Jour. Ind. Soc. Ag. Statistics
45(2), 1993 : 226-229

An Almost Unbiased Product Estimator

Vyas Dubey
Pt. Ravishankar Shukla University, Raipur (M.P.)
(Received : Novemeber, 1991)

Summary

An almost unbiased product estimator has been proposed. Under certain conditions, proposed estimator is more efficient than usual product estimator and the unbiased estimator proposed by Robson [2].

Key Words : Auxiliary Characteristic, Bias, Mean Square Error (MSE), Simple Random Sampling Without Replacement (SRSWOR) procedure, Finite Population Correction (FPC), Relative Efficiency.

Introduction

Product method of estimation, used for estimating the population mean $\bar{Y}$ of a certain characteristic when correlation between the main characteristic y and an auxiliary characteristic x is negative, is well known in the literature. It is defined as

$$\bar{Y}_p = \bar{y}\,\bar{x} / \bar{X}, \tag{1.1}$$

where $\bar{y}$ is the sample mean of y, $\bar{x}$ and $\bar{X}$ are respectively sample and population means of x. $\bar{Y}_p$ is more efficient than the sample mean estimator $\bar{y}$ if

$$(C_{11} / C_{20}) < -(1/2) \tag{1.2}$$

Here $C_{ij} = M_{ij} / \bar{X}^i \bar{Y}^j$, $M_{ij} = E\,(x - \bar{X})^i\,(y - \bar{Y})^j$.

This is a biased estimator. Let a sample of size n be taken from the population of size N by SRSWOR procedure, Bias of $\bar{Y}_p$ is

$$B\,(\bar{Y}_y) = (N - n)\, M_{11} / (N - 1)\, n\bar{X} \tag{1.3}$$

To make $\bar{Y}_p$ unbiased, Robson [2] proposed as follows :

$$t_1 = \bar{Y}_p - \theta \left(s_{yx} / \bar{X}\right) \tag{1.4}$$

$s_{yx} = \sum_{1}^{n} (y - \bar{y})\,(x - \bar{x})/(n - 1)$ is the sample covariance between y and x, $\theta = (N - n)/N\,n$. Let $\hat{B}\,(\overline{Y_p})$ be an estimate of $B(\bar{Y}_p)$. Let us consider a general bias adjusted estimator

$$t_g = \overline{Y}_p - \hat{B}(\overline{Y}_p) \tag{1.5}$$

If $B(\overline{Y}_p)$ is estimated by

$$\hat{B}_1(\overline{Y}_p) = \theta\, s_{yx} / \overline{X}, \tag{1.6}$$

we obtain t_1. But if it is estimated by

$$\hat{B}_2(\overline{Y}_p) = \theta\, s_{yx} / \overline{x}, \tag{1.7}$$

then we obtain another estimator

$$t_2 = \overline{Y}_p - \theta\,(s_{yx} / \overline{x}). \tag{1.8}$$

In section 2, bias and MSE of the proposed estimator t_2, has been obtained; and in section 3, its efficiency has been examined.

2. *Bias and MSE of t_2*

Following Sukhatme *et al* [3], pp.239, the proposed estimator will have bias

$$B(t_2) = \overline{Y}\,\theta^2 C_{11} N\left[\frac{C_{21}}{(N-2)\,C_{11}} - \frac{C_{20}}{(N-1)}\right] \tag{2.1}$$

and MSE

$$M(t_2) = V(t_1) + 2\overline{Y}^{-2}\,\theta^2 C_{11}\,(C_{11} + C_{20}) \tag{2.2}$$

upto $o(n^{-2})$.$V(t_1)$ is the variance of t_1. Its exact expression has been obtained by Robson as

$$\lim_{N \to \infty} V(t_1) = \frac{\overline{Y}^2}{n}\left[(C_{02} + 2C_{11} + C_{20}) + \frac{(C_{20}C_{02} + C_{11}^2)}{(n-1)}\right] \tag{2.3}$$

