On The Use of Coefficient of Variation and β_1, β_2 in Estimating Mean of a Finite Population

[1]B. B. Khare, [1]P. S. Jha and [2]U. Srivastava

[1]Department of Statistics, B.H.U, Varanasi-221005

[2]Statistics Section, MMV, B.H.U, Varanasi-221005

bbkhare56@yahoo.com

Abstract

In this paper the use of coefficient of variation and shape parameters in each stratum, the problem of estimation of population of mean has been considered. The expression of mean squared error of the proposed estimator is derived and its properties are discussed.

Keywords Auxiliary information, MSE, coefficient of variation, stratum, shape parameter.

Introduction

The use of prior information about the population parameters such as coefficient of variation, mean and skewness and kurtosis are very useful in the estimation of the population parameter of the study character. In agricultural and biological studies information about the coefficient of variation and the shape parameters are often available. If these parameters remain essentially unchanged over the time than the knowledge about them in such case it may profitably be used to produce optimum estimates of the parameters (Sen and Gerig (1975)). Searls (1964, 67) and Hirano (1972) have proposed the use of coefficient of variation in the estimation the population mean. Searl and Intarapanich (1990) have suggested the use of kurtosis in the estimation of variance. Sen (1978) has proposed the estimator for population mean using the known value of coefficient of variation.

In Stratified random sampling, the theory has been developed to provide the optimum estimator T_1 of the population mean based on sample mean from each stratum. We extend it by constructing an estimator T_2 using the coefficient of variation C_i and shape parameter β_{1i}, β_{2i} $(i = 1,2,...K)$ from each stratum and discuss its usefulness. We also define estimators T_3 and T_4 when the coefficients of variation are unknown but shape parameters are known and when neither the coefficients of variation are known nor the shape parameters are known.

Estimators and their Mean Square Error

Let N_i denotes the size of the ith stratum and n_i denotes the size of the sample to be selected from the ith stratum and h be the number of strata with

$$\sum_{i=1}^{h} N_i = N \text{ and } \sum_{i=1}^{h} n_i = n \ , \tag{1}$$

where N and n denote the number of units in the population and sample respectively.

Let y_{ij} be the j[th] unit of the i[th] stratum. Then the population mean $\overline{\mathrm{Y}}_{\mathrm{N}}$ can be expressed as

$$\overline{\mathrm{Y}}_{\mathrm{N}} = \sum_{i=1}^{h} p_i \overline{Y}_i \ , \tag{2}$$

where $p_i = \dfrac{N_i}{N}$ and $\overline{Y}_i$ is the population mean for the i[th] stratum.

Let n_i units be selected from the i[th] stratum and the corresponding sampling mean and sample variance be denoted by $\bar{y}_i$ and s_i^2 respectively. Then the estimate of $\overline{Y}_N$ is given by

$$\mathrm{T}_1 = \sum_{i=1}^{h} p_i \bar{y}_i \tag{3}$$

and the

$$\mathrm{V}(\mathrm{T}_1) = \sum_{i=1}^{h} \frac{p_i^2 \sigma_i^2}{n_i} \quad \text{(if f.p.c is ignored),} \tag{4}$$

where σ_i^2 is the population variance of y in the i[th] stratum.

Case 1: Coefficient of variation and the shape parameters are known.

We defined

$$\mathrm{T}_2 = \sum_{i=1}^{h} p_i \{\alpha_i \bar{y}_i + (1-\alpha_i) C_i^{-1} \sqrt{s_i^2}\} \tag{5}$$

and expectation of T_2 is given by

$$\mathrm{E}(\mathrm{T}_2) = \sum_{i=1}^{h} p_i \{\alpha_i \overline{Y}_i + (1-\alpha_i)\overline{Y}_i (1 - \frac{1}{8}\frac{V(s_i^2)}{\sigma_i^4})\}$$

$$= \sum_{i=1}^{h} p_i \{\overline{Y}_i - (1-\alpha_i)(\frac{1}{8}\frac{V(s_i^2)}{\sigma_i^4})\}$$

$$= \sum_{i=1}^{h} p_i \{\overline{Y}_i - \frac{(1-\alpha_i)}{8}(\frac{\beta_{2i}-1}{n_i} + \frac{2}{n(n-1)})\}$$

$$= \overline{Y}_N + O(\frac{1}{n_i})$$

(6)

where β_{2i} is the measure of kurtosis in the i[th] stratum.

