ISSN 0976 – 1411

Available online at www.internationaleJournals.com

International eJournals

INTERNATIONAL eJOURNAL OF MATHEMATICS AND ENGINEERING

www.InternationaleJournals.com

International eJournal of Mathematics and Engineering 37 (2010) 364 - 371

AN ESTIMATOR FOR MEAN OF STRATIFIED POPULATION USING KNOWN COEFFICIENT OF SKEWNESS

B.BHASKARA RAMA SARMA[1] S.V.N.L.LALITHA[2]

1. Faculty of Mathematics, BRS Classes
#2-284, Vivekananda Street, Hanumannagar, Ramavarappadu, Vijayawada, A.P-521108;
Mobile:9441924418;e-mail:bbramasarma@yahoo.co.in
2. Faculty of EEE, K.L.E.F University, Vaddeswaram, Guntur (Dist.) ,A.P

ABSTRACT

In estimating the parameters of a population if one can sacrifice the property of unbiasedness,better estimators in view of minimum mean squared error(MMSE) can be obtained.Estimators for Population mean using known coefficient of variation were proposed by Searles (1964), Srivastava(1974), Upadhyaya and srivastava (1976). Present Author has constructed an estimator for mean of stratified population using Searles approach in a different paper. In the present investigation Srivastava's method is used to construct such an estimator t*.The advantage of this method is that it works without the actual knowledge of population co-efficient of variation. Instead ,it presumes the knowledge of co-efficient of skewness of the data and an interval of c.v. The large sample properties of proposed estimator are studied in detail. Larger gains are observed in efficiencies for small sample sizes.

1. INTRODUCTION & NOTATIONS

Consider a population of N elements divided into k strata so that i^{th} stratum contains N_i elements :

i =1,2,3,4,………..,k .Following notations are adopted.

y_{ij} = j^{th} simple random sampled unit from i^{th} stratum ; j =1,2,3,4,………,n

Sample mean from i^{th} stratum = $\bar{y}_i = n^{-1}\sum_{j=1}^{n} y_{ij}$,

Sample mean square= $s_i^2 = (n-1)^{-1}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_i\right)^2$

i^{th} stratum mean = $\bar{Y}_i = N_i^{-1}\sum_{j=1}^{N_i} Y_{ij}$, i^{th} stratum variance = $\sigma_i^2 = N_i^{-1}\sum_{j=1}^{N_i}\left(Y_{ij} - \bar{Y}_i\right)^2$

It is well known that $E(s_i^2) = \sigma_i^2$ and E($\bar{y}_{st}$) = $\bar{Y}$

$\bar{y}_{st} = \sum_{i=1}^{k} p\bar{y}_i$ =Unbiased estimator of $\bar{Y}$ under stratified random sampling ; $p = \frac{N_i}{N}$ $\forall$ i

Population Mean = $\bar{Y} = N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{N_I} Y_{ij} = \sum_{i=1}^{k} p\bar{Y}_i$; $\sum_{i=1}^{k}\frac{N_i}{N} = kp = 1$, Assume $N_i - 1 \cong N_i$

Population variance= $\sigma^2 = N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{N_i}\left(Y_{ij} - \bar{Y}_i\right)^2$

$\frac{\sigma_i}{\bar{Y}_i} = \frac{\sigma}{\bar{Y}} = \sqrt{c_i} = \sqrt{c}$ = co-efficient of variation for i^{th} stratum (unknown)

Bias and Meansquared error of an estimator t of a population parameter τ are respectively defined as
B(t) =E (t) - τ ,

M(t) = E (t –E(t))^2 where E(t) is the Expectation of t .

