A Note on Efficient Median-Based Ratio Estimator for Population Mean with Applications

U. K. Abdullahi[1]; F. I. Ugwuowo[2]

Department of Statistics,
University of Nigeria, Nsukka, Nigeria.
E-mail: umarkabir9@gmail.com[1]

***Abstract* — Literature in sample survey shows that estimation of mean is one of the challenging aspects and it receives wide attention from the research community with the aim of developing estimators with high precision. So also in agricultural surveys mean/average is one of the parameter of interest. In this research paper we develop an efficient estimator for population mean named median-based ratio exponential ratio estimator using some methods in the literature. The concepts of first order approximation, expectation were used to obtain the bias and mean squared error of the estimator. Real life dataset from two agricultural populations were used to compare the efficiency of the develop estimator with respect to some existing estimators, findings from the empirical analysis shows a significant efficiency gain (more than 20%), hence the estimator is recommended for practical application.**

Keywords: *Finite population mean, Auxiliary parameter, Median-based, exponential ratio Estimator.*

I. Introduction

Sampling is a technique for selecting a sample or subset of the population to make statistical inferences on some characteristics of the whole population. Auxiliary parameter is any parameter which is used at any stage of survey to enhance the efficiency of the estimator of the parameter of interest.

It is also well known that incorporation of auxiliary parameter in the estimation procedure yields better estimators. For example, if the parameter of interest is the mean/average quantity of fruits produced per plot, then auxiliary parameter can be the population mean/average of the production of fruit in the same plot in previous year, another auxiliary parameter can be the population mean/average number of workers in each plot.

Multiplicities of techniques have been used in deriving efficient estimators for population mean under different probability and non-probability sampling schemes, including the exponential method. The aim of the study is to develop median-based ratio exponential ratio estimator for estimating population mean under simple random sampling scheme. The specific objectives were to: (i) develop the median-based ratio exponential ratio estimator (ii) derive the bias and mean square error of the estimator; (iii) derive the efficiency condition of the develop estimator and (iv) evaluate the performance of the estimator in comparison with existing estimators using real life dataset from agricultural studies.

.

.

II. Methodology

The Consider a finite population $P = (P_1, P_2, \ldots, P_N)$ of N units, let a sample be drawn using basic sampling scheme without replacement, let y_i and x_i represent the values of a study variable Y and auxiliary variable X, respectively. For a sample survey problem, when we are interested in estimating the population mean $\bar{Y}$ when the population and sample medians of study variable are known, we assume that the population is finite of size N and a sample of size n is to be selected using simple random sampling scheme. Each unit of the population is identifiable by means of assigning number to the population units from 1 to N, the numbers assign are of nominal scale, the following additional notations are used:

- N – population size
- n – sample size
- $f = n/N$ –sampling fraction
- Y– study variable
- X – auxiliary variable
- $\bar{Y}$, $\bar{X}$ – population means of the study and auxiliary variable respectively
- $\bar{y}, \bar{x}$– sample means of the study and auxiliary variable respectively

- $S_y^2, S_x^2 -$ sample variance of study and auxiliary variables respectively
- $M -$ population median of the auxiliary variable
- $m -$ sample median of the auxiliary variable
- $B(.) -$ bias of the estimator
- $MSE(.) -$ mean square error of the estimator
- $PRE -$ percentage relative efficiency

The existing estimators and their corresponding properties considered in this research are presented in Table 1 below

Table 1: Existing estimators considered in this study

S/No.	Estimators	MSE
1	$\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$ Mean per unit	$\frac{(1-f)}{n}\bar{Y}^2\{C_Y^2\}$
2	$\bar{y}_{L-R} = \bar{y} + \beta(\bar{X} - \bar{x})$ Hansen, Hurwitz & Madow (1953): linear regression	$\frac{(1-f)}{n}\bar{Y}^2 C_Y^2\{1-\rho^2\}$
3	$\bar{y}_{S-med-based} = \bar{y}\left(\frac{M}{m}\right)$ Subramani (2016)	$V(\bar{y}) + R'^2 V(m) - 2R' cov(\bar{y}, m)$
4	$\bar{y}_{exp-ratio} = \bar{y}\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right)$ Bahl and Tuteja (1991)	$V(\bar{y}) + \frac{R^2 V(\bar{x})}{4} - R cov(\bar{y}, \bar{x})$
5	$\bar{y}_{\exp-ratio-median-based} = \bar{y}\exp\left(\frac{M-m}{M+m}\right)$	$V(\bar{y}) + \frac{R'^2 V(m)}{4} - R' cov(\bar{y}, m)$

where $R' = \frac{\bar{Y}}{M}$, $R = \frac{\bar{Y}}{\bar{X}}$, $\beta_m = \frac{cov(\bar{y}, m)}{Var(m)}$,

