**Variance Estimation in Sample Surveys: Utilizing Auxiliary Variable Information**

**Abstract**

This manuscript presents a novel estimator for the assessment of population variance by integrating auxiliary information pertaining to the population median and the population standard deviation. The incorporation of this dual auxiliary information empowers researchers to estimate population variance effectively, even in the presence of skewed data distributions. Utilizing Taylor’s series expansion, the formulations for both bias and mean squared error are derived, extending to the first order of approximations. To evaluate the efficacy of the proposed estimator, a comparative analysis of the mean squared error of several classical estimators is conducted, revealing that these conventional methods exhibit inferior efficiency relative to the proposed estimator. The empirical investigation demonstrates that the recommended estimator yields a lower mean squared error and possesses significant practical relevance in the context of survey sampling.

**Keywords:** Measure of dispersion, auxiliary information, ratio method of estimation, efficiency and mean squared error

**1. Introduction**

The variable we are focusing on, is termed as study variable. Auxiliary information refers to the background knowledge related to the study variable. Integrating auxiliary information into the estimation process will naturally enhance efficiency. For instance, if we want to determine the number of students who have passed in the Statistics course, we need to have an understanding of how many students are enrolled in that course or the subject combinations they have opted for. Similarly, if we aim to estimate the upcoming season's mango production, it is essential to know the regions where mangoes are extensively cultivated. By utilizing such auxiliary information, the estimation process can significantly improve in efficiency.

There are several methods for selecting a random sample from the population for a sample survey. A sampling method from the population is chosen for this manuscript using simple random selection without replacement. Sample surveyors were able to estimate population parameters such as variance, mean and median under this sample method.

There are two estimation techniques that would improve the estimators' efficiency in order to more accurately estimate the population parameters. The ratio method of estimation is used in this research study to estimate the population parameter. The prior knowledge of the variable being studied is used in the ratio method of estimation. Auxiliary information about the study variable is frequently represented by this prior knowledge. Therefore, using a ratio estimate will have enhanced sample estimates.

Population variation is taken into consideration and evaluated throughout this research paper. Additionally, this paper will be scientifically sound due to the inclusion of supplementary data in the form of medians and standard deviations. To estimate population parameters, we employ a double auxiliary information rather than a single prior information. This double auxiliary information consists of the auxiliary variable's median and standard deviation. The median indicates 50% of the data that lie above the middle value and 50% of the data that lie below it, but the standard deviation, which is the positive square root of the variance, indicates the degree of dispersion of the data from its central value.

As a result, averages are employed when dealing with irregular or outlier data. The use of this dual form of dynamic auxiliary data will undoubtedly motivate statisticians, researchers, or scholars for further study in field of survey or policy making.

Das (1978), Isaki (1983) and Upadhyaya (1999) incorporate the auxiliary information for finite population. Garcia et al. (1996), Kadilar (2005, 2006), Singh (2001, 2010, 2014), Subarmani (2012, 2013, 2015), Yadav et al. (2016) gives more precise estimators for population parameters. Kumari et al. (2018, 2019, 2020, 2021) proposed log type ratio estimator using various forms of auxiliary information i.e. kurtosis, coefficient of variation, median for estimating the population parameter. Then Audu et al. and Shahzad (2021) used difference estimators in sample random sampling whereas Zaman and Ashutosh (2022) utilized robust estimator for finite population variance. Recently, Ahmad (2023a, 2023b) uses two auxiliary information for estimation of population variance.

Motivated by the afore mentioned authors, this research work employs a technique that provides a more effective estimator by utilizing the dual form of auxiliary information for population variance estimation. However, the optimal method to estimate the population parameter using its auxiliary information is to use the median alone if the data comprises outliers or is slightly symmetrical. Using the median and standard deviation combined will improve the effectiveness of the suggested work and raise the bar for this paper's efficient estimates.

**2. Terminology used**

Suppose, the population consist of N units. The study variable and its prior information i.e. auxiliary variable is denoted by Y and X respectively. The total of the auxiliary variable is known in advance. So, we have the information about their means also i.e. and . and are population variances of both the variables. Let us take a random sample from simple random sampling without replacement of size . We draw a pair-wise random sample from and respectively and represent it as and . Similarly, their sample means are denoted by and . The unbiased estimates of the population variance is used to estimate the unknown population variance . Since, the two variable are highly positively correlated to themselves, there must be covariance exists between them. It is represent as . Further, the median is denoted by . The combination of this two auxiliary information attracts the work researchers to work in this direction also.

**3. Proposed Estimator**

The estimation of population variance is very important in sample survey to get an idea about the extent of variation present in the data. In this section, the proposed estimator is log-type ratio estimator having auxiliary information in terms of standard deviation and median for estimating the population variance. To estimate the population variance, we propose the following estimator

(11)

Where is the standard deviation and is the median of auxiliary variable x and ‘a’ is the characterizing scalar.

