**Original Research Article**

**Generalized Almost Unbiased Estimator of Finite Population Variance under Stratified Random Sampling**

**Abstract**

This study developed a generalized almost unbiased estimator for obtaining finite population variance under stratified random sampling. The generalized estimator was developed using the approach of almost unbiased. This approach uses filtration parameters in reducing the bias of estimators. The theoretical properties of the generalized estimators were derived along with the expressions of the filtration parameters. The efficiency conditions of the proposed estimator over some existing population variance estimators were obtained theoretically. The performance of the proposed estimator was evaluated empirically using simulated and real-life data sets. The generalized estimator performed better with minimum values of bias, mean square error, and higher percentage relative efficiency. Therefore, the generalized almost unbiased estimator can be utilized to provide better variance estimates for various phenomena such as inflation variation over years, exchange rate variation over years and standard of living variation over for better policymaking.

**Keywords:** Estimator, variance, unbiased, mean square error, efficiency.

1. **Introduction**

Random sampling techniques such simple random sampling, systematic random sampling, stratified random sampling and cluster random sampling are sampling method of selecting a sample from a population where every unit or item has a known probability of being chosen. This approach helps to ensure that the sample is representative of the population and reduce bias [1]. However, when the population units of the variables of interest are often segregated due to a variety of factors which makes the population units heterogeneous in nature. Such circumstances preclude the application of sampling methods such as simple random sampling and systematic random sampling. In order to produce more precise results, stratified random sampling is utilized [1]. Stratification often help in enhancing the estimates accuracy of population characteristics such as mean, variance, total and proportion when the population units are heterogeneous in nature. The utilization of Estimators that relied on stratified random sampling may be helpful to produce more precise results, lowers survey costs and also increased administrative effectiveness. A variety of population characteristics have been estimated under this sampling approach by [2], [3], [4], [5], [6] and [7].

Variance is a vital concept in estimation theory as it provides a numerical value that describes the amount of variation or dispersion in a dataset [8]. Variance as tool can assist in measuring the volatility of financial assets such as stocks and bond [8]. It can be utilized to monitor and control the quality of products, identifying variations that may indicate problems in the manufacturing process. Variance aid in research to understand the variability of outcomes, identify patterns and make predictions [8].

Estimation of population variance for various phenomena such as standard of living, weather condition among others, is a crucial component in developing better policy. Considerable attention has been made in developing efficient estimators under different sampling methods. For example, researchers including, [9], [10], [11], [12], [13], among others contributed in this area. However, some of the existing estimators are biased which may lead to either over- or under-estimation. In addition, the nature of relationship between the study and auxiliary variables which makes the existing variance estimators produce better estimate is another matter of concern. Therefore, in order to address these two problems, the current study developed a generalized almost unbiased estimator for obtaining variance of a finite population under stratified random sampling.

The rest of the paper is structured as follows: Section 2 presents the sample structure, notations and some related estimators, while the suggested variance estimator and its properties are presented in Section 3. The empirical results using simulated and real data sets are presented in Section 4 and Section 5 presents the relevant conclusion.

**1.1 Sample structure, Notations and Some Related Estimators**

Consider $ϑ=\left\{ϑ\_{1}, ,,,,,ϑ\_{N}\right\}$ to be a finite population of size *N* which is split into G non-overlapping strata or homogeneous layers, with each stratum containing $N\_{h}(h=1,2,…,G)$ units, such that $\sum\_{h=1}^{G}N\_{h}=N. $ A sample using the technique of simple random sample of size $n\_{h}$ is drawn *(SRSWOR)* from the population strata $N\_{h}$ without replacement such that$ \sum\_{h=1}^{G}n\_{h}=n$. Assume ($y\_{hi},x\_{hi}$) be pair of sample observations of the study variable *Y* and the auxiliary variable *X* on ith unit$ ϑ\_{i}, i=1,…N$. Then, $\overline{y}\_{h} $and$ \overline{x}\_{h}$ denotes the sample means corresponding to the population means$ \overline{Y}\_{h}$, and$\overline{ X}\_{h}$, respectively, in each stratum. Further, the following notations were defined in [13] as;

$S\_{Y\_{h}}^{2}=\frac{\sum\_{i=1}^{N\_{h}}\left(Y\_{hi}- \overline{Y}\_{h}\right)^{2}}{\left(N\_{h}-1\right)}$: The population variance of the study variable *Y* for the $i^{th}$ stratum

$S\_{X\_{h}}^{2}=\frac{\sum\_{i=1}^{N\_{h}}\left(X\_{hi}- \overline{X}\_{h}\right)^{2}}{\left(N\_{h}-1\right)}$: The population variance of the auxiliary variable *X* for the $i^{th}$ stratum

