International Journal of Statistics and Applied Mathematics 2025; 10(7): 90-97

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(7): 90-97

https://www.mathsjournal.com
Received: 05-09-2025
Accepted: 05-10-2025

Deepak Majhi
Assistant Professor Senior Scale,
PG Department of Mathematics
Veer Kunwar Singh University
Ara, Bihar, India

Corresponding Author:
Deepak Majhi
Assistant Professor Senior Scale,
PG Department of Mathematics
Veer Kunwar Singh University
Ara, Bihar, India

Efficient use of auxiliary information in the estimation of population variance in sample surveys

Deepak Majhi

DOI: https://www.doi.org/10.22271/maths.2025.v10.i10b.2187

Abstract
The present paper we proposed some new estimators to estimate the population variance of study variable using the auxiliary information in simple random sampling without replacement scheme. The expressions for its bias and mean square errors have been obtained. To show the theoretical results we compare our proposed estimators with natural variance and traditional ratio estimator suggested by Isaki (1983) [1]. Empirical studies are carried out to examine the performances of the proposed estimators and consequently the suitable recommendations are made.

Keywords: Population variance, auxiliary information, simple random sampling, bias, mean square error

Introduction

Consider a finite population U = $\left(U_1, U_2, U_3, \ldots, U_N\right)$ of N units. Let y and x stand for the variable under study and an auxiliary variable respectively. Let $\left(y_i, x_i\right)$, i = 1, 2, 3, N, denote the values of the population units for (y, x) respectively. Further let $\left(y_i, x_i\right)$, i =1, 2, 3, n, denote the values of the units included in the sample s_n of size n drawn by simple random sampling without replacement (SRSWOR). To estimate $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2$, it is assume that $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2$ with $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is known. The usual unbiased estimator for population variance S_y^2 is

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \tag{1}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the sample mean of y.
Isaki (1983) [1] presented the ratio estimator for the population variance as

$$t_R = \left(\frac{s_y^2}{s_x^2}\right) S_x^2 \tag{2}$$

where $s_y^2 \text{ and } s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ are unbiased estimators of population variances $S_y^2 \text{ and } S_x^2$ respectively.

Singh *et al.* (2011) [3] propose exponential ratio type estimator for estimating population variance S_y^2 as

$$t_S = s_y^2 \exp\left(\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right) \tag{3}$$

Tailor and Sharma (2012) [4] propose following estimator for estimating population variance S_y^2 as

$$t_1 = fs_y^2 + (1-f)\left(\frac{s_y^2}{s_x^2}\right)S_x^2 \tag{4}$$

$$t_2 = \left[\left(\frac{1-f}{1+2f}\right)s_y^2 + \left(\frac{3f}{1+2f}\right)\left(\frac{s_y^2}{s_x^2}\right)S_x^2\right] \tag{5}$$

$$t_3 = \left[\left(\frac{1-f}{1+3f}\right)s_y^2 + \left(\frac{4f}{1+3f}\right)\left(\frac{s_y^2}{s_x^2}\right)S_x^2\right] \tag{6}$$

Upto the first order of approximations, the variance of s_y^2 and mean square errors of t_R, t_S, t_1, t_2 and t_3 are respectively given by

$$V\left(s_y^2\right) = f_1\left\{\beta_2(y)-1\right\}S_y^4 \tag{7}$$

$$M\left(t_R\right) = f_1\left[\left\{\beta_2(y)-1\right\} + \left\{\beta_2(x)-1\right\}(1-2C)\right]S_y^4 \tag{8}$$

$$M\left(t_S\right) = f_1\left[\left\{\beta_2(y)-1\right\} + \left\{\beta_2(x)-1\right\}\left(\frac{1}{4}-C\right)\right]S_y^4 \tag{9}$$

$$M\left(t_1\right) = f_1\left[\left\{\beta_2(y)-1\right\} + (1-f)\left\{\beta_2(x)-1\right\}\left\{(1-f)-2C\right\}\right]S_y^4 \tag{10}$$

$$M\left(t_2\right) = f_1\left[\left\{\beta_2(y)-1\right\} + A\left\{\beta_2(x)-1\right\}(A-2C)\right]S_y^4 \tag{11}$$

