EFFICIENT FAMILY OF RATIO-TYPE ESTIMATORS FOR MEAN ESTIMATION IN SUCCESSIVE SAMPLING ON TWO OCCASIONS USING AUXILIARY INFORMATION

Nazeema T. Beevi[1], C. Chandran

ABSTRACT

In this paper, we proposed an efficient family of ratio-type estimators using one auxiliary variable for the estimation of the current population mean under successive sampling scheme. This family of estimators have been studied by Ray and Sahai (1980) under simple random sampling using one auxiliary variable for estimation of the population mean. Using these estimators in successive sampling, the expression for bias and mean squared error of the proposed estimators are obtained up to the first order of approximation. Usual ratio estimator is identified as a particular case of the suggested estimators. Optimum replacement strategy is also discussed. The proposed family of estimators at optimum condition is compared with the simple mean per unit estimator, Cochran (1977) estimator and existing other members of the family. Expressions of optimization are derived and results have been justified through numerical study interpretation.

Key words: auxiliary information, bias, mean square error, optimum replacement.

1. Introduction

Nowadays, it is often seen that sample surveys are not limited to one-time inquiries. A survey carried out on a finite population is subject to change over time if the value of a study character of a finite population is subject to (dynamic) change over time. A survey carried out on a single occasion will provide information about the characteristics of the surveyed population for the given occasion only and can not give any information on the nature or the rate of change of the characteristics over different occasions and the average value of the characteristics over all occasions or the most recent occasion. A part of the sample is retained being replaced for the next occasion (or sampling on successive occasions, which is also called successive sampling or rotation sampling).

The successive method of sampling consists of selecting sample units on different occasions such that some units are common with samples selected on previous

---

[1]Department of Statistics, University of Calicut, Kerala - 673 635, India.
E-mail: nazeemathaj@gmail.com.

occasions. If sampling on successive occasions is done according to a specific rule, with replacement of sampling units, it is known as successive sampling. Replacement policy was examined by Jessen (1942), who examined the problem of sampling on two occasions, without or with replacement of part of the sample in which what fraction of the sample on the first occasion should be replaced in order that the estimator of $\bar{Y}$ may have maximum precision. Yates (1949) extended Jessen's scheme to the situation where the population mean of a character is estimated on each of $(h > 2)$ occasions from a rotation sample design. These results were generalized by Patterson (1950) and Narain (1953), among others. Rao and Mudhdkar (1967) and Das (1982), used the information collected on the previous occasions for improving the current estimate. Data regarding changing properties of the population of cities or counties and unemployment statistics are collected regularly on a sample basis to estimate the changes from one occasion to the next or to estimate the average over a certain period. An important aspect of continuous surveys is the structure of the sample on each occasion. To meet these requirements, successive sampling provides a strong tool for generating the reliable estimates at different occasions.

Biradar and Singh (2001) proposed an estimator for the population mean on the second of two successive occasions on an auxiliary variate with an unknown population mean. Singh (2005) developed ratio estimators for the population mean on the current occasion using the information for all the units in successive sampling over two occasions. In many situations, information on an auxiliary variate may be readily available on the first as well as the second occasions; for example, tonnage (or seat capacity) of each vehicle or ship is known in survey sampling of transportation and the number of beds in hospital surveys.

Ray and Sahai (1980) have proposed two parameter family of estimators under SRSWOR scheme using auxiliary variate for estimating the population mean $\bar{Y}$ and assuming population mean $\bar{X}$ of auxiliary variate to be known. The objective of this paper is to develop a two parameter family of estimators that estimate the population mean on the current occasion in successive sampling using auxiliary variables on both occasions.

The paper is divided into nine sections. Sample structure and notations have been discussed in section 2 and section 3 respectively. In section 4, the proposed estimators have been formulated. Optimal choice and minimum mean square error of proposed estimator is derived in section 5 and section 6 respectively. In section 7, the optimum replacement policy is discussed and efficiency comparisons are made in section 8. In secection 9, the results have been justified through real data study with interpretation and we give interpretation and conclusion in section 10.

2. Selection of the sample

Consider a finite population $U = (U_1, U_2 \ldots U_N)$ which has been sampled over two occasions. Let x and y be the study variables on the first and second occasions respectively. We further assumed that the information on the auxiliary variable z_1 and z_2, whose population means are known, which is closely related to x and y on the first and second occasions respectively is available on the first as well as on the second occasion. For convenience, it is assumed that the population under consideration is large enough. Allowing SRSWOR (Simple Random Sampling without Replacement) design in each occasions, the successive sampling scheme is proposed as follows:

- We have n units which constitutes the sample on the first occasion. A random sub sample of $n_m = n\lambda$ $(0 < \lambda < 1)$ units is retained (matched) for use on the second occasion, where λ is the fraction of matched sample on the current occasion.

- In the second occasion $n_u = n\mu$ $(= n - n_m)$ $(0 < \mu < 1)$ units are drawn from the remaining $(N - n)$ units of the population, where μ is the fraction of fresh sample on the current occasion.

Therfore the sample size on the second occasion is also n $(= n\lambda + n\mu)$.

3. Description of Notations

We use the following notations in this paper.

$\bar{X}$: The population mean of the study variable on the first occasion.
$\bar{Y}$: The population mean of the study variable on the second occasion.
$\bar{Z}_1$: The population mean of the auxiliary variable on the first occasion.
$\bar{Z}_2$: The population mean of the auxiliary variable on the second occasion.
S_y^2: Population variance of y.
S_x^2: Population variance of x.
$S_{z_1}^2$: Population variance of z_1.
$S_{z_2}^2$: Population variance of z_2.
n_m: The sample size observed on the second occasion and common with the first occasion.
n_u: The sample size of the sample drawn afresh on the second occasion.
n: Total sample size.

$\bar{z}_n$: The sample mean of the auxiliary variable based on n units drawn on the first occasion.

$\bar{z}_{n_u}$: The sample mean of the auxiliary variable based on n_u units drawn on the second occasion.

