Estimation of Population Mean When the Population Mean of Auxiliary Character is Not Known

Manish Sharma[1,3] and Sharad Bhatnagar[2]
[1]SK University of Agril. Sciences & Technology,Jammu, INDIA
[2]CCSHAU, Hisar, Haryana, INDIA
[3] Corresponding author: Manish Sharma, e-mail: manshstat@gmail.com

Abstract

For estimation of population mean, a general class of estimators is proposed when the population mean of auxiliary character is not known. Its large sample properties have been studied particularly the relative bias and relative mean squared error. The relative efficiencies of proposed estimator are compared with mean, ratio, product and different product type estimators according to mean squared error criterion.

Key Words: Product estimator, Ratio estimator, Relative bias, Relative mean squared error.

1. Introduction

In sampling theory, the auxiliary information X related with the study variable Y is used for (i) stratification (ii) building of estimators such as ratio, regression, product estimator, etc. and (iii) in sample selection (unequal probability sampling). This article focuses on the use of auxiliary information in estimating population parameter of interest. Generally it is assumed that the population mean of the auxiliary variable $\overline{X}$ is known. When $\overline{X}$ is not known, then one employs double sampling in which $\overline{X}$ is first estimated from a larger sample, assuming that it is relatively economical to take a large sample on the auxiliary information. In biological experiments, samplings X and Y may call for the destruction of sampling units. In such situations it may not be feasible to sample auxiliary information for getting an estimate of $\overline{X}$. For such situations, a general class of estimators has been considered and its large sample properties are worked out. The properties of the proposed estimator are compared with the mean, ratio, product and product type estimators of population mean $\overline{Y}$ and conditions have been worked out when the proposed estimator beats them in terms of relative mean squared error.

2. Estimators and their properties

Consider a finite population of size N, the N units being distinct and identifiable. Let Y be the character of interest and X be the auxiliary variable related with Y. Let $\overline{Y}$ and $\overline{X}$ denote the population means of Y and X and $\overline{y}$ and $\overline{x}$ the samples means of Y and X respectively based on a random sample of size n. $\overline{y}$ and $\overline{x}$ are unbiased estimators of $\overline{Y}$ and $\overline{X}$, respectively. Assuming $\frac{N-n}{N} \cong 1$.

When the variables $\overline{y}$ and $\overline{x}$ are positively correlated, ratio estimator of mean is as follows:

$$\overline{y}_R = \frac{\overline{y}}{\overline{x}}\overline{X}. \tag{2.1}$$

When the variables $\overline{y}$ and $\overline{x}$ are negatively correlated, Robson (1957) proposed the product estimator of mean which was rediscovered by Murthy (1964) as

$$\bar{y}_P = \frac{\overline{yx}}{\bar{X}}. \tag{2.2}$$

When the population mean of the auxiliary variable $\bar{X}$ is not known, both the estimators (2.1) and (2.2) cannot be used.

Keeping this in view, a general class of estimators for mean has been considered as

$$t_{\alpha q} = \bar{y}\left[1 + \frac{q s_{xy}}{n\bar{x}\bar{y} + \alpha s_{xy}}\right] \tag{2.3}$$

where

$s_{xy} = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ and α and q are characterizing scalars.

It may be observed that setting q = 0 in the estimator (2.3) yields the conventional estimator ($\bar{y}$) of the population mean $\bar{Y}$ .The relative bias(RB) and relative mean squared error(RM) of the estimator $t_{\alpha q}$ up to order O(n^{-1}) and O(n^{-2}) respectively are worked out as

$$RB(t_{\alpha q}) = \frac{q}{n}C_{11} \tag{2.4}$$

$$RM(t_{\alpha q}) = \frac{C_{02}}{n} + \frac{q}{n^2}\left[qC_{11}{}^2 + 2C_{12} - 2C_{11}{}^2\right] \tag{2.5}$$

where

$$C_{ab} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{x_i - \bar{X}}{\bar{X}}\right)^a\left(\frac{y_i - \bar{Y}}{\bar{Y}}\right)^b$$

for non-negative integer a and b.

