GLOBAL JOURNAL OF PURE AND APPLIED SCIENCES VOL. 14, NO. 1, 2008: 121 - 126
 1SSN 1118-0579

ALTERNATIVE TO SEARLS RATIO TYPE ESTIMATOR IN SAMPLE SURVEYS

A. A. ADEWARA

(Received 3 June 2004; Revision Accepted 6 November 2006)

ABSTRACT

An alternative to Searls (1964) ratio - type estimator is proposed. Based on the mean square error of this proposed ratio - type estimator, it is observed to be more efficient than Searls (1964) ratio - type estimator provided the condition for its supremacy is met. Ten life and hypothetical data sets are used to testify to this claim.

KEYWORDS: alternative, ratio - type, estimator, efficiency, condition

1. INTRODUCTION

Let N and n be the population and sample sizes respectively, $\bar{X}$ and $\bar{Y}$ be the population mean for the auxiliary variable (X) and for the variable of interest (Y), $\bar{x}$ and $\bar{y}$ be the sample means based on the sample drawn, $\bar{x}^*$ and $\bar{y}^*$ be the means of the auxiliary variable and of the variable of interest yet to be drawn (Srivenkataramana and Srinath (1976)), that is the means corresponding to the (N – n) population units and ρ be the coefficient of correlation between X and Y

In this paper, the following estimators are considered for comparison:-

(i). Mean Square.

(ii). Searls (1964) Estimator, $\bar{y}' = W\sum_{i=1}^{n} y_i$, where W $= (\frac{1}{n+c^2_y})$, is chosen so that the mean square error, $E(\bar{y}' - \mu)^2$, is a minimum.

(iii). Alternative Ratio Type Estimator, $\bar{y}_{aa} = \frac{\bar{y}}{\bar{x}^*}\bar{X}$.

Conventionally, $\bar{x} = \bar{X}(1+\Delta_x)$, $\bar{y} = \bar{Y}(1+\Delta_y)$. But, $\bar{x}^* = \bar{X}(1+(\frac{\theta}{1+\theta})\Delta_x)$, where, $\theta = \frac{1}{n}$. Hence, $\bar{x}^* = \bar{X}(1+(\frac{1}{n+1})\Delta_x)$.

The only difference between $\bar{x}^*$ and $\bar{x}$ lies in the coefficient of Δ_x. While that of $\bar{x}^*$ is $(\frac{1}{n+1})$, that of $\bar{x}$ is 1

The bias and mean square error (mse) of $\bar{y}$ and $\bar{y}'$ are:-

$$bias(\bar{y}) = 0 \quad (1)$$

$$\text{var}(\bar{y}) = \frac{s^2_y}{n} = mse(\bar{y}) \quad (2)$$

Searls (1964) gave the bias and MSE of $\bar{y}'$ as:-

$$bias(\bar{y}') = -\frac{\bar{Y}c^2_y}{n+c^2_y} \quad (3)$$

and

$$mse(\bar{y}') = \frac{s^2_y}{n+c^2_y} \quad (4)$$

Also, the bias and mean square error of our alternative ratio - type estimator, $\bar{y}_{aa}$, are:-

$$bias(\bar{y}_{aa}) = \frac{\bar{Y}}{n}[(\frac{1}{n+1})^2 c^2_x - \rho(\frac{1}{n+1})c_x c_y] \quad (5)$$

and

$$mse(\bar{y}_{aa}) = \frac{\bar{Y}^2}{n}[c^2_y - 2\rho(\frac{1}{n+1})c_x c_y + (\frac{1}{n+1})^2 c^2_x] \quad (6)$$

A. A. Adewara, Department of Statistics, University of Ilorin, Ilorin. Kwara State Nigeria

where, $c_x = \frac{s_x}{\bar{x}}$ and $c_y = \frac{s_y}{\bar{y}}$ are the coefficient of variation of x and y respectively. These derivations are shown in the Appendix.

