Estimation of Population Variance in Simple Random Sampling using Auxiliary Information

Sumaira Ajmal Khan
Lahore Garrison University,
Phase 6, DHA Lahore

Mehwish Nawaz
Lahore Garrison University,
Phase 6, DHA Lahore

Kushbikht M. Din
Lahore Garrison University,
Phase 6, DHA Lahore

ABSTRACT
In this paper we propose a new estimator for the population variance using auxiliary information in simple random sampling. We derived a bias and mean square error equation of proposed estimator and compare with the bias and MSE of existing estimator and show that proposed estimator is more efficient than the existing estimators suggested by different authors such that Kadilar and Cingi (2005) [6] , Isaki(1983) [5]. We support this theoretical result with the help of a numerical illustration.

Keywords
Variance estimator, bias, MSE, simple random sampling, auxiliary information, Efficiency

1. INTRODUCTION
Auxiliary information has been used extensively in estimation of parameters like mean and variance since several decades. Isaki (1983) [5] got inspiration from ratio estimator of finite population mean and proposed a ratio estimator, usually called classical estimator, of finite population variance. But Singh and Solanki (2013a) [12] claimed that Isaki (1983) [5] ratio estimator is the member of the class of estimators developed by Das and Tripathi (1978) [3]. Arcos and Rueda (1997) [2] suggested multivariate ratio estimator for population variance. Ahmed et al. (2000) [1] criticized the claim of Arcos and Rueda (1997) [2]. Kadilar and Cingi (2006a) [7] developed an estimator, ratio-type estimator of the mean of population. Kadilar and Cingi (2006b) [8] extended the idea of Isaki (1983) [5] ratio estimator for population variance. Kadilar and Cingi (2006b) [8] involved the information available about coefficient of variation and coefficient of kurtosis of the auxiliary variable to generate these estimators under simple random sampling as well. Gupta and Shabbir (2008) [4] gave a hybrid class of variance estimators for population mean. Subramani and Kumarapandian (2012a) [9] modified the usual ratio-type estimator of Kadilar and Cingi (2006b) [8] for population variance using population median obtained from auxiliary variable. Subramani and Kumarapandian (2012b) [10] further modified the usual ratio-type variance estimators using lower and upper quartiles, inter-quartile range, quartile deviation and quartile average of the auxiliary variable. Subramani and Kumarapandian (2013) [11] developed another more efficient modified ratio-type estimator using median and coefficient of variation of the auxiliary variable.

2. NOTATIONS

\( N \) Population size      \( n \) sample size
\[
\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \quad \text{sample mean of the study variable } y
\]
\[
\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{sample mean of the auxiliary variable } x
\]
\[
S_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{N-1} \quad \text{Population variance of the study variable}
\]
\[
S_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{N-1} \quad \text{Population variance of the auxiliary variable}
\]
\[
S_{xy}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{N-1} \quad \text{sample variance of the study variable}
\]
\[
S_{xy}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1} \quad \text{sample variance of the auxiliary variable}
\]
\[
\lambda' = \frac{1}{n} \quad C_x, C_y = \text{Coefficient of variations}
\]
\[
\beta_1(x) = \frac{\mu_x^3}{\mu_x^2} \quad \text{Skewness of the auxiliary variable}
\]
\[
\beta_2(x) = \frac{\mu_x^4}{\mu_x^2} \quad \beta_2(y) = \frac{\mu_y^4}{\mu_y^2} \quad \text{Kurtosis of the study variable}
\]
\[
E(e_0) = 0 \quad , \quad E(e_1) = 0
\]

3. EXISTING ESTIMATORS IN SIMPLE RANDOM SAMPLING

3.1 Isaki (1983) [5]
Motivated by the estimator of the population mean \( \hat{y}_{PR} = \bar{y} \) Isaki [5] suggested the ratio estimator of the variance of population using auxiliary information.

\[
t_{isaki} = \frac{s_{xy}^2}{s_x^2} \quad \text{.............. (1)}
\]

The MSE of the above estimator using first order approximation is

\[
\text{MSE}(t_{isaki}) \approx \lambda' s_y^2 (\beta_2(y) + \beta_2(x) - 2h)
\]

Or

\[
\text{MSE}(t_{isaki}) \approx \lambda' s_y^2 (\beta_2(y) + \beta_2(x) - 2h')
\]

The bias is

\[
\text{Bias}(t_{isaki}) \approx \lambda' s_y^2 (\beta_2(x) - h)
\]

