Vol.8.Issue.1.2020 (Jan-Mar.)

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

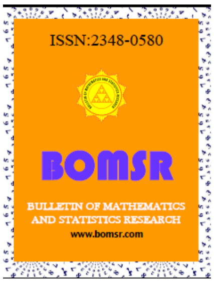
http://www.bomsr.com
Email:editorbomsr@gmail.com

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

RESEARCH ARTICLE

ON SOME ALMOST UNBIASED RATIO TYPE ESTIMATORS

SITANSHU SHEKHAR MISHRA
Lecturer in Statistics, Department of Statistics, Salipur (Autonomous) College, Salipur, Cuttack, Odisha, India
E-mail: sitanshumishra0@gmail.com
DOI: 10.33329/bomsr.8.1.11



ABSTRACT

In this paper three almost unbiased ratio type estimators have been constructed for estimating the population mean $\bar{Y}$ of the study variable *y* using information on an auxiliary variable *x*. Further, the biases and mean square errors (MSE) of these proposed estimators are compared under simple random sampling without replacement (SRSWOR) scheme, both theoretically and with numerical illustrations.

Keywords : Simple Random Sampling, Ratio estimator, Almost unbiased estimator, Bias, Mean square error, Auxiliary variable, Square root transformation,

1. Introduction :

Consider a finite population $U=\{U_1,U_2,...,U_N\}$ of *N* distinct and identificable units. Let *y* and *x* be the variable under study and auxiliary variable *x* respectively. Let (Y_i, X_i), $i=1,2,...,N$ be the paired values indexing the population units. Draw a simple random sample without replacement of size *n*. Let (y_i, x_i), $i=1,2,...,n$ be the paired values attached to the sample units $\{u_1, u_2,...,u_n\}$.

Define $t_i=\bar{y}$ as the unbiased estimator of $\bar{Y}$ of *y* without the use of auxiliary information. Let us consider the problem of estimating the population mean $\bar{Y}$ of the study variable *y*, assuming the prior knowledge of the population mean $\bar{X}$ of the auxiliary variable *x*.

The use of auxiliary information is invariably used in the sample surveys to provide estimators which are more precise over the mean per unit estimator $\bar{y}$ of the study variable. In this connection ratio and regression methods of estimation are used for the purpose. Ratio or ratio type estimators are more popular because of their simplicity.

However, the ratio or ratio type estimators are biased estimators, bias being of $O\left(\frac{1}{n}\right)$, where *n* is the sample size. For large samples the biases may be negligible. But for small sample sizes, the bias may be substantial, so as to make the estimate unreliable. This calls for devising techniques which remove the bias $O\left(\frac{1}{n}\right)$, and the ultimate bias becomes of $O\left(\frac{1}{n^2}\right)$. There are many bias reduction methods available in sampling theory literature. One such method is linear variety estimator proposed by Singh & Singh (1993(a),(b)).

In the following three linear variety of weighted estimators are constructed whose first order biases to $O\left(\frac{1}{n}\right)$ is removed. These weighted estimators may be termed as almost unbiased estimators.

2. Construction of linear variety estimators :

Case-I :

Define $t_1 = \bar{y}$

$$t_2 = \frac{\bar{y}}{\bar{x}}\bar{X} \qquad \text{(ratio estimator)}$$

$$t_3 = \bar{y}\, e^{\frac{1}{2}\left(\frac{\bar{X}-\bar{x}}{\bar{X}}\right)}. \text{ (Swain, 2013)}$$

The linear variety estimator is proposed as

$$T_1 = w_1 t_1 + w_2 t_2 + w_3 t_3 \text{ with } \sum w_i = 1.$$

i.e. $T_1 = w_1\bar{y} + w_2\frac{\bar{y}}{\bar{x}}\bar{X} + w_3\,\bar{y}\, e^{\frac{1}{2}\left(\frac{\bar{X}-\bar{x}}{\bar{X}}\right)}$.

Case-II :

Define $t_1 = \bar{y}$

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

$$t_4 = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^{\frac{1}{2}}. \text{ (Swain, 2013)}$$

The linear variety estimator is proposed as

$$T_2 = w_1 t_1 + w_2 t_2 + w_3 t_4$$

i.e. $T_2 = w_1\bar{y} + w_2 \dfrac{\bar{y}}{\bar{x}}\bar{X} + w_3\,\bar{y}\left(\dfrac{\bar{X}}{\bar{x}}\right)^{\frac{1}{2}}$ with $\Sigma w_i = 1 \ (i = 1, 2, 3)$.

