International Journal of Statistics and Systems
ISSN 0973-2675 Volume 16, Number 1 (2021), pp. 17-36

http://www.ripublication.com

An Improved Estimation Procedure of Finite Population Mean using Information on Median of the Study Variable

[1]Dipika*, [2]Mukti Kant Sukla, [3]Sheela Misra

[1,2] Department of Statistics, Sri Venkateswara College, University of Delhi, India.

[3]Department of Statistics, University of Lucknow, Lucknow-110021, India.

(Corresponding Author)

Abstract

In sampling theory it is a well-established fact that the use of additional information in terms of auxiliary variable increases the precision of estimation procedures. But on the other side it also increases the cost of survey for collecting the information on auxiliary variable. Thus use of median of study variable may be a significant approach in this direction. The present article concerns with the problem of estimating finite population mean of the study variable by using the information available in terms of median of this variable. A ratio type estimator has been developed for this purpose. Bias and mean squared error of the proposed estimator have been derived up to the first order of approximation. The optimum value of the characterizing scalar which minimizes the mean squared error of the proposed estimator has been obtained. The minimum mean squared error for this optimum value of the characterizing scalar is also obtained. The proposed estimator has been compared with the existing estimators of population mean which make use of auxiliary information. To amply corroborate the theoretical findings, an empirical study has also been carried out to judge the performances of the proposed and the competing estimators.

Key words: Estimation, Auxiliary Variable, Bias, Mean Squared Error

1.1 Introduction

Estimation of population parameters is the ultimate purpose of sampling techniques.

Sampling theory deals with obtaining reliable estimates of various parameters under study. Use of auxiliary information is a powerful tool for improving the efficiency of estimators. Sometimes relationship between the study variable and the auxiliary variable decides the type of estimators used for estimation purpose. If there is a positive correlation between the study variable and auxiliary variable then ratio estimator is preferred while in case of negative correlation product estimator is recommended. Generally it is seen that estimators using auxiliary information provides more efficient results than the usual one. But on the other hand cost of survey increases in the collection of information on auxiliary variable. So cost factor discourages the survey statisticians to use the auxiliary information for improving the efficiency of estimators. Now we are looking toward the substitute of auxiliary variable which improves the efficiency of estimator without increasing the survey cost.

In many practical situations it is seen that population mean of the study variable is not known but the population median of the study variable may be known. For example if we ask for the weight or basic salary of a person, it is very hard to get the exact value but we get the information in terms of interval or the pay band. Here we can easily get the median of the study variable which can be utilized for improved estimation of population mean of study variable. Therefore use of median of study variable can be a better option to improve the efficiency without increasing the cost of survey. In the present paper we have proposed an improved estimator of population mean using median of the study variable.

Let us consider a population of N distinct and identifiable units and let $(x_i, y_i), i = 1, 2, ..., n$ be a bivariate sample of size n taken from (*X,* Y) using a simple random sampling without replacement (*SRSWOR)* scheme. Let $\bar{X}$ and $\bar{Y}$ respectively be the population means of the auxiliary and the study variables, and let $\bar{x}$ and $\bar{y}$ be the corresponding sample means. In SRSWOR, It is well established fact that sample means $\bar{x}$ and $\bar{y}$ are unbiased estimators of population means of $\bar{X}$ and $\bar{Y}$ respectively.

To demonstrate the problem in a more efficient way, let us consider an interesting example of mean estimation of study variable using median of study variable given by Subramani (2016)[11].

Example: The estimation of body mass index (BMI) of the 350 patients in a Hospital based on a small simple random sample without replacement has been considered.

Category	**BMI range – kg/m2**	**Number of patients**	**Cumulative total**
Very severely underweight	less than 15	15	15
Severely underweight	from 15.0 to 16.0	35	50
Underweight	from 16.0 to 18.5	67	117
Normal (healthy weight)	from 18.5 to 25	92	209

Category	**BMI range – kg/m2**	**Number of patients**	**Cumulative total**
Overweight	from 25 to 30	47	256
Obese Class I (Moderately obese)	from 30 to 35	52	308
Obese Class II (Severely obese)	from 35 to 40	27	335
Obese Class III (Very severely obese)	over 40	15	350
Total		350	350

Table A: Body mass index of 350 patients in a hospital

The median value will be between 18.5 and 25. So it can be assume that the population median of the BMI is approximately 21.75

1.2 PROPOSED ESTIMATOR

We propose the following ratio type estimator of population mean using known population median of study variable as,

$$\bar{y}_{km} = \bar{y}\left(\frac{m}{km+(1-k)M}\right) \qquad (1.2.1)$$

Where m and M are the sample and population median of the study variable respectively and k is the characterizing scalar to be chosen such that the mean squared error of the proposed estimator is minimum.