Or

\[
\text{Bias}(t_{isaki}) \approx \lambda' s_y^2 (\beta_2(x) - h')
\]

3.2 Kadilar and Cingi (2005) [6]
Kadilar and Cingi [6] proposed the following class of estimators for population variance.

\[
t_{ch1} = s_y^2 [s_x^2 - \beta_2(x)]^{-1} [s_x^2 - \beta_2(x)]
\]
The MSEs of the above estimators are given below respectively:

\[ \text{MSE}(t_{CH1}) \equiv \lambda' S^2 \left[ \beta_2'(y) - 2 \frac{s_y^2}{s_x^2 - \beta_1(x)} h' + \left( \frac{s_y^2}{s_x^2 - \beta_1(x)} \right)^2 \beta_2'(x) \right] \]

\[ \text{MSE}(t_{CH2}) \equiv \lambda' S^2 \left[ \beta_2'(y) - 2 \frac{s_y^2}{s_x^2 - \beta_1(x)} h' + \left( \frac{s_y^2}{s_x^2 - \beta_1(x)} \right)^2 \beta_2'(x) \right] \]

\[ \text{MSE}(t_{CH3}) \equiv \lambda' S^2 \left[ \beta_2'(y) - 2 \frac{s_y^2}{s_x^2 - \beta_1(x)} h' + \left( \frac{s_y^2}{s_x^2 - \beta_1(x)} \right)^2 \beta_2'(x) \right] \]

\[ \text{MSE}(t_{CH4}) \equiv \lambda' S^2 \left[ \beta_2'(y) - 2 \frac{s_y^2}{s_x^2 - \beta_1(x)} h' + \left( \frac{s_y^2}{s_x^2 - \beta_1(x)} \right)^2 \beta_2'(x) \right] \]

The biases are:

\[ \text{Bias}(t_{CH1}) = \lambda' S^2 \left[ \frac{s_y^2}{s_x^2 - \beta_1(x)} \right] \beta_2'(x) - h' \]

\[ \text{Bias}(t_{CH2}) = \lambda' S^2 \left[ \frac{s_y^2}{s_x^2 - \beta_1(x)} \right] \beta_2'(x) - h' \]

\[ \text{Bias}(t_{CH3}) = \lambda' S^2 \left[ \frac{s_y^2}{s_x^2 - \beta_1(x)} \right] \beta_2'(x) - h' \]

\[ \text{Bias}(t_{CH4}) = \lambda' S^2 \left[ \frac{s_y^2}{s_x^2 - \beta_1(x)} \right] \beta_2'(x) - h' \]

3.3 Subramani and Kumarapandiyan (2012b) [10]

Subramani and Kumarapandiyan [10] suggested the following efficient estimator using quartiles and functions of quartiles of the auxiliary variable.

\[ t_{ip} = S^2(y_0 - Q_1) - (s_x^2 + Q_1) \]

\[ t_{it} = S^2(y_0 - Q_3) - (s_x^2 + Q_3) \]

\[ t_{tg} = S^2(y_0 - Q_1) - (s_x^2 + Q_1) \]

\[ t_{tg} = S^2(y_0 - Q_3) - (s_x^2 + Q_3) \]

\[ t_{tg} = S^2(y_0 - Q_1) - (s_x^2 + Q_1) \]

\[ t_{tg} = S^2(y_0 - Q_3) - (s_x^2 + Q_3) \]

Where

\[ Q_R = \frac{Q_1 - Q_3}{2} \]

\[ Q_D = \frac{Q_1 + Q_3}{2} \]

And their respective mean squared errors are:

\[ \text{MSE}(t_{ip}) = \lambda' S^2 \left[ \frac{s_y^2}{s_x^2 + Q_1} \right] \beta_2'(x) - h' \]

\[ \text{MSE}(t_{it}) = \lambda' S^2 \left[ \frac{s_y^2}{s_x^2 + Q_3} \right] \beta_2'(x) - h' \]

\[ \text{MSE}(t_{tg}) = \lambda' S^2 \left[ \frac{s_y^2}{s_x^2 + Q_1} \right] \beta_2'(x) - h' \]

\[ \text{MSE}(t_{tg}) = \lambda' S^2 \left[ \frac{s_y^2}{s_x^2 + Q_3} \right] \beta_2'(x) - h' \]

\[ \text{MSE}(t_{tg}) = \lambda' S^2 \left[ \frac{s_y^2}{s_x^2 + Q_1} \right] \beta_2'(x) - h' \]

\[ \text{MSE}(t_{tg}) = \lambda' S^2 \left[ \frac{s_y^2}{s_x^2 + Q_3} \right] \beta_2'(x) - h' \]

