Journal of the Nigerian Association of Mathematical Physics
Volume 51 (*May, 2019 Issue*), **pp113 – 118**
© J. of NAMP

A MODIFIED ESTIMATOR IN ADAPTIVE CLUSTER SAMPLING

[1]Olayiwola O.M., [1]Ajayi A. O., [1]Aderibigbe T. A., [2]Nurudeen T. S. and [3]Akintola K.A.

1. **Department of Statistics, College of Physical Sciences, Federal University of Agriculture, Abeokuta, Nigeria**
2. **Department of Maths/Statistics, Lagos State Polytechnic, Ikorodu.**
3. **Department of Statistics, Oyo State College of Agriculture, Igboora**

Abstract

Adaptive cluster sampling is an efficient method for estimating rare and hidden clustered population sizes. Researchers have worked on estimator in adaptive cluster sampling using linear combination of coefficient of skewness, kurtosis and variation. There is dearth of information on estimators based on linear combination of median and coefficient of kurtosis to handle problem of outliers in data sets. This study modified the existing estimator in adaptive cluster sampling using linear combination of median and coefficient of skewness. The data on cotton production (metric tonnes) in Nigeria for the year 2005 and 2006 were extracted from 2013 publications of National Bureau of Statistics. Adaptive cluster sampling was used to select ten cotton producing states (Adamawa, Bauchi, Borno, Gombe, Kaduna, Kano, Katsina, Niger, Sokoto, and Zamfara) at random. Cotton production for 2005 and 2006 were considered as the auxiliary variable and variable of interest respectively. Descriptive statistics and box plot were used to describe the data sets. The large sample properties of the proposed estimator were studied up to the first order of approximation. The bias and Mean Square Error (MSE) for the proposed estimator was obtained and compared with that of the existing estimator to determine its efficiency. Data analysis in r-programming was used to analyse the dataset. Descriptive statistics showed that the average production of cottons for 2005 and 2006 were 46.03 and 48.72 metric tonnes respectively. Box plot showed the presence of outliers in the data set. The median and coefficient of skewness for cotton production in 2005 were 25.11 and 0.55 respectively. The biasness and MSE for the proposed estimator are -0.25 and 0.76 respectively, while for the existing estimator were 0.93 and 3.74 respectively. The modified estimator solved the problems of outliers in data sets efficiently with minimum MSE, hence more efficient.

1.0 INTRODUCTION

Adaptive cluster sampling, proposed in [1], is an efficient method for sampling rare and hidden clustered populations. In adaptive cluster sampling, an initial sample of units is selected by simple random sampling. If the value of the variable of interest from a sampled unit satisfies a pre-specified condition C, that is $(i, y_i > c)$ then the unit's neighborhood will also be added to the sample. If any other units that are "adaptively" added also satisfy the condition C, then their neighborhoods are also added to the sample. This process is continued until no more units that satisfy the condition are found. The set of all units selected and all neighboring units that satisfy the condition is called a network. The adaptive sample units, which do not satisfy the condition, are called edge units. A network and its associated edge units are called a cluster. If a unit is selected in the initial sample and does not satisfy the condition C, then there is only one unit in the network.

It is well known that the variable about which we have full information is known as auxiliary variable and the information is known as auxiliary information which is highly (positively or negatively) correlated with the variable under study. Whenever auxiliary variable (information) is known, one would like to use it at the design or estimation stage since it is well known and established that the use of auxiliary information in sampling theory enhances the efficiency of the estimators and it is in use since the use of sampling itself. In this paper, we will study the estimator of population mean in adaptive cluster sampling using an auxiliary variable.

Corresponding Author**:** Ajayi A.O., Email: ajayiabayomi0505@gmail.com, Tel: +2348029105483

2.0 LITERATURE REVIEW

Estimators Under Adaptive Cluster Sampling

Let the population consists of N distinct identifiable units labeled from 1,2, ...,N . Let y_i and $x_i\left(i=1,2,....,N\right)$denote the observation on the characteristic x and y respectively, under study for the i^{th} unit of the population.