$\bar{x}_n$: The sample mean of the study variable based on n units drawn on the first occasion.

$\bar{y}_{n_u}$: The sample mean of the study variable based on n_u units drawn afresh on the second occasion.

$\bar{y}_{n_m}$: The sample mean of the study variable based on n_m units common to both occasions and observed on the first occasion.

ρ_{yx}: The correlation coefficient between the variables y on x.

ρ_{xz_1}: The correlation coefficient between the variables x on z_1.

ρ_{yz_2}: The correlation coefficient between the variables y on z_2.

4. Proposed Family of Estimators in Successive sampling

Ray and Sahai (1980) have proposed a two parameter family of estimators under SRSWOR scheme using auxiliary variate for estimating the population mean $\bar{Y}$. The objective of this paper is to develop a two parameter family of estimators that estimate the population mean on the current occasion in successive sampling using auxiliary variables on the both occasions and it is known. To estimate the population mean $\bar{Y}$ on the second occasion, two different estimators are suggested. The first estimator is a family of estimators based on sample of size $n_u(=n\mu)$ drawn afresh on the second occasion given by:

$$ t_{n_u}^{(RK\theta)} \quad = \quad \bar{y}_{n_u} \left[\frac{K\bar{Z}_2 + \theta\bar{z}_2}{\bar{z}_2 + (K+\theta-1)\bar{Z}_2} \right] \tag{1} $$

The second estimator is a family of estimators based on the sample of size $n_m(=n\lambda)$ common with both the occasions defined as

$$ t_{n_m}^{(RK\theta)} \quad = \quad \bar{y}_{n_m} \left[\frac{K\bar{x}_n + \theta\bar{x}_{n_m}}{\bar{x}_{n_m} + (K+\theta-1)\bar{x}_n} \right] \left[\frac{K\bar{Z}_2 + \theta\bar{z}_2}{\bar{z}_2 + (K+\theta-1)\bar{Z}_2} \right], \tag{2} $$

where K is a non-negative constant and $0 \leq \theta \leq 1$. Combining the estimators t_{n_u} and t_{n_m}, we have the final estimator $t_{RK\theta}$ as follows

$$ t_{RK\theta} \quad = \quad \psi t_{n_u} + (1-\psi)t_{n_m}, \tag{3} $$

where ψ is an unknown constant to be determined such that $MSE(t_{RK\theta})$ is minimum. We prove theoretically that the estimator is more efficient than the proposed

estimator by Cochran (1977) when no auxiliary variables are used at any occasion. Cochran's classical difference estimator is a widely used estimator to estimate the population mean $\bar{Y}$, in successive sampling. It is given by

$$\bar{y}_2' = \phi_2 \bar{y}_{2u}' + (1 - \phi_2)\bar{y}_{2m}',$$

where ϕ_2 is an unknown constant to be determined such that $V(\hat{\bar{Y}})_{opt}$ is minimum and $\bar{y_{2u}}' = \bar{y}_{2u}$ is the sample mean of the unmatched portion of the sample on the second occasion and $\bar{y}_{2m}' = \bar{y}_{2m} + b(\bar{y}_1 - \bar{y}_{1m})$ is based on matched portion. The variance for large N of this estimator is

$$Var(\hat{\bar{Y}})_{opt} = [1 + \sqrt{(1 - \rho^2)}]\frac{S_y^2}{2n}.$$

Similarly, the variance for large N of the mean per unit estimator is given by

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

4.1. Properties of $t_{RK\theta}$

Since t_{n_u} and t_{n_m} both are biased estimators of $t_{RK\theta}$, therefore, resulting estimator $t_{RK\theta}$ is also a biased estimator. The bias and *MSE* up to the first order of approximation are derived as using large sample approximation given below:

$$\bar{y}_{n_u} = \bar{Y}(1 + e_{\bar{y}_{n_u}}), \quad \bar{y}_{n_m} = \bar{Y}(1 + e_{\bar{y}_{n_m}}),$$
$$\bar{x}_{n_m} = \bar{X}(1 + e_{\bar{x}_{n_m}}), \quad \bar{x}_n = \bar{X}(1 + e_{\bar{x}_n}),$$
$$\bar{z}_1 = \bar{Z}_1(1 + e_{\bar{z}_1}), \quad \bar{z}_2 = \bar{Z}_2(1 + e_{\bar{z}_2})$$

where $e_{\bar{y}_{n_u}}, e_{\bar{y}_{n_m}}, e_{\bar{x}_{n_m}}, e_{\bar{x}_n}, e_{\bar{z}_1}$ and $e_{\bar{z}_2}$ are sampling errors and they are of very small quantities. We assume that

$E(e_{\bar{y}_{n_u}}) = E(e_{\bar{y}_{n_m}}) = E(e_{\bar{x}_{n_m}}) = E(e_{\bar{x}_n}) = E(e_{\bar{z}_1}) = E(e_{\bar{z}_2}) = 0$. Then, for simple random sampling without replacement for both the first and second occasions, we write using the occasion wise operation of expectation as:

$$E(e_{\bar{y}_{n_u}}^2) = \left(\frac{1}{n_u} - \frac{1}{N}\right)S_y^2, E(e_{\bar{y}_{n_m}}^2) = \left(\frac{1}{n_m} - \frac{1}{N}\right)S_y^2,$$
$$E(e_{\bar{x}_{n_m}}^2) = \left(\frac{1}{n_m} - \frac{1}{n}\right)S_x^2, E(e_{\bar{x}_n}^2) = \left(\frac{1}{n} - \frac{1}{N}\right)S_x^2,$$
$$E(e_{\bar{z}_1}^2) = \left(\frac{1}{m} - \frac{1}{n}\right)S_{z_2}^2,$$

$$E(e_{\bar{y}_{n_u}} e_{\bar{z}_2}) = \left(\frac{1}{n_u} - \frac{1}{N}\right) S_{yz_2}, \qquad E(e_{\bar{y}_{n_m}} e_{\bar{x}_n}) = \left(\frac{1}{n_m} - \frac{1}{n}\right) S_{yx},$$