The following approximations have been made to study the properties of the proposed estimators as,

$\bar{y} = \bar{Y}(1+e_0)$ and $m = M(1+e_0)$ such that $E(e_0) = 0, E(e_1) = \frac{\bar{M}-M}{M} = \frac{Bias(m)}{M}$ and $E(e_0^2) = \lambda C_y^2, E(e_1^2) = \lambda C_m^2, E(e_0 e_1) = \lambda C_{ym}, where\ \lambda = \frac{1-f}{n}$ and $\bar{M} = \frac{1}{n}\sum_{i=1}^{n} m_i$

Estimator (1.1.1) can be expressed in terms of $e_i's$ (i=1, 2) as,

$$\bar{y}_{km} = \bar{Y}(1+e_0)\left[\frac{M(1+e_1)}{kM(1+e_1)+(1-k)M}\right]$$

$$= \bar{Y}(1+e_0)(1+e_1)(1+ke_1)^{-1}$$

$$\bar{y}_{km} - \bar{Y} = \bar{Y}\,[\,e_0 + e_1(1-k) - k(1-k)e_1^2 + (1-k)e_0e_1 + \cdots] \qquad (1.2.2)$$

Now taking expectation of (1.2.2) and using first order approximation, we have

$$\text{Bias}(\bar{y}_{km}) = E(\bar{y}_{km} - \bar{Y})$$

$$= \bar{Y}\left[(1-k)\frac{Bias(m)}{M} - k(1-k)\lambda C_m^2 + (1-k)\lambda C_{ym} + \cdots\right]$$

$$\text{Bias}(\bar{y}_{km}) = (1-k)\bar{Y}\left[\frac{Bias(m)}{M} - \lambda(kC_m^2 - C_{ym})\right] \quad (1.2.3)$$

Squaring (1.2.2) and taking expectation up to first order of approximation on both sides we get the MSE($\bar{y}_{km}$),

$$\text{MSE}(\bar{y}_{km}) = E(\bar{y}_{km} - \bar{Y})^2$$

$$= \bar{Y}^2[E(e_0^2) + (1-k)^2 E(e_1^2) + 2(1-k)E(e_0e_1)]$$

$$\text{MSE}(\bar{y}_{km}) = \lambda\bar{Y}^2\left[C_y^2 + (1-k)^2C_m^2 + 2(1-k)C_{ym}\right] \quad (1.2.4)$$

Optimum values of k for which mean squared error of proposed estimator is minimum,

$$k_0 = 1 + \frac{C_{ym}}{C_m^2}$$

The minimum mean squared error of the proposed estimator$\bar{y}_{km}$is,

$$MSE(\bar{y}_{km})_{min} = \lambda\bar{Y}^2\left[C_y^2 - \frac{C_{ym}^2}{C_m^2}\right]$$

$$= \lambda\bar{Y}^2C_y^2 - \lambda\bar{Y}^2\frac{C_{ym}^2}{C_m^2} \quad (1.2.5)$$

$$MSE(\bar{y}_{km})_{min} = V(\bar{y}) - \lambda\bar{Y}^2\frac{C_{ym}^2}{C_m^2} \quad (1.2.6)$$

1.3 Bias and MSE of existing and proposed estimators:

Estimator	Expression	Bias	Mean Squared Error
Sample Mean(t_0)	$\frac{1}{n}\sum_{i=1}^{n} y_i$	Unbiased	$\frac{1-f}{n}\bar{Y}^2C_y^2$
Watson Regression Estimator(t_1)(1937)[13]	$\bar{y} + b_{yx}(\bar{X} - \bar{x})$	Unbiased	$\frac{1-f}{n}\bar{Y}^2C_y^2(1-\rho_{yx}^2$
Cochran Ratio Estimator(t_2)(1940)[3]	$\bar{y}\frac{\bar{X}}{\bar{x}}$	$\frac{1-f}{n}\bar{Y}[C_x^2 - C_{yx}]$	$\frac{1-f}{n}\bar{Y}^2[C_y^2 + C_x^2 - 2C_{yx}]$
Bahland TutejaEstimator (t_3)(1991)[2]	$\bar{y}\, exp\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]$	$\frac{1-f}{8n}\bar{Y}[3C_x^2 - 4C_{yx}]$	$\frac{1-f}{n}\bar{Y}^2\left[C_y^2 + \frac{C_x^2}{4} - C_{yx}\right]$