If we assume that the population is sufficiently large, so that f.p.c. can be ignored, 2.1 and (2.2) simplify to

$$B(t_2) = (\overline{Y}/n^2)\,(C_{21} - C_{11}C_{20}) \tag{2.4}$$

and

$$M(t_2) = \frac{\overline{Y}^2}{n}\left[(C_{02} + 2C_{11} + C_{20}) + \frac{(C_{20}C_{02} + C_{11})}{(n-1)} + \frac{2C_{11}\,(C_{11} + C_{22})}{n}\right] \tag{2.5}$$

3. *Efficiency Comparisions:*

From (2.4) it is clear that the proposed estimator is unbiased upto $o(n^{-1})$. Comparing (2.5) with (2.3) we have

$$M(t_2) < V(t_1) \text{ if } C_{11}(C_{11} + C_{20}) < 0. \tag{3.1}$$

It is well known that the usual product estimator is useful for the case of negatively correlated characteristics (*i.e.* $C_{11}<0$).

Therefore condition (3.1) is fulfilled if

$$C_{11} + C_{20} > 0 \tag{3.2}$$

For comparing efficiency of t_2 with $\bar{Y}_p$, we have

$$M(\bar{Y}_p) = \frac{\bar{Y}^2}{n}\left[(C_{02} + 2C_{11} + C_{20}) + \frac{(C_{20}C_{02} + 2C_{11}^2)}{n}\right] \tag{3.3}$$

From (2.3) and 3.3 it can easily be seen that

$$V(t_1) < M(\bar{Y}_p) \text{ if } -1 < \rho < -1/(n-2)^{1/2}, \quad n > 2. \tag{3.4}$$

ρ is correlation coefficient between y and x. Combining (3.2) and (3.4), we have

$$M(t_2) < V(t_1) < M(\bar{Y}_p) \text{ if}$$

$$-(C_{20}/C_{02})^{1/2} < \rho < -1/(n-2)^{1/2}, \quad n > 2. \tag{3.5}$$

In case $C_{20} = C_{02}$, this condition reduces to (3.4) which itself reduces to (1.2) if n=6.

4. *Conclusion*

If a sample of size n>6 is taken from a large population and if a highly negative correlated auxiliary characteristics is selected having c.v. approximately equal to the c.v. of main characteristics, the proposed estimator will be more efficient than the existing estimators.

5. *Numerical Illustration:*

Consider the data from Kortelainen and Mannio [1] where water samples were collected at a depth of one meter from 78 forest lakes located in different parts of Finland and the behaviour of total organic carbon present in the water (y) with the watershed area in lakes (%) (x) was observed. For this data

$$\bar{Y} = 10.9, \quad \bar{X} = 14.0$$
$$C_{02} = 0.38925121, \quad C_{02} = 0.4225$$
$$\rho = -0.7690, \quad C_{11}/C_{2}O = -0.7381$$

Relative efficiency (%) of various estimators with respect to $\bar{y}$ with varying sample sizes has been given in the following table.

Estimator	n=5	n=10	n=25	n=50
$\bar{Y}_p$	149.81	173.82	192.32	199.39
t_1	153.57	179.32	195.66	201.29
t_2	162.42	185.17	198.41	202.74

We see that the proposed estimator is more efficient than existing estimators.

ACKNOWLEDGEMENT

Author is thankful to the referee for this useful suggestions.

REFERENCES

[1] Kortelainen, P. and Mannio, J., 1988. Natural and Anthropogenic acidity sources for finish lakes. *Jour. of Water, Air and Soil Pollution*, **42**, **(3&4)**, 341-352.

[2] Robson, D.S., 1957 : Application of multivariate polykas to the theory of unbiased ratio type estimators. *Jour. Amer. Stat. Assoc.* **52**, 511-522.

[3] Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok C., 1984. Sampling theory of surveys applications. ISAS Publication, New Delhi.