$$E(e_{\bar{y}_{n_m}} e_{\bar{x}_{n_m}}) = \left(\frac{1}{n_m} - \frac{1}{n}\right) S_{yx}, \qquad E(e_{\bar{y}_{n_m}} e_{\bar{x}_n}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_{yx},$$

$$E(e_{\bar{y}_{n_m}} e_{\bar{z}_1}) = \left(\frac{1}{m} - \frac{1}{n}\right) S_{yz_1}, \qquad E(e_{\bar{x}_{n_m}} e_{\bar{x}_n}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_x^2,$$

$$E(e_{\bar{x}_{n_m}} e_{\bar{z}_1}) = \left(\frac{1}{m} - \frac{1}{n}\right) S_{xz_1}, \qquad E(e_{\bar{x}_n} e_{\bar{z}_1}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_{xz_1}$$

$$E(e_{\bar{z}_1} e_{\bar{z}_2}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_z^2.$$

We derive the bias of t_{n_u} and t_{n_m} in lemma 4.3 and lemma 4.5 respectively.

4.2. Lemma

The bias of t_{n_u} denoted by $B(t_{n_u})$ is given by

$B(t_{n_u}) = \bar{Y} \left(\frac{1}{n_u} - \frac{1}{N}\right) \left[Q \left(S_{z_2}^2 (K + \theta)^{-1} - \rho_{yz_2} S_y S_{z2}\right)\right],$

where $Q = \left[\frac{1-\theta}{K+\theta}\right]$, K is a non-negative constant and $0 \le \theta \le 1$.

Expressing (1) in terms of e's, we have

$$t_{n_u} = \bar{Y}(1 + e_{\bar{y}_{n_u}}) \left[\frac{K\bar{Z}_2 + \theta(1 + e_{\bar{z}_2})\bar{Z}_2}{\bar{Z}_2(1 + e_{\bar{z}_2}) + (K + \theta - 1)\bar{Z}_2}\right],$$

$$t_{n_u} = \bar{Y}(1 + e_{\bar{y}_{n_u}}) \left[1 + \frac{\theta}{a} e_{\bar{z}_2}\right] \left[1 + \frac{e_{\bar{z}_2}}{a}\right]^{-1}, \tag{4}$$

where $a = K + \theta$.

Taking expectation of (4) on both sides, we get

$$E(t_{n_u} - \bar{Y}) = \bar{Y} E\left(e_{\bar{y}_{n_u}} + \frac{\theta}{a} e_{\bar{z}_2} - \frac{e_{\bar{z}_2}}{a} + \frac{e_{\bar{z}_2}^2}{a^2} - \frac{\theta}{a^2} e_{\bar{z}_2}^2 + \frac{\theta}{a} e_{\bar{y}_{n_u}} e_{\bar{z}_2} - \frac{e_{\bar{y}_{n_u}} e_{\bar{z}_2}}{a}\right)$$

$$B(t_{n_u}) = \bar{Y} \left(\frac{1}{n_u} - \frac{1}{N}\right) \left[Q \left(S_{z_2}^2 (K + \theta)^{-1} - \rho_{yz_2} S_y S_{z2}\right)\right],$$

where $Q = \left[\frac{1-\theta}{K+\theta}\right]$, K is a non-negative constant and $0 \le \theta \le 1$.

4.3. Remark

The bias of estimator $B(t_{n_u})$ is the same as the bias of $B(t_{RS})$ proposed by Ray and Sahai (1980).

4.4. Lemma

The bias of t_{n_m} denoted by $B(t_{n_m})$ is given by

$$B(t_{n_m}) = \bar{Y}\left[\left(\frac{1}{n_m} - \frac{1}{n}\right)\left[Q\left(S_x^2(K+\theta)^{-1} - \rho_{yx}S_yS_x\right)\right] \right.$$
$$\left. + \left(\frac{1}{n} - \frac{1}{N}\right)\left[Q\left(S_{z_1}^2(K+\theta)^{-1} - \rho_{yz_1}S_yS_{z_1}\right)\right]\right],$$

where $Q = \left[\frac{1-\theta}{K+\theta}\right]$, K is a non-negative constant and $0 \leq \theta \leq 1$.
Expressing (2) in terms of e's, we have

$$t_{n_m} = \bar{Y}(1+e_{\bar{y}_{n_m}})\left[\frac{K(1+e_{\bar{x}_n})\bar{X} + \theta(1+e_{\bar{x}_{n_m}})\bar{X}}{(1+e_{\bar{x}_{n_m}})\bar{X} + (K+\theta-1)(1+e_{\bar{x}_n})\bar{X}}\right]$$

$$\left[\frac{K\bar{Z}_1 + \theta(1+e_{\bar{z}_1})\bar{Z}_1}{(1+e_{\bar{z}_1})\bar{Z}_1 + (K+\theta-1)\bar{Z}_1}\right],$$

$$= \bar{Y}(1+e_{\bar{y}_{n_m}})\left[\frac{a + Ke_{\bar{x}_{n_m}} + \theta e_{\bar{x}_{n_m}}}{a + (a-1)e_{\bar{x}_n} + e_{\bar{x}_{n_m}}}\right]\left[\frac{a + \theta e_{\bar{z}_1}}{a + e_{\bar{z}_1}}\right],$$

$$= \bar{Y}(1+e_{\bar{y}_{n_m}})\left[\frac{1 + \frac{K}{a}e_{\bar{x}_{n_m}} + \frac{\theta}{a}e_{\bar{x}_{n_m}}}{1 + \frac{a-1}{a}e_{\bar{x}_n} + \frac{e_{\bar{x}_{n_m}}}{a}}\right]\left[\frac{1 + \frac{\theta}{a}e_{\bar{z}_1}}{1 + \frac{e_{\bar{z}_1}}{a}}\right],$$

$$t_{n_m} = \bar{Y}(1+e_{\bar{y}_{n_m}})\left[1 + \frac{K}{a}e_{\bar{x}_n} + \frac{\theta}{a}e_{\bar{x}_{n_m}}\right]$$

$$\left[1 + \frac{a-1}{a}e_{\bar{x}_n} + \frac{e_{\bar{x}_{n_m}}}{a}\right]^{-1}[1 + \frac{\theta}{a}e_{\bar{z}_1}][1 + \frac{e_{\bar{z}_1}}{a}]^{-1}. \tag{5}$$

