International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

# Population Variance Estimation Strategy in Two-Occasion Successive Sampling

**Dr. Reba Maji**

Department of Mathematics, Sarojini Naidu College for Women, Dum Dum, Kolkata-700028, India.

**ABSTRACT**
Variation in nature is inescapable; it is present everywhere in our day to day life. In the present manuscript, we are introducing a new class of estimators for estimating the variation phenomenon of a finite population with the help of an additional information under two-occasion rotation sampling scheme. Properties and efficiency of this class of estimators has been illustrated and its optimum replacement strategy has been established. The comparison between the proposed class of estimators with the sample variance estimator through empirical studies, carried over a few natural population datasets, suggests the soundness and usefulness of the proposed estimator in practice.

**Keywords:** Variance estimation, successive sampling, study variable, auxiliary variable, bias, mean square error.

**Mathematics Subject Classification (MSC):**62D05

## 1. Introduction

An inherent behavior of our nature is 'change'. These changes, often, affect the living quality and the human surrounding in some way. This requires the continuous monitoring of the real-life aspects in hand. Thus, the concept of repetitive or successive sampling, a method of surveying same entity at various time-points, has gained considerable attention among survey statisticians. Following pioneering work by Jessen (1942), the study of successive sampling has been expanded by Patterson (1950), Rao and Graham (1964), Gupta (1979), Das (1982), Chaturvedi and Tripathi (1983), Singh and Priyanka (2006), and many others.

Nature continues to vary ceaselessly. Variation exists in our everyday life as well. For instance, a doctor needs to understand and interpret the variations in the human body temperature, blood pressure, pulse rate etc. for proper diagnosis and treatment. Likewise, a number of situations can be encountered where the estimation of population variance of the study character is important. In this regard, works of Singh et al. (2012), Singh et al. (2013), Singh and Singh (2016), Maji et. al. (2018) may be referred.

The present study is an effort in determining a class of estimators of population variance on the second (current) occasion in a two-occasion successive sampling. The properties of the designed estimator have been studied and the efficacy of the proposed work has been examined through empirical studies.

## 2. Sample Structures and Formulation of Estimation Procedure

Sample structure plays an eminent role in formulating the effective estimation procedures of the desired population parameters. To formulate the estimation procedure of finite population variance in two-occasion successive sampling, the related sampling procedure and suggested estimation strategy are described in the following subsections.

**2.1.Sample Structures and Notations**

Let a finite population $U=(U_1, U_2, \ldots, U_N)$ of size N be sampled over two occasions. Let x(y) denote the character under study on the first (second) occasion. We assume that information on an ancillary (auxiliary) variable z, having positive correlation with the study variable x (y), is readily obtainable on first (second) occasion. In this study, for expediency, the population is considered to be sufficiently large. A random sample of size n using simple random sampling without replacement (SRSWOR) is drawn on the first occasion. A sub-sample of size $m(=n\lambda)$, selected at random, is retained from the sample of size n to use it on the second occasion. Another sample of size $u(=n\mu)$ is drawn afresh by the simple random sampling (without replacement) scheme on the second occasion from the entire population in such a way that the total sample size on the current (second) occasion remains n. Here $\mu$ and $\lambda$ $(\mu+\lambda=1)$ are the fractions of fresh and matched samples respectively on the second occasion.