And

$$M\left(t_3\right) = f_1\left[\left\{\beta_2(y)-1\right\} + B\left\{\beta_2(x)-1\right\}(B-2C)\right]S_y^4 \tag{12}$$

where $f_1 = \left(\frac{1}{n}-\frac{1}{N}\right), \beta_2(y) = \frac{\mu_{40}}{\mu_{20}^2}, \beta_2(x) = \frac{\mu_{04}}{\mu_{02}^2}, h = \frac{\mu_{22}}{\mu_{20}\mu_{02}}, C = \frac{(h-1)}{\left\{\beta_2(x)-1\right\}}$ and $\mu_{rs} = E\left\{\left(y_i-\bar{Y}\right)^r\left(x_i-\bar{X}\right)^s\right\}$

2. Proposed Estimators

Following the above argument, we propose some modified exponential ratio type estimators for population variance S_y^2 using different sampling fraction when population variance of S_x^2 is known and are given by

$$T_1 = fs_y^2 + (1-f)s_y^2 \exp\left[\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right] \tag{13}$$

$$T_2=\frac{1-f}{1+2f}s_y^2+\frac{3f}{1+2f}s_y^2\exp\left[\frac{S_x^2-s_x^2}{S_x^2+s_x^2}\right] \tag{14}$$

$$T_3=\frac{1-f}{1+3f}s_y^2+\frac{4f}{1+3f}s_y^2\exp\left[\frac{S_x^2-s_x^2}{S_x^2+s_x^2}\right] \tag{15}$$

3. Properties of the estimators

Theorem 3.1: Bias of the estimators $T_i\,(i=1,2,3)$ defined in equations (13), (14) and (15) up to $o\left(n^{-1}\right)$ are obtained as

$$B(T_1)=\frac{f_1}{2}\left[\left\{\beta_2(x)-1\right\}\left(\frac{3}{4}-C\right)\right]S_y^2$$

$$B(T_2)=\frac{A}{2}f_1\left[\left\{\beta_2(x)-1\right\}\left(\frac{3}{4}-C\right)\right]S_y^2$$

And

$$B(T_3)=\frac{B}{2}f_1\left[\left\{\beta_2(x)-1\right\}\left(\frac{3}{4}-C\right)\right]S_y^2$$

where $f=\frac{n}{N},\ A=\frac{3f}{1+2f}$ and $B=\frac{4f}{1+3f}$

Theorem 3.2: Mean square errors of the estimators $T_i\,(i=1,2,3)$ defined in equations (13), (14) and (15) up to $o\left(n^{-1}\right)$ are derived as

$$M(T_1)=f_1\left[\left\{\beta_2(y)-1\right\}+(1-f)\left\{\beta_2(x)-1\right\}\left(\frac{1-f}{4}-C\right)\right]S_y^4 \tag{16}$$

$$M(T_2)=f_1\left[\left\{\beta_2(y)-1\right\}+A\left\{\beta_2(x)-1\right\}\left(\frac{A}{4}-C\right)\right]S_y^4 \tag{17}$$

And

$$M(T_3)=f_1\left[\left\{\beta_2(y)-1\right\}+B\left\{\beta_2(x)-1\right\}\left(\frac{B}{4}-C\right)\right]S_y^4 \tag{18}$$

4. A class of Estimators

When population variance of the auxiliary variable x is know and w is certain chosen weights, we define a class of estimators to estimate population variance of the study variable y as

$$T_{4w}=ws_y^2+(1-w)s_y^2\exp\left[\frac{S_x^2-s_x^2}{S_x^2+s_x^2}\right] \tag{19}$$

Upto to the first order of approximation, mean square error of the estimator T_{4w} is derived as

$$M(T_{4w})=f_1\left[\left\{\beta_2(y)-1\right\}+(w-1)\left\{\beta_2(x)-1\right\}\left(\frac{w-1}{4}+C\right)\right]S_y^4 \tag{20}$$

