Awakening to Reality

Available online at www.elixirpublishers.com (Elixir International Journal)

Statistics

Elixir Statistics 109C (2017) 47882-47888

Elixir
ISSN: 2229-712X

New Class of Estimators of Population Mean Using Known Median of the Study Variable

S. K. Yadav[1] and Surendra Kumar[2]

[1]Department of Mathematics and Statistics (A Centre of Excellence), Dr.RML Avadh University,Faizabad -24001,U.P.,INDIA.
[2]Department of Mathematics, Govt. Degree College, Pihani- 241406 U.P., INDIA.

ARTICLE INFO

Article history:
Received: 26 April 2017;
Received in revised form: 26 July 2017;
Accepted: 5 August 2017;

Keywords
Study variable,
Bias,
Ratio estimator,
Mean squared error,
Simple random sampling,
Efficiency.

ABSTRACT

In the present manuscript, a new class of estimators of population mean has been proposed. This estimator makes use of the information on the population median of the study variable. The expressions for the bias and the mean squared error of the proposed class of estimators 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 class 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 competing estimators of population mean which make use of auxiliary information. A numerical study is also carried out to judge the performances of the proposed and the competing estimators.

1. INTRODUCTION

To estimate any parameter, the most suitable estimator is the corresponding statistic. The Auxiliary information is used for improved estimation of population parameters but it is collected on additional cost of the survey. Sometimes it is possible to have the information on the population median of the study variable while the population mean of this variable is not known and it is to be estimated. For example if we ask for the exact body weight index or exact blood pressure of a person, it is very hard to get the exact values of above measures but we get the values of these measures in some prescribed intervals. Thus information on the median of the study variable can easily be obtained and can be utilized for improved estimation of population mean of study variable. The use of auxiliary variable which is highly correlated with study variable also improves the efficiency of the estimator but it is collected on additional cost of the survey. In the present paper, we have suggested a generalized ratio type estimator of population mean of the study variable using median of the study variable.

Let us consider the finite population consisting of N distinct and identifiable units and let be a bivariate sample of size n taken from (X, Y) using a simple random sampling without replacement (SRSWOR) scheme. Let and respectively be the population means of the auxiliary and the study variables, and let and be the corresponding sample means. In simple random sampling without replacement, it is well known that sample means and are unbiased estimators of population means of and respectively.

Let us consider two interesting examples of estimation of population mean of study variable using median of study variable given by Subramani (2016) to demonstrate the problem under consideration. The tables of examples and other tables have been used with the permission of the author.

Example1. In the estimation of body mass index (BMI) of the 350 patients of a Hospital, it is reasonable to assume that the population median of the BMI is known based on the information given in Table1.

Table1. Body mass index of 350 patients of a hospital.

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
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**

The median value will be between 18.5 and 25. Approximately one can assume the population median of the BMI value as 21.75.

Tele:
E-mail address: surendra.kumar776@gmail.com

Example2. In the problem of estimating the blood pressure of the 202 patients of a hospital, it is reasonable to assume that the median of the blood pressure is known based on the information available in Table2.

Table2. Blood pressure of 202 patients of a hospital

Category	Systolic, mmHg	Number of patients	Cumulative No. of patients
Hypotension	< 90	10	10
Desired	90–119	112	122
Pre-hypertension	120–139	40	162
Stage 1 Hypertension	140–159	20	182
Stage 2 Hypertension	160–179	13	195
Hypertensive Emergency	≥ 180	7	202
Total		**202**	**202**

The median value will be between 90 and 119. Approximately one can assume the population median value as 104.5.