2. COMPARISON

(a). Efficiency of $\bar{y}_{aa}$ over $\bar{y}'$

Estimator $\bar{y}_{aa}$ is said to be more efficient than Estimator, $\bar{y}'$ if and only if $mse(\bar{y}_{aa}) < mse(\bar{y}')$. That is,

$$\frac{\bar{Y}^2}{n}[c^2{}_y - 2\rho(\frac{1}{n+1})c_x c_y + (\frac{1}{n+1})^2 c^2{}_x] < \frac{s^2{}_y}{n+c^2{}_y}$$. This implies that

$$s^2{}_y > \frac{(n+c^2{}_y)\bar{Y}^2}{n}[c^2{}_y - 2\rho(\frac{1}{n+1})c_x c_y + (\frac{1}{n+1})^2 c^2{}_x]$$. This means that

$$s^2{}_y > Q, \text{ where, } Q = \frac{(n+c^2{}_y)\bar{Y}^2}{n}[c^2{}_y - 2\rho(\frac{1}{n+1})c_x c_y + (\frac{1}{n+1})^2 c^2{}_x]$$

(b). Efficiency of $\bar{y}'$ over $\bar{y}_{aa}$.

Also, Estimator $\bar{y}'$ is said to be more efficient than Estimator, $\bar{y}_{aa}$, if and only if $mse(\bar{y}') < mse(\bar{y}_{aa})$. That is,

$$\frac{s^2{}_y}{n+c^2{}_y} < \frac{\bar{Y}^2}{n}[c^2{}_y - 2\rho(\frac{1}{n+1})c_x c_y + (\frac{1}{n+1})^2 c^2{}_x]$$. This implies that

$$s^2{}_y < \frac{(n+c^2{}_y)\bar{Y}^2}{n}[c^2{}_y - 2\rho(\frac{1}{n+1})c_x c_y + (\frac{1}{n+1})^2 c^2{}_x]$$. This also means that

$$s^2{}_y < Q.$$

3. DATA USED

To demonstrate the efficiency of this alternative ratio - type Estimator, $\bar{y}_{aa}$, with that of **Searls (1964) Estimator**, $\bar{y}'$. The ten life and hypothetical data sets represented below are used to give room for easy comparison.

Table 1:- Data sets used

Population 1:- Singh and Chaudhary (1986). pp 177.

x = number of inhabitants in 1973
y = area under wheat in 1971

$s^2x = 1944.81$, $s^2y = 3974.04$, $\bar{X} = 73.5$, $\bar{Y} = 98.5$, $\rho = 0.88$, $n = 10$

Population 2:- Murthy (1967). pp 422.

x = number of cattles in census
y = number of cattles in surveys

$s^2x = 296850.62$, $s^2y = 268598.61$, $\bar{X} = 514$, $\bar{Y} = 523.5$, $\rho = 0.98$, $n = 12$

Population 3:- Menendez and Reyes (1998). Hypothetical (Shabbir, 2003)
$s^2x = 3.31$, $s^2y = 8.40$, $\bar{X} = 16.6$, $\bar{Y} = 12$, $\rho = 0.20$, $n = 100$

Population 4:- Murthy (1967). pp. 78.
x = geographical area in acres
y = area under winter paddy

$s^2x = 622.35$, $s^2y = 2514.45$, $\bar{X} = 277.19$, $\bar{Y} = 106.69$, $\rho = 0.29$, $n = 16$

Population 5:- Cochran (1977). pp 171 – 172.

x = acres in farm
y = acres in corn

$s^2x = 7619$, $s^2y = 620$, $\overline{X} = 117.28$, $\overline{Y} = 26.30$, ρ=0.67, n=100

Population 6:- Menendez and Reyes (1998). Hypothetical (Shabbir, 2003).

$s^2x = 140415$, $s^2y = 6409$, $\overline{X} = 150$. $\overline{Y} = 90$, ρ = 0.05, n = 100

Population 7:- Tin.(1965). Artificial sampling experiment.

$s^2x = 45$, $s^2y = 500$, $\overline{X} = 5$, $\overline{Y} = 15$, ρ = 0.4, n = 50

Population 8:- Tin.(1965).Artificial sampling experiment.

$s^2x = 45$, $s^2y = 500$, $\overline{X} = 5$, $\overline{Y} = 15$, ρ = 0.6, n = 200

Population 9:- Tin.(1965). Artificial sampling experiment.