Let,

,

, , ,

Now expressing in terms of epsilons ∈ we have,

(12)

where, we assume that , so that is expandable.

Expanding the right hand side of (12) and multiplying out we have

(13)

Taking expectations on both the sides, we get

(14)

Squaring on both the sides of equation (14), we get MSE

Taking expectation on both the sides,

(15)

Further, to evaluate the minimum mean squared error of the advocated estimator we partially differentiate the mean squared error with respect to the unknown constant and equating it equal to zero. Sometimes the equation is very complicated and its explicitly algebraic solution is not available. It is to be solved numerically in that case by the method of iteration, using as a starting value the observed value of some consistent (but inefficient) estimator which can be easily computed. In large samples, such an estimator will tend to be fairly close to the maximum likelihood estimate and higher is its efficiency, the greater is the closeness.

On differentiating with respect to , we get

(16)

Substituting the value of unknown constant i.e. α in the obtained mean squared error of the proposed estimator, we get the required minimum mean squared error of our suggested estimator.

(17)

is the desired optimum mean squared estimator for proposed estimator.

**4. Dominance Condition**

We compare the mean square error of the proposed estimator with the MSE of some conventional estimators. Estimator is the function of random sample values which are used to estimate the unknown population parameter i.e. population mean or population variance etc. Mean squared error indicates about how much is the sample estimate is far from the true value of the estimator. If the estimator has minimum mean squared error than any other estimator, then the considered estimator is said to be efficient estimator. Efficiency of an estimator is measured in terms of variance or mean squared error. It is also one of the good properties of estimator among unbiasedness, consistency and sufficiency. Moving forward to show the supremacy of this research article over some classical estimators, we compare the mean squared error of some known estimators with our suggested one.

1. **Variance estimator having coefficient of variation as auxiliary variable**

The bias and mean squared error are as follows

;

**On comparing**  with proposed estimator, we have

1. **Variance estimator having coefficient of kurtosis as auxiliary variable**

The bias and mean squared error are as follows

;

1. **Variance estimator having coefficient of kurtosis and variation as auxiliary variable**

The bias and mean squared error are as follows

;

1. **Variance estimator having coefficient of kurtosis and variation as auxiliary variable**

The bias and mean squared error are as follows

1. **Variance estimator having median as auxiliary variable**

The bias and mean squared are as follows

;

where

Consequently, the proposed estimator has lesser mean squared error than the conventional estimators. So, the proposed work has gain in efficiency when compared to the estimators present in literature.

**5. Numerical Study**

To compare the effectiveness of the suggested estimator with traditional estimators, we have selected three natural populations. The data summary is provided below:

**Table 1: Summary of the Data**

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristics** | **Population 1** | **Population 2** | **Population 3** |
| N | 22 | 103 | 103 |
| n | 5 | 40 | 40 |
|  | 22.5 | 626.2123 | 62.6212 |
|  | 1467.5 | 557.1909 | 556.5541 |
|  | 0.9022 | 0.9936 | 0.7298 |
|  | 32.8 | 913.5498 | 91.3549 |
|  | 1.45777778 | 1.4588 | 1.4588 |
|  | 2503.2 | 818.1117 | 610.1643 |
|  | 1.705758092 | 1.4683 | 1.0963 |
|  | 13.2 | 37.3216 | 17.8738 |
|  | 5.57 | 37.1279 | 37.1279 |
|  | 7.71 | 37.2055 | 17.2220 |
| Md | 534.5 | 308.05 | 373.82 |

**Table 2: Mean Squared Error of the Estimators**

|  |  |  |  |
| --- | --- | --- | --- |
| **Estimator** | **Population 1** | **Population 2** | **Population 3** |
|  | 775478.55 | 670384402.9 | 35796604.9 |
|  | 775473.88 | 670169790.4 | 35796502.6 |
|  | 775479.19 | 670393032.1 | 35796611.2 |
|  | 775476.10 | 670240637.1 | 35796512.2 |
|  | 775262.47 | 668667060.7 | 35794364.2 |
|  | **543507** | **667490920.7** | **16698592.9** |

The above table shows the mean squared error of all the selected estimators and it is obvious that the proposed estimator (log type ratio estimator using two auxiliary information) performs better than other estimators. Hence, the proposed estimator comes out to be the more efficient estimator when compared to some of theoretical estimators.

**6.Conclusion**

Here, we introduce a new generalized log type estimator that uses the standard deviation and median as auxiliary data. We generate the bias and mean squared error formulations and compare them with several other common estimators. The results of the previous numerical investigation showed that the proposed estimators are the best among all the generalized ratio type variance estimators.

COMPETING INTERESTS DISCLAIMER:

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

Disclaimer (Artificial intelligence)

I had not use any AI Technology.

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.

2.

3.

