VARIANCE ESTIMATION USING ARITHMETIC MEAN, GEOMETRIC MEAN AND HARMONIC MEAN UNDER SIMPLE RANDOM SAMPLING

Prabhakar Mishra and *Rajesh Singh
Department of Statistics, Banaras Hindu University, Varanasi, 221005
*Corresponding author

Abstract
In this paper we have suggested some estimators of population variance using auxiliary information based on arithmetic mean, geometric mean and harmonic mean. We have also suggested an almost unbiased estimator for estimating population variance. The expressions of mean squared error (MSE) have been derived up to the first order of approximation. It has been shown that almost unbiased estimator gives better result than estimators included in the paper. Numerical illustrations are given in support of the theoretical study.

Keywords: Auxiliary information, arithmetic mean, geometric mean, harmonic mean, Mean square error, unbiased estimator and simple random sampling.

Introduction
Use of auxiliary variables is very common in estimating various population parameters with greater efficiency. Let x be the auxiliary variable, highly correlated to the study variable y. If information on an auxiliary variable is readily available then it is a well-known fact that the ratio-type and regression-type estimators can be used for estimation of parameters of interest, due to increase in efficiency of these estimators.

Variations are present everywhere in our daily life. It is the law of nature that no two things or individuals are exactly alike. For instance, a physician needs a full understanding of variations in the degree of human blood pressure, body temperature and pulse rate for adequate prescription (see Singh and Solanki (2012)). Consider the problem of estimating the population variance $\sigma_y^2$ of the study variable y assuming $\sigma_x^2$ is known. Isaki (1983) presented the ratio estimator for the population variance using the auxiliary information. Several authors including Singh and Singh (2001, 2003), Jhajj et al. (2005), Kalidar and Cingi (2007), Shabbir and Gupta (2007), Grover (2010), Singh and Solanki (2012), Singh et al. (2014) and Singh and Singh (2015) suggested improved estimators of $\sigma_y^2$.

Let $U = \{U_1, U_2, \ldots, U_N\}$ be the finite population of size N out of which a sample of size n is drawn according to simple random sampling without replacement (SRSWOR) technique. Let Y and X denote the study variable and auxiliary variable.
taking values \( y_i \) and \( x_i \), respectively on the \( i^{th} \) unit \( U_i \) of the population \( U \). Let \( (\bar{y}, \bar{x}) \) be the sample mean estimator of \( (\bar{Y}, \bar{X}) \), the population means of \( y \) and \( x \) respectively. In order to have a survey estimate of the population variance \( s_y^2 \) of the study character \( Y \) assuming the knowledge of the population variance \( s_x^2 \) of the auxiliary character \( X \), the usual estimator of population variance of the study variable \( Y \) is defined by

\[
P_o = s_y^2
\]

Isaki (1983) suggested a ratio type estimator of population variance, given as

\[
P_{ir} = s_y^2 \left[ \frac{s_x^2}{s_y^2} \right]
\]

Bahl and Tuteja (1991) suggested a ratio type exponential estimator of the population variance as

\[
P_{exp} = s_y^2 \exp \left[ \frac{s_x^2 - s_y^2}{s_x^2 + s_y^2} \right]
\]

The dual to ratio estimator and dual to ratio type exponential estimator of population variance is given by

\[
P_1 = s_y^2 \left[ \frac{s_x^2}{s_y^2} \right]
\]

\[
P_2 = s_y^2 \exp \left[ \frac{s_x^2 - s_y^2}{s_x^2 + s_y^2} \right]
\]

where, \( s_x^2 = \frac{N s_x^2 - n s_x^2}{N - n} \)

Isaki (1983) also suggested the regression type estimator of finite population variance, given by

\[
P_{reg} = s_y^2 + k_o \left[ s_x^2 - s_y^2 \right]
\]

where \( s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \), \( s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \),
\[
S_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2 \quad \text{and} \quad S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{X})^2 .
\]

To obtain the Bias and MSE, we write
\[
s_y^2 = S_y^2 [1 + \omega_o] \quad \text{and} \quad s_x^2 = S_x^2 [1 + \omega_1]
\]
such that \( E(\omega_o) = E(\omega_1) = 0, E(\omega_o^2) = f(\lambda_{40} - 1), E(\omega_1^2) = f(\lambda_{22} - 1) \)

and \( E(\omega_o \omega_1) = f(\lambda_{40} - 1, \lambda_{22} - 1) \).

