

Estimation of Population Variance using Difference Type Estimator

Sheela Misra and R. Karan Singh

Department of Statistics, Lucknow University, Lucknow-226007

(Received September 22, 2003)

Abstract : An estimator for the estimation of population variance using difference type estimator has been proposed, its bias and mean square error (MSE) are found and a comparative study with the conventional unbiased estimator is provided. An empirical example is also included as an illustration.

1. Introduction

Considering a finite population of N unit, let Y_i and X_i ($i = 1, 2, \dots, N$) be the values on the i^{th} unit of the population according to the study variable y and the auxiliary variable x respectively. For simple random sample without replacement of size n with sample values $y_1, y_2, \dots, y_n$ on y and $x_1, x_2, \dots, x_n$ on x . Let $\bar{y}$ and $\bar{x}$ be the usual unbiased sample mean estimators of the population means $\bar{Y}$ and $\bar{X}$ of y and x respectively.

Let

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

be the usual unbiased estimator of population variance

$$\sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$$

of the study variable y . Further, let

$$(1.1) \quad \mu_{rs} = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^r (X_i - \bar{X})^s$$

and

$$(1.2) \quad \beta_2(y) = \frac{\mu_{40}}{\mu_{20}^2} \quad \text{and} \quad \bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$$

and

$$\bar{Z} = \frac{1}{N} \sum_{i=1}^N Z_i, \quad y_i^2 = z_i, \quad Y_i^2 = Z_i.$$

For the estimation of finite population variance σ_y^2 , the proposed estimator using difference type estimator is

$$(1.3) \quad s_D^2 = \frac{1}{(n-1)} \left[\sum_{i=1}^n y_i^2 - n \left\{ \bar{y} + k \left(\frac{\bar{x}}{\bar{X}} - 1 \right) \right\}^2 \right]$$

where k is the characterising scalar determined by minimising the mean square error of s_D^2

2. Bias and MSE of s_D^2

Let $\bar{y} = \bar{Y} + e_0$, $\bar{x} = \bar{X} + e_1$, $\bar{z} = \bar{Z} + e_2$ so that $E(e_0) = E(e_1) = E(e_2) = 0$.

$$E(e_0^2) = \left(\frac{1}{n} - \frac{1}{N} \right) \frac{N}{N-1} \mu_{20}$$

$$E(e_1^2) = \left(\frac{1}{n} - \frac{1}{N} \right) \frac{N}{N-1} \mu_{02}$$

$$E(e_0 e_1) = \left(\frac{1}{n} - \frac{1}{N} \right) \frac{N}{N-1} \mu_{11}$$

$$E(e_2^2) = E \left(\frac{1}{n} \sum_{i=1}^n y_i^2 - \frac{1}{N} \sum_{i=1}^N y_i^2 \right)^2$$

$$= E(\bar{z} - \bar{Z})^2$$

$$= \left(\frac{1}{n} - \frac{1}{N} \right) s_z^2$$

with

$$s_z^2 = \frac{1}{N-1} \sum_{i=1}^N (z_i - \bar{z})^2$$

$$= \left(\frac{1}{n} - \frac{1}{N} \right) \frac{1}{N-1} \left\{ \sum_{i=1}^N z_i^2 - N \bar{z}^2 \right\}$$

$$= \left(\frac{1}{n} - \frac{1}{N} \right) \frac{1}{N-1} \left[\sum_{i=1}^N \left\{ (Y_i - \bar{Y}) + \bar{Y} \right\}^4 - N(\mu_{20} + \bar{Y}^2) \right]$$

$$= \left(\frac{1}{n} - \frac{1}{N} \right) \frac{N}{N-1} (\mu_{40} + 4\bar{Y}\mu_{30} + 4\bar{Y}^2\mu_{20} - \mu_{20}^2)$$

$$\begin{aligned}
 E(e_1 e_2) &= E(\bar{x} - \bar{X})(\bar{z} - \bar{Z}) \\
 &= \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{N-1} \sum_{i=1}^N (Z_i - \bar{z})(X_i - \bar{x}) \\
 &= \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{N-1} \sum_{i=1}^N Y_i^2 (x_i - \bar{x}) \\
 &= \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{N-1} \sum_{i=1}^N \{ (Y_i - \bar{y}) + \bar{y} \}^2 (X_i - \bar{x}) \\
 &= \left(\frac{1}{n} - \frac{1}{N}\right) \frac{N}{N-1} \sum_{i=1}^N (\mu_{21} + 2\bar{y}\mu_{11})
 \end{aligned}$$

Similarly,

$$E(e_0 e_2) = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{N}{N-1} (\mu_{30} + 2\bar{y}\mu_{20})$$