Let n denote the initial sample size and n denote the final sample size. Let ψ_i denote the network that includes unit i and mi be the number of units in that network. The initial sample of units is selected by simple random sampling without replacement.

The estimator of the population mean for the variable of interest under adaptive cluster sampling based on Hansen-Hurwitz estimator as,

$$\bar{y}_{ac}=\frac{1}{n}\sum_{i=1}^{n}\left(w_y\right)_i \qquad (1)$$

Where $\left(w_y\right)_i$ is the average of the variable y under study in the network that includes unit i of the initial sample, that is:

$$\left(w_y\right)_i=\frac{1}{m_i}\sum_{j\in\psi_i}\left(y_j\right) \qquad (2)$$

The variance of $\bar{y}_{ac}$ is given by,

$$V\left(\bar{y}_{ac}\right)=\frac{N-n}{N\left(N-1\right)}\sum_{i=1}^{N}\left[\left(w_y\right)_i-\mu_y\right]^2 \qquad (3)$$

The following ratio estimator in adaptive cluster sampling was proposed by [2] as,

$$\bar{y}_{acR}=\frac{\bar{y}_{ac}}{\bar{x}_{ac}}\mu_x \qquad (4)$$

Where $\bar{x}_{ac}=\frac{1}{n}\sum_{i=1}^{n}\left(w_x\right)_i$ and $\left(w_x\right)_i$ is the average of the auxiliary variable x in the network that includes unit i of the initial sample, that is, $\left(w_x\right)_i=\frac{1}{m_i}\sum_{j\in\psi_i}\left(x_j\right)$

The first order approximated MSE of $\bar{y}_{acR}$ is

$$MSE\left(\bar{y}_{acR}\right)\approx\frac{\left(1-f\right)}{n}\mu_y^2\left[C_{wy}^2+C_{wx}^2-2\rho_{wx.wy}C_{wy}C_{wx}\right] \qquad (5)$$

The following ratio type estimators of population mean proposed by [3] using parameters of the auxiliary information based on [4] estimator and [5] two estimators under adaptive cluster sampling as,

$$\bar{y}_{acR1}=\bar{y}_{ac}\left(\frac{\mu_y+C_{wx}}{\bar{x}_{ac}+C_{wx}}\right) \qquad (6)$$

$$\bar{y}_{acR2}=\bar{y}_{ac}\left[\frac{\mu_y\beta_{2(wx)}+C_{wx}}{\bar{x}_{ac}\beta_{2(wx)}+C_{wx}}\right] \qquad (7)$$

$$\bar{y}_{acR3}=\bar{y}_{ac}\left[\frac{\mu_x+\beta_{2(wx)}}{\bar{x}_{ac}+\beta_{2(wx)}}\right] \qquad (8)$$

Where C_{wx} is the population coefficient of variation of w_x, $\beta_{2(wx)}$ is the population coefficient of kurtosis of w_x Where

$$S_{wy}^2=\frac{1}{N-1}\sum_{i=1}^{N}\left[\left(w_y\right)_i-\mu_y\right]^2, \qquad (9)$$

$$S_{wx}^2=\frac{1}{N-1}\sum_{i=1}^{N}\left[\left(w_x\right)_i-\mu_x\right]^2, \qquad (10)$$

$$S_{wx,wy}=\frac{1}{N-1}\sum_{i=1}^{N}\left[\left(w_x\right)_i-\mu_x\right]\left[\left(w_y\right)_i-\mu_y\right]=\rho_{wx.wy}S_{wx}S_{wy} \qquad (11)$$