Estimator	Expression	Bias	Mean Squared Error
Kadilar Estimator(t_4) (2016)[5]	$\bar{y}\left(\frac{\bar{x}}{\bar{X}}\right)^{\delta} exp\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]$	$\frac{1-f}{n}\bar{Y}\left[\left(\frac{\delta(\delta-1)}{2}+\frac{3}{8}\right)C_x^2 + \left(\delta+\frac{1}{2}\right)C_{yx}\right]$	$\frac{1-f}{n}\bar{Y}^2 C_y^2(1-\rho_{yx}^2$
SubramaniEstimator (2016)(t_5)[11]	$\bar{y}\left(\frac{M}{m}\right)$	$\frac{1-f}{n}\bar{Y}\left[C_m^2 - C_{ym} - \frac{Bias(m)}{M}\right]$	$\frac{1-f}{n}\bar{Y}^2\left[C_y^2+\left(\frac{\bar{Y}}{M}\right)^2 C_y^2 - 2\left(\frac{\bar{Y}}{M}\right)C_{ym}\right]$
Proposed Estimator ($\bar{y}_{km}$)	$\bar{y}\left(\frac{m}{km+(1-k)M}\right)$	$(1-\mathrm{k})\bar{\mathrm{Y}}\left[\frac{\mathrm{Bias(m)}}{\mathrm{M}} - \lambda(\mathrm{kC_m^2} - \mathrm{C_{ym}})\right]$	$\lambda\bar{Y}^2 C_y^2 - \lambda\bar{Y}^2\frac{C_{ym}^2}{C_m^2}$

Table 1.1: Bias and MSE of proposed and existing estimators

1.4 EFFICIENCY COMPARISON

Under this section, a theoretical comparison of the proposed estimator has been made with the competing estimators of population mean. The conditions under which the proposed estimator performs better than the competing estimators have also been given.

Efficiency Comparison of Proposed estimator with usual mean per unit estimator,

$$V(t_0) - MSE(\bar{y}_{km})_{min} > 0 \; if \; \frac{C_{ym}^2}{C_m^2} > 0 \; or \; C_{ym}^2 > 0 \tag{1.4.1}$$

Comparison of Proposed estimator with usual regression estimator proposed by Watson (1937),

$$MSE(t_1) - MSE(\bar{y}_{\mathrm{km}})_{min} > 0 \; if \; \frac{C_{ym}^2}{C_m^2} - C_y^2\rho_{yx}^2 > 0 \tag{1.4.2}$$

Comparison of Proposed estimator with usual ratio estimator given by Cochran (1940),

$$MSE(t_2) - MSE(\bar{y}_{\mathrm{km}})_{min} > 0 \; if \; C_x^2 + \frac{C_{ym}^2}{C_m^2} > 2C_{yx} \tag{1.4.3}$$

Comparison of Proposed estimator with Bahl and Tuteja (1991) ratio type estimator

$$MSE(t_3) - MSE(\bar{y}_{\mathrm{km}})_{min} > 0 \; if \frac{C_x^2}{4} + \frac{C_{ym}^2}{C_m^2} > C_{yx} \tag{1.4.4}$$

Comparison of Proposed estimator with Kadilar's (2016) estimator,

$$MSE(t_4) - MSE(\bar{y}_{km})_{min} > 0, \frac{C_{ym}^2}{C_m^2} - C_y^2 \rho_{yx}^2 > 0 \quad (1.4.5)$$

Comparison of Proposed estimator with Subramani's (2016) estimator

$$MSE(t_5) - MSE(\bar{y}_{km})_{min} > 0, if\ R_5^2 C_m^2 + \frac{C_{ym}^2}{C_m^2} > 2R_5 C_{ym} \quad (1.4.6)$$

