**Incorporating the Auxiliary Information for Estimation of Population Variance**

**Abstract**

The goal of this research project is to develop a more accurate estimate for population variance estimation. Using the supplied data on the median and standard deviation of the auxiliary variable, we obtain the equations for bias and mean squared error of the suggested estimate. Additionally, we examine the suggested estimator's percent relative efficiency in comparison to a few current estimators. We demonstrate numerically that the suggested estimator outperforms traditional estimators.

**Keywords:** Measure of dispersion, ratio method of estimation, bias, mean squared error, percent relative efficiency.

**1 Introduction**

Sampling is the selection of a part of an aggregate of material (or population) to represent the whole population. The part of the population selected by sampling is called a sample. It is from the sample that we make inferences about the population in which we are interested. A survey carried out on a properly selected representative sample is called a sample survey or sample census, as opposed to the complete enumeration or complete census, in which the whole population is enumerated or surveyed. In many cases, we undertake a sample survey in preference to a complete census because of the following considerations:

1. There is a reduction of cost either in terms of money or in term of man-hours. Although the cost per individual may be larger in a sample survey, the total cost is expected to be smaller. In many cases, our resources may be limited or it may be necessary that the results of the survey should be available within a specified time limit. In such cases, it is imperative to adopt a sample survey rather than a complete census.
2. There is generally greater scope in a sample survey than in a census. Some enquiries may require highly trained personnel or specialised equipment for collection of data, thus making a census practically inconceivable. Thus in a sample survey we may have greater coverage both in respect of information collected and in respect of the geographical, demographic or other boundaries taken into account.
3. A sample survey generally gives better quality than a complete census, because in a sample survey it may be possible to employ better trained personnel, effect better supervision or use better equipment than is possible or feasible in a complete census.
4. It should be remembered that in some cases a complete census is ruled out by the nature of the population. If there is population which is infinite or hypothetical, like the population of all the throws that may be made with a coin, sampling is the only course available.
5. Again, if the enumeration is by its very nature destructive example when we want to know the average breaking strength of a type of fibre, we must have recourse to a sample and a rather small sample at that.

However, when time and cost are not important factors for consideration or when detailed information is wanted for all the sub-classes into which the population may be divided or when the population size is not large, a complete enumeration may be more appropriate than any sampling procedure. Again, if basic information is required for every unit of the population, a complete enumeration has got to be undertaken. But even in such situations, sampling methods may be used concurrently to get advance information well ahead of the processing of the complete enumeration data as well as access the quality of these data.

Quite often one comes across situations where the ratio of the variable y to another variable x is considered less variable and so more stable, than y itself. In such a case, it would be better to estimate the ratio of y to x in the population from the sample and then multiply it by the known mean (total) of x to estimate the men (total) of y. This procedure is known as ratio estimation.

The average of a variable gives a general idea as to the whole sets of its values. It is clear, however that for a variable to be really variable, it’s given values will not be all equal to the average. In some cases, they may lie very near to average, while in others they may be widely scattered about it. Thus in order to give a proper idea about the overall nature of the given values of a variable, it is necessary, besides mentioning the average values, to a state how scattered the given values are about the average. Mainly, three different measures are used to determine this feature of a variable, which is called it’s scatter or dispersion. The data scatterings are displayed by the measure of dispersion. It provides a clear picture of the data distribution and explains how the data vary from one another. The dispersion measure indicates whether the distribution of the observations is homogeneous or heterogeneous. Additionally, this research makes use of supplementary data in the form of the median and standard deviation. One of the most widely used methods for estimating population parameters is the ratio method. Numerous researchers use this auxiliary information to increase the efficiency of the parameter, including [1], [2], [3], [4], [5] and [7]. Ratio method of estimation reduces sampling error and it is flexible sampling design. It performs better when the linear relationship between main variable and its prior information is positively correlated to each other. Sampling error is absent in complete enumeration whereas it presents in sampling survey.