Case-III :

Define $t_1 = \bar{y}$

$$t_3 = \bar{y}\, e^{\frac{1}{2}\left(\frac{\bar{X} - \bar{x}}{\bar{X}}\right)}.$$

$$t_4 = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^{\frac{1}{2}}$$

The linear variety estimator is proposed as

$$T_3 = w_1 t_1 + w_2 t_4 + w_3 t_3$$

i.e. $T_3 = w_1\bar{y} + w_2\bar{y}\left(\dfrac{\bar{X}}{\bar{x}}\right)^{\frac{1}{2}} + w_3\,\bar{y}\,e^{\frac{1}{2}\left(\frac{\bar{X} - \bar{x}}{\bar{X}}\right)}$ with $\Sigma w_i = 1$.

3. Biases and Mean square errors of T_1, T_2 and T_3 :

3.1 Bias and MSE of T_1 :

$$T_1 = w_1 t_1 + w_2 t_2 + w_3 t_3$$

i.e. $T_1 = w_1\bar{y} + w_2 \dfrac{\bar{y}}{\bar{x}}\bar{X} + w_3\,\bar{y}\,e^{\frac{1}{2}\left(\frac{\bar{X} - \bar{x}}{\bar{X}}\right)}$.

Define

$$\bar{y} = \bar{Y}(1 + e_0).$$

$$\bar{x} = \bar{X}(1 + e_1)$$

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

$$E(e_0^2) = \theta C_y^2,\ E(e_1^2) = \theta C_x^2,\ E(e_0 e_1) = \theta C_{yx} = \theta\rho C_y C_x.$$

where $\theta = \dfrac{1}{n} - \dfrac{1}{N}$.

Define

$$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2.$$

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2.$$

$$S_{yx} = \frac{1}{N-1}\sum_{i=1}^{N} (y_i - \bar{Y})(x_i - \bar{X}).$$

$$C_y = \frac{S_y}{\bar{Y}} \; , \; C_x = \frac{S_x}{\bar{X}} \; , \; \rho = \frac{S_{yx}}{S_y S_x} \; , \; K = \rho \frac{C_y}{C_x}.$$

Now,

$$T_1 = w_1 \bar{y} + w_2 \frac{\bar{y}}{\bar{x}} \bar{X} + w_3 \bar{y}\, e^{\frac{1}{2}\left(\frac{\bar{X}-\bar{x}}{\bar{X}}\right)}.$$

$$= w_1 \bar{Y}(1+e_0) + w_2\left(\bar{Y}(1+e_0).\frac{\bar{X}}{\bar{X}(1+e_1)}\right) + w_3 \bar{Y}(1+e_0).e^{-e_1/2}.$$

$$= w_1 \bar{Y}(1+e_0) + w_2(\bar{Y}(1+e_0)(1+e_1)^{-1}) + w_3 \bar{Y}(1+e_0)(1 - \frac{e_1}{2} + \frac{e_1^2}{8} - + ...),$$

keeping terms upto second degree.

$$= w_1 \bar{Y}(1+e_0) + w_2(\bar{Y}(1+e_0)(1-e_0)(1-e_1+e_1^2) + w_3 \bar{Y}(1+e_0)(1-\frac{e_1}{2}+\frac{e_1^2}{8} - ...)$$

$$= \bar{Y} + w_2 \bar{Y}\,\theta\left(C_x^2 - C_{yx}\right) + w_3 \bar{Y}\theta\left(\frac{1}{8}C_x^2 - \frac{1}{2}C_{yx}\right).$$

Thus bias of T_1 to $O\left(\frac{1}{n}\right)$, is

$$B(T_1) = E(T_1) - \bar{Y}$$

$$= \theta\bar{Y}\left[w_2(C_x^2 - C_{yx}) + w_3(\frac{1}{8}C_x^2 - \frac{1}{2}C_{yx}) \right]$$

$$= w_2 \; B(t_2) + w_3 \, B(t_3), \text{ since } B(t_1) = 0.$$

where, $B(t_2) = \theta\bar{Y}\,(C_x^2 - C_{yx})$

$$B(t_3) = \theta\bar{Y}\left(\frac{1}{8}C_x^2 - \frac{1}{2}C_{yx}\right).$$

Write

$$T_1 = w_1\bar{Y} + w_1\bar{Y}e_0 + w_2\bar{Y} - w_2\bar{Y}e_1 + w_2\bar{Y}e_0 + w_3\bar{Y} - w_3\bar{Y}\frac{e_1}{2} + w_3\bar{Y}\,e_0$$