The mean square errors of above estimators using Taylor series method up to the first order of approximations respectively are,

$$MSE(\bar{y}_{acR1}) \approx \frac{(1-f)}{n}\mu_y^2\left[C_{wy}^2 + \theta_{w1}^2 C_{wx}^2 - 2\theta_{w1}\rho_{wx.wy}C_{wy}C_{wx}\right] \quad (12)$$

$$MSE(\bar{y}_{acR2}) \approx \frac{(1-f)}{n}\mu_y^2\left[C_{wy}^2 + \theta_{w2}^2 C_{wx}^2 - 2\theta_{w2}\rho_{wx.wy}C_{wy}C_{wx}\right] \quad (13)$$

$$MSE(\bar{y}_{acR3}) \approx \frac{(1-f)}{n}\mu_y^2\left[C_{wy}^2 + \theta_{w3}^2 C_{wx}^2 - 2\theta_{w3}\rho_{wx.wy}C_{wy}C_{wx}\right] \quad (14)$$

Where

$$\theta_{w1} = \frac{\mu_y}{\mu_y + C_{wx}}, \quad \theta_{w2} = \frac{\mu_x\beta_{2(wx)}}{\mu_x\beta_{2(wx)} + C_{wx}}, \quad \theta_{w3} = \frac{\mu_x}{\mu_x + \beta_{2(wx)}} \text{ and } R = \frac{\mu_x}{\mu_y}$$

The problem of estimating the population mean in the presence of an auxiliary variable has been widely discussed in finite population sampling literature. Out of many ratio, product and regression methods of estimation are good examples in this context. A class of estimator of the population mean using one auxiliary variable in the stratified random sampling was suggested by [6] and examined the MSE of the estimators up to the k_{th} order of approximation. Stratified random sampling estimator was proposed by [7], [8], [9] and [10]. Also [11] suggested some ratio cum product estimators in simple random sampling. Some exponential ratio type estimators was suggested by [12] and [13].

3.0 METHODOLOGY

Notations used are:

N- Population Size

n- Sample Size

$f = n/N$ $f = n/N$ Sampling Fraction

Y- Variable of Interest

X- Auxiliary Variable

$\bar{X}, \bar{Y}$ $\bar{X}, \bar{Y}$- Population Means

$\bar{x}, \bar{y}$ $\bar{x}, \bar{y}$- Sample Means

S_x, S_y S_x, S_y- Population Standard Deviations

C_x, C_y C_x, C_y- Co-Efficient Of Variations

ρ ρ - Co-Efficient Of Correlation

Proposed Estimator Using Linear Combination of Median and Kurtosis

$$\bar{y}_{acR7} = \bar{y}_{ac}\left(\frac{\beta_{2(wx)}\mu_x + Md\beta_{2(wx)}}{\beta_{2(wx)}\bar{x}_{ac} + Md\beta_{2(wx)}}\right) \quad (15)$$

Deriving the biasness and Mean Square error of (15)

Where

The estimator of the population mean for the variable of interest under adaptive cluster sampling based on Hansen-Hurwitz estimator as,

$$\bar{y}_{ac} = \frac{1}{n}\sum_{i=1}^{n}(w_y)_i \quad (16)$$

Where $(w_y)_i$ is the average of the variable *y* under study in the network that includes unit *i* of the initial sample, that is:

$$(w_y)_i = \frac{1}{m_i}\sum_{j\in\psi_i}(y_j) \quad (17)$$

The variance of $\bar{y}_{ac}$ is given by,

$$V(\bar{y}_{ac}) = \frac{N-n}{N(N-1)}\sum_{i=1}^{N}\left[(w_y)_i - \mu_y\right]^2 \quad (18)$$

Derivation

Let

$$e_0 = \frac{\bar{y}_{ac} - \mu_y}{\mu_y}, \quad e_1 = \frac{\bar{x}_{ac} - \mu_x}{\mu_x}$$

such that

$$E(e_0) = E(e_1) = 0$$

$$E(e_0^2) = \frac{1}{\mu_y^2} S_{wy}^2 = \frac{1}{\mu_y^2}\left[\frac{1}{(N-1)}\sum_{i=1}^{N}\left[(w_y)_i - \mu_y\right]^2\right],$$

$$E(e_1^2) = \frac{1}{\mu_x^2} S_{wx}^2 = \frac{1}{\mu_x^2}\left[\frac{1}{(N-1)}\sum_{i=1}^{N}\left[(w_x)_i - \mu_x\right]^2\right],$$

$$E(e_0 e_1) = \frac{1}{\mu_x \mu_y} S_{wx,wy} = \frac{1}{N-1}\sum_{i=1}^{N}\left[(w_x)_i - \mu_x\right]\left[(w_y)_i - \mu_y\right] = \rho_{wx.wy} S_{wx} S_{wy} .$$

