| \leq 1$ for $i = 1, 2, 3, 4, 11, 12$ and $|e_{hj}| \leq 1$ for $h = 5, 6, \ldots, 10, j = 1, 2, 3, \ldots, p$.

Under the above transformations, the estimators T_u and T_m take the following forms:

$$T_u = \bar{Y}(1+e_1)\exp\sum_{j=1}^{p}\left[-\frac{e_{5j}}{2}\left(1+\frac{e_{5j}}{2}\right)^{-1}\right] \tag{4}$$

$$T_m = \bar{Y}(1+e_2) - \frac{(1+e_{11})}{(1+e_{12})}\beta_{yx}\bar{X}\left(\begin{array}{c}(1+e_4)\exp\sum_{j=1}^{p}\frac{e_{7j}}{2}\left(1+\frac{e_{7j}}{2}\right)^{-1}\\-(1+e_3)\exp\sum_{j=1}^{p}\frac{e_{6j}}{2}\left(1+\frac{e_{6j}}{2}\right)^{-1}\end{array}\right) \tag{5}$$

Thus, there are the following theorems:

Theorem 1: The bias of the estimator T to the first order of approximations is obtained as

$$B(T) = \varphi B(T_u) + (1-\varphi)B(T_m) \tag{6}$$

where

$$B(T_u) = \left(\frac{1}{u} - \frac{1}{N}\right)\bar{Y}\left\{\frac{3}{8}\sum_{j=1}^{p}\frac{\alpha_{002}}{\bar{z}_j^2} + \frac{1}{8}\sum_{j=1}^{p}\frac{\alpha_{002}}{\bar{z}_j\bar{z}_k} - \frac{1}{2}\sum_{j=1}^{p}\frac{a_{010}\alpha_{001}}{\bar{Y}\bar{z}_j}\right\} \tag{7}$$

191

and

$$B(T_m) = \left(\frac{1}{m} - \frac{1}{n}\right) \left[\begin{array}{l} \sum_{j=1}^{p} \left(\frac{1}{2} \frac{\alpha_{100}\alpha_{001}}{\overline{z}_j} - \frac{3}{8} \overline{X} \frac{\alpha_{002}}{\overline{z}_j^2} - \frac{1}{8} \overline{X} \frac{\alpha_{002}}{\overline{z}_j \overline{z}_k} + \frac{1}{2} \overline{X} \frac{\alpha_{111}}{\overline{z}_j S_{yx}} \right) \\ - \sum_{j=1}^{p} \left(\frac{1}{2} \frac{\alpha_{201}}{\overline{z}_j S_x^2} - \frac{\alpha_{012}}{S_{yz_j}} + \frac{\alpha_{003}}{S_{z_j}^2} \right) - \frac{\alpha_{210}}{S_{yx}} + \frac{\alpha_{200}}{S_x^2} \end{array} \right] \tag{8}$$

where $\alpha_{pqr} = \mathrm{E}[(x_i - \overline{X})^p (y_i - \overline{Y})^q (z_j - \overline{Z}_j)^r]$ for integers $p, q, r \geq 0$ and $j = 1, 2, \ldots, p$.

Proof: The bias of the estimator T is given by

$$\begin{aligned} B(T) = \mathrm{E}\left[T - \overline{Y}\right] &= \varphi \mathrm{E}\left(T_u - \overline{Y}\right) + (1 - \varphi)\mathrm{E}\left(T_m - \overline{Y}\right) \\ &= \varphi B(T_u) + (1 - \varphi)B(T_m) \end{aligned} \tag{9}$$

where $B(T_u) = \mathrm{E}(T_u - \overline{Y})$ and $B(T_m) = \mathrm{E}(T_m - \overline{Y})$.

To derive the $B(T_u)$, proceed as follows:

$$\mathrm{E}\left(T_u - \overline{Y}\right) = \mathrm{E}\left(\overline{Y}(1 + e_1) \exp \sum_{j=1}^{p} \left[-\frac{e_{5j}}{2}\left(1 + \frac{e_{5j}}{2}\right)^{-1} \right] - \overline{Y} \right) \tag{10}$$

Now, expanding the right hand side of (10) binomially and exponentially and taking expectations and retaining the terms up to the first order of approximations, we have the expression of the bias of the estimator T_u as given in (7).

Similarly, the bias of the estimator T_m is written as

$$\mathrm{E}\left(T_m - \overline{Y}\right) =$$

$$\mathrm{E}\left[\overline{Y}e_2 - \frac{(1 + e_{11})}{(1 + e_{12})} \beta_{yx} \overline{X} \left\{ \begin{array}{l} (1 + e_4)\exp\left[\sum_{j=1}^{p} \frac{e_{7j}}{2}\left(1 + \frac{e_{7j}}{2}\right)^{-1} \right] \\ -(1 + e_3)\exp\left[\sum_{j=1}^{p} \frac{e_{6j}}{2}\left(1 + \frac{e_{6j}}{2}\right)^{-1} \right] \end{array} \right\} \right] \tag{11}$$

Expanding (11) binomially and exponentially, taking expectations both sides, and retaining the terms up to the first order of approximations yields the expression of the bias of the estimator T_m as shown in (8).