Expanding (5) the right hand side and neglecting higher terms, we get

$$t_{n_m} = \bar{Y}\left[\left(1 + \frac{K}{a}e_{\bar{x}_n} + \frac{\theta}{a}e_{\bar{x}_{n_m}} - \frac{e_{\bar{x}_{n_m}}}{a} - \frac{a-1}{a}e_{\bar{x}_n}\right.\right.$$

$$+\frac{e_{\bar{x}_{n_m}}^2}{a^2}+2\frac{a-1}{a^2}e_{\bar{x}_{n_m}}e_{\bar{x}_n}+(\frac{a-1}{a})^2-\frac{K}{a^2}e_{\bar{x}_{n_m}}e_{\bar{x}_n}-K\frac{a-1}{a^2}e_{\bar{x}_n}^2$$

$$-\frac{\theta}{a^2}e_{\bar{x}_{n_m}}^2-\frac{a-1}{a^2}\theta e_{\bar{x}_n}e_{\bar{x}_{n_m}}\Bigg)\left(1+\frac{\theta}{a}e_{\bar{z}_1}-\frac{e_{\bar{z}_1}}{a}+\frac{e_{\bar{z}_1}^2}{a^2}-\frac{\theta}{a^2}e_{\bar{z}_1}^2\right)\Bigg]. \qquad (6)$$

Taking expectation of (6) on both sides, we get

$$E(t_{n_m}-\bar{Y})=\bar{Y}\Bigg[\left(\frac{K}{a}e_{\bar{x}_n}+\frac{\theta}{a}e_{\bar{x}_{n_m}}-\frac{e_{\bar{x}_{n_m}}}{a}-\frac{a-1}{a}e_{\bar{x}_n}\right.$$

$$+\frac{e_{\bar{x}_{n_m}}^2}{a^2}+2\frac{a-1}{a^2}e_{\bar{x}_{n_m}}e_{\bar{x}_n}+(\frac{a-1}{a})^2-\frac{K}{a^2}e_{\bar{x}_{n_m}}e_{\bar{x}_n}-K\frac{a-1}{a^2}e_{\bar{x}_n}^2$$

$$-\frac{\theta}{a^2}e_{\bar{x}_{n_m}}^2-\frac{a-1}{a^2}\theta e_{\bar{x}_n}e_{\bar{x}_{n_m}}\Bigg)\left(1+\frac{\theta}{a}e_{\bar{z}_1}-\frac{e_{\bar{z}_1}}{a}+\frac{e_{\bar{z}_1}^2}{a^2}-\frac{\theta}{a^2}e_{\bar{z}_1^2}\right)\Bigg]$$

$$B(t_{n_m})=\bar{Y}\Bigg[\left(\frac{1}{n_m}-\frac{1}{n}\right)\left[Q\left(S_x^2(K+\theta)^{-1}-\rho_{yx}S_yS_x\right)\right]$$

$$+\left(\frac{1}{n}-\frac{1}{N}\right)\left[Q\left(S_{z_1}^2(K+\theta)^{-1}-\rho_{yz_1}S_yS_{z_1}\right)\right]\Bigg],$$

where $Q=\left[\frac{1-\theta}{K+\theta}\right]$. K is a non-negative constant and $0\leq\theta\leq1$.
Using lemma 4.3 and 4.5, the bias of the estimator $t_{RK\theta}$ can be derived as follows.

4.5. Theorem

The bias of the estimator $t_{RK\theta}$ to the first order approximation is,

$$B(t_{RK\theta})\quad=\quad\psi B(t_{n_u})+(1-\psi)B(t_{n_m}), \qquad (7)$$

where

$$B(t_{n_u})=\bar{Y}\left(\frac{1}{n_u}-\frac{1}{N}\right)\left[Q\left(S_{z_2}^2(K+\theta)^{-1}-\rho_{yz}S_yS_{z_2}\right)\right],$$

and

$$B(t_{n_m}) = \bar{Y}\left[\left(\frac{1}{n_m} - \frac{1}{n}\right)\left[Q\left(S_x^2(K+\theta)^{-1} - \rho_{yx}S_yS_x\right)\right]\right.$$
$$\left. + \left(\frac{1}{n} - \frac{1}{N}\right)\left[Q\left(S_{z_2}^2(K+\theta)^{-1} - \rho_{yz}S_yS_{z_2}\right)\right]\right],$$

where
$Q = \left[\frac{1-\theta}{K+\theta}\right]$, K is a non-negative constant and $0 \le \theta \le 1$.
The bias of the estimator $t_{RK\theta}$ is given by

$$B(t_{RK\theta}) = E(t_{RK\theta} - \bar{Y})$$

$$B(t_{RK\theta}) = \psi E(t_{n_u} - \bar{Y}) + (1-\psi)E(t_{n_m} - \bar{Y}), \tag{8}$$

Using lemmas 4.3 and 4.5 into equation (8), we have the expression for the bias of the estimator $t_{RK\theta}$ as shown in (7).

We derive the MSE of t_{n_u} and t_{n_m} in lemma 4.7 and lemma 4.9 respectively.