$s\_{yh}^{2}=\frac{\sum\_{i=1}^{n\_{h}}\left(y\_{hi}-\overline{y}\_{h}\right)^{2}}{\left(n\_{h}-1\right)}$: The sample variance of the study variable *X* for the $i^{th}$ stratum

$s\_{xh}^{2}=\frac{\sum\_{i=1}^{n\_{h}}\left(x\_{hi}-\overline{x}\_{h}\right)^{2}}{\left(n\_{h}-1\right)}$: The sample variance of the auxiliary variable *X* for the $i^{th}$ stratum

$W\_{h}=\frac{N\_{h}}{N}$: The original weight of the $h^{th}$ strata, $h=1,2,…,G$

$f\_{1h}=\left(\frac{1}{n\_{h}}-\frac{1}{N\_{h}}\right)$: The sampling fraction corresponding to the $h^{th} $strata

$μ\_{rs(h)}=\frac{1}{N\_{h}}\sum\_{j=1}^{N\_{h}}\left(Y\_{hj}- \overline{Y}\_{h}\right)^{r}\left(X\_{hj}- \overline{X}\_{h}\right)^{s}$: The moments about the mean for the $i^{th}$ stratum, where r and s will always assume those values (0, 2 and 4).

: The population kurtosis for the $i^{th}$ stratum

$ρ\_{(s\_{yh}^{2},s\_{xh}^{2})}=\frac{\left(λ\_{22(h)}-1\right)}{\sqrt{\left(λ\_{40(h)}-1\right)\left(λ\_{04(h)}-1\right)}}$: Correlation coefficient between the study and auxiliary variable

The usual unbiased population variance estimator under stratified random sampling and its variance are, respectively, given as

 (1)

 (2)

A ratio and product estimators and their properties by [2] are respectively given as

 (3)

 (4)

 (5)

 (6)

 (7)

 (8)

[14] suggested a difference type of estimator of population variance as

 (9)

 (10)

 (11)

where  and

[3] proposed exponential ratio and product-type estimators as

 (12)

 (13)

 (14)

 (15)

 (16)

 (17)

[4] proposed regression-cum-ratio estimator as

 (18)

(19)

The minimum mean square error (MSE) expression of the estimator  up to first order of approximation at optimum values;

 and 

where and , is given as

 (20)

[15] suggested an improved estimators of finite population variance in stratified random sampling as

 ,  (21)

 (22)

The minimum mean square error of the estimator  up to the first order of approximation at the optimum values;

 and 

is given as

 (23)

where

,

, ,

,

.

, . ,  , ,  and 

[5] considered the adoption of multiple and transformed auxiliary information and suggested three ratio cum exponential type estimators as

 (24)

 (25)

 (26)

The bias expression of the estimators up to the first order of approximation are, respectively, given as

 (27)

 (28)

 (29)

where

, . , ,, , ,  and 

The MSE expression of the estimators  ,  and  are, respectively, given as

 (30)

 (31)

 (32)

[6] suggested an improved population variance estimator under stratified random sampling as

 (33)

The bias and MSE expressions of the estimator  at optimum values:

,

 and



are respectively given as

 (34)

 (35)

where 

**2.0 Material and Methods**

**2.1 Proposed Estimator**

The proposed generalized population variance estimator is given as;

 (36)

where, , , and 

By definition, the class *K* will consist of all  of the form;

 (37)

where *K* denotes the set of all possible estimators for estimating the population variance  such that ,  denotes the parameters used for reducing the bias in the class of estimators .

**2.2 Bias and MSE of the Proposed Estimator** $\left(t\_{(st)}\right)$

To derive, the bias and MSE of the proposed estimator $t\_{(st)} $in equation (36), the following error terms are defined:

 and  such that , 

 (38)

Expressing $t\_{(st)}$ in equation (36) in term of error terms, we have

 (39)

By some appropriate simplification, equation (39) becomes:

 (40)

By expanding and  using power series up to the first order approximation, the equation (40) gives:

 (41)

Applying the concept of exponential series to equation (41) and multiplying out the RHS up to the first order approximation, we have

 (42)

where

 (43)

Multiplying the RHS and factorizing the common terms, the (43) becomes;

 (44)

Taking expectation of both sides of equation (44) and applying the results of (38), the bias of the proposed estimator $\left(t\_{(st)}\right)$ is obtained as

 (45)

Squaring and taking expectation of (45) and applying the results of (38), the mean square error of the proposed estimator $\left(t\_{(st)}\right) $is obtained as

 (46)

The optimum values of the filtration parameters  for are obtained by solving the three conditions using Echelon operation. Differentiating (46) partially with respect to  and equating to zero, we have

