MODIFICATION OF RATIO ESTIMATOR FOR POPULATION MEAN

J. O. Muili [1], A. Audu [2]

[1] Department of Mathematics, Kebbi State University of Science and Technology Aliero, Nigeria
[2] Department of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria

Corresponding Author: J. O. Muili, jamiubidemi@gmail.com

ABSTRACT: In this study a ratio type estimator of finite population mean has been proposed under simple random sampling without replacement. The properties of the proposed estimator (Bias and Mean Square Error) have been obtained up to first order of approximation using Taylor's Series Expansion and the condition for its efficiency over some existing estimators have been established. The minimum value of the mean square error of the proposed estimator has also been derived. The efficiency of proposed estimator based on optimal value of the constant, exhibit significant improvement over the existing estimators considered in the study. Empirical studies were conducted with natural populations to demonstrate the performance of the proposed estimator comparing with some existing estimators. The results of the empirical study show that the proposed estimator performs better than the existing estimators.

KEYWORDS: Auxiliary Variable, Efficiency, Mean Square Error.

1. INTRODUCTION

The use of auxiliary information in probability sampling has been commonly used device for improving the precision of the estimator of finite population mean or total under study. If use intelligibly, this information provides us with the sampling strategies better than those in which no auxiliary information is used. ([Coc40]) initiated the use of auxiliary information and proposed a ratio estimator for population mean. It is so far established fact that the ratio type estimator provides better efficiency in comparison to simple mean estimator provided the study and auxiliary variables are positively correlated. If the correlation between the study and auxiliary variables is negative, product estimator is more efficient that sample mean estimator. Modified ratio estimators came into existence and were constructed by using one or more unknown constants. In a class of estimators, the estimator with minimum variance or mean square error is regarded as the most efficient estimator. This concept has been utilized by several researchers to improve the efficiency of ratio and product type estimators for estimating population mean of study variable like ([US99], [STK04], [JMM13], [YMS14]) etc.

Let $U = \{U_1, U_2, U_3, \ldots, U_N\}$ be a finite population having N units and each $U_i = (X_i, Y_i)$, i = 1, 2, 3,…,N has a pair of values. Y is the study variable and X is the auxiliary variable which is correlated with Y. Let $y = \{y_1, y_2, \ldots, y_n\}$ and $x = \{x_1, x_2, \ldots, x_n\}$ be n sample values. $\bar{y}$ and $\bar{x}$ are the sample means of the study and auxiliary variables respectively. Let S_y^2 and S_x^2 be the finite population variance of Y and X respectively and s_y^2 and s_x^2 be respective sample variances based on the random sample of size n drawn without replacement. The following are the other notations used throughout this study.

N : Population size
n : Sample size
Y : Study variable
X : Auxiliary variable
$\bar{y}, \bar{x}$: Sample means of study and auxiliary variables
$\bar{Y}, \bar{X}$: Population means of study and auxiliary variables
ρ : Coefficient of correlation
C_y, C_x : Coefficient of variations of study and auxiliary variables
β_1 : Coefficient of skewness of auxiliary variable
β_2 : Coefficient of kurtosis of auxiliary variable
M_d : Median of the auxiliary variable

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2, \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i,$$

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i, \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \gamma = \frac{1-f}{n}, s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2,$$

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2,$$

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

1.1 Reviewed of Related Literature

([Coc40]) proposed the usual ratio estimator for estimating population mean $(\bar{Y})$ of the study variable (Y) as:

$$\hat{Y}_r = \frac{\bar{y}}{\bar{x}}\bar{X} \quad (1)$$

$$Bias(\hat{Y}_r) = \gamma\bar{Y}(C_x^2 - \rho C_x C_y) \quad (2)$$

$$MSE(\bar{Y}_r) = \gamma\bar{Y}^2(C_y^2 + C_x^2 - 2\rho C_y C_x) \quad (3)$$

([SD81]) proposed ratio type estimator based on the value of coefficient of variation (C_x) of auxiliary variable as:

$$\hat{\bar{Y}}_1 = \bar{y}\left(\frac{\bar{X}+C_x}{\bar{x}+C_x}\right) \quad (4)$$