Theorem 2: The mean square error of the estimator T to the first degree of approximation is obtained as

$$\text{M}(T)=\varphi^2\,\text{M}(T_u)+(1-\varphi)^2\,\text{M}(T_m)+2\varphi(1-\varphi)\text{C}(T_u,T_m) \quad (12)$$

where

$$\text{M}(T_u)=\left(\frac{1}{u}-\frac{1}{N}\right)S_y^2\left[1+\frac{p}{4}-\sum_{j=1}^{p}\rho_{yz_j}+\sum_{j\neq k=1}^{p}\rho_{z_jz_k}\right] \quad (13)$$

$$\text{M}(T_m)=S_y^2\left[\left(\frac{1}{m}-\frac{1}{N}\right)\left[\begin{array}{l}1-\sum_{j=1}^{p}\left(\rho_{yz_j}^2+\rho_{yz_j}\rho_{yz_k}\rho_{z_jz_k}-\rho_{yx}\rho_{yz_j}^2\right)-2\rho_{yx}^2\\+\rho_{yx}^2\left(1+\frac{p}{4}-\sum_{j=1}^{p}\rho_{yz_j}+\sum_{j\neq k=1}^{p}\rho_{z_jz_k}\right)\end{array}\right]\right] \quad (14)$$

$$\text{C}(T_u,T_m)=-\frac{1}{N}S_y^2\left[1-\sum_{j=1}^{p}\rho_{yz_j}^2\right] \quad (15)$$

Proof: The mean square error of the estimator T is given by

$$\begin{aligned}\text{M}(T)&=\text{E}\left[T-\bar{Y}\right]^2=\text{E}\left[\varphi(T_u-\bar{Y})+(1-\varphi)(T_m-\bar{Y})\right]^2\\&=\varphi^2\,\text{M}(T_u)+(1-\varphi)^2\,\text{M}(T_m)+2\varphi(1-\varphi)\text{C}(T_u,T_m)\end{aligned} \quad (16)$$

where $\text{C}(T_u, T_m) = \text{E}[(T_u - \bar{Y})(T_m - \bar{Y})]$, $\text{M}(T_u) = \text{E}(T_u - \bar{Y})^2$, $\text{M}(T_m) = \text{E}(T_m - \bar{Y})^2$. To derive the $\text{M}(T_u)$, proceed as follows:

$$\text{E}(T_u-\bar{Y})^2=\text{E}\left(\bar{Y}(1+e_1)\exp\sum_{j=1}^{p}\left[-\frac{e_{5j}}{2}\left(1+\frac{e_{5j}}{2}\right)^{-1}\right]-\bar{Y}\right)^2 \quad (17)$$

Now, expanding the right hand side of (17) binomially and exponentially and taking expectations and retaining the terms up to the first order of approximations, we have the expression of the mean square error of the estimator T_u as given in (13).

The mean square error of the estimator T_m is written as

$$E\left(T_m - \bar{Y}\right)^2 = E\left[\bar{Y}e_2 - \frac{(1+e_{11})}{(1+e_{12})}\beta_{yx}\bar{X}\left\{\begin{array}{c}(1+e_4)\exp\left[\sum_{j=1}^{p}\frac{e_{7j}}{2+e_{7j}}\right]\\-(1+e_3)\exp\left[\sum_{j=1}^{p}\frac{e_{6j}}{2+e_{6j}}\right]\end{array}\right\}\right]^2 \quad (18)$$

Expanding (18) binomially and exponential, taking expectations both sides, and retaining the terms up to the first order of approximations, the expression is derived for the mean square error of the estimator T_m as shown in (14). Similarly, the expectation of $C(T_u, T_m)$ may be derived in the form shown in (15).

Remark 2: The above results are derived under the assumption that the coefficients of variation of variables x, y, z_j, and z_k are approximately equal. We have also considered the intuitive assumptions $\rho_{xz_j} = \rho_{yz_j}$ $(j = 1, 2, 3,\ldots, p)$, as suggested by Cochran (1984) and Feng and Zou (1997). In the light of these assumptions, the expression of $M(T_m)$ takes the form as shown in (14).

Minimum Mean Square Errors of the Estimator *T*

Because the mean square error of the estimator T in (12) is a function of the unknown constant (scalar) φ, it can be minimized with respect to φ and, subsequently, the optimum value of φ is obtained as

$$\varphi_{opt} = \frac{M(T_m) - C(T_u, T_m)}{M(T_u) + M(T_m) - 2C(T_u, T_m)} \quad (19)$$

From (19), substituting the value of φ_{opt} in (12), we get the optimum mean square error of the estimator *T* as

$$M(T)_{opt} = \frac{M(T_u)M(T_m) - \{C(T_u, T_m)\}^2}{M(T_u) + M(T_m) - 2C(T_u, T_m)} \quad (20)$$