2. REVIEW OF EXISTING ESTIMATORS

The most suitable estimator of population means of the study variable, is given by,

$$t_o = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{1}$$

The above estimator is unbiased for population mean and up to the first order of approximation, its variance is given by,

$$V(t_0) = \frac{1-f}{n} S_y^2 = \frac{1-f}{n}\bar{Y}^2 C_y^2 \tag{2}$$

where, $C_y = \frac{S_y}{\bar{Y}}$, $S_y^2 = \frac{1}{N-1}\sum_{i-1}^{N}(Y_i - \bar{Y})^2 = \frac{1}{^N C_n}\sum_{i-1}^{^N C_n}(\bar{y}_i - \bar{Y})^2$, $f = \frac{n}{N}$.

Watson (1937) proposed the usual regression estimator by using highly correlated auxiliary variable with the study variable and proposed the usual linear regression estimator of population mean as,

$$t_1 = \bar{y} + \beta_{yx}(\bar{X} - \bar{x}) \tag{3}$$

where β_{yx} is the regression coefficient of the line Y on X.

The regression estimator is also unbiased for population mean and its variance up to the first order of approximation, is given by,

$$V(t_1) = \frac{1-f}{n}\bar{Y}^2 C_y^2 (1-\rho_{yx}^2). \tag{4}$$

Cochran (1940) proposed the traditional ratio estimator of population mean using positively correlated auxiliary variable as,

$$t_2 = \bar{y}\frac{\bar{X}}{\bar{x}} \tag{5}$$

This traditional ratio estimator is biased estimator and its bias and mean squared error, up to the first order of approximation are respectively given by,

$$B(t_2) = \frac{1-f}{n}\bar{Y}[C_x^2 - C_{yx}] \text{ and}$$

$$MSE(t_2) = \frac{1-f}{n}\bar{Y}^2[C_y^2 + C_x^2 - 2C_{yx}], \tag{6}$$

where, $C_x = \frac{S_x}{\bar{X}}$, $S_x^2 = \frac{1}{N-1}\sum_{i-1}^{N}(X_i - \bar{X})^2 = \frac{1}{^N C_n}\sum_{i-1}^{^N C_n}(\bar{x}_i - \bar{X})^2$, $\rho_{yx} = \frac{Cov(x,y)}{S_x S_y}$,

$Cov(x,y) = \frac{1}{N-1}\sum_{i-1}^{N}(Y_i - \bar{Y})(X_i - \bar{X})$, and $C_{yx} = \rho_{yx} C_y C_x$.

Bahl and Tuteja (1991) by making use of positively correlated auxiliary variable, suggested the exponential ratio type estimator as,

$$t_3 = \bar{y}\exp\left[\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right] \tag{7}$$

The bias and the mean squared error of this estimator, up to the first order of approximation, are respectively given by,

$$B(t_3) = \frac{1-f}{8n}\bar{Y}[3C_x^2 - 4C_{yx}] \text{ and}$$

$$MSE(t_3) = \frac{1-f}{n}\bar{Y}^2[C_y^2 + \frac{C_x^2}{4} - C_{yx}]. \tag{8}$$

Srivastava (1967) used positively correlated auxiliary variable and proposed the generalized ratio type estimator of population mean as,

$$t_4 = \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right)^{\alpha} \quad (9)$$

where α is a suitably chosen constant such that MSE of above estimator is minimum.

The bias and the mean squared error up to the first order of approximation are respectively given by,

$$B(t_4) = \frac{1-f}{n}\bar{Y}[\frac{\alpha(\alpha-1)}{2}C_x^2 + \alpha C_{yx}] \text{ and}$$

$$MSE(t_4) = \frac{1-f}{n}\bar{Y}^2[C_y^2 + \alpha^2 C_x^2 + 2\alpha C_{yx}]$$

The optimum value of the constant α is, $\alpha = -C_{yx}/C_x^2$.