Expressing $\bar{y}_{acR6}$ in terms of *e's*, we have

$$\bar{y}_{acR2} = \mu_y (1 + e_0)(1 + \theta_{w2h} e_1)^{-1}$$

Expanding to the first degree of approximation

$$\bar{y}_{acR2} \approx \mu_y (1 + e_0)(1 - \theta_{w2} e_1 + (\theta_{w2} e_1)^2) \quad (19)$$

$$\bar{y}_{acR2} \approx \mu_y (1 + e_0 - \theta_{w2} e_1 + (\theta_{w2} e_1)^2 - \theta_{w2} e_0 e_1) \quad (20)$$

Subtracting μ_y and Taking expectation of both sides to derive the biasness

To derive the Mean square error, after subtracting μ_y take square of both sides and take expection

Where $\theta_{w2h} = \dfrac{\beta_{2(wx)}\bar{X}}{\beta_{2(wx)}\bar{X} + Md\beta_{2(wx)}}$.

The mean square error of and biasness of $\bar{y}_{acR2}$ is given by

$$B(\bar{y}_{acR2}) \approx \frac{(1-f)}{n}\mu_y^2\left[\theta_{w2}^2 C_{wx}^2 - \theta_{w2}\rho_{wx,wy} C_{wy} C_{wx}\right] \quad (21)$$

$$MSE(\bar{y}_{acR2}) \approx \left(\frac{(1-f)}{n}\mu_y^2\left[C_{wy}^2 + \theta_{w2}^2 C_{wx}^2 - 2\theta_{w6}\rho_{wx.wy} C_{wy} C_{wx}\right]\right) \quad (22)$$

4.0 RESULTS AND DISCUSSION

Source and Scope of Data (Cotton Data)

Cotton as a major cash crop, is of considerable social and economic importance to Nigeria. Cotton/textile activities are widespread in the country. Cotton production in Nigeria dates back to
1903 with the British Cotton Growers Association taking the lead until 1974, when it was disbanded and replaced by the Cotton Marketing Board to develop, gin and market the produce. Following the deregulation of the Nigerian economy in 1986, the Board was abolished vis-à-vis the economic activities rendered by it. The Cotton Consultative Committee (CCC) was set up in an advisory capacity to the public sector, while a cotton revolving fund scheme with a management Committee (CRFMC) was put in place to ensure the sustainable supply of certified cottonseed to farmers. In 2005, the Cotton Development Committee was established which subsumed both the CCC and the CRFMC, to address the cotton economy in a holistic manner. According Nigeria Bureau of Statistics (NBS) the states are as shown in Table 1.

Table 1. Production of Cotton in Nigeria

States Name	2005	2006
Adamawa	13.00	11.86
Bauchi	29.37	25.42
Borno	85.06	111.16
Gombe	7.81	4.98
Kaduna	.06	.19
Kano	77.65	72.75
Katsina	106.96	131.54
Niger	5.11	4.45
Sokoto	20.85	18.18
Zamfara	114.38	106.65