Further substituting the values from (13)-(15) into (19) and (20), the simplified values of φ_{opt} and $\text{M}(T)_{\text{opt}}$ are obtained as

$$\varphi_{\text{opt}} = \frac{\mu\left(A_9 + \mu A_8\right)}{A_9 - \mu A_{12} + \mu^2 A_{13}} \tag{21}$$

$$\text{M}(T)_{\text{opt}} = \left[\frac{A_{18} - \mu^2 A_{20} - \mu A_{21}}{A_9 + \mu^2 A_{13} - \mu A_{12}}\right]\frac{S_y^2}{n} \tag{22}$$

where

$$A_1 = 1 - \sum_{j=1}^{p}\rho_{yz_j}^2 + \rho_{yz_j}\rho_{yz_k}\rho_{z_jz_k}, \quad A_2 = \rho_{yx}^2\left(1 + \frac{p}{4} - \sum_{j=1}^{p}\rho_{yz_j} + \frac{1}{4}\sum_{j\neq k=1}^{p}\rho_{z_jz_k}\right),$$

$$A_3 = \rho_{yx}\left(2\rho_{yx} - \sum_{j=1}^{p}\rho_{yz_j}\rho_{yz_j}\right), \quad A_4 = 1 - \sum_{j=1}^{p}\rho_{yz_j}^2, \quad A_5 = A_1 + A_2 - A_3, \quad A_6 = A_2 - A_3,$$

$$A_7 = A_1 - A_4, \quad A_8 = A_6 + fA_7, \quad A_9 = A_5 - A_8, \quad A_{10} = 1 + \frac{p}{4} - \sum_{j=1}^{p}\rho_{yz_j} + \frac{1}{4}\sum_{j\neq k=1}^{p}\rho_{z_jz_k},$$

$$A_{11} = A_{10} - A_4, \quad A_{12} = A_{10} - A_9 + fA_{11}, \quad A_{13} = A_8 + fA_{11}, \quad A_{14} = A_{10}A_1 + A_6A_9,$$

$$A_{15} = A_{10}A_1, \quad A_{16} = A_6A_{10}, \quad A_{17} = A_4^2 - A_{15}, \quad A_{18} = A_{14} - A_{16} - fA_{15},$$

$$A_{19} = A_{14} - A_{15} - A_{16} + fA_{17}, \quad A_{20} = fA_{16} - f^2A_{17}, \quad A_{21} = fA_{19} - A_{16}$$

where $f = n/N$.

Optimum Replacement Strategy of the Estimator *T*

The optimum mean square error $\text{M}(T)_{\text{opt}}$ in (22) is a function on μ, the fraction of the sample to be drawn afresh at the second occasion. It is an important factor in reducing the cost of the survey, therefore, to determine the optimum value of μ so that $\bar{Y}$ may be estimated with maximum precision and minimum cost. We thus minimize $\text{M}(T)_{\text{opt}}$ with respect to μ which results in a quadratic equation in μ, which is shown as

$$\mu^2 D_1 - 2\mu D_2 - D_3 = 0 \tag{23}$$

195

where $D_1 = A_{12}A_{20} + A_{13}A_{21}$, $D_2 = A_{10}A_{20} + A_{13}A_{18}$, $D_3 = A_{10}A_{21} - A_{12}A_{18}$.

Solving (23) for μ, the solutions of μ (say $\hat{\mu}$) are given as

$$\hat{\mu} = \frac{D_2 \pm \sqrt{D_2^2 + D_1 D_3}}{D_1} \qquad (24)$$

From (24), it is clear that the real values of $\hat{\mu}$ exist IFF the quantity under the square root is greater than or equal to zero. For any combinations of correlations which satisfy this condition for real solutions, two real values of $\hat{\mu}$ are possible. Hence, while choosing the values of $\hat{\mu}$, it should be remembered that $0 \le \hat{\mu} \le 1$, and that all other values of $\hat{\mu}$ are said to be inadmissible. If both the values of $\hat{\mu}$ are admissible, the lowest one is the best choice as it reduces the cost of the survey. From (24), substituting the admissible value of $\hat{\mu}$ (say μ_0) in (22), we have the optimum value of mean square error of the estimator T, which is shown below:

$$M(T)_{opt}^* = \left[\frac{A_{18} - \mu_0^2 A_{20} - \mu_0 A_{21}}{A_{10} + \mu_0^2 A_{13} - \mu_0 A_{12}} \right] \frac{S_y^2}{n} \qquad (25)$$