The minimum value of $MSE(t_4)$ for optimum value of α is given by,

$$MSE_{\min}(t_4) = \frac{1-f}{n}\bar{Y}^2 C_y^2(1-\rho_{yx}^2) . \quad (10)$$

Reddy (1974) suggested a class of ratio type estimators for population mean as,

$$t_5 = \bar{y}\left[\frac{\bar{X}}{\bar{X}+\alpha(\bar{x}-\bar{X})}\right] \quad (11)$$

The mean squared error of this estimator, up to the first order of approximation are respectively given by,

$$B(t_5) = \frac{1-f}{n}\bar{Y}[\alpha^2 C_x^2 - \alpha C_{yx}] \text{and}$$

$$MSE(t_5) = \frac{1-f}{n}\bar{Y}^2[C_y^2 + \alpha^2 C_x^2 - 2\alpha C_{yx}]$$

The MSE of the above estimator is minimum for optimum value of $\alpha = C_{yx}/C_x^2$ and the minimum MSE is given by,

$$MSE_{\min}(t_5) = \frac{1-f}{n}\bar{Y}^2 C_y^2(1-\rho_{yx}^2) . \quad (12)$$

Kadilar (2016) by making use of auxiliary variable, proposed an exponential type estimator of population mean as,

$$t_6 = \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right)^{\delta} \exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right) \quad (13)$$

Where δ is a constant to be determined such the MSE of above estimator is minimum.
This is biased estimator and its bias and the mean squared error of the above estimator up to the first order of approximation respectively are,

$$B(t_6) = \frac{1-f}{n}\bar{Y}\left[\left\{\frac{\delta(\delta-1)}{2} + \frac{3}{8}\right\}C_x^2 + \left(\delta + \frac{1}{2}\right)C_{yx}\right]$$

$$MSE(t_6) = \frac{1-f}{n}\bar{Y}^2\left[C_y^2 + \left(\delta^2 + \delta + \frac{1}{4}\right)C_x^2 + (2\delta+1)C_{yx}\right] \quad (14)$$

The optimum value of the characterizing scalar δ which minimizes the mean squared error of t_6 is,

$$\delta_{opt} = \left(\frac{1}{2} - \rho_{yx} C_y / C_x\right)$$

The minimum mean squared error of above estimator is,

$$MSE_{\min}(t_6) = \frac{1-f}{n}\bar{Y}^2 C_y^2(1-\rho_{yx}^2) \quad (15)$$

Subramani (2016) proposed the following ratio estimator of population mean of the study variable by making use of median of study variable as,

$$t_7 = \bar{y}\left(\frac{M}{m}\right) \tag{16}$$

where M and m are the population and sample medians of study variable respectively.
The bias and the mean squared error, up to the first order of approximation, are respectively given by,

$$B(t_7) = \frac{1-f}{n}\bar{Y}[C_m^2 - C_{ym} - \frac{Bias(m)}{M}] \text{ and}$$

$$MSE(t_7) = \frac{1-f}{n}\bar{Y}^2[C_y^2 + R_7^2 C_m^2 - 2R_7 C_{ym}], \tag{17}$$

where, $R_7 = \frac{\bar{Y}}{M}$, $C_m = \frac{S_m}{M}$, $S_m^2 = \frac{1}{{}^N C_n}\sum_{i-1}^{{}^N C_n}(m_i - M)^2$, $S_{ym} = \frac{1}{{}^N C_n}\sum_{i-1}^{{}^N C_n}(\bar{y}_i - \bar{Y})(m_i - M)$ and $C_{ym} = \frac{S_{ym}}{\bar{Y}M}$.

Various modified estimators of population mean have been given by various authors in the literature. The latest references can be found in Subramani (2013), Subramani and Kumarapandiyan (2012, 2013), Tailor and Sharma (2009), Yan and Tian (2010), Yadav et al. (2014, 2015), Yadav et al. (2016), and Abid *et al.* (2016).

3. PROPOSED ESTIMATORS

We propose a class of ratio type estimators of population mean using known population median of study variable as,

$$t_p = \bar{y}\left[2 - \left(\frac{M}{m}\right)^{\delta}\right] \tag{18}$$

where δ is a suitably chosen characterizing scalar to be determined such that the MSE of the proposed estimator t_p is minimum.