 (47)

Simplifying equation (47), we obtain

 (48)

The following additional linear restriction is imposed to obtain the unique values of 

 (49)

Expressing the three equations namely; (43), (48) and (49) in matrix form, we have

 (50)

Applying the operation  into equation (50), we have

 (51)

From (51) we obtain the optimum values of the filtration parameters  as:

 (52)

 (53)

 (54)

where:

 and

**2.3 Efficiency Comparisons**

In this subsection, efficiency conditions of the generalized estimator $t\_{(st)}$ over some related estimators were obtained as follows:

1. The proposed estimator $t\_{(st)} $is more efficient than sample variance estimator $ T\_{1(st)} $if





 (55)

where 

Since condition (55) is satisfied, the proposed estimator $t\_{(st)}$ is more efficient than$ T\_{1(st)}$.

1. The proposed estimator $t\_{(st)} $is more efficient than usual ratio estimator $T\_{2(st)} $if



 (56)

Since condition (56) is satisfied, the proposed estimator $t\_{(st)} $is more efficient than$ T\_{2(st)}$.

1. The proposed estimator $t\_{(st)} $is more efficient than usual product variance estimator $T\_{3(st)} $if



 

 (57)

Since condition (57) is satisfied, the proposed estimator $t\_{(st)}$ is more efficient than $T\_{3(st)}$.

1. The proposed estimator $t\_{\left(st\right)} $is more efficient than Rao estimator $T\_{Rao(st)}$ if





 (58)

Since condition (58) is satisfied, the proposed estimator $t\_{(st)}$ is more efficient than$ T\_{Rao(st)}$.

1. The proposed estimator $t\_{\left(st\right)} $is more efficient than Singh *et al*. (2009) ratio estimator $T\_{er(st)}$if





 (59)

Since condition (59) is satisfied, the proposed estimator $t\_{(st)}$ is more efficient than$ T\_{er(st)}$.

1. The proposed estimator $t\_{\left(st\right)} $is more efficient than Singh *et al*. (2009) product estimator $T\_{ep(st)}$ if





 (60)

Since condition (60) is satisfied, the proposed estimator $t\_{(st)}$ is more efficient than$ T\_{ep(st)}$.

1. The proposed estimator $t\_{\left(st\right)} $is more efficient than Shabbir and Gupta (2010) estimator $T\_{SG(st)}$ if





 (61)

Since condition (63) is satisfied, the proposed estimator $t\_{(st)}$ is more efficient than $T\_{SG(st)}$.

1. The proposed estimator $t\_{\left(st\right)} $is more efficient than Koyuncu (2013) estimator $T\_{Ni(st)}$ if





 (62)

Since condition (64) is satisfied, the proposed estimator $t\_{(st)}$ is more efficient than$ T\_{Ni(st)}$.

1. The proposed estimator $t\_{\left(st\right)} $is more efficient than Yasmeen and Noor (2021) ratio estimator $T\_{YN1\left(st\right)}$ if





 (63)

Since condition (63) is satisfied, the proposed estimator $t\_{(st)}$ is more efficient than$ T\_{YN1\left(st\right)}$.

1. The proposed estimator $t\_{\left(st\right)} $is more efficient than Yasmeen and Noor (2021) exponential ratio estimator $T\_{YN2\left(st\right)}$ if





 (64)

Since condition (64) is satisfied, the proposed estimator $t\_{(st)}$ is more efficient than$ T\_{YN2\left(st\right)}$.

1. The proposed estimator $t\_{\left(st\right)} $is more efficient than Yasmeen and Noor (2021) exponential product estimator $T\_{YN3\left(st\right)}$ if





 (65)

Since condition (65) is satisfied, the proposed estimator $t\_{(st)}$ is more efficient than$ T\_{YN3\left(st\right)}$.

1. The proposed estimator $t\_{\left(st\right)} $is more efficient than Ahmad *et al*. (2022) estimator $T\_{α\left(st\right)}$ if





 (66)

Since condition (66) is satisfied, the proposed estimator $t\_{(st)}$ is more efficient than$ T\_{α\left(st\right)}$.

1. **Results and Discussion**

**3.1 Simulation Study**

In this section, a simulation study was carried out to compare the performances of the estimators based on the criteria of bias, mean square error and percentage relative efficiency. The function mvrnorm available in the package MASS was used to generate data from multivariate normal distribution, with given stratum parameters such as mean vector, population sizes, sample sizes, correlation coefficient for the study and the auxiliary variables. Some of the parameters of the simulated data were presented in Table 1 below.