Special Case

When the p auxiliary variates are mutually uncorrelated, i.e., $\rho_{z_j z_k} = 0$ for $j \ne k = 1, 2, 3,\ldots, p$, then the expression of the optimum values of μ and $M(T)_{opt}$ reduce to

$$\hat{\mu} = \frac{D_2^* \pm \sqrt{D_2^{*2} + D_1^* D_3^*}}{D_1^*} \qquad (26)$$

and

$$M(T)_{opt}^* = \left[\frac{A_{18}^* - \mu_0^2 A_{20}^* - \mu_0 A_{21}^*}{A_{10}^* + \mu_0^2 A_{13}^* - \mu_0 A_{12}^*} \right] \frac{S_y^2}{n} \qquad (27)$$

where

$$D_1^* = A_{12}^* A_{20}^* + A_{13}^* A_{21}^*, \quad D_2^* = A_{10}^* A_{20}^* + A_{13}^* A_{18}^*, \quad D_3^* = A_{10}^* A_{21}^* - A_{12}^* A_{18}^*,$$

$$A_1^* = 1 - \sum_{j=1}^{p} \rho_{yz_j}^2, \quad A_2^* = \rho_{yx}^2 \left(1 + \frac{p}{4} - \sum_{j=1}^{p} \rho_{yz_j}\right), \quad A_5^* = A_1^* + A_2^* - A_3, \quad A_6^* = A_2^* - A_3,$$

$$A_7^* = A_1^* - A_4, \quad A_8^* = A_6^* + fA_7^*, \quad A_9^* = A_5^* - A_8^*, \quad A_{10}^* = 1 + \frac{p}{4} - \sum_{j=1}^{p} \rho_{yz_j},$$

$$A_{11}^* = A_{10}^* - A_4, \quad A_{12}^* = A_{10}^* - A_9^* + fA_{11}^*, \quad A_{13}^* = A_8^* + fA_{11}^*, \quad A_{14}^* = A_{10}^* A_1^* + A_6^* A_9^*,$$

$$A_{15}^* = A_{10}^* A_1^*, \quad A_{16}^* = A_6^* A_{10}^*, \quad A_{17}^* = A_4^2 - A_{15}^*, \quad A_{18}^* = A_{14}^* - A_{16}^* - fA_{15}^*,$$

$$A_{19}^* = A_{14}^* - A_{15}^* - A_{16}^* + fA_{17}^*, \quad A_{20}^* = fA_{16}^* - f^2 A_{17}^*, \quad A_{21}^* = fA_{19}^* - A_{16}^*$$

Efficiency Comparison

The percent relative efficiencies of the estimator T with respect to (i) the sample mean estimator $\bar{y}_n$ when there is no matching and (ii) $\hat{\bar{Y}} = \varphi_u^* \bar{y}_u + \left(1 - \varphi^*\right) \bar{y}_m'$ when no additional auxiliary information is used at any occasion, where $\bar{y}_m' = \bar{y}_m + \beta_{yx}\left(\bar{x}_n - \bar{x}_m\right)$, have been obtained for different choices of the correlations involved. Since $\bar{y}_n$ and $\hat{\bar{Y}}$ are unbiased estimators of $\bar{Y}$ following Sukhatme, Sukhatme, Sukhatme, and Asok (1984), the variance of $\bar{y}_n$ and optimum variance of $\hat{\bar{Y}}$ are given by

$$V\left(\bar{y}_n\right) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \tag{28}$$

$$V\left(\hat{\bar{Y}}\right)_{\text{opt}} = \left[1 + \sqrt{1 - \rho_{yx}^2}\right] \frac{S_y^2}{2n} - \frac{S_y^2}{N} \tag{29}$$

The percent relative efficiencies E_1 and E_2 of T (under optimal condition) with respect to $\bar{y}_n$ and $\hat{\bar{Y}}$, respectively, are given by

$$E_1 = \frac{V\left(\bar{y}_n\right)}{M\left(T\right)_{\text{opt}}^*} \times 100, \quad E_2 = \frac{V\left(\hat{\bar{Y}}\right)_{\text{opt}}}{M\left(T\right)_{\text{opt}}^*} \times 100$$

Empirical Study

The expression of the optimum μ (i.e., μ_0) and the percent relative efficiencies E_1 and E_2 are in terms of population correlation coefficients. Therefore, the values of μ_0, E_1, and E_2 have been computed for different choices of positive correlations, while the value of $f (= n/N)$ (sampling fraction) is chosen to be 0.1. For empirical studies, cases of $p = 2$ and 3 have been considered.

Case 1

For $p = 2$ and assuming that the two auxiliary variables are correlated, i.e., $\rho_{z_1 z_2} \neq 0$, the values of A_1, A_2, A_3, A_4, A_9, and A_{10} take the form

$$A_1 = 1 - \left(\rho_{yz_1}^2 + \rho_{yz_2}^2\right) + 2\rho_{yz_1}\rho_{yz_2}\rho_{z_1 z_2}, \quad A_2 = \rho_{yx}^2\left(\frac{3}{2} + 2\rho_{z_1 z_2} - \left(\rho_{yz_1} + \rho_{yz_2}\right)\right),$$

$$A_3 = \rho_{yx}\left(2\rho_{yx} - \left(\rho_{yz_1}^2 + \rho_{yz_2}^2\right)\right), \quad A_4 = 1 - \left(\rho_{yz_1}^2 + \rho_{yz_2}^2\right),$$

$$A_{10} = \frac{3}{2} + \frac{1}{2}\rho_{z_1 z_2} - \left(\rho_{yz_1} + \rho_{yz_2}\right)$$

Substituting these values in (24) and (25) yields the values of optimum $M(T)_{opt}^*$, E_1, and E_2. For different choices of correlations, Tables 1-2 show the optimum values of μ (i.e., μ_0) and percent relative efficiencies E_1 and E_2 of the estimator T (under optimal condition) with respect to $\bar{y}_n$ and $\hat{\bar{Y}}$, respectively.