([STK04]) proposed ratio type estimator by replacing coefficient of variation (C_x) in the work of ([SD81]) by coefficient of kurtosis (β_2) of auxiliary variable as:

$$\hat{\bar{Y}}_2 = \bar{y}\left(\frac{\bar{X}+\beta_2}{\bar{x}+\beta_2}\right) \quad (5)$$

([US99]) proposed ratio type estimator by imposing coefficient of variation (C_x) on the work of ([SK93]) as:

$$\hat{\bar{Y}}_3 = \bar{y}\left(\frac{\bar{X}C_x+\beta_2}{\bar{x}C_x+\beta_2}\right) \quad (6)$$

([ST03]) incorporated coefficient of correlation (ρ) of auxiliary variable on the exiting work and defined new estimator as:

$$\hat{\bar{Y}}_4 = \bar{y}\left(\frac{\bar{X}+\rho}{\bar{x}+\rho}\right) \quad (7)$$

([YT10]) proposed ratio type estimator using coefficient of skewness (β_1) of auxiliary variable as:

$$\hat{\bar{Y}}_5 = \bar{y}\left(\frac{\bar{X}+\beta_1}{\bar{x}+\beta_1}\right) \quad (8)$$

Table 1: Estimators, Constants, Biases and Mean Squared Errors (MSE)

Estimator	Constant	Bias	MSE
$\hat{\bar{Y}}_1 = \bar{y}\left(\frac{\bar{X}+C_x}{\bar{x}+C_x}\right)$ Sisodia and Dwivesdi ([SD81])	$\delta_1 = \left(\frac{\bar{X}}{\bar{X}+C_x}\right)$	$\gamma\bar{Y}(\delta_1^2 C_x^2 - 2\delta_1\rho C_x C_y)$	$\gamma\bar{Y}^2(C_y^2 + \delta_1^2 C_x^2 - 2\delta_1\rho C_y C_x)$
$\hat{\bar{Y}}_2 = \bar{y}\left(\frac{\bar{X}+\beta_2}{\bar{x}+\beta_2}\right)$ Singh *et al.* ([STK04])	$\delta_2 = \left(\frac{\bar{X}}{\bar{X}+\beta_2}\right)$	$\gamma\bar{Y}(\delta_2^2 C_x^2 - 2\delta_2\rho C_x C_y)$	$\gamma\bar{Y}^2(C_y^2 + \delta_2^2 C_x^2 - 2\delta_2\rho C_y C_x)$
$\hat{\bar{Y}}_3 = \bar{y}\left(\frac{\bar{X}C_x+\beta_2}{\bar{x}C_x+\beta_2}\right)$ Upadhyaya and Singh ([US99])	$\delta_3 = \left(\frac{\bar{X}C_x}{\bar{X}+\beta_2}\right)$	$\gamma\bar{Y}(\delta_3^2 C_x^2 - 2\delta_3\rho C_x C_y)$	$\gamma\bar{Y}^2(C_y^2 + \delta_3^2 C_x^2 - 2\delta_3\rho C_y C_x)$
$\hat{\bar{Y}}_4 = \bar{y}\left(\frac{\bar{X}+\rho}{\bar{x}+\rho}\right)$ Singh and Tailor ([ST03])	$\delta_4 = \left(\frac{\bar{X}}{\bar{X}+\rho}\right)$	$\gamma\bar{Y}(\delta_4^2 C_x^2 - 2\delta_4\rho C_x C_y)$	$\gamma\bar{Y}^2(C_y^2 + \delta_5^2 C_x^2 - 2\delta_5\rho C_y C_x)$
$\hat{\bar{Y}}_5 = \bar{y}\left(\frac{\bar{X}+\beta_1}{\bar{x}+\beta_1}\right)$ Yan and Tian ([YT10])	$\delta_5 = \left(\frac{\bar{X}}{\bar{X}+\beta_1}\right)$	$\gamma\bar{Y}(\delta_5^2 C_x^2 - 2\delta_5\rho C_x C_y)$	$\gamma\bar{Y}^2(C_y^2 + \delta_5^2 C_x^2 - 2\delta_5\rho C_y C_x)$
$\hat{\bar{Y}}_6 = \bar{y}\left(\frac{\bar{X}+M_d}{\bar{x}+M_d}\right)$ Subramani and Kumarpandiyan ([SK13])	$\delta_6 = \left(\frac{\bar{X}}{\bar{X}+M_d}\right)$	$\gamma\bar{Y}(\delta_6^2 C_x^2 - 2\delta_6\rho C_x C_y)$	$\gamma\bar{Y}^2(C_y^2 + \delta_6^2 C_x^2 - 2\delta_6\rho C_y C_x)$
$\hat{\bar{Y}}_7 = \bar{y}\left(\frac{\bar{X}+n}{\bar{x}+n}\right)$ Jerajuddin and Kishun ([JK16])	$\delta_7 = \left(\frac{\bar{X}}{\bar{X}+n}\right)$	$\gamma\bar{Y}(\delta_7^2 C_x^2 - 2\delta_7\rho C_x C_y)$	$\gamma\bar{Y}^2(C_y^2 + \delta_7^2 C_x^2 - 2\delta_7\rho C_y C_x)$

