Alternative to Tin Ratio Type Estimator

*Amos A. Adewara

Department of Statistics, University of Ilorin, Ilorin, Nigeria

Received: July 2, 2015; Revised: October, 14, 2015; Accepted: November 9, 2015.

Abstract

In sample surveys, small sample studies are usually known for their sensitivity. Hence, many studies on estimator formulation in ratio estimation are based on moderate or large sample study. In this study, an alternative ratio type estimator capable of solving small sample problems was proposed. The bias and the variance of this alternative ratio type estimator were derived. Two life data sets of small sample sizes and a secondary data of large sample size were used to justify the superiority of the proposed alternative ratio type estimator. It was observed that the proposed alternative ratio type estimator was more efficient than Tin ratio type estimator with respect to its bias and the variance, provided its condition is met.

Keywords: alternative, ratio, estimator, bias, variance, efficiency.

1.0 Introduction

The literature on survey sampling describes a great variety of techniques for using auxiliary information to obtain more efficient estimators. Ratio method of estimation is a good example in this context. If the correlation between the study variable y and the auxiliary variable x is positive (high), then the ratio method of estimation is quite effective. Many authors have used auxiliary information for improved estimation of population mean of study variable y with moderate and large sample sizes [1-5]. In sample surveys, supplementary information is often used for increasing the precision of estimators [6-9]. In this study, an alternative ratio type estimator capable of solving small sample cases in sample surveys is proposed. This estimator can favourably be compared with that of Tin ratio type estimator [10] which was derived from Bearle [11] estimator when we have an infinite population.

Tin [10] defined its ratio type estimator as $t = r[1+\theta(\frac{s_{xy}}{\bar{x}\bar{y}} - \frac{s^2{}_x}{\bar{x}^2})]$ which was derived from Bearle [11] estimator,

$\dfrac{r(1+\theta\frac{s_{xy}}{\bar{x}\bar{y}})}{(1+\theta\frac{s^2{}_x}{\bar{x}^2})}$, where $r = \frac{\bar{y}}{\bar{x}}$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$,

$s^2{}_x = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, $s^2{}_y = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$, $s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ and

$\theta = (\frac{1}{n} - \frac{1}{N}) = \frac{1}{n}$ as $N \to \infty$.

*** Corresponding Author: Tel: +234(0)8033813093; Email: aaadewara@gmail.com**

Tin [10] gave its bias and variance in $O(n^{-3})$ as:

$$bias(t) = R[1 - \frac{3c^2{}_x}{n^2}(c^2{}_x - c_{xy})(1 + \frac{6}{n} + \frac{10c^2{}_x}{n})] \quad (1)$$

and $$\text{var}iance(t) = \frac{R^2}{n}[(c^2{}_x + c^2{}_y - 2c_{xy}) + \frac{1}{n}(1+\frac{1}{n})(2c^2{}_x - 4c^2{}_y c_{xy} + c^2{}_{xy} + c^2{}_x c^2{}_y]$$

$$-\frac{4c^2{}_x}{n^2}(6c^2{}_x - 12c^2{}_y c_{xy} + 5c^2{}_{xy} + c^2{}_x c^2{}_y] \quad (2)$$

respectively, where $c_{xy} = \rho c_x c_y$.

2.0 Materials and Methods

2.1 Data Used

To demonstrate the efficiency of the proposed alternative ratio type estimator, $\bar{y}_{aa}$, over Tin ratio type estimator (t) capable of solving small sample cases, two life data sets are used to give room for easy comparison. The sample sizes considered in this study are: 3, 4, 5, 6, 10, 15, 20 and 25 with the intention of knowing what will happen to the proposed estimator, $\bar{y}_{aa}$, as the sample sizes increases. These two life data sets are the published students' scores in Statistics for Agriculture and Biological Sciences during 2009/2010 and 2010/2011 academic sessions.