Case 2

For $p = 2$ and assuming that the two auxiliary variables are uncorrelated, i.e., $\rho_{z_1 z_2} = 0$, the values of A_1^*, A_2^*, and A_{10}^* take the form

$$A_1^* = 1 - \left(\rho_{yz_1}^2 + \rho_{yz_2}^2\right), \quad A_2^* = \rho_{yx}^2\left(\frac{3}{2} - \left(\rho_{yz_1} + \rho_{yz_2}\right)\right), \quad A_{10}^* = \frac{3}{2} - \left(\rho_{yz_1} + \rho_{yz_2}\right)$$

Using these values in (26) and (27), the optimum values of μ, E_1, and E_2 are shown in Table 3.

Table 1. Optimum values of μ and percent relative efficiencies of T with respect to $\bar{y}_n$ and $\hat{\bar{Y}}$ for $\rho_{yx} = 0.3$

ρ_{yz_2}	$\rho_{z_1 z_2}$	ρ_{yz_1} 0.5			0.6			0.7		
		μ_0	E_1	E_2	μ_0	E_1	E_2	μ_0	E_1	E_2
0.4	0.3	0.7265	132.18	128.80	0.4034	168.19	163.88	0.3533	247.90	241.56
	0.4	0.7101	123.49	120.33	0.3839	152.06	148.16	0.3872	210.05	204.67
	0.5	0.7001	115.87	112.91	0.3700	138.74	135.19	0.3360	181.76	177.10
	0.6	0.6932	109.13	106.34	0.3595	127.57	124.31	0.2952	159.92	155.82
0.6	0.3	0.6022	149.75	145.92	0.3207	193.77	188.81	0.3411	306.04	298.21
	0.4	0.5842	138.08	134.54	0.2827	170.36	166.00	0.2607	243.37	237.15
	0.5	0.5719	128.09	124.81	0.2548	151.91	148.03	0.1979	200.80	195.67
	0.6	0.5630	119.44	116.39	0.2335	137.04	133.53	0.1475	170.24	165.88
0.8	0.3	0.5882	174.09	169.63	0.3120	234.31	228.31	0.2669	418.57	407.85
	0.4	0.5517	157.66	153.63	0.2512	197.91	192.84	0.1760	298.60	290.96
	0.5	0.5252	144.02	140.34	0.2051	170.92	166.55	0.1034	228.96	223.10
	0.6	0.5050	132.53	129.14	0.1690	150.21	146.36	0.0441	184.02	179.31

Table 2. Optimum values of Optimum values of μ and percent relative efficiencies of T with respect to $\bar{y}_n$ and $\hat{\bar{Y}}$ for $\rho_{yx} = 0.5$

ρ_{yz_2}	$\rho_{z_1 z_2}$	ρ_{yz_1} 0.5			0.6			0.7		
		μ_0	E_1	E_2	μ_0	E_1	E_2	μ_0	E_1	E_2
0.4	0.3	*	--	--	0.3809	170.91	158.19	0.3568	249.91	231.31
	0.4	0.7440	123.96	114.73	0.3815	152.39	141.05	0.3875	208.31	192.80
	0.5	0.6739	115.31	106.73	0.3779	137.57	127.33	0.3431	178.37	165.10
	0.6	0.6411	107.74	100.72	0.3738	125.42	116.08	0.3120	155.88	144.27
0.6	0.3	0.6748	152.04	140.72	0.2932	196.4	181.78	0.3407	307.53	284.64
	0.4	0.5913	138.47	128.16	0.2827	170.36	157.68	0.2673	240.90	222.97
	0.5	0.5579	127.09	117.63	0.2740	150.50	139.29	0.2192	197.38	182.69
	0.6	0.5386	117.45	108.71	0.2669	134.83	124.80	0.1851	166.90	154.48
0.8	0.3	0.6209	176.39	163.26	0.3015	236.81	219.19	0.2670	418.02	386.91
	0.4	0.5506	157.54	145.81	0.2567	197.19	182.51	0.1865	294.69	272.76
	0.5	0.5134	142.28	131.69	0.2285	168.91	156.34	0.1319	225.58	208.79
	0.6	0.4896	129.72	120.07	0.2089	147.72	136.73	0.0924	181.87	168.33

Note: "*" indicates μ_0 does not exists and "--" indicates no gain.