$$= \left(w_1 + w_2 + w_3\right)\bar{Y} + w_1\bar{Y}e_0 + w_2\bar{Y}e_0 + w_3\bar{Y}e_0 - w_2\bar{Y}e_1 - w_3\bar{Y}\frac{e_1}{2}$$

$$= \bar{Y} + \bar{Y}e_0\,(w_1 + w_2 + w_3) - \bar{Y}e_1(w_2 + \frac{w_3}{2})$$

$$=\bar{Y}+\bar{Y}e_0-\bar{Y}e_1(w_2+w_3/2)$$

$$=\bar{Y}+\bar{Y}\left[e_0-Ae_1\right],$$

where $A=w_2+\frac{w_3}{2}$.

Now, $T_1-\bar{Y}=\bar{Y}\left(e_0-Ae_1\right)$.

Now, MSE (T_1) = $E(T_1-\bar{Y})^2\cong\bar{Y}^2\,E\left[e_0-Ae_1\right]^2$.

$$\cong\bar{Y}^2\,E\left[e_0^2+A^2e_1^2-2Ae_0e_1\right]$$

$$\cong\bar{Y}^2\left[E(e_0^2)+E(A^2e_1^2)-2AE(e_0e_1)\right]$$

$$=\bar{Y}^2\left[E(e_0^2)+A^2E(e_1^2)-2A\,E(e_0e_1)\right]$$

$$=\bar{Y}^2\left[\theta C_y^2+A^2\theta\,C_x^2-2A\theta C_{yx}\right]$$

$$=\theta\bar{Y}^2\left[C_y^2+A^2\,C_x^2-2A\,C_{yx}\right].$$

MSE (T_1) is minimum, when

$$\frac{\partial\,\mathrm{MSE}\,(T_1)}{\partial A}=0.$$

This implies,

$$\left[\theta\bar{Y}^2\left(C_y^2+A^2C_x^2-2AC_{yx}\right)\right]=0.$$

$$\Rightarrow\quad 2AC_x^2-2\,C_{yx}=0$$

$$\Rightarrow\quad AC_x^2=C_{yx}$$

$$\Rightarrow\quad A=\frac{C_{yx}}{C_x^2}=\frac{\rho C_yC_x}{C_x^2}=\rho\frac{C_y}{C_x}=K,\ \text{say.}$$

Putting this value of $A=\rho\frac{C_y}{C_x}$, we get the minimum MSE of (T_1) as

$$\therefore\ \min\mathrm{MSE}\,(T_1)=\theta\bar{Y}^2C_y^2\left(1-\rho^2\right),$$

which is equal to that of simple linear regression estimator upto terms of $O\left(\frac{1}{n}\right)$.

To get unique values of w_i's, we solve the equations.

$$w_1+w_2+w_3=1. \qquad (1)$$

$$w_2 + \tfrac{1}{2} w_3 = K = \rho \frac{C_y}{C_x} \qquad (2)$$

$$w_1 B(t_1) + w_2 B(t_2) + w_3 B(t_3) = 0. \qquad (3)$$

Equations (1), (2), (3) may be written in the matrix form as

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & \tfrac{1}{2} \\ 0 & B(t_2) & B(t_3) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 1 \\ K \\ 0 \end{bmatrix}$$

Solving, equation by the use of determinants we get

$$w_1 = \frac{\Delta_1}{\Delta} \ ,$$

$$w_2 = \frac{\Delta_2}{\Delta} \ ,$$

$$w_3 = \frac{\Delta_3}{\Delta} \ .$$

where $\Delta = B(t_3) - \frac{1}{2} B(t_2)$

$$\Delta_1 = B(t_3)(1-K) + (K - \frac{1}{2}) B(t_2).$$

$$\Delta_2 = K.B(t_3)$$

$$\Delta_3 = -K.B(t_2).$$

The choice of w_i's (*i* = 1,2,3) remove the bias to terms of $O(\frac{1}{n})$.

3.2 Bias and MSE of T_2 :

$$T_2 = w_1 t_1 + w_2 t_2 + w_3 t_4 .$$

$$T_2 = w_1 \bar{y} + w_2 \frac{\bar{y}}{\bar{x}} \bar{X} + w_3 \bar{y} \left(\frac{\bar{X}}{\bar{x}} \right)^{\frac{1}{2}} .$$

$$B(T_2) = \theta \bar{Y} \left[w_2 \left(C_x^2 - C_{yx} \right) + w_3 \left(\frac{3}{8} C_x^2 - \frac{1}{2} C_{yx} \right) \right].$$

$$MSE(T_2) = \theta \bar{Y}^2 \left[C_y^2 + A^2 C_x^2 - 2A C_{yx} \right].$$

$$Min\ MSE\,(T) = \theta \bar{Y}^2 C_y^2 (1 - \rho^2).$$

which is same as that of variance of traditional linear regression estimator.