where \( f = \frac{1}{n} - \frac{1}{N}, \lambda_{pq} = \frac{\mu_{pq}}{\mu_{20}^{\frac{1}{2}} \mu_{22}^{\frac{1}{2}}} \) and \( \mu_{pq} = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^p (x_i - \bar{X})^q \)

(where \( p, q \) being non-negative integers)

Then, \( s_y^2 = (1 - g\omega_1)S_x^2 ; g = \frac{n}{N-n} . \)

The variance of the usual estimator \( s_y^2 \) under SRSWOR is given as
\[
Var(P_o) = Var\left(s_y^2\right) = fS_y^4 \left(\lambda_{40} - 1\right) \quad (7)
\]

To the first order of approximation, the Bias and MSE expressions of the estimators \( P_1 \) and \( P_2 \) are respectively given by
\[
Bias(P_1) = -S_y^2 gf(\lambda_{22} - 1) \quad (8)
\]
\[
Bias(P_2) = -S_y^2 f \left[ \frac{g^2}{8} (\lambda_{40} - 1) + \frac{g}{4} (\lambda_{22} - 1) \right] \quad (9)
\]
\[
MSE(P_1) = S_y^4 f \left[ (\lambda_{40} - 1) + g^2 (\lambda_{40} - 1) - 2g (\lambda_{22} - 1) \right] \quad (10)
\]
\[
MSE(P_2) = S_y^4 f \left[ (\lambda_{40} - 1) + \frac{g^2}{4} (\lambda_{40} - 1) - g (\lambda_{22} - 1) \right] \quad (11)
\]

To the first order of approximation the MSE’s of the estimators \( P_{IR} \), \( P_{exp} \) and \( P_{reg} \) are respectively given by
\[
MSE(P_{IR}) = S_y^4 f \left[ (\lambda_{40} - 1) + (\lambda_{40} - 1) - 2(\lambda_{22} - 1) \right] \quad (12)
\]
\[
MSE(P_{exp}) = S_y^4 f \left[ (\lambda_{40} - 1) + \frac{1}{4} (\lambda_{40} - 1) - (\lambda_{22} - 1) \right] \quad (13)
\]
Motivated by Singh et al. (2014), we have proposed some estimators of population variance of the study variable $Y$ based on arithmetic mean, geometric mean and harmonic mean ($AM, GM, HM$) of the estimators $(P_o, P_1), (P_o, P_2), (P_1, P_2)$ and $(P_o, P_1, P_2)$. The properties of suggested estimators have been studied up to first order of approximation.

**The estimators based on $P_o$ and $P_1$**

Taking the AM, GM and HM of the estimators $P_o$ and $P_1$, we get the following estimators of the population variance respectively as,

$$P^AM_3 = \frac{P_o + P_1}{2} = \left(\frac{s^2_y}{2}\right) \left(1 + \frac{S^2_x}{S^2_y}\right)$$  \hspace{1cm} (15)

$$P^GM_3 = \left(P_o P_1\right)^{\frac{1}{2}} = s^2_y \left(\frac{S^2_x}{S^2_y}\right)$$  \hspace{1cm} (16)

$$P^HM_3 = \frac{2}{\left(\frac{1}{P_o} + \frac{1}{P_1}\right)} = \frac{2s^2_y}{\left(1 + \frac{S^2_x}{S^2_y}\right)}$$  \hspace{1cm} (17)

To the first degree of approximation, the mean squared errors of $P^j_3$; ($j$= AM, GM, HM) are respectively given by

$$MSE(P^j_3) = S^2_y f \left(\lambda_{40} - 1\right) + \frac{g^2}{4} (\lambda_{44} - 1) - g(\lambda_{22} - 1)$$  \hspace{1cm} (18)

**The estimators based on $P_o$ and $P_2$**

The estimators of $S^2_y$ based on AM, GM and HM of the estimators $P_o$ and $P_2$ are respectively defined as
To the first degree of approximation, the mean squared errors of \( P_4^j \); \( j = AM, GM, HM \) are respectively given by

\[
MSE(P_4^j) = \left[ \lambda_{40} - 1 \right] + \frac{G^2}{16} \left( \lambda_{04} - 1 \right) - \frac{G}{2} \left( \lambda_{22} - 1 \right)
\]

\[ (22) \]

