A REVISIT TO THE USE OF COEFFICIENT OF VARIATION IN ESTIMATING MEAN

By

V.K. Srivastava[1] & Sharad Bhatnagar[2]

(Received : October, 1980)

1. Introduction :

The sample mean $\bar{y}$ is the minimum variance unbiased estimator of $\bar{Y}$, the mean of a population with finite variance σ^2. If one is prepared to sacrifice the unbiasedness property, improved estimators can be obtained. One such estimator was proposed by Searles [1] assuming $C=\sigma^2/\bar{Y}^2$, the relative variance, to be known. When C is not known, a simple alternative is to estimate it from the sample. This led Srivastava [2] to formulate two estimators of $\bar{Y}$. Keeping in view the form of his estimators, we can propose the following family of estimators :

$$t_{kg}=\bar{y}\left(1+\frac{ks^2}{n\bar{y}^2-gs^2}\right)$$

where k and g are the characterizing scalars and s^2 is an unbiased estimator of σ^2.

Srivastava [2] studied two particular cases. viz., t_{-11} and t_{-10}. If we put $g=-k$, we obtain the estimator t_{-kk} considered by Thompson [3] while setting $g=0$ yields the estimator t_{ko} studied by Upadhyaya and Srivastava [4]. Thus the estimator t_{kg} provides a unified type of treatment and the analysis of its properties may help in the development of possibly more efficient estimators for population mean $\bar{Y}$.

When σ^2 is known, we can define the following family of estimators on the pattern of t_{kg} :

$$t^*_{kg}= \bar{y}\left(1+\frac{k\sigma^2}{n\bar{y}^2+g\sigma^2}\right)$$

1. Lucknow University, Lucknow

2. Cotton Research Station, Haryana Agricultural University, Sirsa.

A comparison of the estimators t_{kg} and t^*_{kg} may furnish an idea of the change in properties attributable to lack of knowledge of population variance σ^2. The next Section gives the relative bias, to order $0(n^{-2})$, and relative mean squared error, to order $0(n^{-3})$, of the both the estimators t_{kg} and t^*_{kg}. Some implications and analysis of these results are presented.

2. THE PRINCIPAL RESULTS

Let γ_1 and γ_2 be the Pearson's measures of skewness and kurtosis in the population and $\theta=\gamma_1/C^{\frac{1}{2}}$. Assuming the characterizing scalers k and g to be nonstochastic, it is easy to see[1] that the relative bias of t_{kg} to order $0(n^{-2})$ is

$$RB\,(t_{kg})=\left[1+\frac{(1-\theta-g)C}{n}\right]\frac{kC}{n} \qquad ...(1)$$

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

$$RM(t_{kg})=\frac{C}{n}+[2(\theta-1+k]\frac{KC^2}{n^2}+\delta_{kg}\frac{kC^3}{n^3} \qquad ...(2)$$

$$\delta_{kg}=2g(3-2\theta)+8\theta-2\frac{\gamma_2}{C}-6-k\left(2g+4\theta-3-\frac{\gamma_2+2}{C}\right) \qquad ...(3)$$

It is seen from (2) that both the estimators t_{-11} and t_{-10} have identical mean squared errors to order $0(n^{-2})$ as observed by Srivastava [2]; they differ with respect to terms of order $0(n^{-3})$. Thus we have

$$RM(t_{-10})-RM(t_{-11})=(\delta_{-11}-\delta_{-10})\frac{C^3}{n^3}=4(2-\theta)\frac{C^3}{n^3} \qquad ...(4)$$

which is positive if $\theta<2$. This implies that t_{-11} will have smaller mean squared error than t_{-10} until the population is highly positively skewed. At least for all negatively skewed and symmetrical populations, t_{-11} will definitely be better than t_{-10}.

From (2) we see that the estimator t_{kg} will dominate over the conventional estimator $\bar{y}$ with respect to mean squared error if.

$$0<k<2(1-\theta)\,;\; g<g^* \qquad \text{for } \theta<1 \qquad ...(5)$$

$$2(1-\theta)<k<0\,;\; g>g^* \qquad \text{for } \theta>1 \qquad ...(6)$$

where

$$g^*=1+\frac{2\frac{\gamma_2}{C}-4\theta-k\left(\frac{\gamma_2+2}{C}-4\theta+1\right)}{2\,(3-2\theta-k)} \qquad ...(7)$$

For populations having symmetric distribution ($\theta=0$), all estimators t_{kg} with $0<k<2$ dominate over $\bar{y}$ for all values of g. Larger gain is achieved when

$$g<1+\frac{2\frac{\gamma_2}{C}-k\left(\frac{\gamma_2+2}{C}+1\right)}{2(3-k)} \quad \ldots(8)$$

which reduces to the following condition for mesokurtic populations ($\gamma_2=0$) :

$$g<1-\frac{(C+2)k}{2(3-k)C} \qquad [0<k<2] \quad \ldots(9)$$

The condition (8) is satisfied so long as $g\leqslant 1$ and $0<k<2$. This suggests that for normal populations if we choose g less than 1 and k between 0 and 2, larger gain in efficiency of the estimator t_{kg} over $\bar{y}$ are expected according to mean squared error critarian to order $0(n^{-3})$.