To study the properties of the proposed estimator, following assumptions have been used as, $\bar{y} = \bar{Y}(1+e_0)$ and $m = M(1+e_1)$ **such that** $E(e_0) = 0$, $E(e_1) = \frac{\bar{M} - M}{M} = \frac{Bias(m)}{M}$ **and**

$$E(e_0^2) = \frac{1-f}{n}C_y^2,\ E(e_1^2) = \frac{1-f}{n}C_m^2,\ E(e_0 e_1) = \frac{1-f}{n}C_{ym},$$

where, $\bar{M} = \frac{1}{n}\sum_{i=1}^{n} m_i$

The proposed class of estimators t_p can be expressed in terms of $e_i{}'s\ (i = 1, 2)$ as,

$$t_p = \bar{Y}(1+e_0)[2 - (1-e_1)^{-\delta}]$$

Expanding the right hand side of the above equation and retaining the terms up to the first order of approximation, we get,

$$t_p = \bar{Y}(1+e_0)\left[2 - \left\{1 - \delta e_1 + \frac{\delta(\delta+1)}{2}e_1^2\right\}\right],$$

$$= \bar{Y}(1+e_0)\left[1 + \delta e_1 - \frac{\delta(\delta+1)}{2}e_1^2\right]$$

$$= \bar{Y}\left[1 + e_0 + \delta e_1 + \delta e_0 e_1 - \frac{\delta(\delta+1)}{2}e_1^2\right]$$

$$t_p - \bar{Y} = \bar{Y}\left[e_0 + \delta e_1 + \delta e_0 e_1 - \frac{\delta(\delta+1)}{2}e_1^2\right]. \tag{19}$$

Taking expectation on both sides of (19) and putting the values of various expectations, we get the bias of the proposed estimator t_p, up to the first order of approximation as,

$$B(t_p) = \bar{Y}\left[\frac{1-f}{n}\delta C_{ym} - \frac{\delta(\delta+1)}{2}C_m^2 + \delta\frac{Bias(m)}{M}\right]$$

From equation (19), up to the first order of approximation, we have,

$t_p - \bar{Y} \cong \bar{Y}[e_0 + \delta e_1]$ Squaring both sides and taking expectations on both sides of above equation, we get the mean squared error of the proposed estimator t_p as,

$$MSE(t_p) = \bar{Y}^2 E(e_0^2 + \delta^2 e_1^2 + 2\delta\, e_0 e_1)$$

$$= \bar{Y}^2 [E(e_0^2) + \delta^2 E(e_1^2) + 2\delta\, E(e_0 e_1)]$$

Putting the values of different expectations in above equation, we have,

$$MSE(t_p) = \frac{1-f}{n} \bar{Y}^2 [C_y^2 + \delta^2 C_m^2 + \theta\delta C_{ym}] \quad (20)$$

which is minimum for,

$$\delta_{opt} = -C_{ym} \big/ C_m^2 \, .$$

The minimum mean squared error of the proposed class of estimators t_p for the optimum value of δ_{opt} is,

$$MSE_{\min}(t_p) = \frac{1-f}{n} \bar{Y}^2 \left[C_y^2 - \frac{C_{ym}^2}{C_m^2} \right]. \quad (21)$$

4. EFFICIENCY COMPARISON

The present section is the theoretical comparison with the competing estimator of population mean.

From equation (21) and equation (2), we have,

$V(t_0) - MSE_{\min}(t_p) > 0$ if

$\frac{C_{ym}^2}{C_m^2} > 0$, or if $C_{ym}^2 > 0$

Thus the proposed estimator is better than the usual mean per unit estimator of population mean.

From equation (21) and equation (4), we have,

$MSE(t_1) - MSE_{\min}(t_p) > 0$ if

$$\frac{C_{ym}^2}{C_m^2} - C_y^2 \rho_{yx}^2 > 0$$

The proposed estimator is better than the usual regression estimator of Watson (1937) under above condition.