([SK13]) proposed a ratio estimator using median of auxiliary variable (M_d) as:

$$\hat{\bar{Y}}_6 = \bar{y}\left(\frac{\bar{X}+M_d}{\bar{x}+M_d}\right) \quad (9)$$

([JK16]) was the first to used sample size (n) in order to increase precision of ([Coc40]) estimator as:

$$\hat{\bar{Y}}_7 = \bar{y}\left(\frac{\bar{X}+n}{\bar{x}+n}\right) \quad (10)$$

The above estimators with their properties are summarize in the table 1 which shows the estimators, constants, biases and Mean Squared Errors (MSE) in literature.

2. PROPOSED ESTIMATOR

Motivated by the work of Jerajuddin and Kishun ([JK16]) and Subramani and Kumarpandiyan ([SK13]), we proposed a ratio estimator of finite population mean as:

$$\hat{\bar{Y}}_p = \bar{y}\left(\frac{\bar{X}+n}{\bar{x}+n}\right)^{\alpha} \quad (11)$$

2.1 Assumptions of the Proposed Estimator

i. Population of study is finite.
ii. Sample size is known.
iii. Information about the study variable *(y)* and auxiliary variable *(x)* in sample are completely available.

2.2 Properties of the Proposed Estimator

Defining $e_0 \frac{\bar{y}-\bar{Y}}{\bar{Y}}$ and $e_1 \frac{\bar{x}-\bar{X}}{\bar{X}}$ such that $\bar{y} = \bar{Y}(1 + e0$, and $x=X1+e1$, from the definition of $e0$and $e1$, obtaining

$$\left.\begin{matrix} E(e_0) = E(e_1) = 0,\ E(e_0^2) = \gamma C_y^2 \\ E(e_1^2) = \gamma C_x^2, E(e_0 e_1) = \gamma\rho C_y C_x \end{matrix}\right\} \quad (12)$$

Now expressing (11) in error terms as:

$$\hat{\bar{Y}}_p = \bar{Y}(1+e_0)\left(\frac{\bar{X}+n}{\bar{X}(1+e_1)+n}\right)^{\alpha} \quad (13)$$

