IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online

Comparing Ratio-Type Estimators using Auxiliary Information in Simple Random Sampling

Saja Mohammad Hussein[1], Batool Amer[2]

[1] Department of Statistics, University of Baghdad، Baghdad, Iraq
[2] Department of Statistics, University of Baghdad، Baghdad, Iraq

ARTICLE INFO

Article history:

Received	20/11/2024
Revised	20/11/2024
Accepted	14/1/2025
Available online	15/5/2025

Keywords:
Estimator Ray and Singh
Estimator Sisodia and Dwivedi
Estimator Singh and Kakran
Estimator Upadhyaya and Singh
Mean Square Error

ABSTRACT

Researchers in the subject of simple random sampling used to rely on the well-known classical estimator, the ratio estimator, when estimating the population mean. However, with the development of studies and research, researchers proposed a set of new estimators that increase the efficiency of the estimation, especially when using auxiliary information alongside the study variable Y. In this research, a set of ratio-type estimators was presented using auxiliary information to estimate the arithmetic mean of the simple random sample. These estimators were applied to a set of real data, and a comparison was made between these estimators through the MSE comparison criterion Sampling is a fundamental tool in scientific research, as it is used to select a representative sample of the population and provide accurate estimates of its characteristics. This research focuses on estimating the population mean using Ratio estimators in simple random sampling, and aims to compare traditional estimation methods with modified methods To find out whether the modified estimators are more efficient than the traditional estimator in estimating the arithmetic mean of the population or not.

Modified Ratio estimators based on auxiliary variables, such as the coefficient of variation and the coefficient of square, were used to improve the accuracy of population mean estimates. The method was also applied to data on small industrial establishments in Iraq for the year 2023, where 450 establishments were selected out of 1968 establishments.

The results showed that the Sisodia and Dovi (2012) estimators were the most efficient in providing accurate estimates, as they showed the lowest mean square error (MSE) and the least bias compared to other methods..

1. Introduction

The sampling method is one of the most important tools in scientific research, as it provides the necessary rules and principles for collecting data and information essential to the research process. Sampling plays a critical role across various fields of study. It involves selecting a subset of individuals from a larger population, which is then statistically analyzed to derive conclusions that can be generalized to the entire population. This process is vital for obtaining reliable statistical results that accurately reflect the characteristics of the population from which the sample is drawn.

The mean is a fundamental concept in statistical analysis, particularly when applied to a specific population for basic reductions, especially in the context of simple random sampling. Among the methods used for accurate estimation, proportion estimators are an effective tool for estimating variability. This is particularly true when the auxiliary variable is closely related to the target variable.

* Corresponding author. E-**mail** address: saja@coadec.uobaghdad.edu.iq
https://doi.org/10.62933/j8mqkh03

This method is particularly important in cases where the population is large and heterogeneous.Using an estimated proportion helps reduce assumption errors by incorporating effective information about the population, such as the known properties of the covariate.

In this research, we used simple random sampling, the most basic form of random sampling. This method is considered one of the most accurate for studying a population, as it minimizes sampling errors and enhances the reliability of the results. In simple random sampling, a subset is selected from the population under study, with each element having an equal chance of being chosen. This means that every member of the population has the same opportunity to be included in the sample, regardless of their characteristics. As a result, simple random sampling is less prone to selection bias.

1.1.The research problem is that The process of conducting statistical analysis on an entire population is often difficult, costly, and time-consuming. Collecting complete data for the entire population is not always feasible. Therefore, researchers typically resort to sampling to understand the characteristics of the population. Among the various sampling methods, simple random sampling is considered the best because it is less prone to bias compared to other methods. By using a sample, it is possible to estimate the average of the population, providing valuable insights into its overall characteristics.

