Jour. Ind. Soc. Ag. Statistics
Vol. XXXVIII, No. 3 (1986), pp. 403-409

ON THE USE OF POPULATION VARIANCE IN ESTIMATING MEAN

SHARAD BHATNAGAR
Cotton Research Station, HAU, Sirsa

(Received : January, 1985)

SUMMARY

A general class of estimators for population mean has been discussed alongwith their efficiency, under the assumption of the knowledge of population variance. The extent of bias involved under different situations has also been worked out.

Keywords : Population variance, meansquared error, Bias, negatively skewed population, symmetrical populations.

Introduction

The prior knowledge of population variance plays an important role in planning of a survey and in quality control techniques in industries. Such a knowledge may hail from long association with the experimental material or empirical evidence gathered from repeated experiments or other studies.

Assuming the knowledge of population variance and coming out of the class of unbiased estimators for population mean, Upadhyaya and Srivastava [4] proposed an estimator and studied its large sample properties which led Srivastava and Bhatnagar [1] to forward a family of estimators. Later, Upadhyaya and Singh [3] presented an estimator which had the same mean squared error as that of Upadhyaya & Srivastava [4] but smaller bias in large samples. This paper presents a general class of estimators and discusses their efficiency.

Estimators and their Properties

When σ^2 is known, Srivastava and Bhatnagar [1] proposed the class of estimators t_{kg} of mean $\bar{Y}$:

$$t_{kg} + \bar{y}\left(1 + \frac{k\sigma^2}{n\bar{y}^2 + g\sigma^2}\right) \tag{2.1}$$

where k and g are scalars specifying the estimator.

Setting $g = 0$ and $k = 1$, we obtain the estimator t_{10} of Upadhyaya and Srivastava [4] while putting $g = 1$ and $k = 1$ yields t_{11} proposed by Upadhyaya and Singh [3].

Let γ_1 and γ_2 be the Pearson's measures of skewness and kurtosis in the population and $\theta = \gamma_1/C^{1/2}$ where $C = \sigma^2/\bar{Y}^2$. The relative bias, to order $0(n^{-2})$ and the relative mean squared error, to order $0(n^{-3})$ of t_{kg} are as follows :

$$RB(t_{kg}) = \left[1 + (1-g)\frac{C}{n}\right]\frac{kC}{n} \tag{2.2}$$

$$RM(t_{kg}) = \frac{C}{n} + k(k-2)\frac{C^2}{n^2} + \delta_{kg}\frac{kC^3}{n^3} \tag{2.3}$$

where

$$\delta_{kg} = 2\theta - 6(1-g) + k(3-2g). \tag{2.4}$$

(For Proof see Appendix).

It is observed from (2.3) that both t_{10} and t_{11} have identical mean squared errors to order $0(n^{-2})$ as observed by Upadhyaya and Singh [3]: they differ in terms of order $0(n^{-3})$. Thus

$$RM(t_{11}) - RM(t_{10}) = (\delta_{11} - \delta_{10})\frac{C^3}{n^3} = \frac{4C^3}{n^3} \tag{2.5}$$

or

$$RM(t_{11}) > RM(t_{10})$$

so that the estimator t_{10} has a smaller relative mean squared error than the estimator t_{11}.

From (2.3), we see that the estimator t_{kg} will dominate over the conventional estimator $\bar{y}$ with respect to mean squared error if

$$0 < k < 2; \quad g < \frac{3(2-k) - 2\theta}{2(3-k)} \tag{2.6}$$

For negatively skewed population ($\theta < 0$), this condition holds as long as g is negative. Thus, t_{kg} will provide an improved estimator of population mean if k is positive but less than 2. Larger gain for negatively skewed population is achieved when g is taken negative.