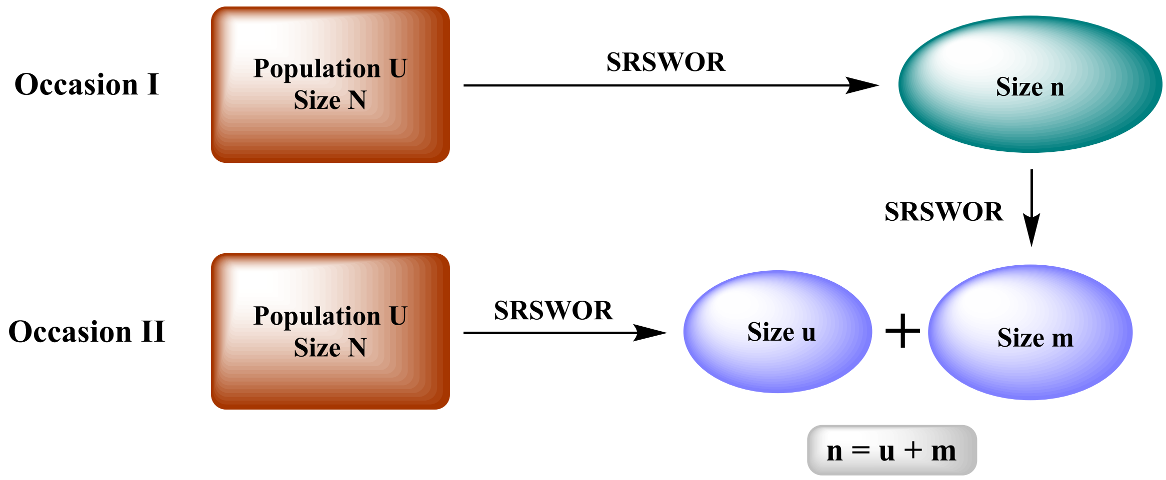


**Figure 1: Sample Structures Under Two-Occasion**

The notations considered throughout this work are as follows.

$S_x^2, S_y^2, S_z^2$ :The population mean squares of the variables x, y, and z, respectively.

$s_x^2(n), s_z^2(n)$ :The sample mean squares of the variables x and z, respectively, based on the sample of size n on the first occasion.

$s_x^2(m), s_y^2(m)$ :The sample mean squares of the variables x and y of the matched sample of size m on the first and current (second) occasions respectively.

$s_y^2(u), s_z^2(u)$ :The sample mean squares of the variables y and z based on the unmatched (fresh) sample of size u on the current (second) occasion.

$\mu_{pqr}=\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\bar{Y}\right)^p\left(x_i-\bar{X}\right)^q\left(z_i-\bar{Z}\right)^r$; p, q, r being non-negative integers.

$$\lambda_{pqr}=\frac{\mu_{pqr}}{\sqrt{\mu_{200}^{p}\mu_{020}^{q}\mu_{002}^{r}}}$$

$$C_0=\sqrt{\lambda_{400}-1},\ C_1=\sqrt{\lambda_{040}-1},\ C_2=\sqrt{\lambda_{004}-1}$$

$$\rho_{01}=\frac{\lambda_{220}-1}{\sqrt{(\lambda_{400}-1)(\lambda_{040}-1)}},\ \rho_{02}=\frac{\lambda_{202}-1}{\sqrt{(\lambda_{400}-1)(\lambda_{004}-1)}},$$

$$\rho_{12}=\frac{\lambda_{022}-1}{\sqrt{(\lambda_{004}-1)(\lambda_{040}-1)}}$$

$$K_{01}=\rho_{01}\frac{C_0}{C_1},\ K_{02}=\rho_{02}\frac{C_0}{C_2},\ K_{12}=\rho_{12}\frac{C_1}{C_2},\ K_{21}=\rho_{12}\frac{C_2}{C_1}$$

$$f_m=\frac{1}{m},\ f_n=\frac{1}{n},\ f=\frac{1}{m}-\frac{1}{n},\ f_u=\frac{1}{u}.$$

**2.2. Proposed Class of Estimators**

In order to estimate the population variance $S_y^2$ on the second (current) occasion, two independent estimators are suggested. One is a ratio estimator based on sample of size $u\,(=n\mu)$ freshly drawn on the second occasion and is given as follows:

$$T_u=s_y^2(u)\frac{S_z^2}{s_z^2(u)} \qquad (2.1)$$

The other estimator $T_{2m}$ is a linear combination of ratio type estimators, based on the matched sample of size $m(=n\lambda)$ common to both the occasions.

$$T_m=P_1s_y^2(m)\frac{s_x^2(n)}{s_x^2(m)}+P_2s_y^2(m)\left(\frac{s_x^2(n)}{s_x^2(m)}\right)^2 \qquad (2.2)$$

where $P_1$ and $P_2$ are constants such that $P_1+P_2=1$ and they are to be determined in such a way that the mean square error of $T_m$ is minimum.

Combining the estimators $T_u$ and $T_m$, the final estimator of $S_y^2$ is given as

$$T=\varphi T_u+(1-\varphi)T_m \qquad (2.3)$$

where, $\varphi$ is an unknown constant to be determined so that the minimum mean square error of the class of estimators T is achieved.

**Remark-1:** In order to estimate the population variance $S_y^2$ of the study variable y on each occasion, the class of estimators $T_u$ is applicable, which means that more belief on $T_u$ could be shown by choosing $\varphi$, in equation (2.3), as 1 (or close to 1), while for estimating changes over occasions, the family of estimators $T_m$ could be more useful and hence $\varphi$ might be chosen as 0 (or close to 0). For asserting both the problems simultaneously, the suitable (optimum) choice of $\varphi$ is required.