In 2004, Cem Kadilar and Hulya Cingi proposed modifications to the ratio estimators introduced by Ray and Singh (1981), specifically for conventional and other types of ratio estimators used in small random sampling. The mean square errors for all the estimators were derived in advance, allowing them to be compared. It was found that, in general, smaller ratio estimators tend to be more dynamic and efficient compared to larger ratio estimators.[1]

In (2012) ,Sisodia, B. and Dwivedi, V. proposed an estimator of the population variance-adjusted ratio of the study variable Y under simple random sampling using the coefficient of variation and the mean of the auxiliary variable X. An experimental study was conducted using real populations to prove the performance of the developed estimator compared to the existing estimators. The proposed estimator, as shown in the experimental study, performs better than the existing estimators i.e. it has the smallest mean square error and the highest relative efficiency.[2]

In 2016, scientists Muhammad Abed, Nasir Abbas, Hafiz Zafar Nazir and Zhengjian Lin proposed a dynamic proportional estimator based on the triplet mean, Hodges-Lehmann and range auxiliary variables. It was also used to share the population correlation, mean coefficient and linear combinations of the auxiliary variable. It is observed that the squares of the errors of the triplet mean, Hodges-Lehmann and range auxiliary variables, coefficient of variation and population correlation coefficient of the auxiliary variable are smaller than those of the existing proportional estimators. The parameters such as linearity, coefficient of variation and kurtosis are affected by extreme values while the triplet mean, mean range and Hodges-Lehmann are immune to extreme values.[3]

In 2019, researchers J.O.Muili, A.Audu, A.B.Odeyale and I.O.Olawoyin proposed improved estimators of population mean ratio using linear combination information of the mean, interquartile range and median values of the auxiliary variable. The mean square error and bias were obtained by Taylor series. An experimental study was conducted and the results showed that the proposed estimators are more efficient than the existing estimators.[4]

In 2023, Isah Muhammad, Yahaya Zakari, Mannir Abdu, Rufai Iliyasu, Mujtaba Suleiman, Samaila Manzo, Samaila Manzo, Ali Muhammad, and Adamu Zakar Adamu proposed a ratio-type estimator for the population mean in simple random sampling without return using information about the auxiliary variable. The proposed estimator was obtained using power transformation strategy and incorporating unknown weight. An experimental study was conducted and the results showed that the proposed ratio estimator is better than the existing estimators considered in this study. Therefore, the proposed estimator is more efficient than the existing estimators based on the criteria of mean square error and relative efficiency ratio.[5]

1.2.The aim of this research

to estimate the mean population by using ratio and compare traditional methods with the proposed modified methods. This comparison will be based on employing these estimators with several statistical measures, including the triple population mean, average range, coefficient of variation, and population correlation coefficient.

2. Ratio Estimators for Estimating the Population Mean in Simple Random Sampling

Suppose we have a finite population $U = \{U_1, U_2, \dots, U_N\}$, consisting of N specified units. Let Y represent the study variable, with the value Y_i measured at each unit U_i , where i = 1 , 2 , 3 , … , N i=1,2,3,…,N. Therefore, the set of values for the variable is $Y = \{Y_1, Y_2, \dots, Y_N\}$.
The population mean, denoted by $\bar{Y}$, is given by the formula:

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i \quad \dots (1)$$

where Y_i represents the value of the study variable at unit U_i , and the sum is taken over all units in the population U, which consists of N selected units.
The simplest estimate of the population mean is the sample mean, which is calculated using simple random sampling without replacement. However, to obtain a more efficient estimator of the population mean, information from an auxiliary variable X can be incorporated. Using this auxiliary variable helps improve the precision of the estimate.