$\beta = \frac{cov(\bar{y}, \bar{x})}{Var(\bar{x})}$, $C'_{yy} = \frac{V(\bar{y})}{\bar{Y}^2}$, $C'_{xx} = \frac{V(\bar{x})}{\bar{X}^2}$,

$C'_{mm} = \frac{V(m)}{M^2}$, $C'_{xy} = \frac{Cov(\bar{y}, \bar{x})}{\bar{X}\bar{Y}}$, $C'_{xm} = \frac{Cov(\bar{x}, m_d)}{\bar{X}M}$,

$C'_{ym} = \frac{Cov(\bar{y}, m_d)}{\bar{Y}M}$

Motivated by the works of Subramani (2016), Bahl and Tuteja (1991) and the fact that the relationship that exists between mean and median of the same variable is extremely strong compared to the correlation between study variable and auxiliary variable. We develop a new median-based ratio estimator of the population mean in simple random sampling without replacement scheme.

$$T_{Exp-Median-Nsukka} = \bar{y}\left(\frac{M}{m}\right)\exp\left(\frac{M-m}{M+m}\right) \quad (1)$$

Properties of the developed estimator (bias and Mean Square Error)

To obtain the bias and mean square error (MSE) of the developed estimator $T_{Exp-Median-Nsukka}$ in (1), we write

Let $K_{\bar{y}} = \frac{(\bar{y} - \bar{Y})}{\bar{Y}}$ and $K_m = \frac{(m - M)}{M}$

Such that $E(K_{\bar{y}}) = 0$ and $E(K_m) = \frac{bias(m)}{M}$

Where,

$$E(K_{\bar{y}}^2) = \frac{Var(\bar{y})}{\bar{Y}^2} = \left(\frac{1-f}{n}\right)C_Y^2,$$

$$E(K_m^2) = \frac{Var(m)}{M^2},$$

$$E(K_{\bar{y}}K_m) = \frac{Cov(\bar{y}, m)}{\bar{Y}M},$$

And expanding (1) in terms of K's, we have

$$T_{Exp-Median-Nsukka} = \bar{Y}(1+K_{\bar{y}})(1+K_m)\exp\left\{\frac{K_m}{2}\left(1+\frac{K_m}{2}\right)^{-1}\right\} \quad (2)$$

We assume that $|K_m| < 1$, so that the expression $(1+K_m)^{-1}$ can be expanded to a convergent infinite series using binomial theorem.

$$(1+K_m)^{-\theta} = 1 - \theta K_m + \theta(\theta+1)\frac{K_m^2}{2}$$

Note: Taylor expansion is given as

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

We also assume that the contribution of terms involving powers in K_m and $K_{\bar{y}}$ higher than the second is negligible, being of order $1/n^v$, where $v > 1$. Thus, from the above expression we write to a first order of approximation. . Hence from (2) we have.

$$T_{Exp-Median-Nsukka} = \bar{Y}(1+K_{\bar{y}})(1+K_m)^{-1} \exp\left\{\frac{K_m}{2}\left(1-\frac{K_m}{2}+\frac{K_m^2}{4}\right)\right\} \quad (3)$$

$$T_{Exp-Median-Nsukka} = \bar{Y}(1+K_{\bar{y}})(1-K_m+K_m^2)\left(1-\frac{K_m}{2}+\frac{3K_m^2}{8}\right) \quad (4)$$

$$T_{Exp-Median-Nsukka} - \bar{Y} = \bar{Y}\left(\frac{13K_m^2}{8}-\frac{3K_m}{2}-\frac{3K_{\bar{y}}K_m}{2}\right) \quad (5)$$