Table 3. Optimum values of Optimum values of μ and percent relative efficiencies of T with respect to $\bar{y}_n$ and $\hat{\bar{Y}}$ for $\rho_{z_1 z_2} = 0.0$

ρ_{yx}	ρ_{yz_2}	ρ_{yz_1} 0.5			0.6			0.7		
		μ_0	E_1	E_2	μ_0	E_1	E_2	μ_0	E_1	E_2
0.5	0.3	0.4203	138.95	128.61	0.4317	158.35	146.57	0.3727	187.50	173.54
	0.4	0.4874	156.18	144.55	0.5084	181.27	167.78	0.4661	219.80	203.44
	0.5	0.5084	181.27	167.78	0.5359	216.08	200.00	0.4938	272.88	252.57
	0.6	0.4661	219.80	203.44	0.4938	272.88	252.57	0.4112	371.39	343.75
0.7	0.3	0.5632	157.93	132.85	0.5865	185.15	155.75	0.6036	228.38	192.11
	0.4	0.6015	182.74	153.71	0.6312	220.87	185.79	0.6662	287.47	241.82
	0.5	0.6312	220.87	185.79	0.6759	281.34	236.66	0.7712	417.32	351.04
	0.6	0.6662	287.47	241.82	0.7712	417.32	351.04	*	--	--
0.9	0.3	0.9388	268.34	184.24	*	--	--	*	--	--
	0.4	*	--	--	*	--	--	*	--	--
	0.5	*	--	--	*	--	--	*	--	--
	0.6	*	--	--	*	--	--	*	--	--

Note: "*" indicates μ_0 does not exists and "--" indicates no gain.

Case 3

For $p = 3$ and assuming that the two auxiliary variables are correlated, i.e., $\rho_{z_j z_k} \neq 0$ for $j \neq k = 1, 2, 3$, the values of $A_1, A_2, A_3, A_4,$ and A_{10} take the form

$$A_1 = 1 - \left(\rho_{yz_1}^2 + \rho_{yz_2}^2 + \rho_{yz_3}^2 \right) + 2\left(\rho_{yz_1}\rho_{yz_2}\rho_{z_1 z_2} + \rho_{yz_1}\rho_{yz_3}\rho_{z_1 z_3} + \rho_{yz_2}\rho_{yz_3}\rho_{z_2 z_3} \right),$$

$$A_2 = \rho_{yx}^2 \left(\frac{7}{4} + 2\left(\rho_{z_1 z_2} + \rho_{z_1 z_3} + \rho_{z_2 z_3} \right) - \left(\rho_{yz_1} + \rho_{yz_2} + \rho_{yz_3} \right) \right),$$

$$A_3 = \rho_{yx}\left(2\rho_{yx} - \left(\rho_{yz_1}^2 + \rho_{yz_2}^2 + \rho_{yz_3}^2 \right) \right), \quad A_4 = 1 - \left(\rho_{yz_1}^2 + \rho_{yz_2}^2 + \rho_{yz_3}^2 \right),$$

$$A_{10} = \frac{7}{4} + \frac{1}{2}\left(\rho_{z_1 z_2} + \rho_{z_1 z_3} + \rho_{z_2 z_3} \right) - \left(\rho_{yz_1} + \rho_{yz_2} + \rho_{yz_3} \right)$$

In this case there are seven different correlations. For a few sets of these seven correlations, the optimum value of μ (i.e., μ_0) and percent relative efficiencies E_1 and E_2 of the estimator T (under optimal condition) with respect to $\bar{y}_n$ and $\hat{\bar{Y}}$ have been computed and shown below:

Set 1:

$$\rho_{yx} = 0.3, \quad \rho_{yz_1} = 0.5, \quad \rho_{yz_2} = 0.6, \quad \rho_{yz_3} = 0.5, \quad \rho_{z_1 z_2} = 0.3, \quad \rho_{z_1 z_3} = 0.4,$$
$$\rho_{z_2 z_3} = 0.6, \quad \mu_0 = 0.3664, \quad E_1 = 104.37, \quad E_2 = 101.37$$

Set 2:

$$\rho_{yx} = 0.3, \quad \rho_{yz_1} = 0.6, \quad \rho_{yz_2} = 0.6, \quad \rho_{yz_3} = 0.5, \quad \rho_{z_1 z_2} = 0.3, \quad \rho_{z_1 z_3} = 0.4,$$
$$\rho_{z_2 z_3} = 0.6, \quad \mu_0 = 0.2900, \quad E_1 = 110.60, \quad E_2 = 107.77$$

Set 3:

$$\rho_{yx} = 0.3, \quad \rho_{yz_1} = 0.7, \quad \rho_{yz_2} = 0.6, \quad \rho_{yz_3} = 0.5, \quad \rho_{z_1 z_2} = 0.3, \quad \rho_{z_1 z_3} = 0.4,$$
$$\rho_{z_2 z_3} = 0.6, \quad \mu_0 = 0.2393, \quad E_1 = 119.33, \quad E_2 = 116.27$$

Set 4:

$$\rho_{yx} = 0.3, \quad \rho_{yz_1} = 0.8, \quad \rho_{yz_2} = 0.6, \quad \rho_{yz_3} = 0.5, \quad \rho_{z_1 z_2} = 0.3, \quad \rho_{z_1 z_3} = 0.4,$$
$$\rho_{z_2 z_3} = 0.6, \quad \mu_0 = 0.2105, \quad E_1 = 131.59, \quad E_2 = 128.22$$