Where $\delta_p = \frac{\bar{X}}{\bar{X}+n}$
Simplifying (13) up to first order approximation, it reduces to (14) as

$$\hat{\bar{Y}}_p = \bar{Y}(1+e_0)(1+\delta_p e_1)^{-\alpha} \quad (14)$$

$$\hat{\bar{Y}}_p = \bar{Y}(1+e_0)\left(1 - \alpha\delta_p e_1 + \frac{\delta_p^2\alpha(\alpha+1)}{2!}e_1^2\right) \quad (15)$$

Taking Expectation of both sides of (15)

$$E\left(\hat{\bar{Y}}_p - \bar{Y}\right) = \bar{Y}E\left(\begin{matrix} e_0 - \alpha\delta_p e_1 + \frac{\delta_p^2\alpha(\alpha+1)}{2}e_1^2 \\ -\alpha\delta_p e_0 e_1 \end{matrix}\right) \quad (16)$$

Applying the results of (12) obtaining the $Bias\left(\hat{\bar{Y}}_p\right)$ as:

$$Bias\left(\hat{\bar{Y}}_p\right) = \gamma\bar{Y}\left[\frac{\alpha(\alpha+1)}{2}\delta_p^2 C_x^2 - \alpha\delta_p\rho C_y C_x\right] \quad (17)$$

Squaring and expanding (16), obtaining (18) as:

$$MSE\left(\hat{\bar{Y}}_p\right) = \bar{Y}^2\left[e_0^2 + \alpha^2\delta_p^2 e_1^2 - 2\alpha\delta_p e_0 e_1\right] \quad (18)$$

Taking expectation and apply the results of (12), obtaining $MSE\left(\hat{\bar{Y}}_p\right)$ as:

$$MSE\left(\hat{\bar{Y}}_p\right) = \gamma\bar{Y}^2\left[\begin{matrix} C_y^2 + \alpha^2\delta_p^2 C_x^2 \\ -2\alpha\delta_p\rho C_y C_x \end{matrix}\right] \quad (19)$$

Obtaining the expression for the value of α, differentiate $MSE\left(\hat{\bar{Y}}_p\right)$ partially with respect to α and equate to zero as:

$$\frac{\left(MSE\left(\hat{\bar{Y}}_p\right)\right)}{\partial\alpha} = \gamma\bar{Y}^2\left[2\alpha\delta_p^2 C_x^2 - 2\delta_p\rho C_y C_x\right] = 0 \quad (20)$$

Simplifying (20) for α, obtaining optimum value of α as:

$$\alpha = \frac{\rho C_y}{\delta_p C_x} \quad (21)$$

Substituting (21) in (19)

$$MSE\left(\hat{\bar{Y}}_p\right)_{min} = \gamma\bar{Y}^2 C_y^2\left[1-\rho^2\right] \quad (22)$$

2.3 Percentage Relative Efficiency (PRE)

Percentage Relative Efficiency (PRE) is the performance measures of estimators of efficiency gain. PRE of the estimators were computed using the formula:

$$PRE = \frac{MSE\left(\hat{\bar{Y}}_r\right)}{MSE\left(\hat{\bar{Y}}_i\right)}\ \text{X}\ 100 \quad (23)$$

Where $MSE\left(\hat{\bar{Y}}_i\right)$ are the existing and proposed estimators in the study.

2.4 Data Set

Population I: X= Data on number of workers and Y= Output for 80 factories in a region

Population II: X= Data on number of workers and Y= Output for 40 factories in a region

Population III: Fixed Capital and Y= Output for 40 factories in a region

Table 2 shows the numerical values of the parameters used in computing Bias, Mean Square Error (MSE) and Percentage Relative Efficiency (PRE) of the proposed and other related estimators considered in the study.

Table 3 shows the bias of the proposed and other related estimators.

Table 4 shows the MSE of the proposed and some related estimators considered in the study for all the data sets I, II, and III. The results revealed that the proposed estimator has the least MSE among the estimators.