4. PROPOSED ESTIMATOR

We proposed a new modified ratio type variance estimator of the auxiliary variable. The modified ratio type variance estimator for population variance is defined as

\[ t_{new} = S^2(y_0 + e_0) \left( \frac{S^2(y_1) + S^2(y_2)}{S^2(y_1) + S^2(y_2) + S^2(y_3)} \right) \]

4.1 Bias of Proposed Estimator

\[ t_{new} = S^2(y_0 + e_0) \left( \frac{S^2(y_1) + S^2(y_2)}{S^2(y_1) + S^2(y_2) + S^2(y_3)} \right)^{-1} \]

\[ t_{new} = S^2(y_0 + e_0) \left( \frac{S^2(y_1) + S^2(y_2)}{S^2(y_1) + S^2(y_2) + S^2(y_3)} \right)^{-1} \]

Where

\[ \Omega = \frac{S^2(y_0 + e_0)}{S^2(y_1) + S^2(y_2) + S^2(y_3)} \]

Expanding the expression by using Taylor’s series

\[ t_{new} = S^2(y_0 + e_0) \left( 1 - \Omega e_1 + (\Omega e_1)^2 - \Omega e_1 e_2 \right) \]

\[ t_{new} = S^2(y_0 + e_0) \left( 1 - \Omega e_1 + (\Omega e_1)^2 - \Omega e_1 e_2 \right) \]

\[ t_{new} = S^2(y_0 + e_0) \left( 1 - \Omega e_1 + (\Omega e_1)^2 - \Omega e_1 e_2 \right) \]

\[ E(t_{new}) = E \left( \frac{S^2(y_0 - \Omega e_1 + (\Omega e_1)^2 - \Omega e_1 e_2)}{S^2(y_0 + e_0)} \right) \]

\[ B(t_{new}) = \Omega \lambda S^2 \left( \frac{\beta_2'(x)}{\beta_2(x)} - h' \right) \]

4.2 Mean Square Error of the proposed Estimator

Now comes MSE of the estimator

\[ \text{MSE}(t_{new}) = E \left( \frac{S^2(y_0 - \Omega e_1)}{S^2(y_0 + e_0)} \right)^2 \]

And to ignore higher order

\[ = S^2 \left( \frac{E \left( y_0 \right) + \Omega^2 E \left( e_1 \right) - 2\Omega E \left( e_0 e_1 \right) \right)}{} \]

\[ = S^2 \left( \frac{\lambda \beta_2'(y_0) + \Omega^2 \beta_2'(x) - 2\Omega h' \right)}{} \]

\[ = S^2 \left( \frac{\beta_2'(y_0) + \beta_2'(x)}{\beta_2(x)} - h' \right)}{} \]

\[ = S^2 \left( \frac{\beta_2'(y_0) + \beta_2'(x)}{\beta_2(x)} - h' \right)}{} \]

\[ = S^2 \left( \frac{\beta_2'(y_0) + \beta_2'(x)}{\beta_2(x)} - h' \right)}{} \]

Where

\[ \varphi = \frac{h'}{\beta_2(x)} \]

In order to minimize MSE to differentiate (i) partially w.r.t $\varphi$ and equating to zero

\[ \frac{\partial}{\partial \varphi} \text{MSE}(t_{new}) = 0 \]

\[ \lambda S^2 + \frac{\partial}{\partial \varphi} \left( \frac{\beta_2'(y_0) + \beta_2'(x)}{\beta_2(x)} - 2\Omega \varphi \right) = 0 \]

\[ \Omega = \varphi \]
\[ \lambda^2 S_\beta^2(y) = \lambda^2 S_\beta^2(y) \left( 1 - \rho^2 \frac{\beta_2(x)}{\beta_2(y)} \right) \]

\[ \text{MSE}(t_{new}) = \lambda^2 S_\beta^2(y) \left( 1 - \left( \rho' \right)^2 \right) \]

Where

\[ \rho' = \frac{\kappa'}{\sqrt{\beta_2(x)\beta_2(y)}} \]

5. PROPOSED ESTIMATOR

To demonstrate the performance of the proposed estimator empirically in comparison to other estimators, I have used five data sets. The description of data sets is given below.