4.6. Lemma

The mean square error of t_{n_u} denoted by $M(t_{n_u})$ is given by

$$MSE(t_{n_u}) = \bar{Y}^2\left(\frac{1}{n_u} - \frac{1}{N}\right)\left[S_y^2 + Q^2S_{z_2}^2 - 2Q\rho_{yz}S_yS_{z_2}\right],$$

where $Q\left[= \frac{1-\theta}{K+\theta}\right]$, $0 \le \theta \le 1$.
Expressing (2) in terms of e's, we have

$$t_{n_u} = \bar{Y}(1+e_{\bar{y}_{n_u}})\left[\frac{K\bar{Z}_2 + \theta(1+e_{\bar{z}_2})\bar{Z}_2}{\bar{Z}_2(1+e_{\bar{z}_2}) + (K+\theta-1)\bar{Z}_2}\right], \tag{9}$$

where $Q = \left[\frac{1-\theta}{K+\theta}\right]$.

$$t_{n_u} = \bar{Y}(1+e_{\bar{y}_{n_u}})\left[K\bar{Z}_2 + \theta(1+e_{\bar{z}_2})\bar{Z}_2\right]$$

$$\left[(1+e_{\bar{z}_2})\bar{Z}_2 + (K+\theta-1)\bar{Z}_2\right]^{-1},$$

$$= \bar{Y}(1+e_{\bar{y}_{n_u}})[a+\theta e_{\bar{z}_2}][a+e_{\bar{z}_2}]^{-1},$$

$$= \bar{Y}(1+e_{\bar{y}_{n_u}})[a+\theta e_{\bar{z}_2}][a-e_{\bar{z}_2}]. \tag{10}$$

Squaring (10) and expectation, the right hand side and neglecting the terms with power two or greater, we get

$$E(t_{n_u}-\bar{Y})^2 = \bar{Y}^2 E\left[e_{\bar{y}_{n_u}}+\frac{\theta}{a}e_{\bar{z}_2}-\frac{e_{\bar{z}_2}}{a}\right]^2.$$

$$MSE(t_{n_u}) = \bar{Y}^2\left(\frac{1}{n_u}-\frac{1}{N}\right)\left[S_y^2+Q^2S_{z_2}^2-2Q\rho_{yz_2}S_yS_{z_2}\right], \tag{11}$$

where $Q = \left[\frac{1-\theta}{K+\theta}\right], 0 \le \theta \le 1$.

4.7. Remark

The mean square of estimator $MSE(t_{n_u})$ is the same as the mean square of $MSE(t_{RS})$ proposed by Ray and Sahai (1980).

4.8. Lemma

The mean square error of t_{n_m} denoted by $MSE(t_{n_m})$ is given by

$$MSE(t_{n_m}) = \bar{Y}^2\left[\left(\frac{1}{n_m}-\frac{1}{N}\right)S_y^2+\left(\frac{1}{n_m}-\frac{1}{n}\right)\left[Q^2S_x^2-2Q\rho_{yx}S_yS_x\right]+\right.$$
$$\left.\left(\frac{1}{n}-\frac{1}{N}\right)\left[Q^2S_{z_1}^2-2Q\rho_{yz_1}S_yS_{z_1}\right]\right],$$

where $Q = \left[\frac{1-\theta}{K+\theta}\right], 0 \le \theta \le 1$.
Expressing (2) in terms of e's, we have

$$t_{n_m} = \bar{Y}(1+e_{\bar{y}_{n_m}})\left[\frac{K(1+e_{\bar{x}_n})\bar{X}+\theta(1+e_{\bar{x}_{n_m}})\bar{X}}{(1+e_{\bar{x}_{n_m}})\bar{X}+(K+\theta-1)(1+e_{\bar{x}_n})\bar{X}}\right]$$

$$\left[\frac{K\bar{Z}+\theta(1+e_{\bar{z}_1})\bar{Z}}{(1+e_{\bar{z}_{1n}})\bar{Z}+(K+\theta-1)\bar{Z}}\right],$$

$$= \bar{Y}(1 + e_{\bar{y}_{n_m}}) \left[\frac{a + K e_{\bar{x}_n} + \theta e_{x_{n_m}}}{a + (a-1)e_{\bar{x}_n} + e_{x_{n_m}}} \right] \left[\frac{a + \theta e_{\bar{z}_1}}{a + e_{\bar{z}_1}} \right],$$

$$= \bar{Y}(1 + e_{\bar{y}_{n_m}}) \left[\frac{1 + \frac{K}{a} e_{\bar{x}_{n_m}} + \frac{\theta}{a} e_{\bar{x}_{n_m}}}{1 + \frac{a-1}{a} e_{\bar{x}_n} + \frac{e_{\bar{x}_{n_m}}}{a}} \right] \left[\frac{1 + \frac{\theta}{a} e_{\bar{z}_1}}{1 + \frac{e_{\bar{z}_1}}{a}} \right],$$

where
$$Q = \frac{1-\theta}{K+\theta}.$$

$$t_{n_m} = \bar{Y}(1 + e_{\bar{y}_{n_m}}) \left[1 + \frac{K}{a} e_{\bar{x}_n} + \frac{\theta}{a} e_{\bar{x}_{n_m}} \right]$$

$$\left[1 - \frac{a-1}{a} e_{\bar{x}_n} - \frac{e_{\bar{x}_{n_m}}}{a} \right] [1 + \frac{\theta}{a} e_{\bar{z}_1}][1 - \frac{e_{\bar{z}_1}}{a}]. \tag{12}$$

Expanding (12) to the right hand side and neglecting the higher terms, we get

$$t_{n_m} = \bar{Y} \left[1 + e_{\bar{y}_{n_m}} + Q e_{\bar{x}_n} - Q e_{\bar{x}_{n_m}} - Q e_{\bar{y}_n} \right]. \tag{13}$$