Under optimum condition, mean square error of the estimator T_{4w} is obtained as

$$M(T_{4w})_{opt} = f_1\left[\left\{\beta_2(y)-1\right\} - C^2\left\{\beta_2(x)-1\right\}\right]S_y^4 \tag{21}$$

Comparison of the Estimators T_{4w}

(a) In this section we compare the proposed estimators T_{4w} with to respect to the estimators $s_y^2, t_R, t_S, t_1, t_2$ and t_3. Different condition are explored under which our proposed estimator is more efficient than these estimators and shown below:

(i) T_{4w} is better than s_y^2 if $M(T_{4w})_{opt} \le V(s_y^2)$, which gives $C \ge 0$, which is always true. Hence the estimator T_{4w} is better than s_y^2.

(ii) T_{4w} is better than t_R if $M(T_{4w})_{opt} \le M(t_R)$, which gives $(1-C)^2 \ge 0$, which is always true. Hence the estimator T_{4w} is more efficient than t_R.

(iii) T_{4w} is better than t_S if $M(T_{4w})_{opt} \le M(t_S)$, which gives $\left(C-\frac{1}{2}\right)^2 \ge 0$, which is always true. Hence the estimator T_{4w} is preferable over t_S.

(iv) T_{4w} is more efficient than T_1 if $M(T_{4w})_{opt} \le M(T_1)$, which gives $\left(C-\frac{1-f}{2}\right)^2 \ge 0$, which is always true. Hence the estimator T_{4w} is better than T_1.

(v) T_{4w} is more efficient than T_2 if $M(T_{4w})_{opt} \le M(T_2)$, which gives $\left(C-\frac{A}{2}\right)^2 \ge 0$, which is always true. Hence the estimator T_{4w} is preferable over T_2.

(vi) T_{4w} is better than T_3 if $M(T_{4w})_{opt} \le M(T_3)$, which gives $\left(C-\frac{B}{2}\right)^2 \ge 0$, which is always true. Hence the estimator T_{4w} is more efficient than T_3.

(b) In this section we compare the proposed estimators T_{4w} with to respect to the estimators $s_y^2, t_R, t_S, t_1, t_2$ and t_3. Different range of w are given under which our proposed estimator is more efficient than these estimators and shown below:

(i) T_{4w} is better than s_y^2 if $1 < w \le (1+4C)$ holds good.

(ii) T_{4w} is more efficient than t_R if $(8C-2) \le w \le 6$ satisfy this range.

(iii) T_{4w} is better than t_S if $2-f \le w \le (4C+f)$.

(iv) T_{4w} is more efficient than T_1 if $2-f \le w \le (4C+f)$.

(v) T_{4w} is more efficient than T_2 if $(1+A) \le w \le (1+4C-A)$.

(vi) T_{4w} is preferable over T_3 if $(1+B) \le w \le (1+4C-B)$.

(5) A class of Estimator of different scalars

When population variance of the auxiliary variable x is know and w_1 and w_2 are two scalars, we define a class of estimators to estimate population variance of the study variable y as

$$T_{4w}^* = w_1 s_y^2 + w_2 s_y^2 \exp\left[\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right] \tag{22}$$

Upto to the first order of approximation, mean square error of the estimator T_{4w}^{*} under optimality condition is derived as

$$M\left(T_{4w}^{*}\right)_{opt}=\left[\left(w_1^{*}+w_2^{*}-1\right)^2+f_1\left\{\beta_2\left(y\right)-1\right\}\left(w_1^{*}+w_2^{*}\right)^2+f_1\left\{\beta_2\left(x\right)-1\right\}\left\{w_2^{*2}\left(\frac{1}{4}-C\right)-w_1^{*}w_2^{*}C\right\}\right]S_y^4 \tag{23}$$