2.2 On the proposed alternative ratio estimator, $(\bar{y}_{aa})$

Let N and n be the population and sample sizes respectively, $\bar{X}$ and $\bar{Y}$ be the population means for the auxiliary variable (X) and 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 variable of interest yet to be drawn [12] (i.e. the means corresponding to the (N-n) population units) and ρ be the coefficient of correlation between X and Y.

The relationship between $\bar{X}$, $\bar{x}$ and $\bar{x}^*$ is $\bar{X} = f\bar{x} + (1-f)\bar{x}^*$ while that of $\bar{Y}$, $\bar{y}$ and $\bar{y}^*$ is $\bar{Y} = f\bar{y} + (1-f)\bar{y}^*$ [12]. The proposed ratio type estimator is given as:

$\bar{y}_{aa} = \frac{\bar{y}^*}{\bar{x}^*}\bar{X}$ [13], where $\bar{x} = \bar{X}(1+\Delta_{\bar{x}})$, $\bar{y} = \bar{Y}(1+\Delta_{\bar{y}})$, $\bar{x}^* = \bar{X}(1+\theta\Delta_{\bar{x}})$, $\bar{y}^* = \bar{Y}(1+\theta\Delta_{\bar{y}})$,

$\theta = \frac{1}{n(n+1)}$, $\Delta_{\bar{y}} = \frac{\bar{y}-\bar{Y}}{\bar{Y}}$, $\Delta_{\bar{x}} = \frac{\bar{x}-\bar{X}}{\bar{X}}$ such that $|\Delta_{\bar{y}}| < 1$ and $|\Delta_{\bar{x}}| < 1$

Then, $\bar{x}^* = \bar{X}(1+(\frac{1}{n(n+1)})\Delta_{\bar{x}})$ and $\bar{y}^* = \bar{Y}(1+(\frac{1}{n(n+1)}\Delta_{\bar{y}})$

Here, the only difference between:-

(i). $\bar{x}^*$ and $\bar{x}$ lies in the coefficient of $\Delta_{\bar{x}}$, while that of $\bar{x}^*$ is $(\frac{1}{n(n+1)})$, that of $\bar{x}$ is 1.

(ii). $\bar{y}^*$ and $\bar{y}$ lies in the coefficient of $\Delta_{\bar{y}}$, while that of $\bar{y}^*$ is $(\frac{1}{n(n+1)})$, that of $\bar{y}$ is 1.

2.3 *Bias and Variance of the alternative ratio estimator,* $(\bar{y}_{aa})$

Using power series expansion,

$$\bar{y}_{aa} = \frac{\bar{y}^*}{\bar{x}^*}\bar{X} = \frac{\bar{Y}(1+(\frac{1}{n(n+1)})\Delta_{\bar{y}})\bar{X}}{\bar{X}(1+(\frac{1}{n(n+1)})\Delta_{\bar{x}})} = \frac{\bar{Y}(1-(\frac{n}{N-n})\Delta_{\bar{y}})}{(1-(\frac{n}{N-n})\Delta_{\bar{x}})}$$

$$= \frac{\bar{Y}(1+(\frac{1}{n(n+1)})\Delta_{\bar{y}})}{(1+(\frac{1}{n(n+1)})\Delta_{\bar{x}})}$$

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

$$= \bar{Y}(1+(\frac{1}{n(n+1)})\Delta_{\bar{y}})[1-(\frac{1}{n(n+1)})\Delta_{\bar{x}} + (\frac{1}{n(n+1)})^2\Delta_{\bar{x}}^{\,2} - (\frac{1}{n(n+1)})^3\Delta_{\bar{x}}^{\,3}...]$$