**The estimators based on \( P_1 \) and \( P_2 \)**

We propose the following estimators of \( S_y^2 \) based on AM, GM and HM of the estimators \( P_1 \) and \( P_2 \) respectively as

\[
P_5^{AM} = \frac{P_1 + P_2}{2} = \frac{s_y^2}{2} \left( \frac{s_x^{*2}}{S_x^2} + \exp \left( \frac{s_x^{*2} - S_x^2}{s_x^{*2} + S_x^2} \right) \right)
\]

\[ (23) \]

\[
P_5^{GM} = \left( P_1 P_2 \right)^{1/2} = s_y^2 \left( \frac{s_x^*}{S_x} \right) \exp \left( \frac{1}{2} \left( \frac{s_x^{*2} - S_x^2}{s_x^{*2} + S_x^2} \right) \right)
\]

\[ (24) \]

\[
P_5^{HM} = \frac{2}{\left( \frac{1}{P_1} + \frac{1}{P_2} \right)} = 2s_y^2 \left( \frac{s_x^2}{S_x^2} + \exp \left( \frac{S_x^2 - s_x^{*2}}{S_x^2 + s_x^{*2}} \right) \right)
\]

\[ (25) \]

To the first degree of approximation, the mean squared errors of \( P_5^j \); \( j = AM, GM, HM \) are respectively given by
\[MSE\left(P^j_3\right) = S_y^4 f \left( \lambda_{40} - 1 \right) + \frac{g^2}{16} \left( \lambda_{04} - 1 \right) - \frac{3g}{2} \left( \lambda_{22} - 1 \right)\]  

(26)

The estimators based on \( P_o, P_1 \) and \( P_2 \)

We suggest the following estimators of \( S_y^2 \) based on AM, GM and HM of the estimators \( P_o, P_1 \) and \( P_2 \) respectively as

\[P_o^{AM} = \frac{P_o + P_1 + P_2}{3} = \left( \frac{s^2_y}{3} \right) \left( 1 + \frac{s^2_x}{S^2_e} + \exp \left( \frac{s^2_x - S^2_e}{s^2_x + S^2_e} \right) \right)\]  

(27)

\[P_o^{GM} = \left( P_o P_1 P_2 \right)^{\frac{1}{3}} = \left( \frac{s^2_y}{S^2_e} \right) \exp \left( \frac{1}{3} \frac{s^2_x - S^2_e}{s^2_x + S^2_e} \right)\]  

(28)

\[P_o^{HM} = \frac{3}{\left( \frac{1}{P_o} + \frac{1}{P_1} + \frac{1}{P_2} \right)} = \frac{3s^2_y}{\left( 1 + \frac{s^2_y}{s^2_x} + \exp \left( \frac{s^2_x - S^2_e}{s^2_x + S^2_e} \right) \right)}\]  

(29)

To the first degree of approximation, the mean squared errors of \( P_o^j ; (j = \text{AM, GM, HM}) \) are respectively given by

\[MSE(P_o^j) = S_y^4 f \left( \lambda_{40} - 1 \right) + \frac{g^2}{4} \left( \lambda_{04} - 1 \right) - g \left( \lambda_{22} - 1 \right)\]  

(30)

**Almost Unbiased Estimator**

Motivated by Singh et al. (2007), we suggest a class of estimator of population variance.

Suppose \( P_o = s_y^2, \quad P_1 = s_y^2 \left[ \frac{s_x^2}{S^2_x} \right] \quad \text{and} \quad P_2 = s_y^2 \exp \left[ \frac{s^2_x - S^2_e}{S^2_x + s^2_x} \right], \)

such that \( (P_o, P_1, P_2) \in \mathbb{T} \), where \( \mathbb{T} \) denotes the set of all possible estimators for estimating the population variance \( S_y^2 \). By definition, the set \( \mathbb{T} \) is a linear variety if
\[ P_t = \sum_{i=0}^{2} t_i P_i \quad (\in T) \]  

(31)

for

\[ \sum_{i=0}^{2} t_i = 1 \quad (t_i \in \mathbb{R}) \]  

(32)

where \( t_i \); \( i = 0, 1, 2 \) denotes the scalar constants used for reducing the bias in the set of real numbers.

