A NOTE ON COMPARISON OF SOME ESTIMATORS USING AUXILIARY INFORMATION IN SAMPLE SURVEYS

L. N. SAHOO and S. MOHANTY

Department of Statistics, Orissa University of Agriculture and Technology, Bhubaneswar 751 003, India.

ABSTRACT

Three estimators w_a, w_g and w_h, based on arithmetic mean, geometric mean and harmonic mean of mean per unit and usual ratio estimators respectively, have been compared for their biases and mean square errors. It is observed that w_h is preferable in most practical situations.

INTRODUCTION

Let $\bar{x}$ and $\bar{y}$ be the sample means of the characters x and y respectively based on a simple random sample without replacement (SRSWOR) of size n drawn from a finite population of size N. To estimate the population mean $\bar{Y}$ of y, when the population mean $\bar{X}$ of x is known, it is well known that the usual ratio estimator $\bar{y}_r = \bar{y}\bar{X}/\bar{x}$ is more efficient than mean per unit estimator $\bar{y}$ when $\rho C_y/C_x > 1/2$, where C_x and C_y are the coefficients of variation of x and y respectively and ρ the correlation coefficient between them.

In this paper, an attempt has been made to generate and compare three estimators, w_a, w_g and w_h, by taking weighted arithmetic, geometric and harmonic means of $\bar{y}$ and $\bar{y}_r$ respectively as follows:

$$w_a = (1-w)\bar{y} + w\bar{y}_r$$

$$w_g = \bar{y}^{1-w}(\bar{y}_r)^w$$

and

$$w_h = \left(\frac{1-w}{\bar{y}} + \frac{w}{\bar{y}_r}\right)^{-1},$$

where w is a constant to be determined such that $0 \leqslant w \leqslant 1$. Henceforth, we call these estimators w-estimators.

These estimators were suggested separately in earlier attempts to reduce bias and mean square error simultaneously. The estimator w_a was considered by Chakrabarty[1] and Ray *et al*[2]; w_g by Srivastava[3]; and w_h by Reddy[4,5]. Further, Walsh[6] studied w_h with w as the sample estimate of $k = \rho C_y/C_x$. However, these authors did not discuss the basis of formulation of the estimators but used them as such. In the present paper an attempt has been made to compare among these w-estimators for fairly good ranges of w. In the last section, two small artificial populations have been used to show the efficiency of these estimators on the basis of bias and mean square error (MSE).

BIAS AND MSE OF O (1/n)

The bias and MSE of w-estimators to terms of $O(1/n)$ are obtained as follows:

$$B(w_a) = \theta_1 \bar{Y} C_x^2 w(1-k),$$

$$B(w_g) = \theta_1 \bar{Y} C_x^2 w\left(\frac{1+w}{2} - k\right),$$

$$B(w_h) = \theta_1 \bar{Y} C_x^2 w(w-k), \tag{1}$$

and

$$M(w_a) = M(w_g) = M(w_h) = \theta_1 \bar{Y}^2 (C_y^2 - 2w\rho C_y C_x + w^2 C_x^2), \tag{2}$$

where

$$\theta_1 = \left(\frac{1}{n} - \frac{1}{N}\right).$$

It is interesting to note that the bias of w_g is equal to the simple arithmetic mean of biases of the estimators w_a and w_h. It is well known that the variance of $\bar{y}$ is $\theta_1 \bar{Y}^2 C_y^2$ and up to $O(1/n)$ the bias and MSE of $\bar{y}_r$ are $\theta_1 \bar{Y} C_x^2 (1-k)$ and $\theta_1 \bar{Y}^2 (C_y^2 - 2\rho C_y C_x + C_x^2)$ respectively.

Comparing square of the biases of different estimators we get the following results:

(i) $|B(w_g)| \leqslant |B(w_a)|$ if $w \geqslant 4k-3$,

(ii) $|B(w_h)| \leqslant |B(w_a)|$ if $w \geqslant 2k-1$,

and, (iii) $|B(w_h)| \leqslant |B(w_g)|$ if $w \geqslant \frac{1}{3}(4k-1)$.

Since $|B(w_a)| \leqslant |B(\bar{y}_r)|$ for $w \leqslant 1$, the above results lead to the following theorem:

Theorem 1

When $k < 1$,

$$|B(w_h)| < |B(w_g)| < |B(w_a) < |B(\bar{y}_r)| \qquad (3)$$

if $\frac{1}{3}(4k-1) < w < 1$.