where

$$w_1^{*}=\frac{f_1\left\{\beta_2\left(y\right)-1\right\}\left(1-2C\right)}{\left[\left\{2+2f_1\left(\beta_2\left(y\right)-1\right)\right\}\left\{2+2f_1\left(\beta_2(y)-1\right)+2f_1\left(\beta_2(x)-1\right)\left(\frac{1}{4}-C\right)\right\}-\left\{2+2f_1\left(\beta_2(y)-1\right)-f_1\left(\beta_2(x)-1\right)C\right\}^2\right]}$$

$$w_2^{*}=\frac{2f_1\left\{\beta_2\left(y\right)-1\right\}C}{\left[\left\{2+2f_1\left(\beta_2\left(y\right)-1\right)\right\}\left\{2+2f_1\left(\beta_2(y)-1\right)+2f_1\left(\beta_2(x)-1\right)\left(\frac{1}{4}-C\right)\right\}-\left\{2+2f_1\left(\beta_2(y)-1\right)-f_1\left(\beta_2(x)-1\right)C\right\}^2\right]}$$

6. Proposed Estimators in Two-Phase Sampling scheme

In certain practical situations when variance of the auxiliary variable x is not known a priori, the technique of two-phase sampling scheme is used. This scheme requires collection of information on auxiliary variable on its estimated value at first phase sample of size n' and on y for second phase sample of size n $\left(n<n'\right)$ from the first phase sample.

The estimators T_1, T_2 and T_3 in the two-phase sampling scheme take the following form respectively

$$T_{1d}=fs_y^2+\left(1-f\right)s_y^2\exp\left[\frac{s_x'^2-s_x^2}{s_x'^2+s_x^2}\right] \tag{24}$$

$$T_{2d}=\frac{1-f}{1+2f}s_y^2+\frac{3f}{1+2f}s_y^2\exp\left[\frac{s_x'^2-s_x^2}{s_x'^2+s_x^2}\right] \tag{25}$$

$$T_{3d}=\frac{1-f}{1+3f}s_y^2+\frac{4f}{1+3f}s_y^2\exp\left[\frac{s_x'^2-s_x^2}{s_x'^2+s_x^2}\right] \tag{26}$$

Theorem 6.1: Bias of the estimators $T_{id}\left(i=1,2,3\right)$ defined in equations (24), (25) and (26) up to $o\left(n^{-1}\right)$ are obtained as

$$B\left(T_{1d}\right)=\frac{f_3}{2}\left[\left\{\beta_2\left(x\right)-1\right\}\left(\frac{3}{4}-C\right)\right]S_y^2;\ f_3=\left(\frac{1}{n}-\frac{1}{n'}\right)$$

$$B\left(T_{2d}\right)=\frac{A}{2}f_3\left[\left\{\beta_2\left(x\right)-1\right\}\left(\frac{3}{4}-C\right)\right]S_y^2$$

And

$$B\left(T_{3d}\right)=\frac{B}{2}f_3\left[\left\{\beta_2\left(x\right)-1\right\}\left(\frac{3}{4}-C\right)\right]S_y^2$$

Theorem 6.2: Mean square errors of the estimators $T_{id}\left(i=1,2,3\right)$ defined in equations (24), (25) and (26) up to $o\left(n^{-1}\right)$ are derived as

$$M\left(T_{1d}\right)=\left[f_1\left\{\beta_2\left(y\right)-1\right\}+f_3\left(1-f\right)\left\{\beta_2\left(x\right)-1\right\}\left(\frac{1-f}{4}-C\right)\right]S_y^4 \tag{27}$$

$$M\left(T_{2d}\right)=\left[f_1\left\{\beta_2\left(y\right)-1\right\}+f_3\left\{\beta_2\left(x\right)-1\right\}A\left(\frac{A}{4}-C\right)\right]S_y^4 \tag{28}$$

and

$$M\left(T_{3d}\right)=\left[f_1\left\{\beta_2\left(y\right)-1\right\}+f_3\left\{\beta_2\left(x\right)-1\right\}B\left(\frac{B}{4}-C\right)\right]S_y^4 \tag{29}$$