It is observed that both relative bias and relative mean squared error of the estimator $t_{\alpha q}$ do not depend upon α for the degree of our approximation. For bivariate normal population, the expression (2.5) is reduced to

$$RM(t_{\alpha q}) = \frac{C_{02}}{n} + \frac{q(q-2)C_{11}{}^2}{n^2} \quad . \tag{2.6}$$

3. Comparison with the mean, ratio, product and product type estimators

When $\bar{X}$ is not known, a simple solution is to ignore the auxiliary information and to estimate $\bar{Y}$ by $\bar{y}$ which is unbiased with relative variance (RV) as

$$RV(\bar{y}) = \frac{C_{02}}{n} \quad . \tag{3.1}$$

Comparing the relative mean squared error of the estimators $t_{\alpha q}$ with the $\bar{y}$ up to order O (n^{-1}), from (2.6) and (3.1), it has been observed that the relative mean squared errors of both the estimators are same for bivariate normal population.

Comparing the relative mean squared errors of the estimators $t_{\alpha q}$ with $\bar{y}$ up to order O (n^{-2}), under bivariate normal population, we have

$$RM(t_{\alpha q}) - RV(\bar{y}) = \frac{q}{n^2}\left[qC_{11}^2 - 2C_{11}^2\right] = q(q-2)C_{11}^2/n^2 \quad . \tag{3.2}$$

From expression (3.2) that the proposed estimator $t_{\alpha q}$ will be more efficient than $\bar{y}$, if

$$0 < q < 2. \tag{3.3}$$

When the auxiliary characteristic is positively correlated with the characteristic understudy, we used the ratio estimator (2.1).
The relative bias up to order O (n^{-1}) and relative mean squared error up to order O(n^{-2}) of the ratio estimator $\bar{y}_R$ are as follows

$$RB(\bar{y}_R) = \frac{1}{n}(C_{20} - C_{11}) \tag{3.4}$$

$$RM(\bar{y}_R) = \frac{1}{n}(C_{02} - 2C_{11} + C_{20}) + \frac{1}{n^2}\left\{2(2C_{21} - C_{12} - C_{30}) + 3(3C_{20}^2 - 6C_{20}C_{11} + 2C_{11}^2 + C_{20}C_{02})\right\}. \tag{3.5}$$

For bivarate normal population, the expression (3.5) reduces to

$$RM(\bar{y}_R) = \frac{1}{n}(C_{02} - 2C_{11} + C_{20}) + \frac{1}{n^2}[9C_{20}^2 - 18C_{20}C_{11} + 6C_{11}^2 + 3C_{20}C_{02}]. \tag{3.6}$$

Comparing the relative bias of the estimators $t_{\alpha q}$ and $\bar{y}_R$ from (2.4) and (3.4) gives

$$RB(t_{\alpha q}) - RB(\bar{y}_R) = \frac{C_{11}}{n}(q+1) - \frac{C_{20}}{n} \quad . \tag{3.7}$$

So, the estimator $t_{\alpha q}$ has smaller bias than $\bar{y}_R$ if

$$\rho\theta < 1 \quad \text{and} \quad q > 0.$$

Now, comparing the relative mean squared errors of estimators $t_{\alpha q}$ and $\bar{y}_R$, from(2.6)and(3.6)gives

$$RM(t_{\alpha q}) - RM(\bar{y}_R) = \frac{1}{n}(C_{20} - 2C_{11}) + \frac{1}{n^2}(q^2C_{11}^2 - 2qC_{11}^2 - 9C_{20}^2 + 18C_{20}C_{11} - 6C_{11}^2 - 3C_{20}C_{02}) \quad . \tag{3.8}$$

Using (3.8) it follows that the estimator $t_{\alpha q}$ will be more efficient than $\bar{y}_R$ up to order O(n^{-1}) if

$$\rho > \frac{1}{2\theta}$$

and O(n^{-2}) if

$$2 < q < 4$$

where $\theta = \left(\frac{C_{02}}{C_{20}}\right)^{1/2}$ and ρ is the coefficient of correlation between the study variable *Y* and auxiliary variable *X*.

When the auxiliary characteristic is negatively correlated with the characteristic understudy, Robson (1957) proposed the product estimator (2.2).

The relative bias to order O (n^{-1}) and relative mean squared error to order O (n^{-2}) of the product estimator $\bar{y}_P$ are given by

$$RB(\bar{y}_p) = \frac{C_{11}}{n} \tag{3.9}$$

$$RM(\bar{y}_p) = \frac{1}{n}(C_{02} + 2C_{11} + C_{20}) + \frac{1}{n^2}[2(C_{21} + C_{12}) + 2C_{11}^2 + C_{20}C_{02}]. \tag{3.10}$$

For bivarate normal population, the expression (3.10) simplifies to

$$RM(\bar{y}_p) = \frac{1}{n}(C_{02} + 2C_{11} + C_{20}) + \frac{1}{n^2}[2C_{11}^2 + C_{20}C_{02}] \quad . \tag{3.11}$$

By comparing the relative bias of the estimators $t_{\alpha q}$ and $\bar{y}_P$, from (2.4) and (3.9) we have

$$RB(t_{\alpha q}) - RB(\bar{y}_P) = \frac{C_{11}(q-1)}{n} \quad . \tag{3.12}$$

The estimator $t_{\alpha q}$ has smaller bias than $\bar{y}_P$ if

$$C_{11} < 0 \text{ and } q > 1.$$

For $q = 1$, the estimators $t_{\alpha q}$ and $\bar{y}_P$ have the same biases.