Let,

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

$$E(\Delta_x\Delta_y) = \frac{s_{xy}}{n\bar{x}\bar{y}} = (\frac{1}{n})\rho c_x c_y$$

$$bias(\bar{y}_{aa}) = E(\bar{y}_{aa} - \bar{Y})$$

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

Hence,

$$bias(\bar{y}_{aa}) = \frac{\bar{Y}}{n}[((\frac{1}{n(n+1)})^2 c^2{}_x - \rho(\frac{1}{n(n+1)})c_x c_y) + \frac{1}{n}((\frac{1}{n(n+1)})^3 c^2{}_x c_y - (\frac{1}{n(n+1)})^3 c^3{}_x$$

$$+\frac{1}{n^2}((\frac{1}{n(n+1)})^4 c^4{}_x - (\frac{1}{n(n+1)})^4 c^3{}_x c_y)] \qquad \ldots \qquad (3)$$

Also,

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

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

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

$$\frac{1}{n}(3(\frac{1}{n(n+1)})^3 c^2{}_x c_y - 2(\frac{1}{n(n+1)})^3 c_x c^2{}_y - 2((\frac{1}{n(n+1)})^3 c^3{}_x) +$$

$$\frac{1}{n^2}((\frac{1}{n(n+1)})^4 c^4{}_x + (\frac{1}{n(n+1)})^3 c^2{}_x c^2{}_y - 2((\frac{1}{n(n+1)})^4 c^3{}_x c_y)], \; O(n^{-3})$$

$$\operatorname{var}(\bar{y}_{aa}) = (\frac{N-n}{Nn})(\frac{n}{N-n})^2 (s^2{}_y - 2Rs_{xy} + R^2 s^2{}_x) \qquad \ldots \qquad (4)$$

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.

2.4 Efficiency comparison of $\bar{y}_{aa}$ over t

Considering the first terms of the variance of $\bar{y}_{aa}$ and t, the proposed alternative ratio type estimator, $\bar{y}_{aa}$, is said to be more efficient than Tin ratio type estimator, t, if and only if,

$$\frac{\bar{Y}^2(\frac{1}{n(n+1)})^2}{n}[c^2{}_x + c^2{}_y - 2\rho c_x c_y] < \frac{R^2}{n}[c^2{}_x + c^2{}_y - 2\rho c_x c_y], \text{ that is, } \quad (\frac{1}{n(n+1)})^2 \le \frac{1}{\bar{x}^2}.$$

Also, Tin ratio type estimator, t, is said to be more efficient than the proposed alternative ratio type estimator, $\bar{y}_{aa}$, if and only if,

$$\frac{R^2}{n}[c^2{}_x + c^2{}_y - 2\rho c_x c_y] < \frac{\bar{Y}^2(\frac{1}{n(n+1)})^2}{n}[c^2{}_x + c^2{}_y - 2\rho c_x c_y], \text{ that is, } \quad (\frac{1}{n(n+1)})^2 > \frac{1}{\bar{x}^2}$$

3.0 Results

The estimates obtained using the two life data sets described in Section 2.1 are presented in Tables 1 and 2.

Table 1: Estimates obtained on t and $\bar{y}_{aa}$ using 2009/2010 academic session life data set

Sample size (n)	3	4	5	6	10	15	20	25
ρ	0.8220	0.6822	0.8019	0.7213	0.5951	0.4569	0.4879	0.3957
$(\frac{1}{n(n+1)})$	0.0833	0.0500	0.0333	0.0238	0.0091	0.0042	0.0024	0.0015
R	0.7543	0.6406	0.6074	0.6631	0.6333	0.5947	0.6181	0.6218
$s^2{}_x$	308.3336	272.25	375.2998	312	202	162.7143	163.3158	169.9167
$s^2{}_y$	48	374.25	423.8	357.7667	248.9	239.9238	200.2605	221.8333
c_x	0.3010	0.3041	0.4003	0.3758	0.2787	0.2551	0.2434	0.2370
c_y	0.1575	0.5567	0.7002	0.6069	0.4884	0.5209	0.4361	0.4355
$\bar{x}$	58.33	54.25	48.4	47	51	50	52.5	55
$\bar{y}$	44	34.75	29.1	31.1667	32.3	29.7333	32.45	34.20
bias(t)	0.1248	-0.0395	-0.0969	-0.0341	-0.0035	0.0037	0.0061	0.0120
bias($\bar{y}_{aa}$)	0.0522	-0.0493	-0.0004	-0.0001	0	0	0	0
var(t)	0.0159	0.0087	0.0069	0.0053	0.0218	0.0018	0.0008	0.0007
var($\bar{y}_{aa}$)	0.1647	0.1285	0.0384	0.0165	0.0013	0.0002	0.0001	0.0001
$(\frac{1}{n(n+1)})^2$	0.0694	0.0025	0.0011	0.0006	0.0001	0.00002	0.00006	0.000002
$\frac{1}{\bar{x}^2}$	0.0001	0.00003	0.0004	0.0005	0.0004	0.0004	0.00036	0.00033