Relative efficiency of an estimator t_1 over another estimator t_2 is defined as

$$\text{REF}(t_1, t_2) = \frac{M(t_2)}{M(t_1)}$$

1.1 Estimator for stratum mean $\bar{Y}_i$ (using Srivastava's (1974) method):

Assuming that the coefficients of variation $\sqrt{C_i} = \sqrt{C}$ for i = 1, 2, ….., k to be unknown, an estimator for i^{th} stratum mean based, on Srivastava's (1974) method is given by,

$$t_i = \frac{\bar{y}_i}{1 + \frac{s_i^2}{n\bar{y}_i^2}} \qquad (1.1.1)$$

for i = 1, 2, ….., k

The large sample properties of (1.1.1) are investigated below.

write $$\bar{y}_i = \bar{y}_i + u_i \qquad (1.1.2)$$

$$s_i^2 = \sigma_i^2 + v_i \qquad (1.1.3)$$

where u_i and v_i are such that

$$\left.\begin{aligned} &E(u_i) = E(v_i) = 0 \\ \text{and} \quad &E(u_i\ u_j) = E(v_i\ v_j) = E(u_i\ v_j) = 0 \end{aligned}\right\} \qquad (1.1.4)$$

for i = 1, 2, ….., k, $\quad i \neq j$

Using (1.1.1), (1.1.2) and (1.1.3),

$$t_i = \left(\bar{Y}_i + u_i\right)^3 \left[\left(\bar{Y}_i + u_i\right)^2 + \frac{1}{n}\left(\sigma_i^2 + v_i\right)^{-1}\right]$$

$$= \bar{Y}_i\left(1 + \frac{u_i}{\bar{Y}_i}\right)^3 \left[1 + \frac{2u_i}{\bar{Y}_i} + \frac{u_i^{\ 2}}{\bar{Y}_i^2} + \frac{c}{n} + \frac{v_i}{n\bar{Y}_i^{\ 2}}\right]^{-1} \qquad (1.1.5)$$

Relating terms to 0 (n^{-2}), basis in t_i ($B(t_i)$) and mean squared error in t_i ($M(t_i)$) are shown to be,

$$B(t_i) = \frac{-c}{n}\left[1 - \frac{1}{n}\sqrt{\beta_{1i} c}\right]\bar{Y}_i \qquad (1.1.6)$$

$$M(t_i) = \frac{\sigma_i^2}{n}\left[1 + \frac{1}{n}\left(3c - 2\sqrt{\beta_{1i} c}\right)\right] \qquad (1.1.7)$$

Assuming that the coefficient of skewness of i[th] stratum $\beta_{1i} = \beta_1$ for all i = 1,2, ……, k, (1.1.6) and (1.1.7) are equal to,

$$B(t_i) = \frac{-c}{n}\left[1 - \frac{1}{n}\sqrt{\beta_1 c}\right]\bar{Y}_i \qquad (1.1.8)$$

$$M(t_i) = \frac{\sigma_i^2}{n}\left[1 + \frac{1}{n}\left(3c - 2\sqrt{\beta_1 c}\right)\right] \qquad (1.1.9)$$

1.2 Proposed estimator for population mean $\bar{Y}$

Using all the k estimators in (1.1.1) an estimator t* is proposed for population mean $\bar{Y}$ as,

$$t^* = q \sum_{i=1}^{k} t_i \tag{1.1.10}$$

where q is to be determined so as to minimize the mean squared error in t* (MSE (t*))

$$\begin{aligned} \text{MSE}(t^*) &= E\,[t^* - \bar{Y}]^2 \\ &= \text{Variance of } t^* + (\text{Bias in } t^*)^2 \\ &= V(t^*) + [B(t^*)]^2 \end{aligned} \tag{1.1.11}$$

From (1.1.10) and (1.1.11) ,

$$MSE(t^*) = q^2 \left[\sum_{i=1}^{k} V(t_i) + \sum_{i \neq j} Cov(t_i, t_j) \right] + [B(t^*)]^2 \tag{1.1.12}$$

Now,

$$\sum_{i=1}^{k} V(t_i) = \sum_{i=1}^{k} M(t_i) + t \sum_{i=1}^{k} [B(t_i)]^2 \tag{1.1.13}$$

Using (1.1.8) and (1.1.9) in (1.1.13) and retaining terms to 0(n^{-2}),

$$\sum_{i=1}^{k} V(t_i) = \frac{(n + 4c - 2\sqrt{\beta_1 c})}{n^2} \sum_{i=1}^{k} \sigma_i^2 \tag{1.1.14}$$