## 3. Properties of the Proposed Estimators T

### 3.1. Bias and MSE of T

From equations (2.1) and (2.2), it may be observed that $T_u$ and $T_m$ are biased for $S_y^2$, therefore, the resulting estimators T defined in Equations (2.3) is also biased for $S_y^2$. The bias B (.) and mean square errors M (.) of the proposed estimators up to the first order of approximations are derived using large sample approximations (ignoring f.p.c.) under the following assumptions:

$s_y^2(u) = S_y^2(1+e_0),\ s_z^2(u) = S_z^2(1+e_1),\ s_y^2(m) = S_y^2(1+e_2),$

$s_x^2(n) = S_x^2(1+e_3),\ s_x^2(m) = S_x^2(1+e_4)$

Such that $E(e_i) = 0$ and $|e_i| < 1\ \forall\ i = 0, 1, 2, 3, 4.$

By the above transformations, the estimators $T_u$ and $T_m$ becomes:

$$T_u = S_y^2(1+e_0)(1+e_1)^{-1} \tag{3.1}$$

$$T_m = P_1 S_y^2(1+e_2)(1+e_3)(1+e_4)^{-1} + P_2 S_y^2(1+e_0)(1+e_3)^2(1+e_4)^{-2} \tag{3.2}$$

Considering expectations on both sides of equation (3.1) and (3.2) and keeping the terms of order $O(N^{-1})$, we find the bias and mean square errors of $T_u$ and $T_m$ as

$$B(T_u) = E(T_u - S_y^2) = f_u S_y^2 C_2^2(1-K_{02}) \tag{3.3}$$

$$M(T_u) = E(T_u - S_y^2)^2 = f_u S_y^4\left[C_0^2 + C_2^2 - 2\rho_{02}C_0C_2\right] \tag{3.4}$$

$$B(T_m) = E(T_m - S_y^2) = fP_1S_y^2\left(C_1^2 - \rho_{01}C_0C_1\right) + fP_2S_y^2\left(3C_1^2 - 2\rho_{01}C_0C_1\right) \tag{3.5}$$

$$M(T_m) = S_y^4\left[f_m C_0^2 + f(1+P_2)^2 C_1^2 - 2f(1+P_2)\rho_{01}C_0C_1\right] \tag{3.6}$$

Minimizing $M(T_m)$ with respect to $P_2$ we get

$P_{2_{opt}} = K_{01} - 1$ and $P_{1_{opt}} = 1 - P_{2_{opt}} = 2 - K_{01}$

Therefore,

$$B(T_m)_{opt} = f(2-K_{01})S_y^2\left(C_1^2 - \rho_{01}C_0C_1\right) + f(K_{01}-1)S_y^2\left(3C_1^2 - 2\rho_{01}C_0C_1\right) \tag{3.7}$$

$$M(T_m)_{opt} = S_y^4\left[f_m C_0^2 + fK_{01}^2C_1^2 - 2fK_{01}\rho_{01}C_0C_1\right] \tag{3.8}$$

Therefore, we have the succeeding theorems.

**Theorem 3.1.** Bias of the estimators T of order $O(N^{-1})$ is given as

$$B(T) = \varphi B(T_u) + (1-\varphi)B(T_m)_{opt} \tag{3.9}$$

where, $B(T_u)$ and $B(T_m)$ are given by the equations (3.3) and (3.7).

**Proof:** The bias of T is given by

$$B(T) = E\left(T - S_y^2\right)$$
$$= \varphi E\left(T_u - S_y^2\right) + (1-\varphi) E\left(T_m - S_y^2\right)$$
$$= \varphi B(T_u) + (1-\varphi) B(T_m)_{opt} \tag{3.10}$$

Using the expansions of $B(T_u)$ and $B(T_m)$ from equations (3.3) and (3.7) in (3.10), we get the aspect for the bias of our proposed estimator T as given in the equation (3.9).

**Theorem 3.2.** Mean square error of T up to the first order of approximations is obtained as

$$M(T) = \varphi^2 M(T_u) + (1-\varphi)^2 M(T_m)_{opt} \tag{3.11}$$

where, $M(T_u)$ and $M(T_m)$ are obtained from equations (3.4) and (3.8).