Table 1: Simulation Parameters

|  |  |
| --- | --- |
| Parameters  | Simulated Dataset I |
|  | Stratum 1 | Stratum 2 | Stratum 3 |
| $$N\_{h}$$ | 1000 | 1500 | 3000 |
| $$n\_{h}$$ | 200 | 250 | 300 |
| $$n\_{h}-r\_{h}$$ | 140 | 150 | 200 |

The biases, mean square errors (MSEs), and percentage relative efficiencies (PREs) of the modified and some existing estimators were computed using (39), (40) and (41) respectively.

$Bias(T\_{i\left(st\right)})=\frac{1}{N\_{h}}\sum\_{i=1}^{N\_{h}}\left(\hat{\overbar{T\_{h}}}-S\_{Y\_{h}}^{2}\right)$ (67)

$MSE(T\_{i\left(st\right)})=\frac{1}{N\_{h}}\sum\_{i=1}^{N\_{h}}\left(\hat{\overbar{T\_{h}}}-S\_{Y\_{h}}^{2}\right)^{2}$ (68)

$ PRE\left(T\_{i\left(st\right)}\right)=\frac{MSE\left(T\_{0\left(st\right)}\right)}{MSE\left(T\_{i\left(st\right)}\right)}×100$ (69)

where $T\_{i\left(st\right)}$ is either of $T\_{1(st)}$, $T\_{2(st)}$, $T\_{3(st)}$, $T\_{Rao(st)},$ $T\_{er(st)}$, $T\_{ep(st)}$, $T\_{N1(st)}$, $T\_{N2(st)}$, $T\_{N3(st)}$, $T\_{N4(st)}$, $T\_{N5(st)}$, $T\_{YN1(st)}$, $T\_{YN2(st)}$, $T\_{YN3(st)}$, $T\_{α\left(st\right)}$ and $t\_{z\left(st\right)}$.

Table 2: Biases, MSEs and PREs of the Proposed and Existing Estimators using Simulated Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| S/N | Estimator  | Bias | MSE | PRE | Rank |
| 1 | $$T\_{1(st)}$$ | -3.060833 | 3.127716 | 100.00 | 13th  |
| 2 | $$T\_{2(st)}$$ | -3.06094 | 3.127940 | 100.0078 | 9th  |
| 3 | $$T\_{3(st)}$$ | -3.060717 | 3.127473 | 99.99282 | 15th  |
| 4 | $$T\_{Rao(st)}$$ | -2.823602 | 3.126670 | 100.0335 | 7th  |
| 5 | $$T\_{er(st)}$$ | -3.060887 | 3.127830 | 99.99634 | 14th  |
| 6 | $$T\_{ep(st)}$$ | -3.060776 | 3.127597 | 100.0038 | 11th  |
| 7 | $$T\_{SG\left(st\right)}$$ | -3.160064 | 3.342929 | 93.56215 | 17th  |
| 8 | $$T\_{N1\left(st\right)}$$ | -3.058716 | 3.123271 | 100.1423 | 5th  |
| 9 | $$T\_{N2\left(st\right)}$$ | -3.058709 | 3.123257 | 100.1428 | 4th  |
| 10 | $$T\_{N3\left(st\right)}$$ | -3.058736 | 3.123314 | 100.1409 | 6th  |
| 11 | $$T\_{N4\left(st\right)}$$ | -3.058667 | 3.123168 | 100.1456 | 2nd  |
| 12 | $$T\_{N5\left(st\right)}$$ | -3.058685 | 3.123206 | 100.1444 | 3rd |
| 13 | $$T\_{YN1\left(st\right)}$$ | -3.06094 | 3.127940 | 99.99281 | 16th  |
| 14 | $$T\_{YN2\left(st\right)}$$ | -3.061007 | 3.128084 | 99.98823 | 12th  |
| 15 | $$T\_{YN3\left(st\right)}$$ | -3.060545 | 3.127135 | 100.0186 | 8th  |
| 16 | $$T\_{α\left(st\right)}$$ | -3.060722 | 3.127477 | 100.0076 | 10th  |
| 17 | $$t\_{\left(st\right)}$$ | -3.05601 | 3.121668 | 100.1937 | 1st |

Table 2 presents the empirical bias, MSE and PRE of the proposed and some existing estimators using simulated data set. The performances of the estimators were indicated by their rank in the fifth column of the table. Based on the results obtained from, the proposed estimator $(t\_{\left(st\right)})$ performed better with minimum values of bias, MSE, and higher PRE (-3.05601, 3.121668 and 100.1937) respectively compared to usual variance ($T\_{1(st)}$); classical ratio and product estimators ($T\_{2(st)}$ and $T\_{3(st)}$); regression type estimator ($T\_{Rao(st)}$); exponential ratio and product type estimators ($T\_{er(st)} $and $T\_{ep(st)}$); ratio-regression-type estimator ($T\_{SG(st)}$); [13] estimators ($T\_{N1(st)}$, $T\_{N2(st)}$, $T\_{N3(st)}$, $T\_{N4(st)}$ and $T\_{N5(st)}$); [4] estimators ($T\_{YN1(st)}$, $T\_{YN2(st)}$ and $T\_{YN3(st)}$); [5] estimator ($T\_{α\left(st\right)}$).