Also, since t_i, t_j are computed from two different strata they are independent, and hence,

$$Cov(t_i, t_j) = 0 \tag{1.1.15}$$

Now, from (1.1.10)

$$\begin{aligned} B(t^*) &= q \sum_{i=1}^{k} E(t_i) - \bar{Y} \\ &= q \sum_{i=1}^{k} [\bar{Y}_i + B(t_i)] - \bar{Y} \\ &= q \sum_{i=1}^{k} \bar{Y}_i + q \sum_{i=1}^{k} B(t_i) - \bar{Y} \end{aligned} \tag{1.1.16}$$

Using (1.1.5), (1.1.8) in (1.1.16) and retaining terms to $0(n^{-2})$

$$B(t^*) = \frac{q}{p}\bar{Y} - \frac{qc}{n}(1 - \frac{1}{n}\sqrt{\beta_1 c})\frac{\bar{Y}}{p} - \bar{Y} \tag{1.1.17}$$

Using (3.2.11) in (3.5.17)

$$[B(t^*)]^2 = \frac{\sigma^2}{c}\left[\frac{q^2}{p^2}\left((1-\frac{c}{n})^2 + \frac{2c}{n^2}\sqrt{\beta_1 c}\right) + \frac{2q}{p}\left(\frac{c}{n} - \frac{c}{n^2}\sqrt{\beta_1 c} - 1\right) + 1\right] \tag{1.1.18}$$

Using (1.1.14), (1.1.15), (1.1.18), in (1.1.12)

$$MSE(t^*) = \frac{q^2(n+4c-2\sqrt{\beta_1 c})}{n^2}\sum_{i=1}^{k}\sigma_i^2 + \left[\frac{q^2}{p^2}\left((1-\frac{c}{n})^2 + \frac{2c}{n^2}\sqrt{\beta_1 c}\right) + \frac{2q}{p}\left(\frac{c}{n} - \frac{c}{n^2}\sqrt{\beta_1 c} - 1\right) + 1\right] \tag{1.1.19}$$

Minimization of MSE (t*) in (1.1.19) is found to result in complicated expressions, it is reduced to $0(n^{-1})$ and given by,

$$MSE\ (t^*) = \frac{q^2}{pn}\sigma^2 + \frac{\sigma^2}{c}\left[\frac{q^2}{p^2}\left(1 - \frac{2c}{n}\right) + \frac{2q}{p}\left(\frac{c}{n} - 1\right) + 1\right] \tag{1.1.20}$$

Differentiating (1.1.20) with respect to q and equating the differential coefficient to zero, for $c < \frac{n}{p-2}$,

$$q = p[1 - (p-1)\frac{c}{n}] \tag{1.1.21}$$

Now from (1.1.1), (1.1.21) and (1.1.10),

$$t^* = p[1 - (p-1)\frac{c}{n}]\sum_{i=1}^{k}\frac{n\bar{Y}_i^3}{(n\bar{Y}_i^2 + s_i^2)}$$

The minimum mean squared error is obtained by using (1.1.21) in (1.1.20), as

$$M\ (t^*) = \frac{p\sigma^2}{n} \tag{1.1.22}$$

Also bias in t* is,

$$B(t^*) = -\frac{cp}{n}\overline{Y} \tag{1.1.23}$$

Comparison of M (t*) with V($\overline{Y}$)

From (1.1.23) and (1.1.1),

$$\frac{M\ (t^*)}{V\ (\overline{Y})} = 1 \tag{1.1.24}$$

It follows from (1.1.24) that to 0 (n^{-1}) , t* is at least as efficient as $\overline{Y}$.

PARTICULAR CASE

In particular, when $q = \frac{1}{k}$ is used in (1.1.1) the estimator T for $\overline{Y}$ is given by

$$T = \frac{1}{k}\sum_{i=1}^{k} t_i \tag{1.1.25}$$

for i = 1, 2, ……,k

Bias in T, B (T) is given by

$$B(T) = E(T) - \overline{Y}$$

$$= \frac{1}{k}\sum_{i=1}^{k} E(t_i) - \overline{Y}$$

$$= \frac{1}{k}\sum_{i=1}^{k} [\overline{Y_i} + B(t_i)] \tag{1.1.26}$$

Using (1.1.5), (1.1.8) in (1.1.26)

$$B(T) = \frac{-c}{n}[1 - \frac{1}{n}\sqrt{\beta_1 c}\,]\overline{Y} \tag{1.1.27}$$