Case 4

For $p = 3$ and assuming that the two auxiliary variables are uncorrelated, i.e., $\rho_{z_1 z_2} = 0$ for $j \neq k = 1, 2, 3$, the values of A_1, A_2, and A_{10} take the form

$$A_1 = 1 - \left(\rho_{yz_1}^2 + \rho_{yz_2}^2 + \rho_{yz_3}^2 \right), \quad A_2 = \rho_{yx}^2 \left(\frac{7}{4} - \left(\rho_{yz_1} + \rho_{yz_2} + \rho_{yz_3} \right) \right),$$
$$A_{10} = \frac{7}{4} - \left(\rho_{yz_1} + \rho_{yz_2} + \rho_{yz_3} \right)$$

For a few sets of the above four correlations, the values of the optimum value of μ (i.e., μ_0) and percent relative efficiencies E_1 and E_2 are shown below:

Set 1:

$$\rho_{yx} = 0.3, \quad \rho_{yz_1} = 0.5, \quad \rho_{yz_2} = 0.6, \quad \rho_{yz_3} = 0.4, \quad \mu_0 = 0.6004, \quad E_1 = 382.42,$$
$$E_2 = 372.64$$

Set 2:

$\rho_{yx} = 0.4$, $\rho_{yz_1} = 0.5$, $\rho_{yz_2} = 0.6$, $\rho_{yz_3} = 0.4$, $\mu_0 = 0.7981$, $E_1 = 397.89$,

$E_2 = 379.44$

Set 3:

$\rho_{yx} = 0.5$, $\rho_{yz_1} = 0.5$, $\rho_{yz_2} = 0.6$, $\rho_{yz_3} = 0.4$, $\mu_0 = 0.3807$, $E_1 = 449.12$,

$E_2 = 415.67$

Set 4:

$\rho_{yx} = 0.6$, $\rho_{yz_1} = 0.5$, $\rho_{yz_2} = 0.6$, $\rho_{yz_3} = 0.4$, $\mu_0 = 0.6317$, $E_1 = 568.19$,

$E_2 = 505.06$

Conclusion

1. From Tables 1-2 it is vindicated that:

    a. For the fixed values of ρ_{yx}, $\rho_{z_1z_2}$, and ρ_{yz_1}, the values of μ_0 decrease and E_1 and E_2 increase with the increasing values of ρ_{yz_2}. Similarly, for fixed values of ρ_{yx}, $\rho_{z_1z_2}$, and ρ_{yz_2}, the optimum value of μ_0 decrease and E_1 and E_2 increase with the increasing values of ρ_{yz_1}. These patterns indicate that a smaller fresh sample on the current occasion is required if highly correlated auxiliary variables are available.

    b. For the fixed values of ρ_{yx}, ρ_{yz_1}, and ρ_{yz_2}, the values of μ_0, E_1, and E_2 decrease with the increasing values of $\rho_{z_1z_2}$; this means that the auxiliary variables are quite sensitive with respect to the relation between them.

2. From Table 3, i.e., when the auxiliary variables are uncorrelated, it has been observed that

    a. For fixed values of ρ_{yz_1} and ρ_{yz_2}, the values of E_1 and E_2 increase with increasing value of ρ_{yx}, while no definite patterns are observed in μ_0.

    b. For fixed values of ρ_{yx} and ρ_{yz_1}, the values of E_1 and E_2 increase with increasing value of ρ_{yz_2}, while no definite patterns are observed in μ_0. Similar patterns are visible for the case when the values of ρ_{yx} and ρ_{yz_2} are fixed and increasing values of ρ_{yz_1} are observed.

3. For $p = 3$ and when the three auxiliary variables are uncorrelated, for fixed values of ρ_{yx}, $\rho_{z_1z_2}$, $\rho_{z_2z_3}$, $\rho_{z_1z_3}$, ρ_{yz_2}, and ρ_{yz_3}, the values of μ_0 decrease while E_1 and E_2 increase with the increasing values of ρ_{yz_1}. Similar patterns are observed if the case for the increasing values of ρ_{yz_2} or ρ_{yz_3} is taken into account.

4. For $p = 3$ and when the three auxiliary variables are mutually correlated, we observed that no specific pattern is seen as for so many combinations of correlations the optimum values of μ_0 do not exist. This behavior suggests that the correlation between the auxiliary variable do not play a significant role in terms of the proposed estimator.

5. It could be seen that the results are more appreciable for one and two auxiliary variables, while when the number of auxiliary variables increases, the expressions become complex due to the increase in the number of correlations. Hence, practically, it is more realistic to use two auxiliary variables out of several available auxiliary variables.

    Thus, it is clear that the use of the auxiliary variables is highly rewarding in terms of the proposed estimator. It is also clear that, if the information on highly correlated auxiliary variables is used, only a relatively small fraction of the sample on the current (second) occasion is desired to be replaced by a fresh sample, which reduces the cost of the survey. Hence, it can be recommended for future use.

References

Biradar, R. S., & Singh, H. P. (2001). Successive sampling using auxiliary information on both occasions. *Bulletin of the Calcutta Statistical Association, 51*(203-204), 243-251.