Squaring and taking expectation (13), we get *MSE* of the estimator t_{n_m} up to first order of approximation as,

$$E(t_{n_m} - \bar{Y})^2 = \bar{Y} E \left[1 + e_{\bar{y}_{n_m}} + Q e_{\bar{x}_n} - Q e_{\bar{x}_{n_m}} - Q e_{\bar{z}_n} \right]^2, \tag{14}$$

$$= \bar{Y}^2 E[e_{\bar{y}_{n_m}}^2 + Q^2 e_{\bar{x}_n}^2 + Q^2 e_{\bar{x}_{n_m}}^2 + Q^2 e_{\bar{z}_1}^2 + 2Q e_{\bar{y}_{n_m}} e_{\bar{x}_n}$$

$$-2Q e_{\bar{y}_{n_m}} e_{\bar{x}_{n_m}} - 2Q e_{\bar{y}_{n_m}} e_{\bar{z}_n} - 2Q^2 e_{\bar{x}_n} e_{\bar{y}_{n_m}} - 2Q^2 e_{\bar{x}_n} e_{\bar{z}_1} + 2Q^2 e_{\bar{x}_{n_m}} e_{\bar{z}_1}]$$

$$MSE(t_{n_m}) = \bar{Y}^2 \left[\left(\frac{1}{n_m} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{n_m} - \frac{1}{n} \right) [Q^2 S_x^2 - 2Q\rho_{yx} S_y S_x)] \right.$$

$$\left. + \left(\frac{1}{n} - \frac{1}{N} \right) [Q^2 S_{z_1}^2 - 2Q\rho_{yz} S_y S_{z_1}] \right], \tag{15}$$

where $Q = \left[\frac{1-\theta}{K+\theta} \right], 0 \le \theta \le 1$.
Using lemma 4.7 and lemma 4.9, we derive the MSE of $t_{RK\theta}$.

4.9. Theorem

The estimators of t_{n_u} and t_{n_m} are based on two independent samples of sizes n_u and n_m respectively. The mean square error of the estimator $t_{RK\theta}$ to the first order approximation is,

$$MSE(t_{RK\theta}) = \psi^2 MSE(t_{n_u}) + (1 - \psi)^2 MSE(t_{n_m}), \qquad (16)$$

where

$$MSE(t_{n_u}) \quad = \quad \bar{Y}^2 \left(\frac{1}{n_u} - \frac{1}{N} \right) \left[S_y^2 + Q^2 S_{z_2}^2 - 2Q\rho_{yz_2} S_y S_{z_2} \right],$$

and

$$MSE(t_{n_m}) \quad = \quad \bar{Y}^2 \left[\left(\frac{1}{n_m} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{n_m} - \frac{1}{n} \right) \left[Q^2 S_x^2 - 2Q\rho_{yx} S_y S_x \right] + \right.$$

$$\left. \left(\frac{1}{n} - \frac{1}{N} \right) \left[Q^2 S_{z_1}^2 - 2Q\rho_{yz_1} S_y S_{z_1} \right] \right],$$

where $Q = \left[\frac{1-\theta}{K+\theta} \right]$, $0 \le \theta \le 1$.
The mean square error of the estimator $t_{RK\theta}$ is given by

$$MSE(t_{RK\theta}) = E(t_{RK\theta} - \bar{Y})^2$$

$$MSE(t_{RK\theta}) = E[\psi(t_{n_u} - \bar{Y}) + (1 - \psi)(t_{n_m} - \bar{Y})]^2, \qquad (17)$$

$$MSE(t_{RK\theta}) = \psi^2 MSE(t_{n_u}) + (1 - \psi)^2 MSE(t_{n_m}),$$

using lemma 4.7 and lemma 4.9 into the equation (17), we have the expression for the *MSE* of the estimator $t_{RK\theta}$ as shown in (16).

5. Optimal Choice for Proposed Family of Ratio-Type Estimators $t_{RK\theta}$

The mean square error of the $t_{RK\theta}$ is a function of Q and also the function of K and θ. We minimize for K and θ. We have $\frac{\partial(MSE(t_{RK\theta}))}{\partial K} = 0$ and $\frac{\partial(MSE(t_{RK\theta}))}{\partial \theta} = 0$. This gives $Q = -\frac{U}{V}$, assuming $\frac{\partial Q}{\partial K} \neq 0$ and $\frac{\partial Q}{\partial \theta} \neq 0$. The differential equations are given

below:

$$\frac{\partial (MSE(t_{KR\theta}))}{\partial K} = 0 \Rightarrow \psi^2 \frac{\partial MSE(t_{n_u})}{\partial K} + (1-\psi)^2 \frac{\partial M(t_{n_m})}{\partial K} = 0, \qquad (18)$$

$$\frac{\partial (MSE(t_{KR\theta}))}{\partial \theta} = 0 \Rightarrow \psi^2 \frac{\partial MSE(t_{n_u})}{\partial \theta} + (1-\psi)^2 \frac{\partial MSE(t_{n_m})}{\partial \theta} = 0, \qquad (19)$$

where ψ is an unknown constant. From (18) and (19), differentiate with respect to K and θ, we get

$$\hat{K} = \frac{1 - \theta(1 + \rho_{yz_1} \frac{S_y}{S_{z_1}})}{\rho_{yz_1} \frac{S_y}{S_{z_1}}}, \quad \theta \quad fixed.$$

$$\hat{\theta} = \frac{1 - K\rho_{yz_1} \frac{S_y}{S_{z_1}}}{1 + \rho_{yz_1} \frac{S_y}{S_{z_1}}}, \qquad K \quad fixed.$$

6. Minimum Mean Square Error of $t_{RK\theta}$

To get the optimum value of ψ, we partially differentiate the expression (16) with respect to ψ, and put it equal to zero and we get

$$\psi_{opt} = \frac{MSE(t_{n_m})}{MSE(t_{n_u}) + MSE(t_{n_m})}. \qquad (20)$$

Substituting the values of $MSE(t_{n_u})_{opt}$ and $MSE(t_{n_m})_{opt}$ from (11) and (15) in (20), we get

$$\begin{aligned}
\psi_{opt} &= \frac{(p_1 + \mu p_2)}{(p_1 + \mu^2 p_2)} \\
&= \frac{\mu[(p_1 + \mu p_2)]}{(p_1 + \mu^2 p_2)}.
\end{aligned}$$

Substitution of ψ_{opt} from (20) into (16) gives optimum value of MSE of $t_{RK\theta}$ as:

$$MSE(t_{RK\theta})_{opt} = \frac{MSE(t_{n_m}) MSE(t_{n_u})}{MSE(t_{n_u}) + MSE(t_{n_m})}. \qquad (21)$$

Substituting the values of $MSE(t_{n_u})$ and $MSE(t_{n_m})$ from (11) and (15) in (21), we get

$$MSE(t_{RK\theta})_{opt} = \frac{1}{n}\left[\frac{p_1^2 + \mu p_1 p_2}{p_1 + \mu^2 p_2}\right], \tag{22}$$

where $p_1 = q_1 + q_3$, $p_2 = q_2 - q_3$, $q_1 = S_y^2$, $q_2 = Q^2 - 2Q\rho_{yx}S_yS_x$, $q_3 = Q^2 - 2Q\rho_{yz_1}S_yS_{z_1}$, and $\mu = \frac{n_u}{n}$.

