ISSN: 0975-766X
CODEN: IJPTFI
Available Online through
www.ijptonline.com
Research Article

AN ESTIMATOR FOR MEAN OF STRATIFIED POPULATION USING KNOWN CO-EFFICIENT OF SKEWNESS

B. Bhaskara Rama Sarma[1], V. Vasanta Kumar*[2]

[1] Research Scholar, Department of Mathematics, K L University, Vaddeswaram, Guntur District, Andhra Pradesh, India.

[2] Department of Mathematics, K L University, Vaddeswaram, Guntur District, Andhra Pradesh, India.

Email: vvkumar@kluniversity.in

Received on: 21-08-2017 *Accepted on: 25-09-2017*

Abstract

In estimating the parameters of a population if one can sacrifice the property of biasedness, better estimators in view of minimum mean squared error (MMSE) can be obtained. In the present paper estimators for population mean of a stratified population using known co-efficient of skewness are constructed. The relative efficiencies of proposed estimator over the conventional estimator are obtained. Numerical results are presented for various values of co-efficient of skew sizes ness and sample sizes Larger gains are observed in efficiencies of proposed estimator for small sample.

Key Words: Minimum mean squared error; Unbiased estimator; Stratum mean; Stratum variance.

1. Introduction

Estimators for Population mean using known coefficient of variation were proposed by Searles.(3), Srivastava(4), Upadhyaya and Srivastava (1976). Bhaskara Rama Sarma.proposed [1]has constructed an estimator for mean of Stratified Population using Searles approach. In the present investigation Srivastava(4) method is used to construct such estimators using known co-efficient of skew ness .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 skew ness of the data and an interval of co-efficient of variation

Following notations are used in this paper

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

s_i^2 : Sample mean squ, i =1,2,3…, k .

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

$\bar{y}_i$: Sample mean from ith stratum, i= 1,2, . . . k.

$$\bar{y}_i = n^{-1}\sum_{j=1}^{n} y_{ij} .$$

are from ith stratum

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

$\bar{Y}_i$: Mean of ith stratum

$$\bar{Y}_i = N_i^{-1}\sum_{j=1}^{N_i} Y_{ij} .$$

σ_i^2 : Variance of ith stratum

$$\sigma_i^2 = N_i^{-1}\sum_{j=1}^{N_i}\left(Y_{ij} - \bar{Y}_i\right)^2 .$$

$\bar{Y}$: Overall 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 $N_i - 1 \cong N_i$; i= 1,2, . . . k.

σ^2 : Population variance

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

$\sqrt{c_i}$: Co efficient of variation for ith stratum (unknown)

$$\sqrt{c_i} = \frac{\sigma_i}{\bar{Y}_i} .$$

$\bar{y}_{st}$: Unbiased estimator of $\bar{Y}$ under stratified random sampling

$$\bar{y}_{st} = \sum_{i=1}^{k} p\bar{y}_i \; ; \quad p = \frac{N_i}{N} \;\; , \;\; i = 1,2, \ldots k.$$

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

Bias and Mean squared 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

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

2. ESTIMATOR FOR STRATUM MEAN $\bar{Y}_i$

(USING SRIVASTAVA'S (4) METHOD)

Assuming that the coefficients of variation $\sqrt{c_i} = \sqrt{c}$, i = 1, 2, …, k to be unknown, an estimator for i^th stratum mean based on Srivastava's (4) method is given by,

$$t_i = \frac{\bar{y}_i}{1 + \frac{s_i^2}{n\bar{y}_i^2}} \quad \text{for } i = 1, 2, ..., k \tag{1}$$

The large sample properties of the estimator t_i are investigated below.

Write $\bar{y}_i = \bar{Y}_i + u_i$, (2)

And $s_i^2 = \sigma_i^2 + v_i$ (3)

where u_i and v_i are such that E(u_i) = E(v_i) = 0 ,

and E($u_i u_j$) = E($v_i v_j$) = E($u_i v_j$) = 0, i = 1,2,… k ,$i \neq j$. (4)

Substituting $\bar{y}_i$ and s_i^2 from (2) and (3) in (1),

$$t_i = \bar{Y}_i\left(1+\frac{u_i}{\bar{Y}_i}\right)^3\left[1+\frac{2u_i}{\bar{Y}_i}+\frac{u_i^2}{\bar{Y}_i^2}+\frac{c}{n}+\frac{v_i}{n\bar{Y}_i^2}\right]^{-1} \tag{5}$$