Now, comparing the relative mean square errors of the estimator $t_{\alpha q}$ and $\bar{y}_P$, from expression (2.6) and (3.11), we have

$$RM(t_{\alpha q}) - RM(\bar{y}_P) = -\frac{1}{n}(C_{20} + 2C_{11}) + \frac{1}{n^2}(q^2C_{11}^2 - 2qC_{11}^2 - 2C_{11}^2 - C_{20}C_{02}) \quad . \tag{3.13}$$

The estimator $t_{\alpha q}$ will be more efficient than $\bar{y}_P$ up to order O (n^{-1}) if

$$\rho > \frac{-1}{2\theta}$$

and O (n^{-2}) if

$$-1 < q < 3.$$

Robson (1957) developed the unbiased product estimator obtained by subtracting the unbiased estimate of the bias of $\bar{y}_P$ as follows

$$t^* = \frac{\overline{yx}}{\bar{X}} - \frac{1}{n}\frac{s_{xy}}{\bar{X}} \; . \tag{3.14}$$

The relative variance, up to order O (n^{-2}), of t^* is given by

$$RV(t^*) = RM(\bar{y}_P) - \frac{1}{n^2}\left[C_{11}^2 + 2(C_{21} + C_{12})\right] . \tag{3.15}$$

From (3.15), it is observed that the unbiased estimator t^* has smaller variance than the relative mean squared error of $\bar{y}_P$ provided

$$\left[C_{11}^2 + 2(C_{21} + C_{12})\right] > 0 \quad . \tag{3.16}$$

The condition (3.16) holds good for all symmetric bivariate distributions with $C_{21} = C_{12} = 0$.

Singh (1989) proposed an almost unbiased product estimator of $\bar{Y}$ as follows

$$t_1^* = \frac{\bar{y}\,\bar{x}}{\bar{X}}\left[1 + \frac{s_{xy}}{n\,\bar{x}\,\bar{y}}\right]^{-1} \quad . \tag{3.17}$$

The relative bias and relative mean squared error of the estimator t_1^*, up to O(n^{-2}) are given by

$$RB(t_1^*) = \frac{C_{11}^2}{n^2} \tag{3.18}$$

$$RM(t_1^*) = RV(t^*) \; . \tag{3.19}$$

So, the estimator t_1^* is unbiased up to O(n^{-1}) and has relative mean squared error equal to relative variance of the estimator t^*. It performs better than estimator $\bar{y}_P$ when the condition (3.16) holds.

Now, comparing the relative mean squared errors of the estimators t^* and $t_{\alpha q}$, from (2.5) and (3.15) under bivariate normal population, we have

$$RM(t_{\alpha q}) - RV(t^*) = -\frac{1}{n}(C_{20} + 2C_{11}) + \frac{1}{n^2}\left[q^2C_{11}^2 - 2qC_{11}^2 - C_{02}C_{20} - C_{11}^2\right] . \tag{3.20}$$

From (3.20), the estimator $t_{\alpha q}$ will be more efficient than t^* up to order O (n^{-1}) if

$$\rho > \frac{-1}{2\theta}$$

and O (n^{-2}) if

$$-1 < q < 3.$$

Further, Dubey (1993) proposed an unbiased product estimator as

$$t_2^* = \frac{\bar{y}\,\bar{x}}{\bar{X}} - \frac{s_{xy}}{n\bar{x}} \tag{3.21}$$

whose relative mean squared error up to order O(n^{-2}), respectively, are given by

$$RV\left(t_2^*\right) = RM\left(\bar{y}_P\right) + \frac{1}{n^2} C_{11}\left(2C_{20} + C_{11}\right) - \frac{2}{n^2}\left(C_{21} + C_{12}\right)$$

$$= RV\left(t^*\right) + \frac{2}{n^2} C_{11}\left(C_{20} + C_{11}\right) \quad . \tag{3.22}$$