Mean squared error in T is given by

$$MSE\ (T) = V(T) + [B(T)]^2 \tag{1.1.27a}$$

But from (1.1.25)

$$V(T) = \frac{1}{k^2}\left[\sum_{i=1}^{k} V(t_i) + \sum_{i \neq j} Cov(t_i, t_j)\right]$$

Now

$$\sum_{i=1}^{k} V(t_i) = \sum_{i=1}^{k} M(t_i) - \sum_{i=1}^{k} [B(t_i)]^2 \quad (1.1.28)$$

Using (1.1.8) and (1.1.9) and (1.1.14) in (1.1.28) and retaining terms to 0 (n^{-2})

$$\sum_{i=1}^{k} V(t_i) = \frac{\sigma^2}{pn}[1+\frac{2c}{n}-\frac{2\sqrt{\beta_1 c}}{n}] \quad (1.1.29)$$

Also since t_i , t_j are independent ,

$$\mathrm{Cov}(t_i , t_j) = 0 \quad (1.1.30)$$

Using (1.1.26), (1.1.27), (1.1.28) and (1.1.30) in (1.1.27a)

$$MSE(T) = \frac{p\sigma^2}{n}[1+\frac{2c}{n}-\frac{2\sqrt{\beta_1 c}}{n}]+\frac{c\sigma^2}{n^2} \quad (1.1.31)$$

From (1.1.31)

$$\frac{MSE(T)}{V(\bar{Y})} = 1+\frac{c}{n}(2+k)-\frac{2}{n}\sqrt{\beta_1 c} \quad (1.1.32)$$

It can be seen that

$$\frac{MSE(T)}{V(\bar{Y})} < 1 \text{ ; if } 0<\sqrt{c}<\frac{2\sqrt{\beta_1}}{k+2} \quad (1.1.33)$$

(1.1.33) gives the range for coefficient of variation $\sqrt{c}$ to lie so that T is more efficient than $\bar{Y}$

The relative efficiencies of T over $\bar{Y}$ for given β_1 and for specified $\sqrt{c}$ in the range are tabulated below.

Case (a) From (1.1.32) and $\beta_1 = 0, \sqrt{c} = 0$

$\frac{MSE(T)}{V(\bar{Y})} = 1,$ i.e., T and $\bar{Y}$ are equally efficient.

Case (b) $\beta_1 \neq 0$, the results are tabulated in the following tables.

Table1.1

β_1 \ C	2	3	4
1	(0, 0.500)	(0, 0.600)	(0, 0.330)
2	(0, 0.707)	(0, 0.566)	(0, 0.470)
3	(0, 0.866)	(0, 0.693)	(0, 0.577)

$REF(T,\bar{Y})$ are tabulated below for specific values of β_1 and $\sqrt{c}$ with different sample sizes.

TABLE 1.2

$$REF(T,\bar{Y}) = [1+\frac{c}{n}(2+k)-\frac{2}{n}\sqrt{\beta_1 c}]^{-1}\,100\%$$

	$\beta_1 = 1, k = 2$			$\beta_1 = 3, k = 4$		
n \ $\sqrt{c}$	0.1	0.3	0.4	0.1	0.3	0.5
5	103.30	105.04	103.30	106.06	111.11	104.86
10	101.63	102.46	101.63	102.94	105.26	102.37
15	101.08	101.63	101.08	101.94	103.45	101.57
20	101.80	101.21	101.80	101.45	102.56	101.17

Conclusions

With increasing sample size, for given coefficient of variation value, relative efficiency decreases.

REFERENCES

[1]. Bhaskara Rama Sarma .B: “A Note on Estimation of Mean in a Stratified Population using known Co-Efficient Of Variation", accepted for publication in’ in **International Journal of Statistics&Systems,** Research India Publications,New Delhi
[2]. Kendall,M.G& Stuart,A.:’The Advanced theory of Statistics’, Vol.1,Charles Griffin&Co,London.
[3]. Searles.D.T :‘The Utilisation of a known co-efficient of variation in the estimation procedure’,Journal of American Stat.Asso.,1964,Vol.30,pp.1225-26.
[4] Srivastava,V.K : “On the use of co-efficient of variation in estimating mean”; Journal of Indian Society for Agri. Stats., Vol.26, 1974, pp.33-36