Table 2: Populations I, II and III: ([JK16])

Parameter	Population I	Population II	Population III
N	80	40	40
n	20	8	8
$\bar{Y}$	51.8264	50.7858	50.7858
$\bar{X}$	11.2646	2.3033	9.4543
ρ	0.9413	0.8006	0.8349
C_y	0.3542	0.3295	0.3295
C_x	0.7507	0.8406	0.6756
β_1	1.0500	0.9740	0.8799
β_2	-0.0634	-0.5344	-0.4622
M_d	7.5750	1.2500	7.0700

Table 3: Bias of the Estimators

Estimator	Population I	Population II	Population III
$\hat{\bar{Y}}_r$	0.6088	2.4624	1.3742
$\hat{\bar{Y}}_1$	0.5066	1.1009	1.1379
$\hat{\bar{Y}}_2$	0.6185	4.6177	1.5697
$\hat{\bar{Y}}_3$	0.6218	5.2908	1.6759
$\hat{\bar{Y}}_4$	0.4839	1.1402	1.0895
$\hat{\bar{Y}}_5$	0.4715	0.9809	1.0763
$\hat{\bar{Y}}_6$	0.1007	0.7777	0.2186
$\hat{\bar{Y}}_7$	0.0331	0.0724	0.1687
$\hat{\bar{Y}}_p$ Proposed	-0.0204	-0.0508	0.0635

Table 4: Mean Square Error (MSE)

Estimator	Population I	Population II	Population III
$\hat{\bar{Y}}_r$	18.9793	95.8641	49.8536
$\hat{\bar{Y}}_1$	15.2581	42.0165	41.0566
$\hat{\bar{Y}}_2$	19.3383	188.0469	57.3218
$\hat{\bar{Y}}_3$	19.4592	217.7027	61.4403
$\hat{\bar{Y}}_4$	14.4503	43.4736	39.2910
$\hat{\bar{Y}}_5$	14.0113	37.6265	38.8113
$\hat{\bar{Y}}_6$	2.7825	30.4301	11.6837
$\hat{\bar{Y}}_7$	1.8389	11.5428	10.6098
$\hat{\bar{Y}}_p$ Proposed	1.4400	10.0540	8.4831

Table 5: Percentage Relative Efficiency (PRE)

Estimator	Population I	Population II	Population III
$\hat{\bar{Y}}_r$	100	100	100
$\hat{\bar{Y}}_1$	124.3884	228.1582	121.4265
$\hat{\bar{Y}}_2$	98.14358	50.97882	86.97145
$\hat{\bar{Y}}_3$	97.53381	44.03441	81.14153
$\hat{\bar{Y}}_4$	131.3419	220.5111	126.883
$\hat{\bar{Y}}_5$	135.4571	254.7781	128.4513
$\hat{\bar{Y}}_6$	682.0952	315.0305	426.6936
$\hat{\bar{Y}}_7$	1032.101	830.5099	469.8826
$\hat{\bar{Y}}_p$ Proposed	1318.007	953.4921	587.6814

Table 5 shows the Percentage Relative Efficiency (PRE) of the proposed and some related estimators considered in the study for all the data sets I, II, and III. The results revealed that the proposed estimator has highest PRE than other estimators.

3. EFFICIENCY COMPARISONS

In this section efficiencies of the proposed estimator are compared with efficiencies of some existing estimators in the literature.