$s^2x = 45$, $s^2y = 500$, $\overline{X} = 5$, $\overline{Y} = 15$, ρ = 0.8, n = 1000

Population 10.- Hypothetical

$s^2x = 0.5$, $s^2y = 0.5$, $\overline{X} = 0.2$, $\overline{Y} = 0.2$, ρ = 0.5, n = 20

4. RESULTS

The estimates obtained using the above life data sets are shown in **Tables 2 and 3 below**:-

Table 2:- Estimates of $\bar{y}$, $\bar{y}'$ and $\bar{y}_{aa}$ when $\bar{y}_{aa}$ is preferred

Estimator	pop 1	pop 2	pop 3
Samp.size(n)	10	12	100
s^2x	1944.810	296850.62	3.3100
s^2y	3974.040	268598.61	8.4000
Q	3546.343	245606.10	8.3899
c^2x	0.3600	1.1236	0.0120
c^2y	0.4096	0.9801	0.0058
$\bar{y}$	98.5000	523.5000	12.0000
$\bar{x}$	73.5000	514.0000	16.6000
ρ	0.8800	0.9800	0.2000
$bias(\bar{y})$	0	0	0
$bias(\bar{y}_{aa})$	-0.2733	-3.1611	-0.00001
$bias(\bar{y}')$	-3.8758	-39.5284	-0.0010
$mse(\bar{y})$	397.4040	22383.2200	0.0840
$mse(\bar{y}_{aa})$	340.6800	18921.7400	0.0838
$mse(\bar{y}')$	381.7668	20693.1100	0.0839

Table 3:- Estimates of $\bar{y}$, $\bar{y}'$ and $\bar{y}_{aa}$ when $\bar{y}'$ i preferred

Estimator	pop 4	pop 5	pop 6	pop 7	pop 8	pop 9	pop 10
Samp.size(n)	16	100	100	50	200	1000	20
s^2x	622.35	7619	140415	45	45	45	0.5
s^2y	2514.45	620	6409	500	500	500	0.5
Q	2532.83	624.7101	6432.37	515.0123	502.84	500.39	0.7757
c^2x	0.0081	27.6492	6.2407	1.8	1.8	1.8	12.5
c^2y	0.2209	4.3056	0.7912	2.222	2.222	2.222	12.5
$\bar{y}$	106.69	12	90	15	15	15	0.2
$\bar{x}$	277.19	16.6	150	5	5	5	0.2
ρ	0.2900	0.67	0.05	0.4	0.6	0.8	0.5
$bias(\bar{y})$	0	0	0	0	0	0	0
$bias(\bar{y}_{aa})$	-0.0046	-0.0086	-0.0004	-0.0045	-0.0004	-0.00002	-0.0027
$bias(\bar{y}')$	-1.4529	-0.4953	-0.7065	-0.6383	-0.1648	-0.0333	-0.0769
$mse(\bar{y})$	157.1531	6.2000	64.09	10	2.5	0.5	0.025
$mse(\bar{y}_{aa})$	156.1464	5.9892	63.8188	9.8619	2.4866	0.4993	0.0238
$mse(\bar{y}')$	155.0130	5.9441	63.5869	9.5725	2.4725	0.4989	0.0154

5. DISCUSSIONS

From the estimates in Table 2 for populations 1 - 3, our alternative ratio -type estimator, $\bar{y}_{aa}$, has the least mean square error (mse) since our condition, $s^2_y > \frac{(n+c^2_y)\bar{Y}^2}{n}[c^2_y - 2\rho(\frac{1}{n+1})c_x c_y + (\frac{1}{n+1})^2 c^2_x]$, for its superiority over Searls (1964) ratio - type estimator, $\bar{y}'$, is satisfied.

Also from Table 3 where the estimates for populations 4 – 10 are shown, Searls (1964) ratio - type estimator, $\bar{y}'$, has the least mean square error (mse) since our condition,

$$s^2_y < \frac{(n+c^2_y)\bar{Y}^2}{n}[c^2_y - 2\rho(\frac{1}{n+1})c_x c_y + (\frac{1}{n+1})^2 c^2_x],$$

for its superiority over our alternative ratio - type estimator, $\bar{y}_{aa}$, is satisfied.

6. CONCLUSION

Therefore, the alternative ratio - type estimator, $\bar{y}_{aa}$, is preferred whenever

$s^2_y > \frac{(n+c^2_y)\bar{Y}^2}{n}[c^2_y - 2\rho(\frac{1}{n+1})c_x c_y + (\frac{1}{n+1})^2 c^2_x]$ while the Searls(1964) ratio - type estimator, $\bar{y}'$, is also preferred whenever $s^2_y < \frac{(n+c^2_y)\bar{Y}^2}{n}[c^2_y - 2\rho(\frac{1}{n+1})c_x c_y + (\frac{1}{n+1})^2 c^2_x]$.