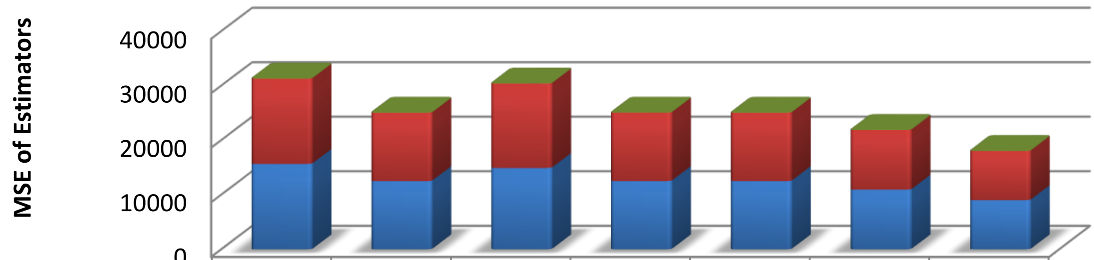
The proposed estimator will perform better than considered existing estimators if the conditions derived under section 3 from (1.4.1)-(1.4.6) will holds for considered population or known sample.

1.5 NUMERICAL STUDY

To judge the theoretical findings, we have considered the natural populations given in Subramani (2016). He has used three natural populations. The population 1 and 2 have been taken from Singh and Chaudhary (1986, page no. 177) and the population 3 has been taken from Mukhopadhyay (2005, page no. 96). In populations 1 and 2, the study variable is the estimate the area of cultivation under wheat in the year 1974, whereas the auxiliary variables are the cultivated areas under wheat in 1971 and 1973 respectively. In population 3, the study variable is the quantity of raw materials in lakhs of bales and the number of labourers as the auxiliary variable, in thousand for 20 jute mills. Tables 3 and 4 represent the parameter values along with constants, along with proposed estimator, variances and mean squared errors of existing and proposed estimator

Parameter	Population-1	Population-2	Population-3	Parameter	Population-1	Population-2	Population-3
N	34	34	20	R_7	1.1158	1.1158	1.0247
n	5	5	5	C_y^2	0.125014	0.125014	0.008338
${}^N C_n$	278256	278256	15504	C_x^2	0.088563	0.096771	0.007845
$\bar{Y}$	856.4118	856.4118	41.5	C_m^2	0.100833	0.100833	0.006606
$\bar{M}$	736.9811	736.9811	40.0552	C_{ym}	0.07314	0.07314	0.005394
M	767.5	767.5	40.5	C_{yx}	0.047257	0.048981	0.005275
$\bar{X}$	208.8824	199.4412	441.95	ρ_{yx}	0.4491	0.4453	0.6522

Graphical and tabular presentation of MSE's of Estimators for population 1, population 2 and population 3

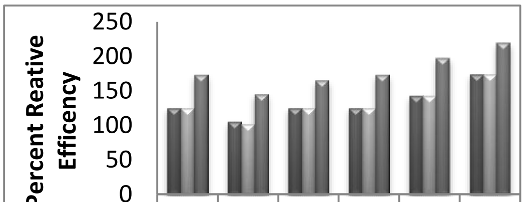


	t_0	t_1	t_2	t_3	t_4	t_5	ȳ_km
Popln-3	2.15	1.24	1.48	1.3	1.24	1.09	0.98
Popln-2	15640.97	12539.3	15492.08	12539.3	12539.3	10926.53	9002.22
Popln-1	15640.97	12486.75	14895.27	12498.01	12486.75	10926.53	9002.22

Graph 1.1: Bar graph and MSE's of Proposed and existing estimator

1.6 Percent Relative Efficiency

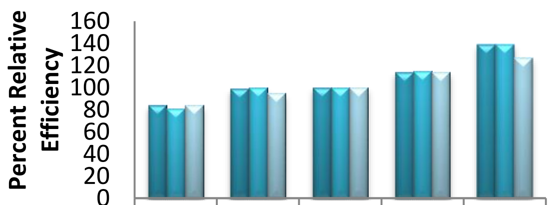
(a) Graphical and tabular presentation of Percent relative efficiency of t_1, t_2,t_3,t_4, t_5 and Proposed estimator



	t_1	t_2	t_3	t_4	t_5	ȳ_km
Popln-1	125	105	125	125	143	174
Popln-2	125	101	125	125	143	174
Popln-3	173	145	165	173	197	219