**Proof:** It is obvious that the mean square error of our suggested estimator T is given by

$$M(T) = E\left(T - S_y^2\right)^2 = E\left[\left\{\varphi T_u + (1-\varphi) T_m\right\} - S_y^2\right]^2$$
$$= E\left[\varphi\left(T_u - S_y^2\right) + (1-\varphi)\left(T_m - S_y^2\right)\right]^2$$
$$= \varphi^2 M(T_u) + (1-\varphi)^2 M(T_m)_{opt} + 2\varphi(1-\varphi) E\left[\left(T_u - S_y^2\right)\left(T_m - S_y^2\right)\right] \tag{3.12}$$

Substituting the expressions of $M(T_u)$ and $M(T_m)$ from equations (3.4) and (3.8) in equation (3.12), we find the formula for the mean square error of T as presented in equation (3.11).

It is to be noted that the estimators $T_u$ and $T_m$ are based on two non-overlapping samples of sizes u and m respectively. The covariance type terms $\left(\text{i.e. } E\left[\left(T_u - S_y^2\right)\left(T_m - S_y^2\right)\right]\right)$ are of order $N^{-1}$, hence for large population, they are ignored.

## 3.2. Minimum Mean Square Error of the Proposed Estimator T

It can be noted from Remark-1 and equation (3.11) that mean square error of T in equation (3.11) is a function of constant $\varphi$, therefore, it is to be minimized with respect to $\varphi$ and subsequently the optimal value of $\varphi$ is obtained as

$$\varphi_{opt} = \frac{M(T_m)_{opt}}{M(T_u) + M(T_m)_{opt}} \tag{3.13}$$

and putting the value of $\varphi_{opt}$ from equation (3.16) in equation (3.14), we obtain the optimal mean square error of T as

$$M(T)_{opt} = \frac{M(T_u) \times M(T_m)_{opt}}{M(T_u) + M(T_m)_{opt}} \tag{3.14}$$

Further, substitution of the values from equations (3.4) and (3.8) in equation (3.14) yields the simplified form of $M(T)_{opt}$ as

$$M(T)_{opt} = \frac{d(a+b-b\mu)}{n\left(d+(a+b-d)\mu - b\mu^2\right)} S_y^4 \qquad (3.15)$$

where, $a = C_0^2 + K_{01}^2 C_1^2 - 2K_{01}\rho_{01}C_0C_1$, $b = 2K_{01}\rho_{01}C_0C_1 - K_{01}^2C_1^2$, $d = C_0^2 + C_2^2 - 2\rho_{02}C_0C_2$ and $\mu\left(=\frac{u}{n}\right)$ is the fraction of sample drawn afresh on the current (second) occasion.

### 3.3. Optimum Replacement Strategy

The key design parameter affecting the estimates of change is the overlap between successive samples. Maintaining high overlap between repeats of a survey is operationally convenient, since many sampled units have been located and have some experience in the survey. Hence, to determine the optimum value of $\mu$ so that population variance $S_y^2$ may be estimated with maximum precision, we minimize $M(T)_{opt}$ in equation (3.18) with respect to $\mu$ and hence we get

$$\hat{\mu} = \frac{(a+b) \pm \sqrt{ad}}{b} \qquad (3.16)$$

The real values of $\hat{\mu}$ exist, iff $ab \geq 0$. For any situation, which satisfies this condition, two real values of $\hat{\mu}$ are possible, hence, to choose a 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, lowest one will be the best choice as it reduces the cost of the survey. Substituting the admissible value of $\hat{\mu}$ say $\mu_0$ from equation (3.16) in equation (3.15), we have the optimum values of the mean square error of estimator T, which is shown as

$$M(T)_{opt^*} = \frac{d(a+b-b\mu_0)}{n\left(d+(a+b-d)\mu_0 - b\mu_0^2\right)} S_y^4 \qquad (3.17)$$

## 4. Efficiency Comparison

It is important to investigate situations under which our proposed estimation strategy succeeds better than the usual. Thus, to elucidate the performance of the proposed class of estimators T, we have computed the percent relative efficiency (PRE) of the proposed class of estimator T under its respective optimality condition with respect to the natural estimator $s_y^2(n)$. The estimator $s_y^2(n)$ is considered as the sample estimator of the population variance $S_y^2$, when there is no matching. Since, $s_y^2(n)$ is unbiased estimator of $S_y^2$, so for large $N$ (i.e., $N \to \infty$), the variance of $s_y^2(n)$ is given by

$$V\left(s_y^2(n)\right) = \frac{1}{n} S_y^4 C_0^2 \qquad (4.1)$$

The percent relative efficiency (PRE) of T with respect to $s_y^2(n)$ is designated as

$$PRE = \frac{V\left(s_y^2(n)\right)}{M(T)_{opt^*}} \times 100. \qquad (4.2)$$

### 4.1.1. Numerical Illustration

Three natural population datasets has been chosen to illustrate the efficiency of our proposed methodology. The source of the populations, the nature of the variables y, x, z and the values of the various parameters are given by Table-1 and In Table-2 and Table-3, we have examined the efficacy of the proposed work through the data set of natural populations given as follows:

**Population I-Source: Murthy (1967)**
y: Area under wheat in 1964.
x: Area under wheat in 1963.
z: Cultivated area in 1961.