Comparing MSE we have the following theorem:

Theorem 2

The w-estimators are more efficient than both $\bar{y}$ and $\bar{y}_r$ when

$$\frac{w}{2} < k < \frac{1+w}{2},$$

$$\text{i.e.}\quad 2k-1 < w < 2k. \qquad (4)$$

Thus we find that when k lies between $w/2$ and $1/2$, $\bar{y}_r$ is less efficient than $\bar{y}$, whereas w-estimators are more efficient than $\bar{y}$.

It can be easily shown that MSE of w-estimators up to $O(1/n)$ in (2) are minimum when $w=k$ and for that case, (i) MSE of w-estimators are equal to that of linear regression estimator, (ii) w_h is almost unbiased, and (iii) w_a is more biased than w_g, followed by the usual ratio estimator $\bar{y}_r$.

We find that the efficiency of w-estimators depends on the value of w, which in turn depends on k. But in practice k is not known. However, a good guess of k can be made from a pilot survey or past data or experience to get an estimated value of w (cf. Srivastava[3], Reddy[7]).

HIGHER ORDER COMPARISON OF MSE OF w-ESTIMATORS

Following Tin[8], we obtain the MSE of the w-estimators to $O(1/n^2)$ as follows:

$$\begin{aligned}M(w_a) = \bar{Y}^2 [&\theta_1 (C_{02} - 2wC_{11} + w^2 C_{20}) \\ &- 2w\left(\theta_2 - \frac{3\theta_1}{N}\right)\{wC_{30} - (1+w)C_{21} \\ &+ C_{12}\} + \theta_1^2 w \{9wC_{20}^2 - 6(1+2w)C_{20}C_{11} \\ &+ 2(2+w)C_{11}^2 + (2+w)C_{20}C_{02}\}],\end{aligned}$$

$$\begin{aligned}M(w_g) = \bar{Y}^2 [&\theta_1 (C_{02} - 2wC_{11} + w^2 C_{20}) \\ &- w\left(\theta_2 - \frac{3\theta_1}{N}\right)\{w(1+w)C_{30} \\ &- (1+3w)C_{21} + 2C_{12}\} \\ &+ \theta_1^2 w \{\tfrac{1}{4} w(1+w)(11+7w)C_{20}^2 \\ &- (1+w)(2+7w)C_{20}C_{11} \\ &+ 2(1+2w)C_{11}^2 + (1+2w)C_{20}C_{02}\}],\end{aligned}$$

$$\begin{aligned}M(w_h) = \bar{Y}^2 [&\theta_1 (C_{02} - 2wC_{11} + w^2 C_{20}) \\ &- 2w\left(\theta_2 - \frac{3\theta_1}{N}\right)(w^2 C_{30} - 2wC_{21} + C_{12}) \\ &+ 3\theta_1^2 w^2 (3w^2 C_{20}^2 - 6wC_{20}C_{11} \\ &+ 2C_{11}^2 + C_{20}C_{02})],\end{aligned}$$

where $\theta_2 = (n^{-2} - N^{-2})$ and $C_{ij} = K_{ij}/\bar{X}^i \bar{Y}^j$, K_{ij} being the (i, j)th cumulant of x and y.

Now we have

$$\begin{aligned}M(w_a) - M(w_g) = \bar{Y}^2 w(1-w)\Bigg[&\left(\theta_2 - \frac{3\theta_1}{N}\right) \\ &(C_{21} - wC_{30}) + \theta_1^2 C_{20}C_{02}(1-\rho^2) \\ &+ \frac{\theta_1^2}{4} C_{20}^2 \{7(w-k)^2 + 11(w-k) \\ &+ (1-k)(14w - 5k)\}\Bigg]. \qquad (5)\end{aligned}$$

Hence, from (5), w_g is more efficient than w_a when

$$w \leqslant \frac{C_{21}}{C_{30}} \text{ and } k \leqslant w \leqslant 1. \qquad (6)$$

Comparing MSE of w-estimators in a similar manner we get similar conditions as in (6) and hence, the following theorem:

Theorem 3

Among the w-estimators,

$$M(w_h) \leqslant M(w_g) \leqslant M(w_a), \qquad (7)$$

when $k \leqslant w \leqslant C_{21}/C_{30}$ or 1 (whichever is less).

In situations where the joint distribution of x and y is bivariate normal, (7) holds good whenever $w \geqslant k$. A similar condition was obtained by Reddy[4] for the case $w=k$ while comparing w_g with w_h.