Graph 1.2: Bar graph of Percent relative efficiency

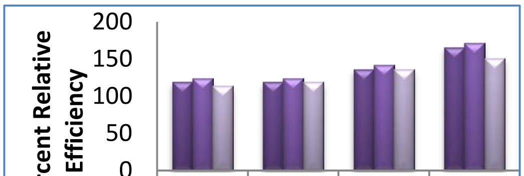
(b) Graphical and tabular presentation of percent relative efficiency of t_2,t_3,t_4,t_5 and Proposed estimator over t_1



	t_2	t_3	t_4	t_5	ȳ_km
Popln-1	84	99	100	114	139
Popln-2	81	100	100	115	139
Popln-3	84	95	100	114	127

Graph 1.3: Bar graph of Percent relative efficiency

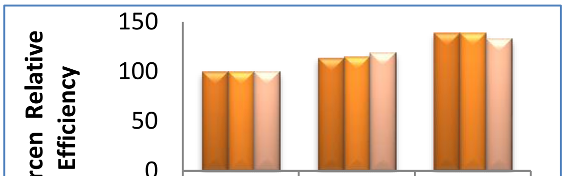
(c) Graphical and tabular presentation of Percent relative efficiency of t_3,t_4,t_5 and Proposed estimator over t_2,



	t_3	t_4	t_5	ȳ_km
Popln-1	119	119	136	166
Popln-2	124	124	142	172
Popln-3	114	119	136	151

Graph 1.4: Bar graph of Percent relative efficiency

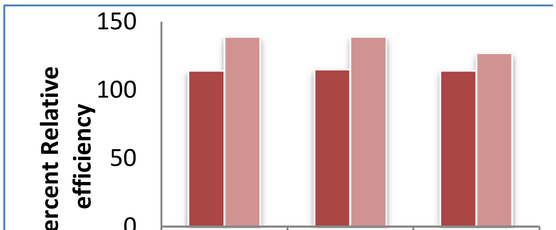
(d) Graphical and tabular presentation of percent relative efficiency of t_4, t_5 and Proposed estimator over t_3



	t_4	t_5	ȳ_km
Popln-1	100	114	139
Popln-2	100	115	139
Popln-3	100	119	133

Graph 1.5: Bar graph of Percent relative efficiency

(e) Graphical and tabular presentation of Percent relative efficiency of t_5 and Proposed estimator over t_4



	Popln-1	Popln-2	Popln-3
t_5	114	115	114
ȳ_km	139	139	127

Graph 1.6: Bar graph of Percent relative

(f) Graphical and tabular presentation of Percent relative efficiency of proposed estimator over Subramani Estimator (t_5)

	Popln-1	Popln-2	Popln-3
ȳ_km	121	121	111

Graph 1.7: Bar graph of Percent relative efficiency

1.7 Concluding Remarks

1 The minimum value of mean squared error of proposed estimator for the optimum value of k $\left(k_0 = 1 + \frac{C_{ym}}{C_m^2}\right)$ is $MSE(\bar{y}_{km})_{min} = \lambda \bar{Y}^2 C_y^2 - \lambda \bar{Y}^2 \frac{C_{ym}^2}{C_m^2}$

2 From (1.2.6) it is clearly observed that proposed estimator $\bar{y}_{km}$ will always have lesser mean squared error than per unit sample mean.

3 Efficiency conditions under which proposed estimator perform better than existing estimator are shown under section 1.4 from (1.3.1)-(1.3.6).

4 From graph 1.1 it is clear that proposed estimator have lesser mean squared

error among all competing estimators.

5 From graph 1.2 it is shown that proposed estimator is 174% efficient for population 1 and 2 and 219 % efficient for population 3 over mean per unit estimator.

6 Graph1.2 shows that proposed estimator has highest percent relative efficiency than Watson estimator, Cochran estimator, Bahl and Tuteja estimator, Kadilar estimator and Subramani estimator over mean per unit estimator.

7 From graph 1.5 it is clear that proposed estimator is 139% more efficient for population 1 and 2 and 127 % efficient for population 3 over Watson estimator.