7. Replacement Policy of $t_{RK\theta}$

In order to estimate $t_{RK\theta}$ with maximum precision an optimum value of μ should be determined so as to know what fraction of the sample on the first occasion should be replaced. We minimize, $MSE(t_{RK\theta})_{opt}$ in (22) with respect to μ, the optimum value of μ is obtained as,

$$\hat{\mu} = \frac{-p_1 \pm \sqrt{p_1^2 + p_1 p_2}}{p_2}, \tag{23}$$

where $p_1 = q_1 + q_3$, $p_2 = q_2 - q_3$. From (23) it is obvious that for $\rho_{yz_1} \neq \rho_{yx}$ two values of $\hat{\mu}$ are possible, therefore to choose the value of $\hat{\mu}$, it should be remembered that $0 \leq \hat{\mu} \leq 1$. All other values of $\hat{\mu}$ are inadmissible. If both the real values of $\hat{\mu}$ are admissible, the lowest one will be the best choice as it reduces the total cost of the survey. Substituting the value of $\hat{\mu}$ from (23) in (22), we get

$$MSE(t_{\hat{R}\hat{K}\theta})_{opt} = \frac{1}{n}\left[\frac{p_1^2 + \hat{\mu} p_1 p_2}{p_1 + \hat{\mu}^2 p_2}\right], \tag{24}$$

μ is the fraction of fresh sample drawn on the current occasion. To estimate the population mean on each occasion 1 is a better choice of μ. However, to estimate the change in the mean from one occasion to the other, μ should be 0.

8. Efficiency Comparisons

In this section, to compare $t_{RK\theta}$ with respect to $\bar{y}$, sample mean of y, when a sample units are selected at second occasion without any matched portion. Since $\bar{y}$ and is unbiased estimators of $\bar{Y}$, its variance for large N is respectively given by

$$Var(\bar{y}) = \frac{S_y^2}{n},$$

The variance of Cochran's (1977) estimator is

$$Var(\hat{\bar{Y}})_{opt} = [1 + \sqrt{1 - \rho_{yx}^2}]\frac{S_y^2}{2n},$$

the mean square error of ratio estimator is

$$MSE(\bar{y}_R)_{opt} = \frac{1}{n}\bar{Y}^2[S_y^2 + S_x^2 - 2\rho_{yx}S_yS_x].$$

and the variance of Biradar and Singh's estimator is

$$V(\hat{\bar{Y}}_2)_{opt} = \frac{S_{y_2}^2}{n_1}[\frac{(1 - \lambda\rho_{x_2.y_2}^2)(1 - \mu R_{y_2.y_1x_1x_2}^2)}{1 - \mu^2 R_{y_2.y_1x_1x_2}^2 - \lambda^2\rho_{x_2.y_2}^2}].$$

The relative efficiencies of $t_{RK\theta}$ with respect to $\bar{y}$, $\hat{\bar{Y}}$ and $\bar{y}_R$ are given by

$$R_1 = \frac{Var(\bar{y})}{MSE(t_{RK\theta})_{opt}} \times 100,$$

$$R_2 = \frac{Var(\hat{\bar{Y}})_{opt}}{MSE(t_{RK\theta})_{opt}} \times 100$$

$$R_3 = \frac{MSE(\bar{y}_R)_{opt}}{MSE(t_{RK\theta})_{opt}} \times 100,$$

and

$$R_4 = \frac{V(\hat{\bar{Y}}_2)_{opt}}{MSE(t_{RK\theta})_{opt}} \times 100$$

the estimator $t_{RK\theta}$ (at optimal conditions) is also compared with respect to the estimator $V(\bar{y})$, where

$$MSE(t_{RK\theta})_{opt} = \frac{1}{n}\left[\frac{p_1^2 + \hat{\mu}p_1p_2}{p_1 + \hat{\mu}^2p_2}\right] \tag{25}$$

and

$$\hat{\mu} = \frac{-p_1 \pm \sqrt{p_1^2 + p_1p_2}}{p_2}, \tag{26}$$

where $p_1 = q_1 + q_3$, $p_2 = q_2 - q_3$.

8.1. Numerical Illustration

The results obtained in previous sections are examined with the help of two natural population sets of data.

Population Source:1 [Free access to the data from the Statistical Abstracts of the United States ($https://www.census.gov/history/www/reference/publications/statistical_abstracts.html$).]

The first population presents $N = 41$ states of the United States. Let y_i (study variable on the second occasion) be the corn production (in million bushels) during 2009 in the ith state of the United States, x_i (study variable on the first occasion) be the corn production (in million bushels) during 2008 in the ith state of the United States and z_i (auxiliary variable) be the corn production (in million bushels) during 2006 in the ith state of the United States. Consider the population, $N = 41$ states of sample size $n = 30$ states were selected using SRSWOR in the year 2008. From each one of the sample drawn (matched samples), $n_m = 25$ states were retained and fresh sample (unmatched samples) $n_u = 5$ states were selected from the remaining $N - n = 41 - 30 = 11$ states using SRSWOR in the year 2009. The results in Table 6.9. show the comparison of the suggested estimator $t_{RK\theta}$ with respect to the estimators $\bar{y}$, $\hat{\bar{Y}}$ and $\bar{y}_R$ respectively for different selection of matched and unmatched samples. For convenience, the different selections of n_m and n_u are considered as different sets in the population, which are given below:

Population Source 2 [Free access data from the data.gov.in.(https://data.gov.in/)]