**3.2 Empirical Study**

In this section, four real-life data sets were used to assess the performance of the proposed generalized estimator over some existing estimators base on the criteria of bias, mean square error and percentage relative efficiency.

**Dataset I [16]**

*Y* = Leaf area for the newly developed strain of wheat, *X* = Weight of leaves, Strata = Farms, $N=39$, $n=14$.

Table 3: Parameters of the Dataset I

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | $$N\_{h}$$ | $$n\_{h}$$ | $$C\_{xh}$$ | $$S\_{y\_{h}}^{2}$$ | $$S\_{x\_{h}}^{2}$$ | $$ρ\_{xy(h)}$$ | $$λ\_{40(h)}$$ | $$λ\_{04(h)}$$ | $$λ\_{22(h)}$$ |
| Stratum 1 | 12 | 4 | 0.1119 | 6.0664603 | 11.57158 | 0.6293 | 1.9394547 | 2.274823 | 1.912346 |
| Stratum 2 | 13 | 5 | 0.0734 | 5.2915229 | 8.139014 | 0.5843 | 2.9819269 | 3.436904 | 2.970998 |
| Stratum 3 | 14 | 5 | 0.1194 | 6.4961301 | 12.44946 | 0.6512 | 2.3448986 | 2.895550 | 2.513438 |

**Dataset II [17]**

*Y* = Area under wheat in the region in 1974, *X* =Area under wheat in the region in 1973, Strata = Regions, $=34,$ $ n=10$.

Table 4: Parameters of the Dataset II

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | $$N\_{h}$$ | $$n\_{h}$$ | $$C\_{xh}$$ | $$S\_{y\_{h}}^{2}$$ | $$S\_{x\_{h}}^{2}$$ | $$ρ\_{xy(h)}$$ | $$λ\_{40(h)}$$ | $$λ\_{04(h)}$$ | $$λ\_{22(h)}$$ |
| Stratum 1 | 9 | 3 | 0.1118928 | 31978.25 | 11.5715804 | 0.865 | 2.9286 | 2.07 | 2.38 |
| Stratum 2 | 10 | 3 | 0.0733753 | 37629.39 | 8.139014 | 0.932 | 1.511 | 1.42 | 1.42 |
| Stratum 3 | 15 | 4 | 0.11937840 | 6893.067 | 12.4494617 | 0.755 | 2.20 | 2.42 | 2.28 |

**Dataset III [17]**

*Y* = Output for 80 factories in a region, *X =*Fixed capital, Strata = Regions, $N=80$, $n=41$.

Table 5: Parameters of the Dataset III

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | $$N\_{h}$$ | $$n\_{h}$$ | $$S\_{y\_{h}}^{2}$$ | $$S\_{x\_{h}}^{2}$$ | $$ρ\_{xy(h)}$$ | $$C\_{xh}$$ | $$λ\_{40(h)}$$ | $$λ\_{04(h)}$$ | $$λ\_{22(h)}$$ |
| Stratum 1 | 19 | 10 | 583977.5 | 441743.5 | 0.9122 | 1.24 | 3.28 | 1.45 | 1.46 |
| Stratum 2 | 32 | 16 | 456563.3 | 653762.4 | 0.9325 | 0.98 | 1.56 | 3.09 | 1.74 |
| Stratum 3 | 14 | 7 | 195208.8 | 661122.6 | 0.8955 | 1.16 | 1.62 | 1.62 | 1.80 |
| Stratum 4 | 15 | 8 | 437923.5 | 331978.2 | 0.9566 | 1.08 | 2.22 | 1.90 | 2.02 |

**Dataset IV [18]**

Tobacco: Area (hectares), yield and production (metric tons) in specified countries during 1998.