The bias in T_2 is of order $\frac{1}{n_i}$ and will be negligible for large n_i's.

The mean square error of the estimator is

$$\text{MSE}(\text{T}_2/\alpha_i) = \sum_{i=1}^{h} \frac{p_i^2\sigma_i^2}{n_i}\left[\{\alpha_i^2 + \alpha_i(1-\alpha_i)C_i^{-1}\sqrt{\beta_{1i}} + \frac{(1-\alpha_i)^2}{4}C_i^{-2}(\beta_{2i}-1)\} + O(\frac{1}{n^{3/2}})\right]. \quad (7)$$

Minimising (7) with respect to α_i, we get the optimum value of α_i is given by

$$\alpha_{\text{i}opt} = \frac{\beta_{2i} - 2C_i\sqrt{\beta_{1i}} - 1}{4C_i^2 + 4C_i\sqrt{\beta_{1i}} + \beta_{2i} - 1}\ , \quad (8)$$

where β_{1i} is the measure of kurtosis in the i[th] stratum.

On putting the optimum value of $\alpha_{i\,opt}$ from (8) in (7) and on simplification we get

$$\text{MSE}(\text{T}_2)_{\min} = \sum_{i=1}^{h} \frac{p_i^2\sigma_i^2}{n_i}\left[\{\frac{\beta_{2i} - \beta_{1i} - 1}{(\beta_{2i} - \beta_{1i} - 1) + (\sqrt{\beta_{1i}} - 2C_i)^2}\} + O(\frac{1}{n^{3/2}})\right]. \quad (9)$$

The value of $\alpha_{i\,opt}$ will be less than one for $\sqrt{\beta_{1i}} < 2C_i$, which implies that the distribution is near normal, poison, negative binomial and Neyman type I. The value of $\alpha_{i\,opt}$ will be equal to one for $\sqrt{\beta_{1i}} = 2C_i$, which is true for gamma and exponential distribution. The value of $\alpha_{i\,opt}$ will be greater than one for $\sqrt{\beta_{1i}} > 2C_i$, which is likely to the distribution of lognormal or inverse Gaussian. It is easy to see that T_2 will always be more efficient than T_1 if $\sqrt{\beta_{1i}} < 2C_i$ or $\sqrt{\beta_{1i}} > 2C_i$, justifying the use of T_2 in the case of near normal, poison, negative binomial, Neyman type I and lognormal or inverse Gaussian distribution. T_2 is equally efficient T_1, if $\sqrt{\beta_{1i}} = 2C_i$ and so for in gamma or exponential distribution one may use T_1 or T_2. This shows that proposed estimator T_2 is uniformly superior to the estimator T_1, though a comparatively high efficiency may be seen in near normal, poison, negative binomial than lognormal or inverse Gaussian distribution.

Case 2: C_i's are unknown, β_{1i}'s and β_{2i}'s are known.

When C_i's are unknown, we use their estimates c_i based on a larger sample of size n_i' from a previous occasion. Now we define an estimator T_3 for $\bar{Y}_N$ given by

$$\mathrm{T}_3 = \sum_{i=1}^{h} p_i \{\alpha_i' \bar{y}_i + (1-\alpha_i') c_i^{-1} \sqrt{s_i^2}\} \tag{10}$$

The mean square error of the estimator T_3 as given by

$$\mathrm{MSE}(\mathrm{T}_2 / \alpha_i') = \sum_{i=1}^{h} \frac{p_i^2 \sigma_i^2}{n_i} \left[\alpha_i'^2 + \alpha_i'(1-\alpha_i') C_i^{-1} \sqrt{\beta_{1i}} + (1-\alpha_i')^2 C_i^{-2} \{(\beta_{2i}-1) + 4 n_i C_i^{-2} V(c_i)\} \right], \tag{11}$$

where $V(c_i) = \dfrac{C_i^2}{n_i'} \{(\beta_{2i}-1) - \dfrac{\mu_3}{\mu_2 \mu_1'} + \dfrac{\mu_2}{\mu_1^2}\} \cong \dfrac{C_i^2}{n_i'}(\beta_{2i}-1)$.