4.0 Discussion

The estimates for sample sizes 3, 4, 5, 6, 10, 15, 20 and 25 represents the paired scores for 3, 4, 5 up to 25 students. For samples of sizes 10, 15, 20 and 25 in the two data sets (Tables 1 and 2), the estimated variances of t and $\bar{y}_{aa}$ indicate that the alternative ratio type estimator, $\bar{y}_{aa}$ is preferred because it has the least variance and also satisfies the condition $(\frac{1}{n(n+1)})^2 \le \frac{1}{\bar{x}^2}$, as stated. However, the estimated variances of t and $\bar{y}_{aa}$ for sample sizes 3, 4, 5 and 6; the reverse is the case in which Tin ratio type estimator, t, is preferred since $(\frac{1}{n(n+1)})^2 > \frac{1}{\bar{x}^2}$.

Of interest with this alternative ratio type estimator is that as the sample size increases (i.e. as the number of students offering these two courses); both its biases and the variances tend towards zero. To buttress this claim further using secondary and hypothetical data [14, 15];

$s^2{}_x = 3.31,\ s^2{}_y = 8.40,\ \bar{x} = 16.6,\ \bar{y} = 12.0,\ \rho = 0.20,\ \ s_{xy} = 1.0546$, n = 100, $(\frac{1}{n(n+1)})^2 = 9.8030x10^{-09}$,

$\frac{1}{\bar{x}^2} = 3.6290x10^{-03},\ (\frac{1}{n(n+1)})^2 \le \frac{1}{\bar{x}^2}$, bias(t) = 0.0052, bias($\bar{y}_{aa}$) = 0, var(t) = 0.0001 and var($\bar{y}_{aa}$) = 0 respectively. Based on these estimates, it was observed that the alternative ratio type estimator $\bar{y}_{aa}$, has the least variance since the condition $(\frac{1}{n(n+1)})^2 \leq \frac{1}{\bar{x}^2}$ is satisfied; hence, it is preferred.

Table 2: Estimates obtained on t and $\bar{y}_{aa}$ using 2010/2011 academic session life data set

Sample size (n)	3	4	5	6	10	15	20	25
ρ	0.9042	0.7852	0.6448	0.7095	0.7497	0.7912	0.7813	0.7549
$(\frac{1}{n(n+1)})$	0.0833	0.05	0.0333	0.0238	0.0091	0.0042	0.0024	0.0015
R	0.8993	0.8343	0.8565	0.8745	0.8473	0.8270	0.8786	0.8847
$s^2{}_x$	12.3333	12.9167	11.8002	11.3665	61.6001	133.4953	130.5686	127.6234
$s^2{}_y$	8.3333	66.9167	51.1999	51.10	105.3777	114.4094	181.5237	176.6101
c_x	0.0758	0.0794	0.0770	0.0746	0.1933	0.2884	0.2747	0.2692
c_y	0.0693	0.2167	0.1873	0.1810	0.2984	0.3228	0.3686	0.3580
$\bar{x}$	46.3333	45.25	44.60	45.1667	40.60	40.0667	41.60	41.96
$\bar{y}$	41.6667	37.75	38.20	39.5	34.40	33.1333	36.55	37.12
bias(t)	0.0027	-0.0151	-0.0064	-0.0070	-0.0081	0.0114	-0.0043	-0.0003
bias($\bar{y}_{aa}$)	0	-0.0002	0	0	0	0	0	0
var(t)	0.0014	0.0038	0.0022	0.0017	0.0013	0.0005	0.0006	0.0004
var($\bar{y}_{aa}$)	0.0042	0.0233	0.0073	0.0028	0.0004	0.0001	0	0
$(\frac{1}{n(n+1)})^2$	0.0694	0.0025	0.0011	0.0006	0.0001	0.00002	0.00006	0.000002
$\frac{1}{\bar{x}^2}$	0.0005	0.00049	0.0005	0.0005	0.00061	0.00062	0.00058	0.00057