*Y* = Yield (Metric tons), *X* = Area (Hectare), *Z* = Production (Metric tons), Strata = Continents, N=103, n=75 and m=16

Table 6: Parameters of the Dataset IV

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | $$N\_{h}$$ | $$n\_{h}$$ | $$S\_{y\_{h}}^{2}$$ | $$S\_{x\_{h}}^{2}$$ | $$ρ\_{xy(h)}$$ | $$C\_{xh}$$ | $$λ\_{40(h)}$$ | $$λ\_{04(h)}$$ | $$λ\_{22(h)}$$ |
| Stratum 1 | 12 | 9 | 0.204663 | 306708635 | -0.6845 | 3.84849 | 3.08317 | 6.2136 | 4.1019 |
| Stratum 2 | 18 | 13 | 0.520352 | 392241621 | -0.1382 | 1.35059 | 3.00399 | 3.5242 | 0.5206 |
| Stratum 3 | 34 | 25 | 0.329574 | 675871615 | 0.2102 | 5.88771 | 2.73314 | 13.2358 | 1.0322 |
| Stratum 4 | 39 | 28 | 0.847973 | 7566352438 | -0.1597 | 5.26203 | 10.7998 | 14.9648 | 0.3680 |

Table 7: Bias, MSE and PRE of the Proposed and Existing Estimators using Dataset I & II

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Dataset I |  |  |  | Dataset II |  |  |
| Estimator  | Bias | MSE | PRE | Rank | Bias | MSE | PRE | Rank |
| $$T\_{1(st)}$$ | - | 0.004715021 | 100 | 11th  | - | 188354.1 | 100 | 13th  |
| $$T\_{2(st)}$$ | 0.015219 | 0.001138751 | 414.05 | 2nd  | -27.9889 | 25526.83 | 737.87 | 2nd  |
| $$T\_{3(st)}$$ | 0.053052 | 0.018496170 | 25.50 | 16th  | 284.8535 | 567019.20 | 33.22 | 17th  |
| $$T\_{Rao(st)}$$ | 1.338503 | 0.013796470 | 34.18 | 13th  | 16941.3200 | 207164.4 | 90.92 | 14th  |
| $$T\_{er(st)}$$ | -0.000924 | 0.001382557 | 341.04 | 3rd  | -46.1025 | 74397.76 | 253.17 | 9th  |
| $$T\_{ep(st)}$$ | 0.017992 | 0.011217740 | 42.03 | 12th  | 110.3187 | 367929.90 | 51.19 | 16th  |
| $$T\_{SG\left(st\right)}$$ | 0.190679 | 0.001668142 | 282.65 |  6th  | 563.1601 | 32856.03 | 573.27 | 6th  |
| $$T\_{N1\left(st\right)}$$ | 0.093267 | 0.001562525 | 301.76 | 4th  | 194.0670 | 31653.72 | 595.05 | 5th  |
| $$T\_{N2\left(st\right)}$$ | 0.101063 | 0.002177900 | 216.50 | 9th  | 321.3616 | 38421.44 | 490.23 | 7th  |
| $$T\_{N3\left(st\right)}$$ | 0.012408 | 0.001587906 | 296.93 | 5th  | 71.7311 | 29575.85 | 636.85 | 3rd  |
| $$T\_{N4\left(st\right)}$$ | 0.101913 | 0.002108672 | 223.60 | 8th  | 319.6517 | 41267.78 | 456.42 | 8th  |
| $$T\_{N5\left(st\right)}$$ | 0.077566 | 0.001680353 | 280.60 | 7th  | 264.5368 | 30688.25 | 613.77 | 4th  |
| $$T\_{YN1\left(st\right)}$$ | 0.036763 | 0.004523937 | 104.22 | 10th  | -65.4400 | 81767.74 | 230.35 | 10th  |
| $$T\_{YN2\left(st\right)}$$ | -0.090308 | 0.014054670 | 33.55 | 14th  | -461.8832 | 284960.50 | 66.10 | 15th  |
| $$T\_{YN3\left(st\right)}$$ | -0.012422 | 0.018385110 | 25.65 | 15th  | 459.7782 | 140516.90 | 134.04 | 12th  |
| $$T\_{α\left(st\right)}$$ | 0.021920 | 0.038204130 | 12.34 | 17th  | 3.8890 | 89857.83 | 209.61 | 11th  |
| $$t\_{\left(st\right)}$$ | 0 | 0.000890827 | 529.29 | 1st  | 0 | 20959.36 | 898.66 | 1st  |