Now from (1.3), we have

$$\begin{aligned}
 s_D^2 &= \frac{n}{(n-1)} \left[\frac{1}{n} \sum Y_1^2 - \left\{ \bar{y} + k \left(\frac{\bar{x}}{\bar{X}} - 1 \right) \right\}^2 \right] \\
 &= \left(1 - \frac{1}{n} \right)^{-1} \left[\bar{z} - \left\{ \bar{y} + k \left(\frac{\bar{x}}{\bar{X}} - 1 \right) \right\}^2 \right] \\
 &= \left(1 - \frac{1}{n} \right)^{-1} \left[\bar{z} + e_2 - \left\{ (\bar{y} + e_0) + k \frac{e_1}{\bar{X}} \right\}^2 \right] \\
 &= \left(1 - \frac{1}{n} \right)^{-1} \left[\bar{z} + e_2 - \left\{ (\bar{y} + e_0)^2 + \frac{k^2 e_1^2}{\bar{X}^2} + \frac{2ke_1(\bar{y} + e_0)}{\bar{X}} \right\} \right] \\
 &= \left(1 - \frac{1}{n} \right)^{-1} \left[\bar{z} - \bar{y}^2 + e_2 - 2\bar{y}e_0 - e_0^2 - \frac{k^2}{\bar{X}^2} e_1^2 - \frac{2k\bar{y}}{\bar{X}} e_1 - \frac{2k}{\bar{X}} e_1 e_0 \right] \\
 (2.1) \quad &= \left(1 - \frac{1}{n} \right)^{-1} \left[\sigma_y^2 + e_2 - 2\bar{y}e_0 - \frac{2k\bar{y}}{\bar{X}} e_1 - e_0^2 - \frac{k^2}{\bar{X}^2} e_1^2 - \frac{2ke_1 e_0}{\bar{X}} \right]
 \end{aligned}$$

Bias up to order $O\left(\frac{1}{n}\right)$ is

$$\begin{aligned}
 (2.2) \quad \text{Bias } s_D^2 &= E(s_D^2) - \sigma_y^2 \\
 &= \frac{\sigma_y^2}{n} - E\left(e_0^2 + \frac{k^2}{\bar{X}^2} e_1^2 + \frac{2ke_1e_0}{\bar{X}}\right) \\
 &= \frac{\sigma_y^2}{n} - \left(\frac{1}{n} - \frac{1}{N}\right) \frac{N}{N-1} \left(\mu_{20} + \frac{k^2}{\bar{X}^2} \mu_{02} + \frac{2k}{\bar{X}} \mu_{11}\right)
 \end{aligned}$$

Further, mean square error of s_D^2 to terms of order of $O\left(\frac{1}{n}\right)$ is

$$\begin{aligned}
 \text{MSE } s_D^2 &= E\left(s_D^2 - \sigma_y^2\right)^2 \\
 &= E\left\{e_0^2 + 2\bar{Y}e_0 + \frac{2\bar{Y}k}{\bar{X}}e_1\right\}^2 \\
 &= \left(\frac{1}{n} - \frac{1}{N}\right) \frac{N}{N-1} \left[\mu_{40} + 4\bar{Y}\mu_{30} + 4\bar{Y}^2\mu_{20} - \mu_{20}^2 + 4\bar{Y}^2\mu_{20} + \frac{4\bar{Y}^2k^2}{\bar{X}^2}\mu_{02} \right. \\
 &\quad \left. - 4\bar{Y}(\mu_{30} + 2\bar{Y}\mu_{20}) - \frac{4\bar{Y}k}{\bar{X}}(\mu_{21} + 2\bar{Y}\mu_{11}) + \frac{8\bar{Y}^2k}{\bar{X}}\mu_{11}\right] \\
 &= \left(\frac{1}{n} - \frac{1}{N}\right) \frac{N}{N-1} \left[\mu_{40} - \mu_{20}^2 + \frac{4\bar{Y}^2k^2}{\bar{X}^2}\mu_{02} - \frac{4\bar{Y}k}{\bar{X}}\mu_{21}\right] \\
 (2.3) \quad &= \left(\frac{1}{n} - \frac{1}{N}\right) \frac{N}{N-1} \left[\mu_{20}^2\{\beta_2(y) - 1\} + \frac{4\bar{Y}k}{\bar{X}}\left(\bar{Y}\frac{k}{\bar{X}}\mu_{02} - \mu_{21}\right)\right]
 \end{aligned}$$

which is minimised for the optimum value of k to be

$$(2.4) \quad k_0 = \frac{\mu_{21}\bar{X}}{2\bar{Y}\mu_{02}}$$

and the minimum mean square error is given by

$$\begin{aligned}
 \text{MSE } (s_D^2)_{\min} &= \left(\frac{1}{n} - \frac{1}{N}\right) \frac{N}{N-1} \left[\{\beta_2(y) - 1\}\mu_{20}^2 - \frac{\mu_{21}^2}{\mu_{02}}\right] \\
 (2.5) \quad &= \text{MSE } (s_y^2) - \left(\frac{1}{n} - \frac{1}{N}\right) \frac{N}{N-1} \frac{\mu_{21}^2}{\mu_{02}}
 \end{aligned}$$