Expressing the estimator \( P_i \) in terms of \( \omega \)'s, we have

\[ P_i = S_y^2 \left( 1 + \omega_o - \omega_i g \left( t_1 + \frac{t_2}{2} \right) - \frac{1}{8} \omega_i^2 g^2 t_2 - \omega_o \omega_i g \left( t_1 + \frac{t_2}{2} \right) \right) \]  

(33)

Subtracting \( S_y^2 \) from equation (33), we get

\[ P_i - S_y^2 = S_y^2 \left( \omega_o - \omega_i g t - \frac{1}{8} \omega_i^2 g^2 t_2 - \omega_o \omega_i g t \right) \]  

(34)

where

\[ t = t_1 + \frac{t_2}{2} \Rightarrow 2t = 2t_1 + t_2 \]  

(35)

After squaring both sides of equation (34) and taking expectations, we get MSE of the estimator \( P_i \) up to the first order of approximation as

\[ \text{MSE}(P_i) = S_y^4 \left[ \lambda_{4+1} + g^2 \mu^2 (\lambda_{0+1} - 1) - 2g t (\lambda_{2+1} - 1) \right] \]  

(36)

Differentiating equation (36) with respect to \( t \), we get minimum MSE(\( P_i \)) at

\[ t = \left( \frac{\lambda_{2+1} - 1}{g (\lambda_{0+1} - 1)} \right) = L \text{(say)} \]  

(37)
Thus the minimum MSE of $P_i$ is given by

$$MSE_{\min}(P_i) = S_j^4 \left[ (\lambda_{40} - 1) - \frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)} \right]$$  \hspace{1cm} (38)

From equation (35) and (37), we have

$$2t_1 + t_2 = 2L$$  \hspace{1cm} (39)

From equation (32) and (39), we have only two equations in three unknowns, so we cannot find unique values of $t_i$'s, $i = 0, 1, 2$. In order to get unique values of $t_i$'s, we shall impose the linear restriction:

$$\sum_{i=0}^{2} t_i B(P_i) = 0$$  \hspace{1cm} (40)

where $B(P_i)$ denotes the bias in the $i^{th}$ estimator.

Equations (32), (39) and (40) can be written in the matrix form as

$$\begin{bmatrix}
1 & 1 & 1 \\
0 & 2 & 1 \\
0 & B(P_1) & B(P_2)
\end{bmatrix} \begin{bmatrix}
t_o \\
t_1 \\
t_2
\end{bmatrix} = \begin{bmatrix}
1 \\
2L \\
0
\end{bmatrix}$$  \hspace{1cm} (41)

Using equation (41), we get the unique values of $t_i$'s, $i = 0, 1, 2$ as

$$\begin{cases}
t_o = 1 - L + 4L^2 \\
t_1 = L + 4L^2 \\
t_2 = -8L^2
\end{cases}$$  \hspace{1cm} (42)

Use of these $t_i$'s, $i = 0, 1, 2$ remove the bias of the estimator $P_i$ up to terms of first order at (31).

**Efficiency Comparisons**

From equations (7), (10), (11), (12), (13), (14), (18), (22), (26), (30) and (38), we have

1. $\text{MSE}(P_0) \geq \text{MSE}(P_i)$ if $g \leq \frac{2(\lambda_{22} - 1)}{(\lambda_{04} - 1)}$

2. $\text{MSE}(P_0) \geq \text{MSE}(P_2)$ if $g \leq \frac{4(\lambda_{22} - 1)}{(\lambda_{04} - 1)}$
3. \( \text{MSE}(P_o) \geq \text{MSE}(P_{IR}) \) if \( \frac{\lambda_{22} - 1}{\lambda_{04} - 1} \geq \frac{1}{2} \)

4. \( \text{MSE}(P_o) \geq \text{MSE}(P_{\exp}) \) if \( \frac{\lambda_{22} - 1}{\lambda_{04} - 1} \geq \frac{1}{4} \)

5. \( \text{MSE}(P_{IR}) \geq \text{MSE}(P_1) \) if

Case 1 \( g \geq 1 \) then \( g \leq \frac{2(\lambda_{22} - 1)}{\lambda_{04} - 1} - 1 \)

Case 2 \( g \leq 1 \) then \( g \geq \frac{2(\lambda_{22} - 1)}{\lambda_{04} - 1} - 1 \)

6. \( \text{MSE}(P_2) \leq \text{MSE}(P_{\exp}) \) if

Case 1 \( g \geq 1 \) then \( g \leq \frac{4(\lambda_{22} - 1)}{\lambda_{04} - 1} - 1 \)

Case 2 \( g \leq 1 \) then \( g \geq \frac{4(\lambda_{22} - 1)}{\lambda_{04} - 1} - 1 \)