Taking the expectation of both side of (5), we obtained the bias of $\left(T_{Exp-Median-Nsukka}\right)$ to the first degree of approximation as,

$$bias\left(T_{Exp-Median-Nsukka}\right) = \bar{Y}\left(\frac{\lambda 13C_m^2}{8} - \frac{3bias(m)}{2M} - \frac{3\lambda C_{\bar{y}m}}{2}\right) \quad (6)$$

Where, $\lambda = \frac{1-f}{n}$

Squaring both side of the equation (5), and neglecting the terms of $K's$ having power greater than two we have,

$$\left(T_{Exp-Median-Nsukka} - \bar{Y}\right)^2 = \bar{Y}^2\left\{K_{\bar{y}}^2 + \frac{9K_m^2}{4} - 3K_{\bar{y}}K_m\right\} \quad (7)$$

Taking the expectation of both sides of (7), we get the MSE of $\left(T_{Exp-Median-Nsukka}\right)$ as,

$$MSE\left(T_{Exp-Median-Nsukka}\right) = \bar{Y}^2\lambda\left(C_{\bar{y}}^2 + \frac{9C_m^2}{4} - 3C_{\bar{y}m}\right) \quad (8)$$

III APPLICATIONS

In this section, Algebraic efficiency comparison of the developed estimator with some existing estimators is presented.

The mean per unit unbiased estimator

The developed estimator is more efficient than mean per unit estimator in SRSWOR if $MSE\left(T_{Exp-Median-Nsukka}\right) \le MSE\left(\bar{y}\right)$,i.e

$$\bar{Y}^2\lambda\left(C_{\bar{y}}^2 + \frac{9C_m^2}{4} - 3C_{\bar{y}m}\right) \le \frac{(1-f)}{n}\bar{Y}^2\left\{C_Y^2\right\}$$

$$\Rightarrow 9C_m^2 > 12C_{\bar{y}m}$$

Hansen, Hurwitz and Madow (1953) linear regression estimator

The developed estimator is more efficient than linear regression estimator if $MSE\left(T_{Exp-Median-Nsukka}\right) \le MSE\left(\bar{y}_{L-R}\right)$,i.e

$$\bar{Y}^2\lambda\left(C_{\bar{y}}^2 + \frac{9C_m^2}{4} - 3C_{\bar{y}m}\right) \le \frac{(1-f)}{n}\bar{Y}^2 C_Y^2\left\{1-\rho^2\right\}$$

$$\Rightarrow \frac{9C_m^2}{4} - 3C_{\bar{y}m} > -\rho C_{\bar{y}}^2$$

Subramani (2016) estimator

The developed estimator is more efficient than median based ratio estimator, if $MSE\left(T_{Exp-Median-Nsukka}\right) \le MSE\left(\bar{y}_{S-median-based}\right)$,i.e

$$\bar{Y}^2\lambda\left(C_{\bar{y}}^2 + \frac{9C_m^2}{4} - 3C_{\bar{y}m}\right) \le V(\bar{y}) + R'^2 V(m) - 2R'\mathrm{cov}(\bar{y}, m)$$

$$\Rightarrow 5\bar{Y}\frac{Var(m)}{M} < 4Cov(\bar{y}, m)$$

3.1 Data Description

Dataset 1: The populations considered in this study are real-life dataset taken from Singh and Chaudhary (1986) and Mukhopadhyay(2005). The dataset were also used by Subramani (2016), Srija and Subramani (2018) and Abdullahi and Ugwuowo (2020).

The populations 1 pertain to estimate the area of cultivation under wheat in the year 1974, whereas the auxiliary information is the cultivated areas under wheat in 1973.

The population 2 pertains to the quantity of raw materials (study variable, in lakhs of bales) for 20 jute mills and the number of labourers (the auxiliary variable, in thousand).

Table 2 below summarized the dataset from two (2) populations for the study and the corresponding functions to be used for numerical analysis.