The optimum value of α_i' is given by

$$\alpha'_{\mathrm{i}opt} = \frac{(\beta_{2i} - \beta_{1i} - 1) + 4 n_i C_i^{-2} V(c_i)}{(\beta_{2i} - \beta_{1i} - 1 + 4 n_i C_i^{-2} V(c_i)) + (\sqrt{\beta_{1i}} - 2C_i)^2} . \tag{12}$$

It is easy to see that

$$\mathrm{MSE}(\mathrm{T}_3 / \alpha'_{\mathrm{i}opt})_{\min} = \sum_{i=1}^{h} \frac{p_i^2 \sigma_i^2}{n_i} \left[\frac{(\beta_{2i} - \beta_{1i} - 1) + 4 n_i C_i^{-2} V(c_i)}{(\beta_{2i} - \beta_{1i} - 1 + 4 n_i C_i^{-2} V(c_i)) + (\sqrt{\beta_{1i}} - 2C_i)^2} \right]. \tag{13}$$

It may be remarked that (13) differs from (9) by a single term $4 n_i C_i^{-2} V(c_i)$ both in numerator and denominator. The nature of the estimator T_3 is similar to T_2 and its MSE will converge to $MSE(T_2)$ for $\dfrac{V(c_i)}{C_i^2} \to 0$.

Case 3: C_i's, β_{1i}'s and β_{2i}'s are unknown:

When C_i's , β_{1i}'s and β_{2i}'s are not known then they can be estimated on the basis of a larger sample of size $n_i' >> ...n_i$ from the past data and we may have the estimator for the population mean $\bar{Y}_N$ given by

$$\mathrm{T}_4 = \sum_{i=1}^{h} p_i \{\hat{\alpha}_i \bar{y}_i + (1-\hat{\alpha}_i) c_i^{-1} \sqrt{s_i^2}\} , \tag{14}$$

where $\hat{\alpha}_{\mathrm{i}opt} = \dfrac{\hat{\beta}_{2i} - 2\hat{C}_i\sqrt{\hat{\beta}_{1i}} - 1}{4\hat{C}_i^2 + 4\hat{C}_i\sqrt{\hat{\beta}_{1i}} + \hat{\beta}_{2i} - 1}$.

It is easy to see that the $\mathrm{MSE(T_4)}$ will be same as $\mathrm{MSE(T_3)}$ because after estimating the unknown parameters in the constant α_{iopt}, the MSE will remains unchanged up to the terms of O (n^{-1}) (Srivastava and Jhajj (1983)).

References

1. Searls, D. T. (1964): The utilization of coefficient of variation in the estimation procedure. Jour. of Amer. Stat. Assoc., 59, 1125-1126.

2. Searls, D. T. (1967): A note on the use of a approximately known coefficient of variation. The Amer. Statistician, 21, 20-21.

3. Sen, A. R. and Gerig, T. M. (1975): Estimation of a population mean having equal coefficient of variation on succession occasions. Bull. Int. Stat. Inst., 46, 314-22.

4. Sen, A. R. (1978): Estimation of the population mean when the coefficient of variation is known. Commu. Stat. Theory Meth., A7, 1, 657-672.

5. Srivastava, S.K. and Jhajj, H.S. (1983): A class of estimators of the population means using multi- auxiliary information. Calcutta Stat. Assoc. Bull, 32, 47-56.

6. Searls, D. T. and Intarapanich R (1990): A note on an estimator for variance that utilized the kurtosis. Amer. Stat., 44(4), 295-296.

7. Hirano K (1972): Using some approximately known coefficient of variation in estimating mean. Proc. Inst. Stat. Math, 20(2), 61-64.