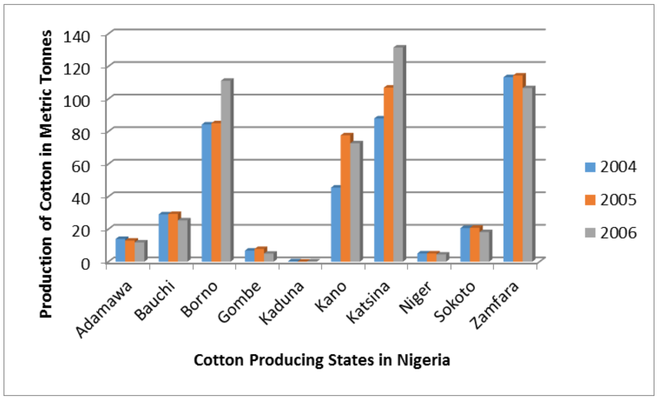

Fig. 1: Cotton Producing states in Nigeria

Table 2: Descriptive statistics for Cotton Data

Year	Minimum	Maximum	Mean	Std Deviation	Median	Skewness
2005(X)	.06	114.38	46.03	44.92	25.11	.550
20 06(Y)	.19	131.54	48.72	51.37	21.80	.660

Table 3: Table of Estimators

	N	n	Existing $B(\bar{y}_R)$	Proposed $B(\bar{y}_{acR})$	Existing $MSE(\bar{y}_R)$	Proposed $MSE(\bar{y}_{acR})$	Efficiency
Y, X	37	10	0.93	-0.25	3.74	0.76	4.88

5.0 CONCLUSION

This work proposed estimator in adaptive cluster sampling, the mean square error was derived. The mean square error was used to compare existing ratio estimator in adaptive cluster sampling and it was found to be more efficient.

The study used a derived ratio estimator using the interactive effects of median and kurtosis with population mean of the auxiliary variable in adaptive cluster sampling and compared the precision of both the modified estimator and the conventional one and discovered that there is gain in precision in the derived ratio estimator. Hence, we strongly recommend that the new modified estimator may be preferred over the existing modified ratio estimators for the use of practical applications.

REFERENCES

[1] Thompson, S.K. 1990. Adaptive cluster sampling, *J. Am. Statist. Assoc.*, 85, 1050-1059

[2] Dryver, A.L. and C. Chao (2007). Ratio estimators in adaptive cluster sampling, Environmetrics,**18**, 607-620.

[3] Chutiman. N., (2013). Adaptive cluster sampling using auxiliary variable, J. Math. Stat., **9**, 249-255

[4] Sisodia B.V.S. and V.K. Dwivedi. 1981. A modified ratio estimator using coefficient of variation of auxiliary variable, *J. Indian Soc. Agric. Statist.*, 33, 13-18.

[5] Upadhyaya, L.N. and H. P. Singh. 1999. Use of transformed auxiliary variable in estimating the finite population mean, *Biometri. J.*, 41, 627-636.

[6] Diana, G. (1993). A class of estimators of the population mean in stratified random sampling. Statistica 53(1), 59–66.

[7] Kadilar, C. and H. Cingi. 2003. A study on the chain ratio-type estimator, *Hacettepe Journal of Mathematics and Statistics*, 32,105-108.

[8] Singh, R., Kumar, M., Singh, R.D., and Chaudhary, M.K. (2008). Exponential Ratio Type Estimators in Stratified Random Sampling. (Proceedings of International Symposium on Optimisation and Statistics (I.S.O.S.), 2008 held at A.M.U., Aligarh, India, during 29-31 Dec 2008).

[9] Singh, H. P., Vishwakarma, G. K. (2008). A family of estimators of population mean using auxiliary information in stratified sampling. Commun. Statist. Theo. Meth.37, 1038–1050.

[10] Koyuncu, N. and Kadilar, C. (2009). Family of Estimators of Population Mean Using Two Auxiliary Variables in Stratified Random Sampling. *Commun. in Statist. Theo. and Meth.*, 38: 2398–2417. 6.

[11] Perri, P. F. (2007): Improved ratio-cum-product type estimators. Statist. Trans. 8:51–69. 7.

[12] Bahl. S. and Tuteja, R.K. (1991): Ratio and Product Type Exponential Estimator. Infrm. and Optim. Sci., XIII, 159-163. 2.

[13] Singh, R., Chauhan, P., Sawan, N., & Smarandache, F. (2007). Improvement in estimating the population mean using exponential estimator in simple random sampling. *Auxiliary Information and a priori Values in Construction of Improved Estimators*, 33.