REFERENCES

Cochran, W.G., 1977. Sampling Techniques, 3rd Edition. James Wiley, New York.

Menedez, E. and Reyes, A., 1998. On an efficient regression type estimator. Biom. J. 40 : 79 – 84

Murthy, M.N., 1967. Sampling Theory and Methods. Statistical Publishing Society, Calcutta, India.

Srivenkataramana, T. and Srinath, K.P., 1976. Ratio and Product Method of estimation in Sample Surveys when the two variables are moderately correlated. Vignana Bharathi. 2: 54 - 58

Searles, D.T., 1964. The utilization of a known coefficient of variation in the estimation procedure. J. Amer. Stat. Assoc. 59: 1225 - 1226.

Shabbir, J., 2003. On an efficient ratio type estimator. Proc. Pakistan Acad. Sci. 40(2): 179 -182

Singh, D. and Chandhary, F.S., 1986. Theory and Analysis of sample survey Design. Wiley Eastern Ltd.

Tin, M., 1965. Comparison of some Estimators. J. Amer. Stat. Asso. 60: 295 – 307.

APPENDIX

DERIVATION OF THE BIAS AND MEAN SQUARE ERROR OF $\bar{y}_{aa}$

Let,

$\bar{y}_{aa} = \frac{\bar{y}}{\bar{x}^*}\bar{X}$, where

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

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

(iii). $\bar{x}^* = \bar{X}(1+(\frac{\theta}{1+\theta})\Delta_{\bar{x}})$, $\theta = \frac{1}{n}$. Then, $\bar{x}^* = \bar{X}(1+(\frac{1}{n+1})\Delta_{\bar{x}})$,

(iv). $\Delta_{\bar{x}} = \frac{\bar{x}-\bar{X}}{\bar{X}}$ such that $|\Delta_{\bar{x}}| < 1$ and

(v). $\Delta_{\bar{y}} = \frac{\bar{y}-\bar{Y}}{\bar{Y}}$ such that $|\Delta_{\bar{y}}| < 1$.

Therefore using power series expansion,

$$
\begin{aligned}
\bar{y}_{aa} = \frac{\bar{y}}{\bar{x}^*}\bar{X} &= \frac{\bar{Y}(1+\Delta_{\bar{y}})\bar{X}}{\bar{X}(1+(\frac{1}{n+1})\Delta_{\bar{x}})} \\
&= \frac{\bar{Y}(1+\Delta_{\bar{y}})}{(1+(\frac{1}{n+1})\Delta_{\bar{x}})} \\
&= \bar{Y}(1+\Delta_{\bar{y}})(1+(\frac{1}{n+1})\Delta_{\bar{x}})^{-1} \\
&= \bar{Y}(1+\Delta_{\bar{y}})(1-(\frac{1}{n+1})\Delta_{\bar{x}}+(\frac{1}{n+1})^2\Delta_{\bar{x}}^2)+o(n^{-1}) \\
&= \bar{Y}[1+\Delta_{\bar{y}}-(\frac{1}{n+1})\Delta_{\bar{x}}+(\frac{1}{n+1})^2\Delta_{\bar{x}}^2-(\frac{1}{n+1})\Delta_{\bar{x}}\Delta_{\bar{y}}+(\frac{1}{n+1})^2\Delta_{\bar{x}}^2\Delta_{\bar{y}}]
\end{aligned}
$$

$$= \bar{Y}[1+\Delta_{\bar{y}}-(\frac{1}{n+1})\Delta_{\bar{x}}+(\frac{1}{n+1})^2\Delta_{\bar{x}}^2-(\frac{1}{n+1})\Delta_{\bar{x}}\Delta_{\bar{y}}], \text{ ignoring higher orders.}$$