As in case I, the equations can be written as in the matrix form as

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & \frac{1}{2} \\ 0 & B(t_2) & B(t_4) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 1 \\ K \\ 0 \end{bmatrix}$$

Using, we get unique values of w_2's (i=1,2,3) as

$$w_1 = \frac{\Delta_1}{\Delta} \;,$$

$$w_2 = \frac{\Delta_2}{\Delta} \;,$$

$$w_3 = \frac{\Delta_3}{\Delta} \;.$$

where $\Delta = B(t_4) - \frac{1}{2} B(t_2)$

$$\Delta_1 = B(t_4)(1-K) + (K - \frac{1}{2}) B(t_2).$$

$$\Delta_2 = K.B(t_4)$$

$$\Delta_3 = -K.B(t_2).$$

3.3 Bias and MSE of T_3 :

$$T_3 = w_1 t_1 + w_2 t_4 + w_3 t_3 .$$

i.e., $$T_3 = w_1 \bar{y} + w_2 \bar{y} \left(\frac{\bar{X}}{x} \right)^{\frac{1}{2}} + w_3 \bar{y}\, e^{\frac{1}{2}\left(\frac{\bar{X} - \bar{x}}{\bar{X}} \right)}.$$

$$B(T_3) = \theta \bar{Y} \left[w_2 \left(\frac{3}{8} C_x^2 - \frac{1}{2} C_{yx} \right) + w_3 \left(\frac{1}{8} C_x^2 - \frac{1}{2} C_{yx} \right) \right].$$

$$\text{MSE}(T_3) = \theta \bar{Y}^2 \left[C_y^2 + A^2 C_x^2 - 2A C_{yx} \right].$$

This is minimum when $\frac{\partial \text{MSE}(T_3)}{\partial A} = 0.$

i.e. $A = \rho \frac{C_y}{C_x} = K$

Min MSE $(T_3) = \theta \bar{Y}^2 C_y^2 (1 - \rho^2).$

which is same as that of traditional linear regression estimator.

The equations for solving w_1, w_2 and w_3 can be written as in the matrix form as :

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & B(t_4) & B(t_3) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 1 \\ K \\ 0 \end{bmatrix}$$

We have,

$$w_1 = \frac{\Delta_1}{\Delta}, \qquad w_2 = \frac{\Delta_2}{\Delta}, \qquad w_3 = \frac{\Delta_3}{\Delta}.$$

where $\Delta = \frac{1}{2} B(t_3) - \frac{1}{2} B(t_4)$

$$\Delta_1 = B(t_3)(\frac{1}{2} - K) + (K - \frac{1}{2}) B(t_4).$$

$$\Delta_2 = K.B(t_3)$$

$$\Delta_3 = -K.B(t_4).$$

4. Efficiencies of T_1, T_2 and T_3 :

To $O(\frac{1}{n})$,

MSE $(t_1) = V(t_1) = \bar{Y}^2 \theta C_y^2$

MSE $(t_2) = \theta \bar{Y}^2 \left(C_y^2 + C_x^2 - 2C_{yx} \right)$

MSE $(t_3) = \theta \bar{Y}^2 \left(C_y^2 + \frac{1}{4} C_x^2 - C_{yx} \right)$

MSE $(t_4) = \theta \bar{Y}^2 \left(C_y^2 + \frac{1}{4} C_x^2 - C_{yx} \right)$

Where $\theta = \left(\frac{1}{n} - \frac{1}{N} \right)$.

MSE (T_1) , MSE (T_2) and MSE (T_3) being equal to that of the linear regression estimator always less than those of MSE (t_1), MSE (t_2), MSE (t_3) and MSE (t_4) upto terms of $O(\frac{1}{n})$.

5. Numerical Illustrations :

Consider four natural populations given in Table 1 along with the computed parameters given in Table 2.