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

$$B(t_i) = \frac{-c}{n}\left[1-\frac{1}{n}\sqrt{\beta_{1i}c}\right]\bar{Y}_i \tag{6}$$

$$M(t_i) = \frac{\sigma_i^2}{n}\left[1+\frac{1}{n}\left(3c-2\sqrt{\beta_{1i}c}\right)\right] \tag{7}$$

Assuming that the coefficient of skew ness of i^th stratum $\beta_{1i} = \beta_1$ for all i = 1,2…, k, (6) and (7) are equivalent to,

$$B(t_i) = \frac{-c}{n}\left[1-\frac{1}{n}\sqrt{\beta_1 c}\right]\bar{Y}_i \tag{8}$$

$$M(t_i) = \frac{\sigma_i^2}{n}\left[1+\frac{1}{n}\left(3c-2\sqrt{\beta_1 c}\right)\right] \tag{9}$$

3. PROPOSED ESTIMATOR FOR POPULATION MEAN $\bar{Y}$

Using all the k estimators in (1) an estimator t* is proposed for population mean $\bar{Y}$ as,

$$t^* = q\sum_{i=1}^{k} t_i \qquad (10)$$

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

E [t* - $\bar{Y}$]^2

= Variance of t* + (Bias in t*)^2

= V(t*) + [B(t*)]^2 (11)

From (10) and (11) ,

$$MSE(t^*) = q^2\left[\sum_{i=1}^{k} V(t_i) + \sum\sum_{i\neq j}^{k} Cov(t_i, t_j)\right] + [B(t^*)]^2 \qquad (12)$$

Now, $\sum_{i=1}^{k} V(t_i) = \sum_{i=1}^{k} M(t_i) + t\sum_{i=1}^{k}[B(t_i)]^2$ (13)

Using (8) and (9) in 13) and retaining terms to 0(n^{-2}),

$$\sum_{i=1}^{k} V(t_i) = \frac{(n+4c-2\sqrt{\beta_1 c})}{n^2}\sum_{i=1}^{k}\sigma_i^2 \qquad (14)$$

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

Cov (t_i, t_j) = 0 (15)

Now, from (10)

$$B(t^*) = q\sum_{i=1}^{k} E(t_i) - \bar{Y}$$

$$= q\sum_{i=1}^{k}[\bar{Y}_i + B(t_i)] - \bar{Y}$$

$$= q\sum_{i=1}^{k}\bar{Y}_i + q\sum_{i=1}^{k} B(t_i) - \bar{Y} \qquad (16)$$

Substituting t_i from (5) and B(t_i) from (8) in (16) and retaining terms to 0(n^{-2})

$$B(t^*) = \frac{q}{p}\bar{Y} - \frac{qc}{n}(1-\frac{1}{n}\sqrt{\beta_1 c})\frac{\bar{Y}}{p} - \bar{Y} \qquad (17)$$

Using (11) in (17)

$$[B(t^*)]^2 = \frac{\sigma^2}{c}\left[\frac{q^2}{p^2}\left((1-\frac{c}{n})^2 + \frac{2c}{n^2}\sqrt{\beta_1 c}\right) + \frac{2q}{p}\left(\frac{c}{n} - \frac{c}{n^2}\sqrt{\beta_1 c} - 1\right) + 1\right] \qquad (18)$$

Using (14), (15) and (18) in (12)

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

(19)

As the differentiation of MSE (t*) with respect to q leads to a complicated expression we consider terms in MSE (t*) up to 0(n^{-1}). Then

$$MSE(t^*) = \frac{q^2}{pn}\sigma^2 + \frac{\sigma^2}{c}\left[\frac{q^2}{p^2}\left(1-\frac{2c}{n}\right) + \frac{2q}{p}\left(\frac{c}{n} - 1\right) + 1\right] \qquad (20)$$

Differentiating (20) with respect to q and equating the differential coefficient to zero, for $c < \frac{n}{p-2}$

$$q = p[1-(p-1)\frac{c}{n}] \qquad (21)$$

Now from (1), (21) and (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 (21) in (20), as

$$M(t^*) = \frac{p\sigma^2}{n} \qquad (22)$$

Also bias in t* is, $B(t^*) = -\frac{cp}{n}\bar{Y}$

(23)

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

From (23) and (1),

$$\frac{M(t^*)}{V(\bar{Y})} = 1 \qquad (24)$$

It follows from (24) that to 0 (n^{-1}) , t* is at least as efficient as $\bar{Y}$.