5.1 Data 1

Source: Gupta and Shabbir (2008) [4]
The available statistics are based on the data obtained from 104 villages existing in East Anatolia (Turkey) in 1999. \( Y \), the variable of interest represents apple’s production level per 100 tons. \( X \), the auxiliary variable shows number of apple trees (in 100s).

5.2 Data 2

Source: Gupta and Shabbir (2008) [4]
The data collected from 278 villages/Towns/wards under the control of Gajole polstation (Malda district, west Bengal) India. \( Y \), indicates number of agricultural labourers for the year 1971. \( X \), the number of agricultural labourers for the year 1961.

5.3 Data 3

Source: Rohini et al (Jan 2012)

Data collected from 142 Indian cities with population 0.1 million and above. \( Y \): 1971 population census. \( X \): 1961 population census.

5.4 Data 4

Source: Kadilar and Cingi (2005) [6] ratio estimators for the population variance in simple random sampling.

The data obtained from 94 villages in Mediterranean (Turkey) in 1999. \( Y \): the variable of interest represents apple’s production level per 100 tons. \( X \): the auxiliary variable shows number of apple trees (in 100 sec).

5.5 Data 5

Source: Kadilar and Cingi (2005) [6] ratio estimators for the population variance in simple random sampling.

The available statistics obtained from 173 villages existing in East and southeast Turkey. \( Y \): the variable of interest represents apple’s production level per 100 tons. \( X \): the auxiliary variable shows no of trees (1 unit shows 100 apple trees).

| Estimators | Data 1 | Data 2 | Data 3 | Data 4 | Data 5 |
|------------|--------|--------|--------|--------|--------|
| Isaki (1983) | 21.23184951 | 1282.9966 | 7900281.139 | 12569.0763 | 224.623916 |
| Kadilar & Cingi (2005) | 21.23185393 | 1413.1212 | 7900391.028 | 12569.346 | 225.768988 |
| Subramani & Kumarapandiyam (2012b) | 21.23185393 | 1413.1212 | 7900391.028 | 12569.346 | 225.768988 |
| Proposed Estimator | 21.13030158 | -350.02315 | 7210041.419 | 11436.3159 | 58.1996276 |

| \( \beta_2(x) \) | 16.523 | 25.896 | 40.8536 | 24.1 | 27.9 |
| \( \beta_2(y) \) | 16.523 | 25.896 | 40.8536 | 24.1 | 27.9 |
| \( \beta_2(x) \) | 16.523 | 25.896 | 40.8536 | 24.1 | 27.9 |
| \( \beta_2(y) \) | 16.523 | 25.896 | 40.8536 | 24.1 | 27.9 |
| \( \beta_2(y) \) | 16.523 | 25.896 | 40.8536 | 24.1 | 27.9 |

Table 1 Bias of the Existing and the Proposed Estimator
Table 2: MSE of the Existing and the Proposed Estimators

| Estimators                | Data 1          | Data2          | Data3          | Data4          | Data5          |
|--------------------------|-----------------|----------------|----------------|----------------|----------------|
| Isaki (1983)             | t_{Isaki}       | 4862.205231    | 3778793.028    | 1.92215E+14    | 1827374379    | 39643.29503   |
| Kadilar & Cingi (2005)   | t_{CH1}         | 4862.205422    | 3983111.977    | 1.92216E+14    | 1827397122    | 39675.37353   |
|                          | t_{CH2}         | 4862.885344    | 4862.885344    | 4862.885344    | 4862.885344    | 4862.885344   |
|                          | t_{TCH3}        | 4862.205422    | 4862.205422    | 4862.205422    | 4862.205422    | 4862.205422   |
|                          | t_{CH4}         | 4862.885467    | 4862.885467    | 4862.885467    | 4862.885467    | 4862.885467   |
| Subramani & Kumarapandiyan (2012b) | t_{fB1} | 4862.204945    | 3468925.171    | 1.92214E+14    | 1827353716    | 39594.00714   |
|                          | t_{fB2}         | 4862.204372    | 3022246.855    | 1.92212E+14    | 1827312394    | 39496.7733    |
|                          | t_{fB3}         | 4862.204659    | 3220379.202    | 1.92213E+14    | 1827333054    | 39545.16751   |
|                          | t_{fB4}         | 4862.204945    | 3468925.171    | 1.92214E+14    | 1827353716    | 39594.00714   |
|                          | t_{fB5}         | 4862.204659    | 3220379.202    | 1.92213E+14    | 1827333054    | 39545.16751   |
| Proposed Estimator       | t_{new}         | 4857.818785    | 2846652.986    | 1.84037E+14    | 1792753628    | 36249.92917   |