Table 7 displayed the empirical biases, mean square errors and percentage relative efficiencies of the proposed estimator and some existing estimators considered using datasets I and II. The performances of the estimators is indicated by their ranks in the fourth column of the table. The results computed from Dataset I showed that the proposed generalized estimator $(t\_{\left(st\right)})$ peformed better with minimum values of bias, MSE and higher PRE (0, 0.000890827 and 529.29, respectively) compared to the usual variance ($T\_{1(st)}$); classical ratio and product estimators ($T\_{2(st)}$ and $T\_{3(st)}$); regression type estimator ($T\_{Rao(st)}$); exponential ratio and product type estimators ($T\_{er(st)} $and $T\_{ep(st)}$); ratio-regression-type estimator ($T\_{SG(st)}$); [13] estimators ($T\_{N1(st)}$, $T\_{N2(st)}$, $T\_{N3(st)}$, $T\_{N4(st)}$ and $T\_{N5(st)}$); [4] estimators ($T\_{YN1(st)}$, $T\_{YN2(st)}$ and $T\_{YN3(st)}$); [5] estimator ($T\_{α\left(st\right)}$). The results obtained from Dataset II also revealed that the proposed generalized estimator $(t\_{z\left(st\right)})$ performed better with minimum values of bias, MSE and higher PRE (0, 20959.36 and 898.66, respectively) compared to existing estimators considered.

Table 8: Bias, MSE and PRE of the Proposed and Existing Estimators using Dataset III & IV

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Dataset III |  |  |  | Dataset IV |  |  |
| Estimator  | Bias | MSE | PRE | Rank | Bias | MSE | PRE | Rank |
| $$T\_{1(st)}$$ | - | 419607.1 | 100.00 | 13th  | - | 1.559826e-07 | 100.00 | 3rd  |
| $$T\_{2(st)}$$ | 83.7904 | 398344.8 | 105.34 | 12th  | 0.000255519 | 4.130966e-07 | 37.76 | 16th  |
| $$T\_{3(st)}$$ | 164.6257 | 2298883.0 | 18.25 | 17th  | -4.495743e-06 | 3.526359e-07 | 44.23 | 15th  |
| $$T\_{Rao(st)}$$ | 2153.9200 | 1407059.0 | 29.82 | 16th  | 0.001072130 | 7.958634e-07 | 19.60 | 17th  |
| $$T\_{er(st)}$$ | 10.8432 | 298307.4 | 140.66 | 3rd  | 9.638156e-05 | 2.253442e-07 | 69.22 | 14th  |
| $$T\_{ep(st)}$$ | 51.2609 | 762244.1 | 55.05 | 15th  | -3.362577e-05 | 2.050118e-07 | 76.08 | 12th  |
| $$T\_{SG\left(st\right)}$$ | 879.5920 | 265977.3 | 157.76 | 2nd  | 0.176261600 | 2.190612e-07 | 71.21 | 13th  |
| $$T\_{N1\left(st\right)}$$ | 898.3932 | 313597.1 | 133.80 | 4th  | 0.001006607 | 1.641558e-07 | 95.02 | 6th  |
| $$T\_{N2\left(st\right)}$$ | 540.9849 | 341597.4 | 122.84 | 7th  | 0.001073261 | 1.661652e-07 | 93.87 | 8th  |
| $$T\_{N3\left(st\right)}$$ | 540.9844 | 352324.6 | 119.10 | 10th  | 0.001073219 | 1.640338e-07 | 95.09 | 5th  |
| $$T\_{N4\left(st\right)}$$ | 540.9847 | 348518.2 | 120.40 | 8th  | 0.001073214 | 1.654352e-07 | 94.29 | 7th  |
| $$T\_{N5\left(st\right)}$$ | 540.9854 | 388491.5 | 108.01 | 11th  | 0.001073223 | 1.621856e-07 | 96.18 | 4th  |
| $$T\_{YN1\left(st\right)}$$ | 331.1033 | 348774.2 | 120.31 | 9th  | 0.003319547 | 1.864132e-07 | 83.68 | 10th  |
| $$T\_{YN2\left(st\right)}$$ | -443.9922 | 431050.4 | 97.35 | 14th  | 0.002770780 | 1.779673e-07 | 87.65 | 9th  |
| $$T\_{YN3\left(st\right)}$$ | -915.2366 | 335591.1 | 125.04 | 6th  | 0.000306262 | 1.977179e-07 | 78.89 | 11th  |
| $$T\_{α\left(st\right)}$$ | 203.6395 | 318347.8 | 131.81 | 5th  | 4.262074e-05 | 1.557866e-07 | 100.13 | 2nd  |
| $$t\_{\left(st\right)}$$ | -1.433e-15 | 254877.6 | 164.63 | 1st  | 4.345755e-23 | 1.550209e-07 | 100.62 | 1st  |