3. Concluding Remarks

(a) The optimum value minimising $MSE(s_D^2)$ is

$$(3.1) \quad k_0 = \frac{\mu_{21} \bar{X}}{2\bar{Y} \mu_{02}}$$

and the minimum mean square error of s_D^2 for the optimum value k_0 is given by

$$(3.2) \quad MSE(s_D^2)_{\min} = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{N}{N-1} \left[\mu_{20}^2 \beta_2(v) - 1 \right] - \frac{\mu_{21}^2}{\mu_{02}}$$

which, when N is large relative to n so that finite population correction terms may be ignored, reduces to

$$(3.3) \quad \frac{1}{n} \left[\mu_{20}^2 \beta_2(v) - 1 \right] - \frac{\mu_{21}^2}{\mu_{02}}$$

In case the optimum value k_0 in (3.1) or its guessed value being unknown, the alternative is to replace it by its estimate from sample values. Thus, for large N or simple random sampling with replacement, replacing μ_{21} , μ_{02} and $\bar{Y}$ by their unbiased estimators

$$\hat{\mu}_{21} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 (x_i - \bar{X}),$$

$$\hat{\mu}_{02} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2$$

and $\hat{\bar{Y}} = \bar{y}$ respectively, we get the estimated optimum value k_0 to be

$$(3.4) \quad \hat{k}_0 = \frac{\hat{\mu}_{21} \bar{X}}{2\bar{Y} \hat{\mu}_{02}}$$

in which case the estimator depending on estimated optimum value is

$$s_D^{*2} = \frac{1}{n-1} \left[\sum_{i=1}^n y_i^2 - n \left\{ \bar{y} + \hat{k}_0 \left(\frac{\bar{x}}{\bar{X}} - 1 \right) \right\}^2 \right]$$

which, following the method as employed for s_D^2 in section 2, attains the minimum mean square error given by (3.3).

(b) We know that

$$(3.5) \quad MSE(s_y^2) = \left(\frac{1}{n} - \frac{1}{N}\right) \left(\frac{N}{N-1}\right) \mu_{20}^2 \left\{ \beta_2(v) - 1 \right\}$$

which, for large N , reduces to

$$(3.6) \quad \frac{\mu_{20}^2}{n} \left\{ \beta_2(v) - 1 \right\}$$

Comparing (3.5) with (3.2), we see that the estimator s_D^2 has less mean square error than that of s_y^2 , that is, s_D^2 is more efficient than the conventional estimator s_y^2 in the sense of having lesser mean square error. Similar remarks follow for s_D^{*2} also.

4. An Empirical Study

From the data given in Cochran¹ dealing with paralytic polio cases in "placebo" (y) group and paralytic polio cases in "not included" group (x), the required value μ_{rs} are calculated and relative efficiency of estimator s_D^2 with respect to s_y^2 for estimating population variance σ_y^2 is worked out as follows :

We have,

$$\mu_{20} = 9.9528$$

$$\mu_{21} = 35.84$$

$$\mu_{02} = 10.84$$

$$\mu_{40} = 425.9$$

Using above values, we get

$$MSE(s_y^2) = \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{N}{N-1} \right) (326.8417)$$

$$MSE(s_D^2) = \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{N}{N-1} \right) (208.3449)$$

The relative efficiency of s_D^2 (or s_D^{*2}) with respect to s_y^2 is

$$\frac{MSE(s_y^2)}{MSE(s_D^2)} \times 100 \% = 156.8753 \%$$

showing that the proposed estimator s_D^2 (or s_D^{*2}) is better than s_y^2 in the sense of having lesser mean square error.

References

1. W. G. Cochran : Sampling Techniques (3rd edn.), Wiley Eastern Pvt. Ltd., New Delhi, (1977).
2. P. V. Sukhatme, B. V. Sukhatme, S. Sukhatme and C. Ashok : Sampling Theory of Surveys and Applications, Iowa State University Press, Iowa USA, (1984).