The population presents, $N = 28$ states in India. Let y_i (study variable on the second occasion) be the infant mortality rate (per 1000 live births) during 2011 in the ith state in India, x_i (study variable on the first occasion) be the infant mortality rate (per 1000 live births) during 2010 in the ith state in India and z_i (auxiliary variable) be the infant mortality rate (per 1000 live births) during 2009 in the ith state in India. Consider the population, $N = 28$ states of sample size $n = 20$ states were selected using SRSWOR in the year 2010. From each one of the sample drawn (matched samples), $n_m = 17$ states were retained and fresh sample (unmatched samples) $n_u = 3$ states were selected from remaining $N - n = 28 - 20 = 8$ states using SRSWOR in the year 2011. The results in Table 6.10. show the comparison of the suggested estimator $t_{RK\theta}$ with respect to the estimators $\bar{y}$, $\hat{\bar{Y}}$ and $\bar{y}_R$ respectively for different selection of matched and unmatched samples. For convenience, the different selections of n_m and n_u are considered as different sets in the population, which are given below:

	$Bias(t_{R\hat{K}\hat{\theta}})$	MSE/Variance	Relative Efficiencies of $t_{RK\theta}$ with respect to $\bar{y}, \hat{\bar{Y}}, \bar{y}_R$ and $\hat{\bar{Y}}_2$
$n=30, n_m=25, n_u=5,$ $\hat{K}=1.5, \hat{\theta}=0.33$	1.0031	$MSE(t_{R\hat{K}\hat{\theta}})$=0.1197 $V(\bar{y})$= 35.03 $V(\hat{\bar{Y}})_{opt}$= 33.97 $MSE(\bar{y}_R)_{opt}$= 32.12 $V(\hat{\bar{Y}}_2)_{opt}$ = 15.02	R_1 =292.64 R_2 =283.79 R_3 =268.33 R_4 =125.73
$n=30, n_m=20, n_u=10$ $\hat{K}=2.31, \hat{\theta}=0.17$	3.7410	$MSE(t_{R\hat{K}\hat{\theta}})$=0.2489 $V(\bar{y})$= 39.61 $V(\hat{\bar{Y}})_{opt}$= 37.33 $MSE(\bar{y}_R)_{opt}$= 36.19 $V(\hat{\bar{Y}}_2)_{opt}$ = 17.14	R_1 =159.14 R_2 =148.77 R_3 =145.39 R_4 =68.89
$n=30, n_m=17, n_u=13$ $\hat{K}=2.84, \hat{\theta}=0.12$	3.9971	$MSE(t_{R\hat{K}\hat{\theta}})$=0.3141 $V(\bar{y})$= 40.73 $V(\hat{\bar{Y}})_{opt}$= 39.88 $MSE(\bar{y}_R)_{opt}$= 38.64 $V(\hat{\bar{Y}}_2)_{opt}$ = 19.25	R_1 =129.67 R_3 =126.96 R_3 =123.01 R_4 =61.29

Table 8.1. The suggested estimators, $t_{RK\theta}$ is compared to $\bar{y}, \hat{\bar{Y}}, \bar{y}_R$ and $\hat{\bar{Y}}_2$ for the population by using Real data results

9. Interpretations of Empirical Results of $t_{RK\theta}$

The following conclusions can be made from Table 8.1.

1. The values of R_1, R_2, R_3 and R_4 are increasing while K is increasing with the decreasing values of θ for fixed values of sample size n.

2. The values of R_1, R_2, R_3 and μ are increasing with increasing values of ρ_{yz}. The behaviour indicates that an agreement with the Sukhatme et.al (1984) results, which explains that more the value of ρ_{yx}, more the fraction of fresh sample is required at the second (current) occasion.

3. For the fixed values of ρ_{yx}, ρ_{yz}, $\hat{K}$ and $\hat{\theta}$ there is an appreciable gain in the performance of the proposed estimator $t_{RK\theta}$ over $\bar{y}$ and $\hat{\bar{Y}}$ with the increasing value of μ.

10. Conclusions

Tables 8.1. clearly indicates that the proposed estimators is more efficient than the simple arithmetic mean estimator, Cochran (1977) estimator and ratio estimator. The following conclusion can be drawn from Tables 8.1. For fixed K, θ, the values of R_1, R_2, R_3 and R_4 are increasing. This phenomenon indicates that a smaller fresh sample at current occasion is required if a highly positively correlated auxiliary characters is available. This means the performance of the precision of the estimates also reduces the cost of the survey.

REFERENCES

BIRADAR, R. S., SINGH, H. P., (2001). Successive sampling using Auxiliary Information on both the Occasions, Calcutta Statistical Association Bulletin, 51, 243–251.

COCHRAN, W.G., (1977). Sampling Techniques. 3rd edition.

DAS, K., (1982). Estimation of population ratio on two occasions. Jour. Ind. Soc. Agri. Stat. 34 (2), 1–9.

JESSEN, R. J., (1942). Statistical Investigation of a sample survey for obtaining farm facts. Iowa Agricultural Experiment Statistical Research Bulletin, 304.

NARAIN, R. D., (1953). On the recurrence formula in sampling on successive occasions. Jour. Ind. Soc. Agri. Stat., 5, 96–99.

PATTERSON, H. D., (1950). Sampling on successive occasions with partial replacement of units. Jour. Roy. Stat. Soc., $B(12)$, 241–255.

RAO, J. N. K., MUDHOLKAR, G. S., (1967). Generalized multivariate estimators for the mean of finite population parameters. Jour. Amer. Stat. Asso., 62, 1008–1012.

RAY, S. K., SAHAI, A., (1980). Efficient families of ratio and product type estimators. Biometrika, 67(1), 211–215.

SEN, A. R., (1971). Successive sampling with two auxiliary variables. Sankhya, Series B, (33), 371–378.

SEN, A. R., (1973). Theory and application of sampling on repeated occasions. Jour. Amer. Stat. Asso., 59, 492–509.

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

YATES, F., (1949). Sampling Methods for Censuses and Surveys. Charles Griffln and Co., London.