Table 8 displayed the empirical bias, mean square error and percentage relative efficiency of proposed estimator and some existing estimators considered using datasets III and IV. The performances of the estimators is indicated by their ranks in the fourth column of the table. The results computed from Dataset III showed that the proposed generalized estimator $(t\_{\left(st\right)})$ performed better with minimum values of bias, MSE and higher PRE (-1.433e-15, 254877.6 and 164.63, respectively) compared to usual variance ($T\_{1(st)}$); classical ratio and product estimators ($T\_{2(st)}$ and $T\_{3(st)}$); regression type estimator ($T\_{Rao(st)}$); exponential ratio and product type estimators ($T\_{er(st)} $and $T\_{ep(st)}$); ratio-regression-type estimator ($T\_{SG(st)}$); [13] estimators ($T\_{N1(st)}$, $T\_{N2(st)}$, $T\_{N3(st)}$, $T\_{N4(st)}$ and $T\_{N5(st)}$); [4] estimators ($T\_{YN1(st)}$, $T\_{YN2(st)}$ and $T\_{YN3(st)}$); [5] estimator ($T\_{α\left(st\right)}$). The results obtained from Dataset IV also revealed that the proposed generalized estimator $(t\_{z\left(st\right)})$ performed better with minimum values of bias, MSE and higher PRE (4.345755e-23, 1.550209e-07 and 100.62, respectively) compared to existing estimators considered.

1. **Conclusion**

A generalized almost unbiased estimator for estimating variance of a finite population under stratified random sampling was developed in this paper. The properties of the estimator such as bias and mean square error (MSE) were derived. Empirical study was carried out using real life data sets. The generalized estimator performed better with least values of bias, MSE and higher PRE than some related estimators considered. The generalized estimator were more suitable to estimate variation characteristics of data sets when the study and auxiliary variables are either positively or negatively correlated. Therefore, the generalized almost unbiased estimator can be utilized to provide reliable and accurate variation estimates in the presence of complete response under stratified random sampling.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1. Option 1 was declared

2.

3.

**References**

1. Cochran WG. Sampling Theory when the Sampling Units are of Unequal Sizes. Journal of American Statistical Association. 1942;37:199-212.
2. Isaki CT. Variance estimation using auxiliary information. Journal of American Statistics Association. 1983;78:117–123.
3. Singh R, Chauhan P, Sawan N, Smarandache F. Improvement in estimating the population mean using exponential estimator in stratified random sampling. International Journal of Statistics & Economics. 2009;3(9):13–18.
4. Shabbir J, Gupta S. Some estimators of finite population variance of stratified sample mean. Communications in Statistics - Theory and Methods. 2010;39(16):3001–3008.
5. Yasmeen U, Noor M. Estimation of finite population variance under stratified sampling technique. Journal of Reliability and Statistical Studies. 2021;14(2):565-584.
6. Ahmad S, Hussain S, Shabbir J, Zahid E, Aamir M, Onyango R. Improve Estimation of finite population variance using dual supplementary information under stratified random sampling. Mathematical Problems in Engineering. 2022;3(2):12-19.
7. Zakari Y, Muhammad I. Modified estimator of finite population variance under stratified random sampling. Engineering Proceedings of the 4th International Electronic Conference on Applied Sciences of Science Forum. Published by Multidisciplinary Digital Publishing Institute (MDPI); 2023.
8. Montgomery DC. Introduction to statistical quality control. 7th ed. John Wiley & Sons; 2013.
9. Das AK, Tripathi TP. Use of auxiliary information in estimating the finite population variance. Sankhya C. 1978;40(2):139–148.
10. Kadilar C, Cingi H. Ratio estimators for the population variance in simple and stratified random sampling. Applied Mathematics and Computation. 2006;173(2):1047–1059.
11. Muili JO, Audu A, Singh RVK, Yunusa IA. Improved estimators of finite population variance using unknown weight of auxiliary variable. Annals. Computer Science Series. 2019;17(1):148-153.
12. Muhammad I, Zakari Y, Audu A. Generalized estimators for finite population variance using measurable and affordable auxiliary character. Asian Research Journal of Mathematics. 2022;18(1):14-30.
13. Muhammad I. Generalized Ratio-Product cum Regression Variance Estimator in Two-Phase Sampling. CBN Journal of Applied Statistics. 2023;14(2):73-101. DOI: 10.33429/Cjas.14223.4/5
14. Rao TJ. On certail methods of improving ratio and regression estimators. Communication in Statistics-Theory and Methods. 1991;20(10), 3325–3340.
15. Koyuncu N. Improved Estimators of Finite Population Variance in Stratified
Random Sampling. World Applied Sciences Journal. 2013;23(12):130-137.
16. Singh R, Mangat NS. Elements of survey sampling. London: Kluwer Academic Publishers; 1996.
17. Murthy MN. Sampling theory and methods. Statistical Publication Society, Calcutta, OH, US; 1967.
18. Sarjinder S. Advanced sampling theory with applications Springer Science Business Media, B.V; 2003.