**Population II-Source: Anderson (1958)**
y: Head length of second son.
x: Head length of first son.
z: Head breadth of first son.

| Population | N | $C_0$ | $C_1$ | $C_2$ | $\rho_{01}$ | $\rho_{02}$ | $\rho_{12}$ |
|---|---|---|---|---|---|---|---|
| **I** | 80 | 1.1255 | 1.6065 | 1.3662 | o.7319 | 0.7940 | 0.9716 |
| **II** | 25 | 1.3512 | 1.4295 | 1.2853 | 0.5057 | 0.5683 | 0.4213 |

**Table 1: Parametric values of different populations**

**Table 2: PRE of T with respect to $s_y^2(n)$**

| Population | $s_y^2(n)$ | T |
|---|---|---|
| **I** | 100 | 147.97 |
| **II** | 100 | 118.03 |

**Table 3: PRE of T with respect to $s_y^2(n)$**

| Population | $\mu_0$ | PRE |
|---|---|---|
| **I** | 0.97 | 184.33 |
| **II** | 0.85 | 122.36 |

(under optimum replacement strategy)

## 5. Discussion and Conclusion

Following conclusions may be established from the present study:

(a) In Table 2, we have calculated the percent relative efficiency (PRE) of the estimator T with respect to the natural estimator $s_y^2(n)$ of the population variance and we see that the PREs' are greater than 100, which means that under different natural population set up, the suggested estimator T is superior to $s_y^2(n)$, the sample variance estimator.

(b) In Table 3, the percent relative efficiency (PRE) of the estimator T with respect to the natural estimator $s_y^2(n)$ of the population variance have been calculated under optimum replacement strategy and also in this case we see that the PREs' are higher than 100, which concludes that under optimum replacement strategy, the suggested estimator T is superior to $s_y^2(n)$, the sample variance estimator.

Therefore, the proposed estimator T is justified and may be recommended to the survey practitioners for their use in real life problems.

## 6. References

1. Anderson T. W., 1958, An introduction to multivariate statistical analysis. John Wiley & sons, Inc., New York.
2. Chaturvedi D. K. and Tripathi T. P., 1983, Estimation of population ratio on two occasions using multivariate auxiliary information. Journal of Indian Statistical Association, **21**, 113-120.
3. Das A. K., 1982, Estimation of population ratio on two occasions. Journal of the Indian Society of Agricultural Statistics, **34**, 1-9.
4. Gupta P. C., 1979, Sampling on two successive occasions. Journal of Statistical Research, **13**, 7-16.
5. Jessen R. J., 1942, Statistical investigation of a sample survey for obtaining farm facts. Iowa Agricultural Experiment Station Research, Bulletin No. **304**, Ames, Iowa, USA, 1-104.
6. Maji, R., Singh,G. N. and Bandyopadhyay,A. (2018): Efficient Variance Estimation Strategy in Two-Occasion Successive Sampling. International Journal of Applied Mathematics and Statistics, 2018, Ceser Publication, **57 (3)**, 194-202.
7. Murthy M. N., 1967, Sampling theory and methods. Statistical Publishing Society. Calcutta, India.
8. Patterson H. D., 1950, Sampling on successive occasions with partial replacement of units. Journal of the Royal Statistical Society, **12**, 241-255.
9. Rao J. N. K. and Graham J. E., 1964, Rotation design for sampling on repeated occasions.American Statistical Association, **59**, 492-509.
10. Singh G. N. and Singh A. K., 2016, On estimation of finite population variance in two-occasion successive sampling. International Journal of Applied Mathematics and Statistics, **55(3)**,28-40.
11. Singh G. N. and 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.
12. Singh G. N., Priyanka K., Prasad S., Singh S. and Kim J. M., 2013, A class of estimators for population variance in two occasion rotation patterns. Communications for Statistical Applications and Methods, **20(4)**, 247-257.
13. Singh G. N., Singh V. K., Priyanka K., Prasad S. and Karna J. P., 2012, Rotation patterns under imputation of missing data over two-occasion. Communications in Statistics-Theory and Methods, **41(4)**, 619-637.