8 Graph1.4 shows that PRE of proposed estimator is 166% over Watson estimator, which is higher among all.

9 Graph 1.6 shows that PRE of proposed estimator over Cochran estimator(t_2) for population one and two is 172% and for population three it is 151%, which is higher than Bahl and Tuteja estimator, Kadilar estimator and Subramani estimator over Cochran estimator.

10 Graph 1.4 shows the graphical representation of percent relative efficiency of Bahl and Tuteja estimator, Kadilar estimator, Subramani estimator and proposed estimator over Cochran estimator.

11 Graph 1.5 shows that proposed estimator is more efficient than Bahl and Tutej aestimator.

12 From graph 1.3 it is clear that for population 1 and 2, proposed estimator is 139% more efficient than Kadilar's estimator and for population 3 it is 127% more efficient.

13 Graph 1.6 shows the representation of percent relative efficiency of Subramani and proposed estimator over the Kadilar estimator.

14 From graph 1.7 we conclude that proposed estimator is more efficient than Subramani estimator, which is a ratio estimator based on the utilization of median of study variable. So for survey practitioners it is suggested to use proposed estimator if they have the information on median of study variable to obtain the more reliable estimates without increasing the cost of survey.

Reference

1. ABID, M., ABBAS, N. SHERWANI, R.A.K. AND NAZIR, H. Z. (2016). Improved Ratio Estimators for the population mean using non-conventional measure of dispersion. *Pakistan Journal of Statistics and Operations Research*, XII (2), 353-367.
2. BAHL, S. AND TUTEJA, R.K. (1991). Ratio and product type exponential estimator, *Information and Optimization Sciences*, XII (I), 159-163.

3. COCHRAN, W. G. (1940). The Estimation of the Yields of the Cereal Experiments by Sampling for the Ratio of Grain to Total Produce. *The Journal of Agricultural Science, 30*, 262-275.
4. JERAJUDDIN, M. AND KISHUN, J. (2016). Modified Ratio Estimators for Population Mean Using Size of the Sample, Selected From Population, *IJSRSET*, 2, 2, 10-16.
5. KADILAR, G.O. (2016). A New Exponential Type Estimator for the Population Mean in Simple Random Sampling, Journal of Modern Applied Statistical Methods, 15, 2, 207-214.
6. REDDY, V.N. (1974): On a transformed ratio method of estimation. Sankhya, C, 36(1), 59-70.
7. SRIVASTAVA, S.K. (1967): An estimator using auxiliary information in sample surveys. Cal. Statist. Assoc. Bull., 16, 62-63.
8. SUBRAMANI, J. (2013). Generalized modified ratio estimator of finite population mean, Journal of Modern Applied Statistical Methods, 12 (2), 121–155.
9. SUBRAMANI, J., KUMARAPANDIYAN, G. (2012). Estimation of population mean using coefficient of variation and median of an auxiliary variable, International Journal of Probability and Statistics, 1 (4), 111–118.
10. SUBRAMANI, J., KUMARAPANDIYAN, G. (2013). A new modified ratio estimator of population mean when median of the auxiliary variable is known, Pakistan Journal of Statistics and Operation Research, Vol. 9 (2), 137–145.
11. SUBRAMANI, J. (2016).A new median based ratio estimator for estimation of the finite population mean, *Statistics in Transition New Series,* 17, 4, 1-14.
12. TAILOR, R., SHARMA, B. (2009). A modified ratio-cum-product estimator of finite population mean using known coefficient of variation and coefficient of kurtosis, Statistics in Transition-New Series, 10 (1), 15–24.
13. WATSON, D.J. (1937). The estimation of leaf area in field crops, The Journal of Agricultural Science, 27, 3, 474-483.
14. YADAV, S.K; MISHRA, S.S. AND SHUKLA, A.K. (2014). 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.
15. YADAV, S.K; MISHRA, S.S. AND SHUKLA, A.K. (2015). Estimation Approach to Ratio of Two Inventory Population Means in Stratified Random Sampling, *American Journal of Operational Research*, 5, 4, 96-101.
16. YADAV, S. K; GUPTA, SAT; MISHRA, S. S. AND SHUKLA, A. K. (2016). Modified Ratio and Product Estimators for Estimating Population Mean in Two-Phase Sampling, *American Journal of Operational Research,* 6, 3, 61-68.
17. YADAV, S.K., KADILAR C.(2013): Improved Exponential Type ratio estimator of population variance. Rev. Colom. Estad., 36, 145-152.