From equation (21) and equation (6), we have,

$MSE(t_2) - MSE_{\min}(t_p) > 0$ if

$C_x^2 - 2C_{yx} + \frac{C_{ym}^2}{C_m^2} > 0$, if

$$C_x^2 + \frac{C_{ym}^2}{C_m^2} > 2C_{yx}$$

The proposed estimator is better than the usual ratio estimator given by Cochran (1940) with the above condition.

From equation (21) and equation (8), we have,

$MSE(t_3) - MSE_{\min}(t_p) > 0$ if

$\frac{C_x^2}{4} - C_{yx} + \frac{C_{ym}^2}{C_m^2} > 0$, or

$$\frac{C_x^2}{4} + \frac{C_{ym}^2}{C_m^2} > C_{yx}$$

The proposed estimator performs better than Bahl and Tuteja (1991) ratio type estimator of population mean under above condition.

From equation (21) and equation (10), we have,

$MSE(t_4) - MSE_{\min}(t_p) > 0$, if

$$\frac{C_{ym}^2}{C_m^2} - C_y^2 \rho_{yx}^2 > 0$$

With the above condition, the proposed estimator performs better than the Srivastava (1967) estimator.

It is also better than Reddy (1974) and Kadilar (2016) estimators of population mean using auxiliary information under the above condition as for Srivastava (1967) estimator in above equation.

From equation (21) and equation (17), we have,

$MSE(t_7) - MSE_{\min}(t_p) > 0$, if

$$R_7^2 C_m^2 - 2R_7 C_{ym} + \frac{C_{ym}^2}{C_m^2} > 0$$, or

$$R_7^2 C_m^2 + \frac{C_{ym}^2}{C_m^2} > 2R_7 C_{ym}$$

The proposed estimator performs better than the Subramani (2016) competing estimator of population mean using information on median of the study variable with the above condition.

5. NUMERICAL STUDY

In this section, the numerical illustration has been done. We have considered the natural populations given in Subramani (2016). The three populations are taken as; 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. Following Tables 3-5 represent the parameter values for three populations along with constants, biases of proposed and competing estimators and variances and mean squared errors of existing and proposed estimators respectively.

Table-3. Parameter values and constants for three natural populations.

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

Table-4. Bias of various estimators.

Estimator	Popln-1	Popln-2	Popln-3
t_2	35.3748	40.9285	0.1067
t_3	1.39995	1.72380	0.0019
t_4	-1.60997	1.76775	0.0054
t_5	2.07309	1.85541	0.0167
t_6	27.4137	27.4137	0.3743
t_7	57.7705	57.7705	0.5061

Table-5. Mean squared error of various estimators.

Estimator	Popln-1	Popln-2	Popln-3
t_0	15640.97	15640.97	2.15
t_1	12486.75	12539.30	1.24
t_2	14895.27	15492.08	1.48
t_3	12498.01	12539.30	1.30
t_4	12486.75	12539.30	1.24

t_5	12486.75	12539.30	1.24
t_6	12486.75	12539.30	1.24
t_7	10926.53	10926.53	1.09
t_p	9002.22	9002.22	0.98

6. RESULTS AND CONCLUSION

In the present manuscript, we have proposed a general class of estimators of population mean using additional information on population median of the study variable without increasing cost of the survey. The bias and the mean squired error of the proposed estimator have been obtained up to the first order of approximation. The proposed estimator is theoretically compared with the competing estimators of population mean using auxiliary variables. The conditions under which proposed estimator perform better than the competing estimators have been given. The theoretical findings have been justified through some natural populations. It has been shown that proposed estimator is better than the competing estimators of population mean as it has lesser mean squared error among all mentioned estimators. Thus it is recommended to use the proposed estimator for improved estimation of population mean.

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 Agric. 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.