7. \( \text{MSE}(P_o) \geq \text{MSE}_{\text{min}}(P_{\text{reg}}) \) if \( \frac{(\lambda_{22} - 1)^2}{\lambda_{04} - 1} \geq 0 \)

8. \( \text{MSE}(P_o) \geq \text{MSE}(P_j) \) if \( g \leq \frac{4(\lambda_{22} - 1)}{\lambda_{04} - 1} \); \( j = (\text{AM, GM, HM}) \)

9. \( \text{MSE}(P_o) \geq \text{MSE}(P_j) \) if \( g \leq \frac{8(\lambda_{22} - 1)}{\lambda_{04} - 1} \); \( j = (\text{AM, GM, HM}) \)

10. \( \text{MSE}(P_o) \geq \text{MSE}(P_j) \) if \( g \leq \frac{8(\lambda_{22} - 1)}{3(\lambda_{04} - 1)} \); \( j = (\text{AM, GM, HM}) \)

11. \( \text{MSE}(P_o) \geq \text{MSE}(P_j) \) if \( g \leq \frac{4(\lambda_{22} - 1)}{\lambda_{04} - 1} \); \( j = (\text{AM, GM, HM}) \)

12. \( \text{MSE}(P_o) \geq \text{MSE}_{\text{min}}(P_1) \) if \( \frac{(\lambda_{22} - 1)^2}{\lambda_{04} - 1} \geq 0 \)
Empirical Study

For empirical study, we consider three natural population data sets.

**Population 1**: Murthy (1967, p.226)

Y: Output
X: Number of workers.

\[ N = 80, \ n = 10, \ \lambda_{40} = 2.2667, \lambda_{04} = 3.65 \lambda_{22} = 2.3377. \]

**Population 2**: Murthy (1967, p.127)

Y: Cultivated area (in acres).
X: Area in square miles.

\[ N = 80, \ n = 10, \lambda_{40} = 2.373, \lambda_{04} = 2.0193 \lambda_{22} = 1.6757. \]

**Population 3**: Sukhatme and Sukhatme (1970, p.185)

Y: Area under wheat in 1937 (in acres).
X: Cultivated area in 1931.

\[ N = 80, \ n = 10, \lambda_{40} = 3.5469, \lambda_{04} = 3.2816 \lambda_{22} = 2.6601. \]

**Table 1**: Value of \( t_i^{'},(i = 0,1,2) \)

<table>
<thead>
<tr>
<th>Scalars</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>( t_0 )</td>
<td>47.410275</td>
</tr>
<tr>
<td>( t_1 )</td>
<td>53.477369</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>-99.887644</td>
</tr>
</tbody>
</table>

Using these values of \( t_i^{'},(i = 0,1,2) \) given in Table 1, one can reduce the bias to the first order approximation in the estimators \( P_i \) at (31).
Table 2: Percent relative efficiency of different estimators with respect to $S^2_γ$

<table>
<thead>
<tr>
<th>Scalars</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>$P_o$</td>
<td>100.0000</td>
</tr>
<tr>
<td>$P_1$</td>
<td>134.9590</td>
</tr>
<tr>
<td>$P_2$</td>
<td>116.3049</td>
</tr>
<tr>
<td>$P_{IR}$</td>
<td>102.0462</td>
</tr>
<tr>
<td>$P_{exp}$</td>
<td>214.1504</td>
</tr>
<tr>
<td>$P_{reg}$</td>
<td>214.1725</td>
</tr>
<tr>
<td>$p^j_3$</td>
<td>116.3049</td>
</tr>
<tr>
<td>$p^j_4$</td>
<td>107.8474</td>
</tr>
<tr>
<td>$p^j_5$</td>
<td>125.3574</td>
</tr>
<tr>
<td>$p^j_6$</td>
<td>116.3049</td>
</tr>
<tr>
<td>$P_t$</td>
<td>214.1725</td>
</tr>
</tbody>
</table>

**Conclusion**

From Table 2, we observe that the estimators constructed by taking AM, GM and HM of the usual estimator $S^2_γ$, dual to Isaki (1983) estimator and dual to exponential ratio type variance estimator performs better than $S^2_γ$. From Table 2, we also observe that dual to exponential ratio type variance estimator performs inferior than exponential ratio type estimator. Reason is that the condition (6) in efficiency comparison is rarely met in practice. It is not advisable to construct dual to exponential ratio type variance estimators. The unbiased estimator $P_t$ performs better than all other estimators considered here.
References