7. A class of Estimators in Two-Phase Sampling Scheme

When population variance of the auxiliary variable S_x^2 is not known and w is certain chosen weights, we define a class of estimators to estimate population variance of the study variable S_y^2 as

$$T_{4wd}=ws_y^2+\left(1-w\right)s_y^2\exp\left[\frac{s_x'^2-s_x^2}{s_x'^2+s_x^2}\right] \tag{30}$$

Upto to the first order of approximation, mean square error of the estimator T_{4wd} is derived as

$$M\left(T_{4wd}\right)=\left[f_1\left\{\beta_2\left(y\right)-1\right\}+f_3\left(1-w\right)\left\{\beta_2\left(x\right)-1\right\}\left(\frac{1-w}{4}-C\right)\right]S_y^4 \tag{31}$$

Under optimum condition, mean square error of the estimator T_{4wd} is obtained as

$$M\left(T_{4wd}\right)_{opt}=\left[f_1\left\{\beta_2\left(y\right)-1\right\}-f_3C^2\left\{\beta_2\left(x\right)-1\right\}\right]S_y^4 \tag{32}$$

(8) A class of Estimator of different scalars in Two-Phase Sampling

When population variance of the auxiliary variable x is not know and w_1 and w_2 are two scalars, we define a class of estimators to estimate population variance of the study variable y as

$$T_{4wd}^*=w_1s_y^2+w_2s_y^2\exp\left[\frac{s_x'^2-s_x^2}{s_x'^2+s_x^2}\right] \tag{33}$$

Upto to the first order of approximation, mean square error of the estimator T_{4wd}^* under optimality condition is derived as

$$M\left(T_{4w}^*\right)_{opt}=\left[\left(w_1^*+w_2^*-1\right)^2+f_1\left\{\beta_2\left(y\right)-1\right\}\left(w_1^*+w_2^*\right)^2+f_1\left\{\beta_2\left(x\right)-1\right\}\left\{w_2^{*2}\left(\frac{1}{4}-C\right)-w_1^*w_2^*C\right\}\right]S_y^4 \tag{34}$$

where

$$w_1^*=\frac{f_3\left\{\beta_2\left(y\right)-1\right\}\left(1-2C\right)}{\left[\left\{2+2f_1\left(\beta_2\left(y\right)-1\right)\right\}\left\{2+2f_1\left(\beta_2(y)-1\right)+2f_3\left(\beta_2(x)-1\right)\left(\frac{1}{4}-C\right)\right\}-\left\{2+2f_1\left(\beta_2(y)-1\right)-f_3\left(\beta_2(x)-1\right)C\right\}^2\right]}$$

$$w_2^*=\frac{2f_3\left\{\beta_2\left(y\right)-1\right\}C}{\left[\left\{2+2f_1\left(\beta_2\left(y\right)-1\right)\right\}\left\{2+2f_1\left(\beta_2(y)-1\right)+2f_3\left(\beta_2(x)-1\right)\left(\frac{1}{4}-C\right)\right\}-\left\{2+2f_1\left(\beta_2(y)-1\right)-f_3\left(\beta_2(x)-1\right)C\right\}^2\right]}$$

9 A. Empirical Study

To analyse the performance of the proposed estimators of population variance S_y^2 in comparison to the other estimators, we consider the data given in Murthy (1967) [2], p-226). The variables and data set is given as
y: output, x = number of workers

N= 80, n' = 25, n = 10, $\beta_2(y)=2.2667$, $\beta_2(x)=3.65$, h = 2.3377 and C = 0.5048.

The percent relative efficiencies (PRE) of the various estimators of population variance with respect to the estimators $s_y^2, t_R, t_S, t_1, t_2$ and t_3 have been computed and presented in the Table 1-4.