The estimator t_2^* has smaller relative mean squared error than product estimator as well as the estimator proposed by Robson (1957) for bivariate normal population if

$$0 < \Theta < -2\,\rho.$$

Now, comparing the relative mean squared errors of the estimators t_2^* and $t_{\alpha q}$, from (2.5) and (3.22) for bivariate normal population, we have

$$RM\left(t_{\alpha q}\right) - RV\left(t_2^*\right) = -\frac{1}{n}\left(C_{20} + 2C_{11}\right) + \frac{1}{n^2}\left[q^2 C_{11}^2 - 2qC_{11}^2 - C_{02}C_{20} - 2C_{11}C_{20} - 3C_{11}^2\right]. \tag{3.23}$$

From (3.23), we conclude that the estimator $t_{\alpha q}$ is more efficient than t_2^* up to order O (n^{-1}) if

$$\rho > \frac{-1}{2\theta}$$

and O(n^{-2}) if

$$-1 < q < 3.$$

Sharma and Bhatnagar (2008) proposed a general class of biased product estimators as

$$T_{q^*} = \bar{y}\left[\frac{\bar{x}}{\bar{X}} + \frac{q^* s_x^2}{\bar{x}^2}\right] \quad . \tag{3.24}$$

Taking $q^*=2$ in (3.24), we find an estimator as

$$T_2 = \bar{y}\left[\frac{\bar{x}}{\bar{X}} + \frac{2s_x^2}{\bar{x}^2}\right] \; . \tag{3.25}$$

The relative bias and relative mean squared error of the estimator T_2 upto order O(n^{-1}) and O (n^{-2}), respectively, are given as:

$$RB\ (T_2) = RB\ (\bar{y}_P) + 2\ \frac{C_{20}}{n} \tag{3.26}$$

$$RM\ (T_2) = RM\ (\bar{y}_P) + \frac{1}{n^2}\left[4C_{02}C_{20} + 4C_{30} + 4C_{21} - 4C_{20}^2\ \right]. \tag{3.27}$$

Now, if we compare the relative bias of estimators $t_{\alpha q}$ with T_2, we have

$$RB\ (t_{\alpha q}) - RB\ (T_2) = \frac{1}{n}\ [\,C_{11}\ (q-1) - 2\,C_{20}\,] \quad . \tag{3.28}$$

From (3.28) it can be concluded that the estimator $t_{\alpha q}$ has smaller relative bias than T_2 provided $q>1$ and $\rho\Theta < 2$.

Similarly, by the comparing the relative mean squared errors of the estimators $t_{\alpha q}$ and T_2, from (2.5) and (3.27) under a bivariate normal population, we have

$$RM(t_{\alpha q})\text{-}RM(T_2) = -\frac{1}{n}\left[2C_{11} + C_{20}\right] + \frac{1}{n^2}\left[C_{11}^2\ (q^2 - 2q - 2) - 5C_{02}C_{20} + 4C_{20}^2\ \right] \tag{3.29}$$

the estimator $t_{\alpha q}$ will be more efficient than T_2 upto order $O(n^{-1})$ if

$$\rho > \frac{-1}{2\theta}$$

and Upto $O(n^{-2})$ if

$$0<q<3.$$

4. Conclusion

The proposed estimator $t_{\alpha q}$ can perform better than ratio estimator if $\rho > \frac{1}{2\theta}$ and 2 < q <4, whereas it has been found to be more efficient than product, Robson (1957), Singh (1989), Dubey(1993) & Sharma and Bhatnagar (2008) estimators if $\rho > \frac{-1}{2\theta}$ and $0<q<3$.

So, the proposed estimator is more efficient than the mean, ratio, product and product type estimators according to relative mean squared error criterion under different conditions.

References

Dubey, V. (1993) "An almost unbiased product estimator", *Journal of Indian Society of Agricultural Statistics,* 45(2), 226-29.

Murthy, M.N. (1964), "Product method estimation", *Sankhya,* 26, 69-74.

Robson, D.S. (1957), "Application of multivariate polykas to the theory of unbiased ratio type estimators", *Journal of American Statistical Assoc*iation 52, 511-522**.**

Sharma M. and Bhatnagar S. (2008), "A general class of estimators for estimation of population mean through auxiliary information"*, Journal of Indian Statistical Association,* 46(2), 155-161.

Singh, H.P. (1989), "A class of unbiased product estimators of population means", *Journal of Indian Society of Agricultural Statistics*, 41(1), 113-118.