Chaturvedi, D. K., & Tripathi, T. P. (1983). Estimation of population ratio on two occasions using multivariate auxiliary information. *Journal of the Indian Statistical Association, 21*, 113-120.

Cochran, W. G. (1984). *Sampling techniques* (3rd ed.). New Delhi, India: Wiley Eastern.

Das, A. K. (1982). Estimation of population ratio on two occasions. *Journal of the Indian Society of Agricultural Statistics, 34*(2), 1-9.

Feng, S., & Zou, G. (1997). Sample rotation method with auxiliary variable. *Communications in Statistics – Theory and Methods, 26*(6), 1497-1509. doi: 10.1080/03610929708831996

Gupta, P. C. (1979). Sampling on two successive occasions. *Journal of Statistical Research, 13*(13), 7-16.

Jessen, R. J. (1942). Statistical Investigation of a Sample Survey for obtaining farm facts. *Agricultural Research Bulletins, 304*, 1-104. Retrieved from http://lib.dr.iastate.edu/ag_researchbulletins/27/

Patterson, H. D. (1950). Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society. Series B (Methodological), 12*(2), 241-255. Available from http://www.jstor.org/stable/2983982

Rao, J. N. K., & Graham, J. E. (1964). Rotation design for sampling on repeated occasions. *Journal of the American Statistical Association, 59*(306), 492-509. doi: 10.1080/01621459.1964.10482175

Sen, A. R. (1971). Successive sampling with two auxiliary variables. *Sankhyā: The Indian Journal of Statistics, Series B, 33*(3/4), 371-378. Available from http://www.jstor.org/stable/25051764

Sen, A. R. (1972). Successive sampling with p ($p \geq 1$) auxiliary variables. *The Annals of Mathematical Statistics, 43*(6), 2031-2034. Available from http://www.jstor.org/stable/2240218

Sen, A. R. (1973). Note: Theory and application of sampling on repeated occasions with several auxiliary variables. *Biometrics, 29*(2), 381-385. doi: 10.2307/2529401

Singh, G. N. (2003). Estimation of population mean using auxiliary information on recent occasion in h-occasion successive sampling. *Statistics in Transition, 6*(4), 523-532.

Singh, G. N. (2005). On the use of chain-type ratio estimator in successive sampling. *Statistics in Transition, 7*(1), 21-26.

Singh, G. N., & Homa, F. (2013). Effective rotation patterns in successive sampling over two occasions. *Journal of Statistical Theory and Practice, 7*(1), 146-155. doi: 10.1080/15598608.2012.755484

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

Singh, G. N., Karna, J. P., & Prasad, S. (2011). On the use of multiple auxiliary variables in estimation of current population mean in two-occasion successive sampling. *Sri Lankan Journal of Applied Statistics, 12*, 101-116. doi: 10.4038/sljastats.v12i0.4970

Singh, G. N., & Prasad, S. (2010). Some estimator of population mean in two-occasion rotation patterns. *Association for the Advancement of Modelling & Simulation Techniques in Enterprises, 47*(2), 1-18.

Singh, G. N., & Prasad, S. (2013). Best linear unbiased estimators of population mean on current occasion in two-occasion successive sampling. *Statistics in Transition – New Series, 14*(1), 57-74.

Singh, G. N., & Priyanka, K. (2006). On the use of chain-type ratio to difference estimator in successive sampling. *International Journal of Applied Mathematics and Statistics, 5*(S06), 41-49.

Singh, G. N., & Priyanka, K. (2007). On the use of auxiliary information in search of good rotation patterns on successive occasions. *International Journal of Statistics and Economics, 1*(A07), 42-60.

Singh, G. N., & Priyanka, K. (2008). On the use of several auxiliary variates to improve the precision of estimates at current occasion. *Journal of the Indian Society of Agricultural Statistics, 62*(3), 253-265. Retrieved from http://isas.org.in/jisas/jsp/volume/vol62/7-G.N.Singh.pdf

Singh, G. N., & 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, G. N., & Singh, V. K. (2001). On the use of auxiliary information in successive sampling. *Journal of the Indian Society of Agricultural Statistics, 54*(1), 1-12. Retrieved from http://www.isas.org.in/jsp/volume/vol54/G.N.%20Singh%20.pdf

Singh, H. P., Tailor, R., Singh, S., & Kim, J.-M., (2011). Estimation of population variance in successive sampling. *Quality & Quantity, 45*(3), 477-494. doi: 10.1007/s11135-009-9309-9

Singh, H. P., & Vishwakarma, G. K. (2009). A general procedure for estimating population mean in successive sampling. *Communications in Statistics – Theory and Methods, 38*(2), 293-308. doi: 10.1080/03610920802169594

Singh, V. K., Singh, G. N., & Shukla, D. (1991). An efficient family of ratio-cum-difference type estimators in successive sampling over two occasions. *Journal of Scientific Research, 41*(C), 149-159.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., & Asok, C. (1984). *Sampling theory of surveys with applications*. Ames, IA: Iowa State University Press.