Table 1: Percent relative efficiencies of the estimators T_1, T_2 and T_3 w.r.to. $s_y^2, t_R, t_S, t_1, t_2$ and t_3

Estimators	s_y^2	t_R	t_S	t_1	t_2	t_3
T_1	209.9135	205.7043	-	158.1985	116.4293	106.7615
T_2	136.9383	134.1924	-	103.2017	-	-
T_3	145.9556	143.0288	-	109.9974	-	-

Table 2: Percent relative efficiency of the estimator T_{4w} w.r.to. $s_y^2, t_R, t_S, T_1, T_2$ and T_3

Estimators	s_y^2	t_R	t_S	T_1	T_2	T_3
T_{4w}	214.1725	209.8779	100.0103	102.0289	156.4007	146.7382

Table 3: Percent relative efficiency of the estimator T_{4w}^* w.r.to. $s_y^2, t_R, t_S, T_1, T_2$ and T_3

Estimators	s_y^2	t_R	t_S	T_1	T_2	T_3	T_{4w}
T_{4w}^*	225.2561	220.7393	105.1859	107.3090	164.4946	154.3320	105.1751

Table 4: Percent relative efficiency of the estimator T_{4w}^* for different values of w_1 and w_2 w.r.to. $s_y^2, t_R, t_S, T_1, T_2$ and T_3

T_{4w}^*		Estimators						
w_1	w_2	s_y^2	t_R	t_R	T_1	T_2	T_3	T_{4w}
0.10	0.90	211.3298	207.0922	-	100.6747	154.3248	144.7905	-
0.15	0.85	208.2322	204.0567	-	-	152.0628	142.6682	-
0.20	0.80	204.1287	200.0355	-	-	149.0661	139.8567	-
0.25	0.75	199.1412	195.1480	-	-	145.4240	136.4396	-

Table 5: Percent relative efficiencies of the estimators T_{1d}, T_{2d} and T_{3d} w.r.to. s_y^2 and t_R

Estimators	s_y^2	t_R
T_{1d}	156.0182	152.8897
T_{2d}	122.6945	120.2342
T_{3d}	127.5354	124.9780

Table 6: Percent relative efficiency of the estimator T_{4wd} w.r.to. $s_y^2, t_R, t_S, T_{1d}, T_{2d}$ and T_{3d}

Estimators	s_y^2	t_R	t_S	T_{1d}	T_{2d}	T_{3d}
T_{4wd}	157.6156	154.4551	-	101.0239	128.4619	123.5858

Table 7: Percent relative efficiency of the estimator T_{4wd}^{*} w.r.to. $s_y^2, t_R, t_S, T_{1d}, T_{2d}$ and T_{3d}

Estimators	s_y^2	t_R	t_S	T_{1d}	T_{2d}	T_{3d}	T_{4wd}
T_{4wd}^{*}	168.6992	165.3165	-	163.3993	165.9689	165.5123	163.3034

Table 8: Percent relative efficiency of the estimator T_{4wd}^{*} for different values of w_1 and w_2 w.r.to. $s_y^2, t_R, t_S, T_{1d}, T_{2d}$ and T_{3d}

T_{4wd}^{*}		Estimators						
w_1	w_2	s_y^2	t_R	t_R	T_{1d}	T_{2d}	T_{3d}	T_{4wd}
0.10	0.90	211.3298	207.0922	-	100.6747	154.3248	144.7905	-
0.15	0.85	208.2322	204.0567	-	-	152.0628	142.6682	-
0.20	0.80	204.1287	200.0355	-	-	149.0661	139.8567	-
0.25	0.75	199.1412	195.1480	-	-	145.4240	136.4396	-

10. Conclusions
It is observed from the Table 1-8 indicate that there is a significant gain in efficiency by using proposed variance estimators in comparison to various well-known estimators for population variance. Therefore, suggested estimators are recommended to survey statisticians for their real life use.

References
1. Isaki CT. Variance estimation using auxiliary information. J Am Stat Assoc. 1983;78:117-123.
2. Murthy MN. *Sampling Theory and Methods*. Calcutta (India): Statistical Publishing Society; 1967.
3. Singh R, *et al*. Improvement exponential estimators for population variance using two-auxiliary variables. Ital J Pure Appl Math. 2011;28:101-107.
4. Tailor R, Sharma B. Modified estimators of population variance in presence of auxiliary information. Stat Transit New Ser. 2012;13(1):37-46.