EMPIRICAL STUDY

The efficiencies of the w-estimators have been compared with those of $\bar{y}$ and $\bar{y}_r$ by using two small artificial populations, based on SRSWOR of size 3, by enumerating all possible samples.

The first population consists of 5 units having values of (x, y) as (2, 4), (5, 8), (7, 6), (10, 12) and

Table 1 *Regions of w, $V(\bar{y})$, $|B(\bar{y}_r)|$ and $M(\bar{y}_r)$*

Population no.	Regions of w: Based on theorem 2	Regions of w: Based on theorem 3	$V(\bar{y})$	$\|B(\bar{y}_r)\|$	$M(\bar{y}_r)$
1	$0.30 < w < 1.0$	$0.65 < w < 0.667$	1.3333	0.1110	0.7435
2	$0.0 < w < 0.738$	$0.369 < w < 1.0$	0.7999	0.4090	3.4915

Table 2 *Bias and mean square error of w-estimators*

Population no.	w	w_a: $\|B(w_a)\|$	w_a: $M(w_a)$	w_g: $\|B(w_g)\|$	w_g: $M(w_g)$	w_h: $\|B(w_h)\|$	w_h: $M(w_h)$
1	0.30	0.0333	0.6294	0.0007	0.6347	0 0308	0.6400
	0.40	0.0444	0.4951	0 0071	0.4970	0.0294	0.5031
	0.50	0.0555	0.4110	0.0165	0.4081	0.0221	0.4119
	0.60	0.0666	0.3772	0.0290	0.3689	0.0087	0.3678
	0.65	0.0722	0.3790	0.0365	0.3685	0.0003	0.3642
	0.66	0.0733	0.3809	0.0380	0.3699	0.0023	0.3649
	0.667	0.0741	0.3825	0.0392	0.3713	0.0038	0.3658
	0.75	0.0833	0.4204	0.0537	0.4068	0.0232	0.3961
	0.85	0.0944	0.5120	0.0742	0.4989	0.0529	0.4852
2	0.20	0.0818	0.1818	0.0291	0.1882	0.0190	0.1999
	0.30	0.1227	0.0896	0.0529	0.0646	0.0117	0.0654
	0.369	0.1509	0.1102	0.0732	0.0549	0.0000	0.0381
	0.45	0.1840	0.2222	0.1009	0.1265	0.0209	0.0785
	0.60	0.2454	0.6802	0.1638	0.5139	0.0815	0.3857
	0.738	0.3018	1.3887	0.2353	1.1933	0.1644	0.9964
	0.85	0.3477	2.1662	0.3034	2.0008	0.2535	1.7995
	0.90	0.3681	2.5718	0.3367	2.4428	0.3004	2.2725

(11, 10). The second population reported by Ray *et al*[2] consists of 6 units with values of (x, y) as (1, 7), (3, 8), (4, 10), (6, 11), (7, 11) and (9, 13). The values of k for these populations are 0.65 and 0.369 respectively. Regions of w based on theorems 2 and 3, and exact values of $V(\bar{y})$, $|B(\bar{y}_r)|$ and $M(\bar{y}_r)$ are given in table 1. In table 2, the exact values of biases and MSE of the w-estimators for different values of w are given. It can be seen from table 2 that results of the theorems 2 and 3 are satisfied by both the populations.

CONCLUSION

From the above study we may conclude that it is better to use w-estimators, preferably w_h, instead of $\bar{y}$ and $\bar{y}_r$, when advance knowledge on k is available and satisfies the conditions given in the text.

ACKNOWLEDGEMENT

The authors are grateful to the referee for some helpful suggestions for the improvement of the paper.

10 August 1987; Revised 12 August 1988

1. Chakrabarty, R. P., *J. Indian Soc. Agric. Stat.*, 1979, **31**, 49.
2. Ray, S. K., Sahai, A. and Sahai, A., *Ann. Inst. Stat. Math.*, 1979, **31**, 141.
3. Srivastava, S. K., *Bull. Calcutta Stat. Assoc.*, 1967, **16**, 121.
4. Reddy, V. N., *Sankhya*, 1973, **B35**, 307.
5. Reddy, V. N., *Sankhya*, 1974, **C36**, 59.
6. Walsh, J. E., *Sankhya*, 1970, **A32**, 99.
7. Reddy, V. N., *Sankhya*, 1978, **C40**, 29.
8. Tin, M., *J. Am. Stat. Assoc.*, 1965, **60**, 294.