Some researchers have proposed a set of relative estimators for estimating the population mean, including:

1- In 1981, researchers Ray and Singh proposed estimators for the population mean in a simple random sample, utilizing auxiliary information (mean of the population) . These estimators are as follows:[6]

$$\bar{Y}_{RS} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}}\bar{X} \quad \dots (2)$$

$$\bar{Y}_{RS} = \hat{R}_{pr}\bar{X} \quad \dots (3)$$

Where:
$\bar{Y}_{RS}$: Suggested estimator for researchers Ray and Singh (1981).

$$b = \frac{S_{xy}}{S_x^2} \quad \dots (4)$$

S_x^2: Sample variance for auxiliary variation.
S_{xy}: Sample covariance between the auxiliary variable and the study variable.

$$\hat{R}_{RS} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}} \quad \dots (5)$$

We observe that when b = 0, the proposed estimator reduces to the ratio estimator previously proposed in (5).

$$MSE(\bar{Y}_{RS}) \cong \frac{1-f}{n}\left[R^2 S_x^2 + S_y^2(1-\rho^2)\right] \quad \dots (6)$$

$$B(\bar{Y}_{RS}) = \frac{(1-f)}{n}\frac{S_x^2}{\bar{Y}}R^2 \quad \dots (7)$$

Where:
$f = \frac{n}{N}$, It is called the sampling fraction.
S_y^2: The population variance for the study variable.

$$R = \frac{\bar{Y}}{\bar{X}}$$

2-In 1981, researchers Sisodia and Dwivedi proposed a modified estimator for estimating the population mean when the coefficient of variation of the auxiliary variable C_x is known. By applying the method introduced by Ray and Singh (1981) to the ratio estimator of Sisodia and Dwivedi, a new ratio estimator $\bar{Y}_{PSD}$ is obtained, which is expressed as follows:[7]

$$\bar{Y}_{PSD} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x} + C_x}(\bar{X} + C_x) \quad \dots (8)$$

$$\bar{Y}_{PSD} = \hat{R}_{PSD}\bar{X}_{SD} \quad \dots (9)$$

Where:

$\hat{R}_{PSD} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}+C_x}$ …(10)

C_x:The coefficient of variation of an auxiliary variable.

$MSE(\bar{Y}_{pSD}) = \frac{1-f}{n}\left[R_{SD}^2 S_x^2 + S_y^2(1-\rho^2)\right]$ (11)

$B(\bar{Y}_{pSD}) = \frac{(1-f)}{n}\frac{S_x^2}{\bar{Y}} R_{pSD}^2$ …(12)

$R_{pSD}^2 = \frac{\bar{Y}}{(\bar{X}+C_x)}$ …(13)

3-For the ratio estimator in Singh and Kakran, in 1993, it is similar to the estimator proposed by Sisodia and Dwivedi, but instead of C_x in the equation, $\beta_2(x)$ is used It is the kurtosis coefficient of the auxiliary variable. This results in the same form for the proposed MSE equation. Therefore, the proposed estimator of Singh and Kakran, along with its MSE equation, is as follows:[8]

$$\bar{Y}_{pSK} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{X}+\beta_2(x)}[\bar{X}+\beta_2(x)] = \hat{R}_{pSK}\bar{X}_{SK} \quad \text{…(14)}$$

And,

$MSE(\bar{Y}_{pSK}) \cong \frac{1-f}{n}$ ...(15)

$B(\bar{Y}_{pSK}) = \frac{(1-f)}{n}\frac{S_x^2}{\bar{Y}} R_{pSK}^2$ …(16)

$R_{pSK}^2 = \frac{\bar{Y}}{\bar{X}+\beta_2(x)}$ …(17)

Where:

$\beta_2(x)$: It refers to the kurtosis coefficient of the auxiliary variable.

4- in 1999, researchers Upadhyaya and Singh proposed a new estimator based on the kurtosis coefficient $\bar{Y}_{pUS1}$, as follows:[9]

$$\bar{Y}_{pUS1} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}\beta_2(x)+C_x}[\bar{X}\beta_2(x)+C_x] = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}_{US1}}\bar{X}_{US1} = \hat{R}_{pUS1}\bar{X}_{US1} \quad \text{... (18)}$$

And,

$$MSE(\bar{Y}_{pUS1}) \cong \frac{1-f}{n}\left[R_{US1}^2\beta_2(x)S_x^2 + S_y^2(1-\rho^2)\right] \quad \text{... (19)}$$

Where:

$\hat{R}_{pUS1} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}\beta_2(x)+C_x}$ ... (20)