Then,

$$
\begin{aligned}
bias(\bar{y}_{aa}) &= E(\bar{y}_{aa}-\bar{Y}) \\
&= E[\bar{Y}(1+\Delta_{\bar{y}}-(\frac{1}{n+1})\Delta_{\bar{x}}+(\frac{1}{n+1})^2\Delta_{\bar{x}}^2-(\frac{1}{n+1})\Delta_{\bar{x}}\Delta_{\bar{y}})-\bar{Y}] \\
&= \bar{Y}E[(1+\Delta_{\bar{y}}-(\frac{1}{n+1})\Delta_{\bar{x}}+(\frac{1}{n+1})^2\Delta_{\bar{x}}^2-(\frac{1}{n+1})\Delta_{\bar{x}}\Delta_{\bar{y}})-1] \\
&= \bar{Y}E[(\Delta_{\bar{y}}-(\frac{1}{n+1})\Delta_{\bar{x}}+(\frac{1}{n+1})^2\Delta_{\bar{x}}^2-(\frac{1}{n+1})\Delta_{\bar{x}}\Delta_{\bar{y}})], \text{ where}
\end{aligned}
$$

$E(\Delta_{\bar{x}}) = E(\Delta_{\bar{y}}) = 0$, $E(\Delta_{\bar{x}}^2) = \frac{s^2_x}{n\bar{x}^2} = (\frac{1}{n})c^2_x$, $E(\Delta_{\bar{y}}^2) = \frac{s^2_y}{n\bar{y}^2} = (\frac{1}{n})c^2_y$ and

$E(\Delta_{\bar{x}}\Delta_{\bar{y}}) = \frac{s_{yx}}{n\bar{y}\bar{x}} = (\frac{1}{n})\rho c_x c_y$.

Therefore,

$$bias(\bar{y}_{aa}) = \bar{Y}[0 - (\frac{1}{n+1})0 + (\frac{1}{n(n+1)})^2 c^2{}_x - (\frac{1}{n(n+1)})\rho c_x c_y]$$

$$= \frac{\bar{Y}}{n}[(\frac{1}{n+1})^2 c^2{}_x - (\frac{1}{n+1})\rho c_x c_y].$$ As in eq. (5) in the text.

Also,

$$mse(\bar{y}_{aa}) = E(\bar{y}_{aa} - \bar{Y})^2$$

$$= E[\bar{Y}(1 + \Delta_{\bar{y}} - (\frac{1}{n+1})\Delta_{\bar{x}} + (\frac{1}{n+1})^2 \Delta_{\bar{x}}{}^2 - (\frac{1}{n+1})\Delta_{\bar{x}}\Delta_{\bar{y}}) - \bar{Y}]^2$$

$$= \bar{Y}^2 E[(1 + \Delta_{\bar{y}} - (\frac{1}{n+1})\Delta_{\bar{x}} + (\frac{1}{n+1})^2 \Delta_{\bar{x}}{}^2 - (\frac{1}{n+1})\Delta_{\bar{x}}\Delta_{\bar{y}}) - 1]^2$$

$$= \bar{Y}^2 E[(\Delta_{\bar{y}} - (\frac{1}{n+1})\Delta_{\bar{x}} + (\frac{1}{n+1})^2 \Delta_{\bar{x}}{}^2 - (\frac{1}{n+1})\Delta_{\bar{x}}\Delta_{\bar{y}})]^2$$

$$= \bar{Y}^2 E[(\Delta_{\bar{y}}{}^2 - 2(\frac{1}{n+1})\Delta_{\bar{x}}\Delta_{\bar{y}} + (\frac{1}{n+1})^2 \Delta_{\bar{x}}{}^2 + 4(\frac{1}{n+1})^2 \Delta_{\bar{x}}{}^2 \Delta_{\bar{y}} -$$

$$2(\frac{1}{n+1})^3 \Delta_{\bar{x}}{}^3 + (\frac{1}{n+1})^4 \Delta_{\bar{x}}{}^4 - 2(\frac{1}{n+1})^3 \Delta_{\bar{x}}{}^3 \Delta_{\bar{y}} -$$

$$2(\frac{1}{n+1})\Delta_{\bar{x}}\Delta_{\bar{y}}{}^2 + (\frac{1}{n+1})^2 \Delta_{\bar{x}}{}^2 \Delta_{\bar{y}}{}^2)]$$

$$= \bar{Y}^2 E[(\Delta_{\bar{y}}{}^2 - 2(\frac{1}{n+1})\Delta_{\bar{x}}\Delta_{\bar{y}} + (\frac{1}{n+1})^2 \Delta_{\bar{x}}{}^2)],$$ ignoring higher orders.

$$= \frac{\bar{Y}^2}{n}[c^2{}_y - 2(\frac{1}{n+1})\rho c_x c_y + (\frac{1}{n+1})^2 c^2{}_x].$$ As in eq. (6) in the text.