Table 3: PRE of the Existing and the Proposed Estimators

| Estimators                | Data 1          | Data2          | Data3          | Data4          | Data5          |
|--------------------------|-----------------|----------------|----------------|----------------|----------------|
| Isaki (1983)             | t_{Isaki}       | 4862.205231    | 3778793.028    | 1.92215E+14    | 1827374379    | 39643.29503   |
| Kadilar & Cingi (2005)   | t_{CH1}         | 4862.205422    | 3983111.977    | 1.92216E+14    | 1827397122    | 39675.37353   |
|                          | t_{CH2}         | 4862.885344    | 4862.885344    | 4862.885344    | 4862.885344    | 4862.885344   |
|                          | t_{TCH3}        | 4862.205422    | 4862.205422    | 4862.205422    | 4862.205422    | 4862.205422   |
|                          | t_{CH4}         | 4862.885467    | 4862.885467    | 4862.885467    | 4862.885467    | 4862.885467   |
| Subramani & Kumarapandiyan (2012b) | t_{fB1} | 4862.204945    | 3468925.171    | 1.92214E+14    | 1827353716    | 39594.00714   |
|                          | t_{fB2}         | 4862.204372    | 3022246.855    | 1.92212E+14    | 1827312394    | 39496.7733    |
|                          | t_{fB3}         | 4862.204659    | 3220379.202    | 1.92213E+14    | 1827333054    | 39545.16751   |
|                          | t_{fB4}         | 4862.204945    | 3468925.171    | 1.92214E+14    | 1827353716    | 39594.00714   |
|                          | t_{fB5}         | 4862.204659    | 3220379.202    | 1.92213E+14    | 1827333054    | 39545.16751   |
| Proposed Estimator       | t_{new}         | 4857.818785    | 2846652.986    | 1.84037E+14    | 1792753628    | 36249.92917   |

6. CONCLUSION

In this article we have proposed a modified ratio-type variance estimator using known value of kurtosis of the auxiliary variable. The bias and mean square error of the proposed modified ratio-type variance estimator are obtained and compared with that of existing modified ratio type variance estimator and show that proposed estimator is more efficient than the existing estimator. We have also assessed the performances of the proposed estimator for known population.

7. REFERENCES

[1] Ahmed, M.S. and Hossain, M.I. (2000) . Some competitive estimators of finite Population variance Multivariate Auxiliary Information, Information and Management Sciences, Volume 11 (1), 49-54

[2] Arcos , C.A. and Rueda , G.M , variance estimation using auxiliary information an almost unbiased multivariate ratio estimator. Matrika , Vol.45, pp.171-178, 1997.

[3] Das, A.K, & Tripathi, T.P. (1978). Use of auxiliary information in estimating the finite population variance . Sankhya , 40, 139-148.

[4] Gupta , S , & Shabbir, J. (2008). Variance estimation in simple random sampling using auxiliary information . Hacettepe Journal of Mathematics and Statistics, 37, 57-67.

[5] Isaki , C.T . (1983) . Variance estimation using auxiliary information . Journal of the American Statistical Association, 78, 117-123.

[6] Kadilar , C , Cingi ,H. (2005) . A new ratio estimator in stratified sampling . Comm. Statist. Theory Meth. 34:1-6.

[7] Kadilar , C , Cingi ,H. (2006a). Improvement in variance estimation using auxiliary information . Hacettepe Journal of Mathematics and Statistics , 35(1) , 111-115.

[8] Kadilar , C , Cingi ,H. (2006b) . Ratio estimators for population variance in simple and stratified sampling . Applied Mathematics and Computation , 173, 1047-1058.
[9] Subramani, J, & Kumarapandiyan, G. (2012a). Estimation of population mean using coefficient of variation and median of an auxiliary variable. International Journal of Probability and Statistics, 1(4), 111-118.

[10] Subramani, J, & Kumarapandiyan, G. (2012b). Variance Estimation using median of an auxiliary variable. International Journal of Probability and Statistics, 1(3), 36-40.

[11] Subramani, J, & Kumarapandiyan, G. (2013). Estimation of variance using known coefficient of variation and Median of an auxiliary variable. International Journal of Modern Applied Statistical Methods, 1(12), 58-64.