$B(\bar{Y}_{pUS1}) = \frac{(1-f)}{n}\frac{S_x^2}{\bar{Y}} R_{pUS1}^2$ ... (21)

$R_{pUS1}^2 = \frac{\bar{Y}\beta_2(x)}{\bar{X}\beta_2(x)+C_x}$ ... (22)

5- in 1999, researchers Upadhyaya and Singh proposed a new estimator based on the kurtosis coefficient $\bar{Y}_{pUS2}$, as follows:[9]

$$\bar{Y}_{pUS2} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}C_x+\beta_2(x)}[\bar{X}C_x+\beta_2(x)] = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}_{US2}}\bar{X}_{US2} = \hat{R}_{pUS2}\bar{X}_{US2} \quad \text{... (23)}$$

And

$$MSE(\bar{Y}_{pUS2}) \cong \frac{1-f}{n}\left[R_{US2}^2 C_x^2 S_x^2 + S_y^2(1-\rho^2)\right] \quad \text{... (24)}$$

Where:

$\hat{R}_{pUS2} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}C_x+\beta_2(x)}$ ... (25)

$B(\bar{Y}_{pUS2}) = \frac{(1-f)}{n}\frac{S_x^2}{\bar{Y}} R_{pUS2}^2$ ... (26)

$R_{pUS2}^2 = \frac{\bar{Y}C_x}{\bar{X}C_x+\beta_2(x)}$ ... (27)

$$MSE(\bar{Y}_{pUS2}) \cong \frac{1-f}{n}\left[R_{US2}^2 C_x^2 S_x^2 + S_y^2(1-\rho^2)\right] \quad \text{... (28)}$$

Where:

$$\beta_2 = \frac{N(N+1)\sum_{i=1}^{N}(x_i-\bar{x})^4}{(N-1)(N-2)(N-3)S^3} - \frac{3(N-1)^2}{(N-2)(N-3)}$$

2.1 Application aspect

Industrial establishments play a major role in the economy as they are one of the basic pillars that contribute to the development of the national economy, and these establishments constitute a large part of economic activity in most countries, and their work has developed significantly over time, and these establishments refer to all economic units in which goods or services are produced or manufactured, in order to understand the size and impact of industrial activity in society, the analysis of statistical data related to industrial establishments is considered essential in it, production can be improved and industrial performance evaluated using statistical tools. When applying simple random sampling to small industrial establishments, the goal is to collect data representing activities, and we use this sampling to understand several aspects, such as productivity, costs, and other important variables related to small industrial establishments. In this paragraph, we will address the applied aspect, in which the population average estimate was found using the ratio estimators based on the coefficient of variation on real data for the survey of small industrial establishments in 2023 AD. We use two variables in the data, the first variable (Y) is the quantity of production and the second is the number of workers (X). The data included (1968) industrial establishments for all governorates of Iraq except the Kurdistan Region. This data was obtained from the Ministry of Planning, the Statistics Authority and Geographic Information Systems. To analyze the data and extract the results in the applied aspect, the MATLAB program was used.

The simple random sample size was estimated using Cochran's law (1940) as follows:[10]

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

Whereas:

$$n_0 = \frac{(t\sigma)^2}{d^2}$$

Whereas:

t: It is the tabular value of the normal distribution that corresponds to the level of significance. The level was chosen 0.05

σ: is the standard deviation of the population Y.

d: It is the permissible error and 0.1 was chosen from the sample size.

After implementing the program, 450 views were randomly selected.

The following results were obtained

Table 1: $\hat{y}$ and mse and The biases of ratio estimators.