Let the population consists of N units. Yi and Xi denote the ith characteristics of the population. The population mean of the study variable is denoted by and population mean of the auxiliary variable is denoted by . The population variance of the study variable and the auxiliary variable is denoted by and . Let µpq = be the population product moment between x and y.

Assume that there are N units in the population. Yi and Xi stand for the population's ith features. represents the population mean of the research variable, while represents the population mean of the auxiliary variable. and represent the population variance of the study variable and the auxiliary variable, respectively.

For x and y, let µpq = be the population product moment. be the auxiliary variable's median. and be the skewness and kurtosis coefficients, respectively. Let and represent the research variable's and the auxiliary variable's respective coefficients of variation. As a result, let ρ be the correlation coefficient between the variable being studied and the auxiliary variable. This coefficient, which is expressed as ρ = Cov(x,y)/σxσy, indicates the degree of linear link between two variables.

After that, we take a random sample of size *n* from population of sizedrawn without replacement. Let yi be the ith characteristics of the study variable of the sample and xi be the ith characteristics of the auxiliary variable of the sample. The sample mean of study variable for estimating the population mean is denoted by . The sample mean of auxiliary variable is denoted by .

**2. Review of Literature**

1. **Kadilar & Cingi (2006) variance estimator**

The bias and mean squared error are as follows

(1)

; (2)

1. **Kadilar & Cingi (2006) variance estimator**

The bias and mean squared error are as follows

(3)

; (4)

1. **Kadilar & Cingi (2006) variance estimator**

The bias and mean squared error are as follows

(5)

; (6)

1. **Kadilar & Cingi (2006) variance estimator**

The bias and mean squared error are as follows

(7)

(8)

1. **Subramani & Kumarapandiyan (2012) variance estimator**

The bias and mean squared are as follows

(9)

; (10)

where

**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 the section, the proposed estimator is log-type ratio estimator having auxiliary information in terms of standard deviation and median for estimating the population variance. Standard deviation is one of the widely used measure of dispersion whereas median is one of the measure of central tendency. The combination of both the statistical measures are enable us to estimate the population variance of an estimator effectively.

To estimate the population variance, we propose the following estimator

(11)

Where is the standard deviation and is the coefficient of variation 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 (MSE) 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.

1. **Comparison with variance estimator**
2. **Comparison with variance estimator**
3. **Comparison with variance estimator**
4. **Comparison with variance estimator**
5. **Comparison with variance estimator**

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**

In this work, we present a novel generalized log type estimator that makes use of the median and standard deviation as auxiliary data. The mean squared error and bias expressions are derived and contrasted with a few other standard estimators. The preceding numerical study's outcome demonstrated that, out of all the generalized ratio type variance estimators, the suggested estimators are the best.

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.

**References**

1. Kumari, CandThakur, R. K**.** (2020). An Improved Estimation of Population Variance Using the Coefficient of Kurtosis and Median of an Auxiliary Variable, *International Journal of Engineering Sciences & Research Technology*, Volume 9, Issue 4, pg. 168-175
2. 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
3. Kumari, CandThakur, R. K**.** (2020), The Utilization of Auxiliary information in the estimation procedure of Population Variance. *International Journal of Multidisciplinary Educational Research* (ISSN 2277-7881), 9,5(5), 36-45.
4. Murthy, M. N. (1967). Sampling Theory and Methods, Statistical Publishing Societ Calcutta, India.
5. 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>
6. 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>
7. 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>
8. Subarmani, J. (2015). Generalized Modified Ratio Type Estimator for Estimation of PopulationVariance, SriLankan Journal of Applied Statistics, Vol.16(1):69-9
9. Sukhatme, P. (1944) Moments and Product Moments of Moment-Statistics for Samples of the Finite and Infinite Populations. Sankhy a: The Indian Journal of St atistics , 6, 363-382.
10. Sukhatme, P. and Sukhatme, B. (1970) Sampling Theory of Surveys with Applicationsrome. Asia Publishing House, Bombay.