The biases of individual estimators and weighted estimators are shown in Table 3

Table 1: Description of Populations

Population	Description	N	n	$\bar{Y}$	$\bar{X}$	C_y	C_x	ρ	k
I	Murthy 1967 y = output for 80 factories in a region x=fixed capital	80	20	51.8264	11.2646	0.3542	0.7507	0.9413	0.4441
II	Murthy 1967 y = output for 80 factories in a region x = Data on number of workers	80	20	51.8264	2.8513	0.3542	0.9484	0.9150	0.3416
III	Sukhatme and Sukhatme (1970) y = Number of villages in the circles. x = A circle consisting more than five villages.	89	12	3.360	0.1236	0.60400	2.1901	0.766	0.2111
IV	Kadilar and Cingi (2006) y = the levels of apple production. x = the number of apple trees..	104	20	625.37	13.93	1.866	1.653	0.865	0.9764

Table 2: Values of w_i's (i = 1,2,3) for almost unbiased estimators

Estimator T_1 / Population	w_1	w_2	w_3
I	0.2264	0.1149	0.6587
II	0.3587	0.0416	0.5997
III	0.5669	-0.0107	0.4438
IV	-0.0070	0.9457	0.0613
Estimator T_2 / Population	w_1	w_2	w_3

I	-0.4317	-0.5427	1.9744
II	-0.2411	-0.5574	1.7988
III	0.1233	-0.4546	1.3319
IV	-0.0684	0.8843	0.1841
Estimator T_3 / Population	w_1	w_2	w_3
I	0.1112	0.3451	0.5437
II	0.3163	0.1251	0.5586
III	0.5778	-0.0326	0.4548
IV	-0.9528	2.8372	-0.8844

Table 3: Comparison of Biases

Estimators	$t_1 = \bar{y}$	$t_2 = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)$	$t_3 = \bar{y}e^{\frac{1}{2}\left(\frac{\bar{X}-\bar{x}}{\bar{X}}\right)}$	$t_4 = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^{\frac{1}{2}}$	T_1	T_2	T_3
Bias / Population	$B(t_1)$	$B(t_2)$	$B(t_3)$	$B(t_4)$	$B(T_1)$	$B(T_2)$	$B(T_3)$
I	0	0.6088	-0.1063	0.1675	0.000005	0.0004	0.000006
II	0	1.1506	-0.0800	0.3568	-0.0001	0.0005	0.000006
III	0	0.9174	0.0225	0.3133	0.0001	0.0002	0.000007
IV	0	1.6270	-25.0754	-7.8169	0.0015	0.0004	-0.0015

Comments : As evident from Table 3 the biases of the weighted estimators are very insignificant compared to the biases of individual estimators.

6. Conclusion

The linear variety estimators T_1, T_2 and T_3 are more efficient than the individual estimators and also the biases are negligibly small as per the numerical illustrations. The unknown parameters in the computation of w_i's may be substituted by their consistent estimates to be used in practice.

Acknowledgement

The author is grateful to Prof. A.K.P.C. Swain, retired Professor of Statistics, Utkal University for his valuable suggestions and criticisms, which improved the presentation of results incorporated in this paper.

References :

[1]. Murthy, M.N. (1967): Sampling theory and methods. Statistical Publishing Society, Calcutta, India.

[2]. Singh, S. and Singh, R. (1993a): A new method: Almost separation of bias precipitates in sample surveys. Jour. Ind. Stat. Assoc., 31, 99-105.

[3]. Singh, R. Kumar, M. and Smarandache, F. (2008): Almost unbiased estimators for estimating population mean using known value of some population parameters. Pakistani Journal of Statistics and Operations Research, 4(2), 63-76.

[4]. Cochran, W.G. (1977): Sampling Techniques, 3rd Ed. John Wiley and Sons, Inc. London.

[5]. Kadilar, C. and Cingi, H. (2006). Improvement in estimating the population means in simple random sampling. Appl. Math. Comput., 19, 78-79.

[6]. Sukhatme, P.V. and Sukhatme, B.V. (1970): Sampling theory of surveys with applications. Asia Publishing House, New Delhi.

[7]. Singh, S. and Singh, R. (1991): Almost bias precipitate filteration: A new technique. Aligarh J. Statistics, 11, 5-8.

[8]. Singh, S. and Singh, R. (1993b): A class of almost unbiased ratio and product estimators. Jour. Indian Soc. Agricultural Statist., 14, 35-39.

[9]. Swain, A.K.P.C. (2013): An alternative ratio type exponential estimator. Unpublished communication.

[10]. Swain, A.K.P.C. (2014) : On an improved ratio type estimator of finite population mean in sample surveys. Revista Investigacion Operacional. 35(1), 49-57.

[11]. Swain, A.K.P.C. and Mishra, S.S. (2016): On Comparison of two generalized class of ratio type estimators in sample surveys. Revista investigacion Operacional, 37(3), 272-280.