	$\hat{\hat{y}}$	mse	bias
$\bar{Y}_{RS}$	1445.400000	15446.000000	5.317600
$\bar{Y}_{pSD}$	1459.500000	13812.000000	4.215400
$\bar{Y}_{pSK}$	1447.700000	15174.000000	5.134100
$\bar{Y}_{US1}$	1446.300000	15344.000000	5.249300
$\bar{Y}_{US2}$	1445.500000	15444.000000	5.316200

3. Results and discussion

From the results of Table (1), we find that the best estimate is the estimate proposed by Sisodia and Davidi because it has the lowest mean square error and the lowest bias,the second best estimator is the estimator proposed by the researchers Singh and Kakran because it has the second lowest mean square error and the second lowest bias, and the estimator proposed by the researchers Upadhyaya and Singh, which uses kurtosis ($\bar{Y}_{US1}$) comes in third place as the best estimator because it has the third lowest mean square error and the third lowest bias , and that the estimator proposed by the researchers Upadhyaya and Singh, which uses kurtosis ($\bar{Y}_{US2}$) comes in fourth place as the best estimator because it has the fourth lowest mean square error and the fourth lowest bias , and the estimator proposed by the researchers Ray and Singh comes in fifth place because it has the fifth lowest mean square error and the fifth lowest bias.

In this research, we used bias and MSE together because MSE alone may not be sufficient to fully assess the efficiency of estimates. Although MSE shows the overall performance of the estimator in terms of variance and the overall presence of error, bias helps to understand the direction in which the estimate is trending relative to the true value. Therefore, it is necessary to compare bias along with MSE to get a comprehensive analysis of the accuracy and reliability of the provided estimate.

4. Conclusions

This research shows that the use of modified ratio estimators can significantly improve the accuracy of population mean estimates, especially when using auxiliary variables that provide important additional information. The main recommendation is to adopt these modern methods in future research to obtain more accurate and efficient estimates, in addition to adopting simple random sampling as an effective sampling method, and paying attention to training in the use of advanced software tools for statistical data analysis. It is recommended to use Sisodia and Dwee estimators for data analysis in industrial studies, as they provide accurate estimates with lower mean square errors.

References

[1] Kadilar, C. & Cingi, H. (2004), Ratio estimators in simple random sampling, *Applied Mathematics and Computation* 151, 893–902.

[2] Sisodia, B. V. S. & Dwivedi, V. K. (2012), A modified ratio estimator using coefficient of variation of auxiliary variable, *Journal of the Indian Society of Agricultural Statistics* 33(1), 13–18.

[3] [14] Abid M., Abbas N., Nazir Z.A. & Lin Z. (2016a). ‘Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters’. Revista Colombiana de Estadistica 39(1): 63 – 79.

[4] J.O. Muili, A. Audu , A.B. Odeyale , and I.O. Olawoyin.(2019).’ Ratio Estimators for Estimating Population Mean Using Tri-mean, Median and Quartile Deviation of Auxiliary Variable’. Journal of Science and Technology Research 1(1) 2019 pp. 91-102.

[5] Isah Muhammad,Yahaya Zakari, Mannir Abdu, Rufai Iliyasu, Mujtaba Suleiman, SamailaManzo, Ali Muhammad and Adamu Zakar Adamu.(2023).’ EnhancedRatio-Type Estimator forFinite Population Mean using Auxiliary Variable in SimpleRandom Sampling’. NIPES Journal of Science and Technology Research 5(1) 2023pp. 242-252.

[6] S.K. Ray, R.K. Singh, (1981), Difference-cum-ratio type estimators, *Journal of Indian Statistical Association* 19 147–151.

[7] B.V.S. Sisodia, V.K. Dwivedi, (1981), A modified ratio estimator using coefficient of variation of auxiliary variable, *Journal of Indian Society Agricultural Statistics* 33 13–18.

[8] H.P. Singh, M.S. Kakran, (1993), A Modified Ratio Estimator Using Known Coefficient of Kurtosis of an Auxiliary Character, (unpublished).

[9] Upadhyaya, L. N. & Singh, H. (1999), Use of transformed auxiliary variable in estimating the finite population mean, *Biometrical Journal* 41(5), 627–636.

[10] Cochran, W. G. (1940), Sampling Techniques, 3 edn, Wiley Eastern Limited, New York.