**References**

1. Ahmad, S. et al. (2023a), An enhanced estimator for finite population variance using two auxiliary variables under simple random sampling, *Scientific Reports*, 13, 21444.
2. Ahmad, S. et al. (2023), A new improved generalized class of estimators for population distribution function using auxiliary variable under simple random sampling, *Scientific Reports*, 13(1), 5415.
3. Ashutosh, A, et al. (2022), Cabibration approach for variance estimation in small domain, *Mathematical Problem in Engineering*, 1-8.
4. Zaman, T. and Bulut, H. (2022), A new class of robust ratio estimators for finite population variance, *Scientia Iranica*, 1,1.
5. Kumari, C. and Thakur, R. K. (2021). “An Efficient Log-type class of estimators using Auxiliary information under Double Sampling”, *Journal of Statistics Application & Probability* (ISSN 2090-8431),10(1), 1-6.
6. Shahzad U et el. (2021)|, A novel family of variance estimators based on L-moments and calibration approach under stratified random sampling, *Communication in Statistics- Simulation and Computation*, 1, 1-14
7. Audu, A. et al. (2021). Difference-com-ratio estimators for estimating finite population coefficient of variation in sample random sampling, *Asian Journal of Probability and Statistics*, 13(3), 13-29.
8. Kumari, CandThakur, R. K**.** (2020), Estimation of Population Variance Using Known Coefficient of Kurtosis and Median of an Auxiliary Variable, *International Journal of Advanced Research*, Volume 8, Issue 4, pg. 803-810.
9. Kumari, CandThakur, R. K**.** (2020), An Advanced Class of Log-type estimator for Population Variance using an attribute and a variable, *International Journal of Industrial Engineering Research and Development* (ISSN 0976-6987), 11(1), 1-7.
10. Bhushan, S. and Kumari, C (2019). “Double Sampling Log-type Estimator Using Auxiliary Attribute for Population Variance”, *Journal of Statistics Applications & Probability* (ISSN NO. 2090-8431), 6(3), 1-6*.*
11. Bhushan, S. and Kumari, C. (2018). “A new log type estimator for estimating the population variance”, *International Journal of Computational and Applied Mathematics* (ISSN 1819-4966), 13(1), 43-54.
12. Amir, M., Shabri, A. & Ishaq, M. (2018). Improvement on estimating median for finite population using auxiliary variable in double sampling, Jurnal Teknologi, 80(5), 135-143.
13. Yadav, S. K., Shukla, A. K. (2016). Use of correlation coefficient and quartiles of auxiliary variable for improved estimation of population variance, American Journal Research, 6(2), 33-38.
14. Subarmani, J. (2015). Generalized Modified Ratio Type Estimator for Estimation of PopulationVariance, *Sri Lankan Journal of Applied Statistics*, 16(1):69-9
15. Singh, H. P. et al. (2014). Estimation of finite population variance using auxiliary information in sample surveys, *Statistics, anno LXV*, 1, 99-116.
16. Subramani, J. and Kumarapandiyan, G. (2013) Estimation of Variance Using Known Coefficient of Variation and Median of an Auxiliary Variable. Journal of Modern Applied Statistical Methods, 12, 11. <https://doi.org/10.22237/jmasm/1367381400>
17. Subramani, J. and Kumarapandiyan, G. (2012) Variance Estimation Using Median of the Auxiliary Variable. International Journal of Probability and Statistics, 1, 6-40. <https://doi.org/10.5923/j.ijps.20120103.02>
18. Subramani, J. and Kumarapandiyan, G. (2012) Variance Estimation Using Quartiles and Their Functions of an Auxiliary Variable. International Journal of Statistics and Applications, 2, 67-72. <https://doi.org/10.5923/j.statistics.20120205.04>
19. Singh, H. P. and R. Tailor. (2010). On ratio and product methods with certain known population parameters of auxiliary variable in sample surveys, *SORT*, 34(2), 157-180
20. Kadilar, C. & Cingi, H. (2006). Ratio estimators for the population variance in simple and stratified random sampling. *Applied Mathematics and Computations*, 173(2), 1047-1059.
21. Kadilar, C. & Cingi, H. (2005). A new estimator using two auxiliary variables, *Applied Mathematics and Computations*, 162, 901-908.
22. Singh, H. P. and R. Singh. (2001). Improved ratio type estimator for variance using auxiliary information, *Journal of Indian Society Agriculture Statistics*, 54(3), 276-287.
23. Garcia, M. R. and Cebrain, A. A. (1996). Repeated substitution method: The ratio estimator for the population variance*, Metrika* 43, 101-105
24. Upadhaya, L. N. & Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 41(5), 627-636.
25. Isaki, C. T. (1983). Variance estimation using auxiliary information. *Journal of the American Statistical Association*, 78(381), 117-123.
26. Das, A. K. (1978). Use of auxiliary information in estimating the finite population variance. *Sankhya*, c, 40, 139-148.