We propose the following class of estimators of mean

$$t^*_{pq} = \bar{y} + \frac{p\bar{y}\,\sigma^2}{(n\bar{y}^2 + \sigma^2)} + \frac{q\bar{y}\,\sigma^4}{(n\bar{y}^2 + \sigma^2)^2} \tag{2.7}$$

Supposing p and q to be fixed scalars, the relative bias to order $0(n^{-2})$ and the relative mean squared error to order $0(n^{-3})$ of the estimator t^*_{pq} are as follows :

$$RB(t^*_{pq}) = \left(p + q\frac{C}{n}\right)\frac{C}{n} \tag{2.8}$$

$$RM(t^*_{pq}) = \frac{C}{n} + p(p-2)\frac{C^2}{n^2} + \delta^*_{pq}\frac{PC^3}{n^3} \tag{2.9}$$

where

$$\delta^*_{pq} = 2\theta + p + 2q - 6q/p. \tag{2.10}$$

From (2.8), we observe that the relative bias to order $0(n^{-2})$ vanishes if $p = -q(C/n)$ while from (2.9), we see that the estimator t^*_{pq} dominates over the conventional estimator $\bar{y}$ if

$$0 < p < 2 \tag{2.11}$$

$$\text{and} \quad q > \frac{(2\theta + p)}{2(3-p)}\,p \tag{2.12}$$

Thus, we observe that with respect to mean squared error criterion to order $0(n^{-2})$, the estimator t^*_{pq} with $0 < p < 2$ dominates over $\bar{y}$ for all values of q according to inequality (2.11). Larger gain is found when (2.12) hold. For instance, for symmetrical populations we may choose q to exceed $p^2/2(3-p)$. The same continues to hold true even when the population is negatively skewed.

It may be pointed out that the second term in the relative mean squared error of the estimator t^*_{pq} attains its minimum at $p = 1$ so that the estimator t^*_{10} of Upadhyaya and Singh [3] is an optimum estimator. Substituting $p = 1$ in (2.9) yields the following expression

$$\frac{C}{n} - \frac{C^2}{n^2} + (2\theta + 1 - 4q)\frac{C^3}{n^3} \tag{2.13}$$

implying that large reduction in the relative mean squared error is obtained when

$$q > \tfrac{1}{4}(1 + 2\theta) \tag{2.14}$$

which reduces to the following condition for symmetric distribution $(\theta = 0)$:

$$q > \tfrac{1}{4} \tag{2.15}$$

If we compare the relative mean squared errors of the estimators t^*_{po} and t^*_{pq}, we find

$$RM(t^*_{p0}) - RM(t^*_{pq}) = (\delta^*_{p0} - \delta^*_{pq}) \frac{pC^3}{n^3} = 2q(3-p) \frac{C^3}{n^3} \tag{2.16}$$

whence it follows that for the positive values of the scalars the estimator t^*_{pq} has the smaller relative mean squared error than the estimator t^*_{p0} because the optimum value of $p = 1$. So the addition of the term $\bar{y}\sigma^4/(n\bar{y}^2 + \sigma^2)^2$ in the estimator t^*_{10} of Upadhyaya and Singh [3] is more useful for reducing the relative mean squared error.

ACKNOWLEDGEMENT

Author is thankful to the referees for their useful suggestions.

REFERENCES

[1] Srivastava, V. K. and Bhatnagar, S. (1983) : A revisit to the use of coefficient of variation in estimating mean, *Jour. Ind. Soc. Agri. Stat.*, 35 (1) : 67-71.

[2] Sukhatme, P. V. and Sukhatme, B. V. (1970) : *Sampling Theory of Surveys with Application*, Asia Publishing House.

[3] Upadhyaya L. N. and Singh, H. P. (1984) : Estimation of mean when population variance is known, *Jour. Ind. Soc. Agric. Stat.*, 36 (2) : 19-22.

[4] Upadhyaya L. N. and Srivastava, S. R. (1976) : An efficient estimators of mean when population variance is known, *Jour. Ind. Soc. Agri. Stat.*, 28 (1) : 99-102.

