ON THE USE OF COEFFICIENT OF VARIATION IN ESTIMATING MEAN

By
V. K. Srivastava
Department of Mathematics, Banaras Hindu University, Varanasi.

(Received in March, 1973; Accepted in January, 1974)

Searles [2] considered the use of information pertaining to the coefficient of variation in developing an estimator for population mean with smaller mean squared error. Suppose we have a random sample of size n drawn from a population with mean $\bar{Y}$ and variance σ^2. The estimator proposed by Searles [2] is

$$t = \frac{\bar{y}}{1+\dfrac{c}{n}} \qquad \ldots(1)$$

where $\bar{y} = \dfrac{1}{n} \sum\limits_{i} y_i$ and $c = \dfrac{\sigma^2}{\bar{Y}^2}$.

The bias of the estimator t is

$$B(t) = -\frac{c}{n+c}\bar{Y} \qquad \ldots(2)$$

and the mean squared error

$$M(t) = \frac{\sigma^2}{n+c} = \frac{c}{n+c}\bar{Y}^2 \qquad \ldots(3)$$

which will be smaller than the sampling variance of the conventional unbiased estimator $\bar{y}$.

In practice, c is rarely known so that the estimator t is of no practical utility. However, one may have a figure fairly close to the true value of the coefficient of variation from one's long association with experimental material or from other empirical investigations or from some extraneous source. Often, the coefficient of variation may exhibit a stability in repeated experiments, and it may be possible to make a reasonable guess. Suppose this apriori or guessed value of c is c_g. Searles [2] seems to recommend the replacement of c in

(1) by c_g so that the estimator is now

$$t_g = \frac{\bar{y}}{1 + \frac{c_g}{n}} \qquad \text{.. (4)}$$

but this disturbs the optimal property. For instance, the bias of t_g is

$$B(t_g) = -\frac{c_g}{n + c_g}\,\bar{Y} \qquad \text{...(5)}$$

and the mean squared error

$$M(t_g) = \frac{(n_c + c_g^2)}{(n + c_g)^2}\,\bar{Y}^2 \qquad \text{...(6)}$$

which will be larger than the mean squared error of t.

Interesting to note is that the mean squared error of t_g will be smaller than the sampling variance of $\bar{y}$ only when

$$c > \frac{c_g}{2 + \frac{c_g}{n}} \text{ or } c_g < \frac{2c}{1 - \frac{c}{n}} \qquad \text{.. (7)}$$

which may not always hold.

It is equally illuminating to examine the case when no apriori information about the coefficient of variation is available and a consistent estimator is used for the purpose. One simple choice is then

$$t_e = \frac{\bar{y}}{1 + \frac{c_e}{n}} \qquad \text{...(8)}$$

where $c_e = \frac{s^2}{\bar{y}^2}$ and $s^2 = \frac{1}{n-1}\sum_i (y_i - \bar{y})^2$.

Let us investigate the large-sample properties of this estimator.

Write

$$\begin{aligned} \bar{y} &= \bar{Y} + u \\ s^2 &= \sigma^2 + v \end{aligned} \qquad \text{...(9)}$$

where u and v are of order $O_p(n^{-1/2})$ with $E(u) = E(v) = 0$.

Now, we have

$$\begin{aligned} t_e &= \frac{\bar{y}^3}{\bar{y}^2 + \frac{s^2}{n}} \\ &= (\bar{Y} + u)^3 \left[(\bar{Y} + u)^2 + \frac{1}{n}(\sigma^2 + v) \right]^{-1} \\ &= \bar{Y}\left(1 + \frac{u}{\bar{Y}}\right)^3 \left[1 + \frac{2u}{\bar{Y}} + \frac{u^2}{\bar{Y}^2} + \frac{c}{n} + \frac{v}{n\bar{Y}^2}\right]^{-1} \end{aligned} \qquad \text{...(10)}$$

Expanding the right hand side and retaining terms to order $O_p(n^{-2})$ we find the sampling error

$$t_e - \bar{Y} = \bar{Y}\left[\frac{u}{\bar{Y}} - \frac{c}{n} + \frac{1}{n}\left(\frac{cu}{\bar{Y}} - \frac{v}{\bar{Y}^2}\right) + \frac{1}{n}\left(\frac{c^2}{n} - \frac{cu^2}{\bar{Y}^2} + \frac{uv}{\bar{Y}^3}\right)\right] \quad ...(11)$$

Thus, the bias and mean squared error, to order $O(n^{-2})$, of t_e are

$$B(t_e) = -\frac{c}{n}\left(1 - \frac{1}{n}\sqrt{\beta_1 c}\right)\bar{Y}$$

$$M(t_e) = \frac{\sigma^2}{n}\left[1 + \frac{1}{n}\left(3c - 2\sqrt{\beta_1 c}\right)\right] \quad ...(12)$$

where $\sqrt{\beta_1} = \frac{\mu_3}{\sigma^3}$, μ_3 being the third central moment.

In order to compare t and t_e, we observe that to order $O(n^{-2})$

$$B(t) = -\frac{c}{n}\left(1 - \frac{c}{n}\right)\bar{Y}$$

$$M(t) = \frac{\sigma^2}{n}\left(1 - \frac{c}{n}\right) \quad ...(13)$$

The expressions (12) and (13) may furnish an idea of the change in bias and mean squared error, to the order of our approximation, attributable to replacement of c in (1) by c_e.

Further, the mean squared error (12) will be smaller than the sampling variance of $\bar{y}$ provided

$$\sqrt{\frac{\beta_1}{c}} > \frac{3}{2} \quad ..(14)$$

For normal population, $\beta_1 = 0$ so that the estimator (8) will not only be biased but will also have larger mean squared error. This is in fact true for all populations having a symmetrical distribution.

Another way of obtaining an estimator of c is to put unbiased estimators in numerator and denominator, viz.,

$$c_e' = \frac{s^2}{\bar{y}^2 - \frac{s^2}{n}} \quad ..(15)$$

which when replaces c in (1) leads to the estimator

$$t_e' = \bar{y} - \frac{s^2}{n\bar{y}}. \quad ..(16)$$

The bias and mean squared error, to order $O(n^{-2})$, of t_e' can be derived in the same way as indicated above :

$$B(t_e')=-\frac{c}{n}\left[1+\frac{1}{n}\left(c-\sqrt{\beta_1 c}\right)\right]\bar{Y}$$

$$M(t_e')=\frac{\sigma^2}{n}\left[1+\frac{1}{n}\left(3c-2\sqrt{\beta_1 c}\right)\right] \qquad \text{...(17)}$$

These expressions can be compared with (12), (13) and variance of $\bar{y}$.

References

1. Kendall, M.G. and A. Stuart : *The Advanced Theory of Statistics*, Vol. I, Charles Griffin & Co., London.
2. Searles, D.T. : "The Utilization of a Known Coefficient of Variation in Estimation Procedure" Journ. Amer. Stat. Assn,, Vol. 59, 1964, pp. 1225-6.