The $\hat{\bar{Y}}_p$of estimator of population mean is more efficient than $\hat{\bar{Y}}_r$ if,

$$MSE\left(\hat{\bar{Y}}_p\right) < MSE\left(\hat{\bar{Y}}_r\right)$$
$$\gamma\bar{Y}^2C_y^2[1-\rho^2] < \gamma\bar{Y}^2\left(C_y^2 + C_x^2 - 2\rho C_y C_x\right) \quad (24)$$

The $\hat{Y}_p$of proposed estimator of the population mean is more efficient than $\hat{Y}_i$ if,

$$MSE\left(\hat{\bar{Y}}_p\right) < MSE\left(\hat{\bar{Y}}_i\right) \qquad i=1, 2, 3,\dots,7$$
$$\gamma\bar{Y}^2C_y^2[1-\rho^2] < \gamma\bar{Y}^2\left(C_y^2 + \delta_i^2C_x^2 - 2\delta_i\rho C_y C_x\right) \quad (25)$$

When conditions (24) and (25) are satisfied, we can conclude that the proposed estimator is more efficient than some selected existing estimators.

4. RESULT AND DISCUSSION

In this study, we proposed a ratio type population mean estimator. The performance of the proposed estimator over the usual ratio estimator and some selected existing estimators using three natural populations in which their properties (Bias and Mean Square Errors (MSEs)) were established and comparing their PREs. Tables 4 and 5 show the results of Mean Square Error (MSE) and Percentage Relative Efficiency (PRE) of the proposed and some related estimators considered in the study for all the data sets I, II, and III. The results revealed that the proposed estimator has minimum MSE and highest PRE than other estimators. This results implies that the average of dispersion of the proposed estimators from the population mean is smaller and this indicate that the proposed estimator give better estimates on the average than other estimators in the study.

5. CONCLUSION

Based on the results in Table 5, it is clear that the proposed estimator $\left(\hat{\bar{Y}}_p\right)$ works better than the other existing estimators having the minimum Mean Square Error (MSE) and the highest Percentage Relative Error (PRE). We therefore recommend for use in estimating population mean.

REFERENCES

[Coc40] **Cochran W. G.** – *Sampling Techniques*, 3rd Edition. Wiley Eastern Limited, India, 1940.

[JK16] **Jerajuddin M., Kishun J.** – *Modified Ratio Estimators for Population Mean Using Size of the Sample, Selected From Population*, IJSRSET, 2, 2, 10-16, 2016.

[JMM13] **Jeelani M. I., Maqbool S., Mir S. A.** – *Modified Ratio Estimators of Population Mean using Linear Combination of Estimators of Population Mean using Linear Combination of Coefficient of Skewness and Quartile Deviation*. International Journal of Modern Mathematical Sciences 6: 174 – 183, 2013.

[SD81] **Sisodia B. V. S., Dwivedi V. K.** – *A Modified Ratio Estimator using Coefficient of Variation of Auxiliary Variable*; Jour. of Indian. Soc. of Agri. Stat., Vol. 33(1). Pp. 13– 18, 1981.

[SK13] **Subramani J., Kumarpandiyan G.** – *New Modified Ratio Estimation of Population Mean when Median of Auxiliary Variable is known*; Pak.j.stat.oper.res. Vol. IX No.2, pp 137-145, 2013.

[ST03] **Singh H. P., Tailor R.** – *Use of known Correlation Coefficient in Estimating the Finite Population Means*; Statistics in Transition 6 (4), 555-560, 2003.

[STK04] **Singh H. P., Tailor R., Kakran M. S.** – *An Improved Estimation of Population Mean using Power Transformation*, Journal of the Indian Society of Agricultural Statistics, 8 (2): 223–230, 2004.

[US99] **Upadhyaya L. N., Singh H. P.** – *Use of Transformed Auxiliary Variable in Estimating the Finite Population Mean*. Biometrical Journal, 41: 627-636, 1999.

[YT10] **Yan Z., Tian B.** – *Ratio Method to the Mean Estimation Using Coefficient of Skewness of Auxiliary Variable*. ICICA, Part II, CCIS 106, 103-110, 2010.

[YMS14] **Yadav S. K., Mishra S. S., Shukla A. K.** – *Improved Ratio Estimators for Population Mean Based on Median Using Linear Combination of Population Mean and Median of an Auxiliary Variable*; American Journal of Operational Research, 4(2): 21-27, 2014.