3.1 Particular case

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

$$T^* = \frac{1}{k}\sum_{i=1}^{k} t_i, \quad i = 1, 2, \ldots, k \qquad (25)$$

Bias in T* is given by

$$B(T^*) = E(T^*) - \bar{Y}$$

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

$$= \frac{1}{k}\sum_{i=1}^{k} [\bar{Y}_i + B(t_i)] \qquad (26)$$

Using t_i from (5) and $B(t_i)$ from (8) in (26),

$$B(T^*) = \frac{-c}{n}[1 - \frac{1}{n}\sqrt{\beta_1 c}]\bar{Y} \qquad (27)$$

Mean squared error in T* is given by

$$MSE(T^*) = V(T^*) + [B(T^*)]^2 . \qquad (28)$$

But from (25),

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

Now $\sum_{i=1}^{k} V(t_i) = \sum_{i=1}^{k} M(t_i) - \sum_{i=1}^{k} [B(t_i)]^2$ (29)

Using (8) and (9) and (14) in (29) and retaining terms to 0 (n^{-2})

$$\sum_{i=1}^{k} V(t_i) = \frac{\sigma^2}{pn}[1 + \frac{2c}{n} - \frac{2\sqrt{\beta_1 c}}{n}] \qquad (30)$$

Since t_i , t_j are independent ,

Cov (t_i , t_j) = 0 (31)

Using (26), (27), (29) and (31) in (28)

$$MSE(T^*) = \frac{p\sigma^2}{n}[1 + \frac{2c}{n} - \frac{2\sqrt{\beta_1 c}}{n}] + \frac{c\sigma^2}{n^2} \qquad (32)$$

From (32)

$$\frac{MSE(T^*)}{V(\bar{Y})} = 1 + \frac{c}{n}(2 + k) - \frac{2}{n}\sqrt{\beta_1 c} \qquad (33)$$

It can be seen that

$$\frac{MSE(T^*)}{V(\bar{Y})} < 1 \text{ , if } 0 < \sqrt{c} < \frac{2\sqrt{\beta_1}}{k+2}. \qquad (34)$$

Equation (34) gives the range for coefficient of variation $\sqrt{c}$ 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 i .

When $\beta_1 = 0$ and $\sqrt{c} = 0$

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

Case ii .

When $\beta_1 \neq 0$, the results are tabulated in the following tables.

Table 1:

$\beta_1 \downarrow$ k $\rightarrow$	1	2	3
1	(0, 0.500)	(0, 0.600)	(0, 0.330)
2	(0, 0.707)	(0, 0.566)	(0, 0.470)
3	(0, 866)	(0, 0.693)	(0, 0.577)

The 100% values of

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

are presented in Table 2 for specified values of β_1 and $\sqrt{c}$ with different sample sizes.

Table 2:

	$\beta_1 = 1,\ k = 2$			$\beta_1 = 3,\ k = 4$		
n $\downarrow$ $\sqrt{c}$ $\rightarrow$	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

4. Conclusions

Estimators for mean of a stratified population using known co-efficient of skewness are constructed. The relative efficiencies of proposed estimators over the conventional estimators are obtained. It is observed that the proposed estimator t* for population mean $\bar{Y}$ is at least as efficient as $\bar{Y}$ and the estimator T* is more efficient than $\bar{Y}$ if $0 < \sqrt{c} < \frac{2\sqrt{\beta_1}}{k+2}$. For specified values of coefficient of variation relative efficiency of T* decreases.

5. References

1. Bhaskara Rama Sarma B. (2010). A Note On Estimation Of Mean In A Stratified Population using known Co-efficient Of Variation, *International Journal of Statistics & Systems***,** 5 (1), 17-21.
2. Kendall M.G. & Stuart,A. *The Advanced theory of Statistics*, Vol.1,Charles Griffin & Co, London.
3. Searles D.T. (1964). The Utilisation of a known co-efficient of variation in the estimation procedure, *Journal of American Stat.Asso.*, 30, 1225-1226.
4. Srivastava V.K. (1974). On the use of co-efficient of variation in estimating mean, *Journal of Indian Society for Agri. Stats.*, 26, 33-36.

Corresponding Author:

V. Vasanta Kumar*,

Email: vvkumar@kluniversity.in