Table 2: Summary of the Dataset

Population 1				Population 2			
Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
N	34	R	4.2941	N	20	R	0.0939
n	5	$R^{\prime}$	1.1158	n	5	$R^{\prime}$	1.0247
$\bar{Y}$	856.42	$V(\bar{y})$	91690.37	$\bar{Y}$	41.5	$V(\bar{y})$	14.3605
$\bar{X}$	208.88	$V(\bar{x})$	3849.25	$\bar{X}$	441.95	$V(\bar{x})$	1532.2812
$\bar{M}$	736.98	$V(m)$	59396.28	$\bar{M}$	40.0552	$V(m)$	10.8348
M	767.50	ρ	0.4453	M	40.5	ρ	0.6522
$Cov(\bar{y},m)$	48074.95	$Cov(\bar{y},\bar{x})$	8366.06	$Cov(\bar{y},m)$	9.0665	$Cov(\bar{y},\bar{x})$	96.7461

Percentage Relative Efficiency (PRE)

The Percentage relative efficiency (PRE) of different estimators $T_{.}$ in respect to $T_{Exp-Median-Nsukka}$ is defined as,

$$PRE\left(T_{Exp-Median-Nsukka}, T.\right) = \frac{MSE(T.)}{MSE\left(T_{Exp-Median-Nsukka}\right)} \times 100$$

Tables 3 and 4 below gives the MSE and PRE of the existing and developed estimators with respect to mean per unit ($\bar{y}$), linear regression ($\bar{y}_{L-R}$) and median-based $\bar{y}_{S-median-based}$ estimators respectively for the two populations.

Table 3

MSE/ Variance of some selected existing estimators and that of proposed estimator		PRE with respect to mean per unit estimator ($\bar{y}$)	PRE with respect to linear regression estimator ($\bar{y}_{L-R}$)	PRE with respect to Subramani Median-based estimator ($\bar{y}_{S-median-based}$)
Estimator	MSE	PRE	PRE	PRE
$\bar{y}$	91690.37	100	<100	<100
$\bar{y}_{L-R}$	73195.5841	125.27	100	<100
$\bar{y}_{S-median-based}$	58356.923	157.12	125.43	100
$T_{Exp-Median-Proposed}$	16573.955	553.22	441.63	352.10

Table 4

MSE/ Variance of some selected existing estimators and that of proposed estimator		PRE with respect to mean per unit estimator ($\bar{y}$)	PRE with respect to linear regression estimator ($\bar{y}_{L-R}$)	PRE with respect to Subramani Median-based estimator ($\bar{y}_{S-median-based}$)
Estimator	MSE	PRE	PRE	PRE
$\bar{y}$	14.3605	100	<100	<100
$\bar{y}_{L-R}$	8.2521	174.02	100	<100
$\bar{y}_{S-median-based}$	7.1563	200.67	115.31	100
$T_{Exp-Median-Nsukka}$	1.8134	791.91	455.06	394.63

Table 3 Results from Tables 3 and 4 reveals that the developed estimator $T_{Exp-Median-Nsukka}$ has the least mean square error compared to some existing estimators and it also show significant efficiency gain in respect of percentage relative efficiency.

IV. Conclusion

In this study, we have developed efficient median-based ratio estimator named ratio exponential ratio estimator for population mean with known population median of the study variable as auxiliary parameter. The properties of the proposed estimator (i.e., bias and MSE) were derived. Algebraically we demonstrated the efficiency condition of the proposed estimator compared to some existing estimators in the literature.

Results from the real life dataset show significance efficiency gain for the developed estimator that incorporate median of study variable under exponential, while for the other estimators the result shows efficiency loss. With the significant performance of this estimator which is function of both medians and mean per unit estimator of the study variable, the estimator is recommended for practical application when the efficiency condition is satisfied.

References

Abdullahi U. K. and Ugwuowo F. I. (2020). On efficient median based linear regression estimator for population mean. *Communication in Statistics: Theory and Method*, 1-13. http://doi:10.1080/03610926.2020.1831540.

Bahl S. and Tuteja R. K. (1991). Ratio and product type exponential estimator. *Journal of Information and Optimization Sciences, **12***(1): 159-163.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953). *Sample survey methods and theory*. New York: John wiley.

Mukhopadhyay P. (2005). *Theory and methods of survey sampling*. PHI Learning, 2nd edition, New Delhi.

Singh, D. and Chaudhary, F.S. (1986). *Theory and Analysis of Sample Survey Designs*. New Age International Publication.

Srija, R. and Subramani, J. (2018). Ratio cum median based modified ratio estimators with known first and third quartiles. *International Journal of Communication and Media Sciences*, **5**(1):1-5.

Subramani J. (2016). A New Median Based Ratio Estimator for Estimation of the Finite Population Mean. *Statistics in Trans*ition, **17** (4):591-604