If we restrict our attention to the class of estimators t_{-kk} considered by Thompson [3], we observed that t_{-kk} dominates over $\bar{y}$ for symmetric population when $k>0$. Larger gains are expected if

$$2k^2+k\left(3-\frac{\gamma_2+2}{C}\right)-6-2\frac{\gamma_2}{C}>0 \quad \ldots(10)$$

which reduces to the following conditions for normal populations :

$$2k^2+k\left(3-\frac{2}{C}\right)-6>0\ ;\ k>0. \quad \ldots(11)$$

The above condition holds so long as C is greater than ${}^{2k}/(2k^2+3k-6)$.

Similarly, confining attention to the class of estimators t_{ko} envisaged by Upadhyaya and Srivastava [4], it is easy to verify that the estimator t_{ko} dominates over $\bar{y}$ for normal populations when k lies between 0 and 2. In this context, it may be pointed out that t_{ko} does not possess any finite moment for normal population and therefore the expressions for relative bias and relative mean squared error of t_{ko} as obtained from (1) and (2) by setting $g=0$ should be interpreted carefully. They are asymptotic by nature and hence are subject to the usual qualifications as to their value and interpretation. However, the probability associated with the negativeness of $\bar{y}$ is usually negligible in many practical situations and therefore the approximations are reasonably good. They may be poor when the probability of $\bar{y}$ being negative is appreciable.

The relative bias, to order $0(n^{-2})$, and relative mean squared error, to order $0(n^{-3})$, of t^*_{kg} are given by

$$RB(t^*_{kg}) = \frac{kC}{n}\left[1 + \frac{(1-g)C}{n}\right] \quad ...(12)$$

$$RM(t^*_{kg}) = \frac{C}{n} + \frac{k(k-2)C^2}{n^2} + \delta^*_{kg}\frac{kC^3}{n^3} \quad ...(13)$$

where

$$\delta^*_{kg} = 2\theta - 6(1-g) + k(3-2g). \quad ...(14)$$

Comparing t_{kg} and t^*_{kg}, it is observed from (1) and (12) that both the estimators have identical bias to the order of our approximation for symmetrical populations.

From (2) and (13), we find

$$RM(t_{kg}) - RM(t^*_{kg}) = \frac{2k\theta C^2}{n^2} + (\delta_{kg} - \delta^*_{kg})\frac{kC^3}{n^3} \quad ...(15)$$

which may furnish an idea of the change in mean squared error, to the order of our approximation, attributable to replacement of σ^2 in t^*_{kg} by its unbiased estimator s^2 to yield t_{kg}.

From the first leading term on the right hand side of (15), it is interesting to note thak k can be so suitably chosen for asymmetric populations that t^*_{kg} is better than t_{kg}. For instance, if we take k to be nega tive for negatively skewed populations and k to be positive for positively skewed populations than t^*_{kg} will have smaller mean squared error, at least to order $0(n^{-2})$, than t_{kg}. This implies that there could be situations, viz., ($k<0$; $\theta>0$) or ($k>0$; $\theta<0$) where it may be fruitful to use s^2 despite the availability of σ^2. However, for symmetric populations we have

$$RM(t_{kg}) - RM(t^*_{kg}) = k[(\gamma_2+2)k - 2\gamma_2]\frac{C^2}{n^3} \quad ...(16)$$

Since (γ_2+2) is always positive, the right hand side of (16) is positive implying that lack of knowledge of population variance σ^2 increases the mean squared error at least for symmetrical populations and the increase precipitates in the term of order $0(n^{-3})$ when k lies between 0 and $2\gamma_2/(\gamma_2+2)$; it is positive for platykurtic populations ($\gamma_2<0$) and negative for leptokurtic populations ($\gamma_2>0$). No such constraint on k is needed for mesokurtic populations ($\gamma_2=0$).

REFERENCES

[1] Searles, D.T. (1964) : 'The Utilization of a Known Coefficient of Variation in Estimation Procedure' *Journ. Amer. Stat. Assn.*, Vol. 59, pp. 1225-6.

[2] Srivastava, V.K. (1974) : 'On the Use of Coefficient of Variation in Estimating Mean' *Jour. Ind. Soc. Agri. Stat.* Vol. 26, pp. 33-36.

[3] Thompson, J.R. (1968) : 'Some Shrinkage Techniques for Estimating the Mean'. *Journ. Amer. Stat. Assn.*, Vol. 63, pp. 113-23.

[4] Upadhyaya, L.N. & S.R. Srivastava (1976) : 'A Note on the Use of Coefficient of Variation in Estimating Mean' *Journ. Ind. Soc. Agri. Stat.*, Vol. 28, pp. 97-9.