APPENDIX

In order to derive the expressions for the relative bias and relative mean squared error of the estimators t_{kg}, we write

$$u = \frac{\bar{y} - \bar{Y}}{\bar{Y}} \tag{A.1}$$

It is easy to verify that

$$\begin{aligned} E(u) &= 0 \\ E(u^2) &= \frac{C}{n} \\ E(u^3) &= \frac{C^2\theta}{n^2} \\ E(u^4) &= 3\,\frac{C^2}{n^2} + 0(n^{-3}) \end{aligned} \tag{A.2}$$

From (2.1) and (A.1) we have

$$\frac{t_{kg} - \bar{Y}}{\bar{Y}} = u + \frac{kC}{n}\,(1 + u)\left[1 + 2u + u^2 + g\,\frac{C}{n}\right]^{-1} \tag{A.3}$$

Expanding the expression in square brackets on the right hand side and retaining terms to order $0(n^{-5/2})$, we find

$$\frac{t_{kg} - \bar{Y}}{\bar{Y}} = e_{-1/2} + e_{-1} + e_{-3/2} + e_{-2} + e_{-5/2} \tag{A.4}$$

where

$$\begin{aligned} e_{-1/2} &= u \\ e_{-1} &= k\,\frac{C}{n} \\ e_{-3/2} &= -k\,\frac{C}{n}\,u \\ e_{-2} &= k\,\frac{C}{n}\left(u^2 - g\,\frac{C}{n}\right) \\ e_{-5/2} &= k\,\frac{C}{n}\left(-u^3 + 3g\,\frac{C}{n}\,u\right) \end{aligned} \tag{A.5}$$

Here the suffices of e indicate the order in magnitude. From (A.2) and (A.5), we observe that

$$E(e_{-1/2}) = 0 \tag{A.6}$$

$$E(e_{-1}) = k\,\frac{C}{n} \tag{A.7}$$

$$E(e_{-3/2}) = 0 \tag{A.8}$$

$$E(e_{-2}) = k(1-g)\,\frac{C^2}{n^2} \tag{A.9}$$

Similarly, we have

$$E(e^2_{-1/2}) = \frac{C}{n} \tag{A.10}$$

$$E(e_{-1/2}\, e_{-1}) = 0 \tag{A.11}$$

$$E(e^2_{-1}) = k^2\,\frac{C^2}{n^2} \tag{A.12}$$

$$E(e_{-1/2}\, e_{-3/2}) = -k\,\frac{C}{n}\,E(u^2) = k\,\frac{C^2}{n^2} \tag{A.13}$$

$$E(e_{-1/2}\, e_{-2}) = k\,\frac{C}{n}\left[E(u^3) - g\,\frac{C}{n}\,E(u)\right] = k\,\frac{C^3\theta}{n^3} \tag{A.14}$$

$$E(e_{-1}\, e_{-3/2}) = -k^2\,\frac{C^2}{n^2}\,E(u) = 0 \tag{A.15}$$

$$E(e^2_{-3/2}) = k^2\,\frac{C^2}{n^2}\,E(u^2) = k^2\,\frac{C^3}{n^3} \tag{A.16}$$

$$E(e_{-1/2}\, e_{-5/2}) = k\,\frac{C}{n}\left[-E(u^4) + 3g\,\frac{C}{n}\,E(u^2)\right] = 3k(g-1)\,\frac{C^3}{n^3} \tag{A.17}$$

$$E(e_{-1}\, e_{-2}) = k^2\,\frac{C^2}{n^2}\left[E(u^2) - g\,\frac{C}{n}\right] = k^2(1-g)\,\frac{C^3}{n^3} \tag{A.18}$$

Where terms of higher order of smallness than $0(n^{-3})$ have been neglected.

Thus, the relative bias of t_{kg} to order $0(n^{-2})$ is

$$RB(t_{kg}) = E(e_{-1/2}) + E(e_{-1}) + E(e_{-3/2}) + E(e_{-2}) \tag{A.19}$$

while the relative mean squared error to order $0(n^{-3})$ is

$$RM(t_{kg}) = E(e^2_{-1/2}) + 2E(e_{-1/2}\, e_{-1}) + E(e^2_{-1} + 2e_{-1/2}\, e_{-3/2}) + 2E(e_{-1/2}\, e_{-2} + e_{-1}\, e_{-3/2}) + E(E^2_{-3/2} + 2e_{-1}\, e_{-5/2} + 2e_{-1}\, e_{-2}) \qquad (A.20)$$

Substituting (A.6) – (A.9) in (A.19) and (A.10) – (A.18) in (A.20), we get the results (2.2) and (2.3) after a little algebraic manipulations.

Proceeding in the same way as indicated for t_{kg}, the results (2.8) and (2.9) can be easily derived.