5.0 Conclusion

The alternative ratio type estimator ($\bar{y}_{aa}$) employed in this study is preferred whenever $(\frac{1}{n(n+1)})^2 \leq \frac{1}{\bar{x}^2}$, while the ratio type estimator (t) is preferred to $\bar{y}_{aa}$ whenever $(\frac{1}{n(n+1)})^2 > \frac{1}{\bar{x}^2}$. The justification for studying small sample sizes involving an infinite population is that the two derived equations stated in section 2.4 must be satisfied.

6.0 Acknowledgement

The author would like to thank the referees for their useful suggestions on the previous draft of the paper.

References

[1] Kadilar, C. and Cingi, H. (2004). Ratio estimators in simple random sampling. Applied Mathematics and Computation, Vol. 151, pp. 893-902.

[2] Ogunyinka, P.I. and Sodipo, A.A. (2013). Efficiency of ratio and regression estimators using double Sampling. Journal of Natural Sciences Research, Vol. 3, No.7, pp. 202–207.

[3] Onyeka A.C (2012). Estimation of population mean in post-stratified sampling using known value of some population parameter(s). Statistics in Transition-New Series Vol.13, pp. 65-78.

[4] Pandey, H., Yadav, S. K. and Shukla, A. K. (2011). An improved general class of estimators estimating population mean using auxiliary information. International Journal of Statistics and Systems, Vol. 6, pp. 1-7.

[5] Solanki, R. S., Singh, H. P. and Rathour, A. (2012). An alternative estimator for estimating the finite population mean using auxiliary information in sample surveys. International Scholarly Research Network: Probability and Statistics, Vol. 2012, pp. 1-14.

[6] 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, Vol. 4, No. 2, pp. 21-27.

[7] Subramani, J. and Kumarapandiyan, G. (2012). Estimation of population mean using co-efficient of variation and median of an auxiliary variable. International Journal of Probability and Statistics, Vol.1, No. 4, pp. 111-118.

[8] Yadav, S. K. (2011). Efficient estimators for population variance using auxiliary information. Global Journal of Mathematical Sciences: Theory and Practical, Vol. 3, pp.369-376.

[9] Yadav, S. K. and Kadilar, C. (2013). Improved class of ratio and product estimators. Applied Mathematics and Computation, Vol. 219, pp. 10726–10731.

[10] Tin, M. (1965). Comparison of some estimators. Journal of American Statistical Association, Vol. 60, pp. 295-307.

[11] Bearle, E. M. L. (1962). Some use of computers in operational research. Industrielle Organization, Vol. 31, pp. 27–28.

[12] 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, Vol. 2, pp. 54–58.

[13] Adewara, A. A. (2006). Effect of improving both the auxiliary and variable of interest in ratio and product estimators. Proceeding Pakistan Academy Science, Vol. 43. No. 4, pp. 275–278.

[14] Menedez, E. and Reyes, A. (1998). On an efficient regression type estimator. Biometrical Journal, Vol. 40, pp. 79-84.

[15] Shabbir, J. (2003). On an efficient ratio type estimator. Proceeding Pakistan Academy Science, Vol. 40. No. 2, pp. 179–182.