Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 73(1) 2019 53–57

Unbiased Class of Product Estimators in Circular Systematic Sampling (C.S.S.) Scheme

Kuldeep Rajpoot[1] and K.S. Kushwaha[2]
[1]AKS University, Satna
[2]Jawaharlal Nehru Krishi Vishwavidyalaya, Jabalpur

Received 31 July 2017; Revised 08 November 2018; Accepted 18 February 2019

SUMMARY

An unbiased class of product type estimators in circular systematic sampling (C.S.S.) scheme is proposed to estimate the population mean $\overline{y}_N$ of a response variables y. Jack-knife technique pioneered by Quenouille (1949,1956) has been applied to make the class unbiased. An explicit expression for the sampling variance of the class $T_{\lambda PU}$is derived to the terms of order o(n^{-1}). An empirical study is provided to examine the applied usefulness of the result derived.

Key Words: Product estimator, Jack-knife technique, Circular systematic sampling, M.V.U. estimator.

1. INTRODUCTION

The classical product estimator under linear systematic sampling (LSS) scheme was proposed by Shukla (1971) and its properties were studied. In general, the product estimator is biased. A weighted class of product type estimators was proposed and made exactly unbiased by Kushwaha and Singh (1989) using Jack-knife technique in linear systematic sampling (LSS) scheme.

A serious demerit of linear systematic sampling (LSS) scheme is that it is not possible to estimate the sampling variance of the proposed estimator but it can be estimated unbiasedly with the use of interpenetrating systematic sampling with independent random start.

A simple modification of LSS scheme makes it possible to ensure afixed sample size n and makes the sample mean $\overline{y}_N$ to estimate the population mean $\overline{y}_N$ unbiasedly even in case if N$\neq$nk. This sampling scheme is known as "circular systematic sampling (C.S.S.)" scheme.

Murthy (1977) and Sukhatme (1970) have suggested to use the C.S.S. scheme, in situation when N $\neq$ nk, where k is a positive integer.

The main steps involved in selecting a sample under C.S.S. scheme, are as follows :

i. Select a random number 'r' from 1 to N and name it as random strata,
ii. Choose some integer value of k = (N/n) or take integer nearest to (N/n) and name it as skip or sampling span and
iii. Select all unit in the sample with serial numbers
   r+jk, if (r+jk) $\leq$ N, {j=0,1,2, …,(n-1)},1$\leq$ r $\leq$N
   r+jk-N if (r+jk) > N, {j=0, 1, 2, …, (n-1)},1$\leq$ r $\leq$N

Sudhkar (1978) pointed out that the use of skip or span of sampling as an integer nearest to (N/n) in C.S.S. scheme does not draw a sample of desired size. Sudhakar (1978) has also mentioned that if we take the span of sampling as nearest to $\leq$ (N/n), we

Corresponding author: Kuldeep Rajpoot
E-mail address: kuldeep2rajpoot@gmail.com

do not encounter the above cited difficulty, although it depends on n.

Jack-knife technique has been profitably employed in several estimation and testing problems. In this study, we have proposed the usual product estimator proposed in C.S.S. scheme and have derived a general class of exactly unbiased product type estimators.

An empirical illustration has been provided to examine the performance of the derived estimators with respect to efficiency point of view with other estimators existing in the literature.

2. CLASS OF UNBIASED PRODUCT TYPE ESTIMATORS

Let the population consists of N units U= (U_1, U_2… U_N) numbered from 1 to N. let $N \neq nk$ where k is an integer nearest to (N/n). We select one sample using C.S.S. scheme and observe both the values (y,x) for each and every unit included in the sample. Let (y_{r+jk}, x_{r+jk}, j = 0, 1, 2,…, n-1 and r = 1, 2,…, N) be the sample units.

The circular systematic sample means $(\bar{y}, \bar{x})$ are defined as

$$\bar{y} = \frac{1}{n}\sum_{j=0}^{n-1} y_{r+jk} \text{ , j = 0,1, 2,……, n-1}$$

$$\bar{x} = \frac{1}{n}\sum_{j=0}^{n-1} x_{r+jk} \text{ , k is the sampling span} \qquad (2.1)$$

Sample means $(\bar{y}, \bar{x})$ are unbiased estimators of population means $(\bar{Y}_N, \bar{X}_N)$ respectively. The population mean $\bar{X}_N$ of covariate x is assumed to be known in prior.

The usual product estimator $\bar{y}_p$ for $\bar{Y}_N$ based on a circular systematic sample of size n in defined as

$$\bar{y}_p = \frac{\bar{y}\,\bar{x}}{\bar{X}_N} \qquad (2.2)$$

To reduce the bias of $\bar{y}_p$, we take n=gm and split the selected sample into g sub-samples each of size m in a systematic manner as this avoid the need for selecting the sample in the form of sub-samples of smaller sample size m, and there by retaining the efficiency generally obtained by taking a larger circular systematic sample of size n.

Let ($\bar{y}_t$, $\bar{x}_t$, t = 1, 2…, g) be the unbiased estimators of population means $(\bar{Y}_N, \bar{X}_N)$ based on circular systematic subsample each of size m. With this background, we propose another product type estimator $\bar{y}_{pt}$ written as

$$\bar{y}_{pt} = \frac{\bar{y}_t\,\bar{x}_t}{\bar{X}_N} \qquad \text{(t = 1, 2,3,……..,g)}$$

With its jack knife version written as

$$\bar{y}_p = \frac{1}{g}\sum_{t=1}^{g}\frac{\bar{y}_t\,\bar{x}_t}{\bar{X}_N} \qquad (2.3)$$

The expressions for biases of $(\bar{y}_p, \bar{y}_p)$ to the terms of order o(n^{-1}) can be respectively written as

$$B_1(\bar{y}_p) = \frac{\bar{y}_N}{n}\left\{1+(\text{n}-1)\rho_w\right\}k^* c_x^2 ,$$

where $K^* = \rho\frac{c_y}{c_x}$

$$B_1(\bar{y}_p) = \frac{\bar{y}_N}{n}\left\{g+(\text{n}-g)\rho_w\right\}k^* c_x^2 \qquad (2.4)$$

By definition, we have

$$V(\bar{y})_{css} = \frac{1}{N}\sum_{r=1}^{n}(\bar{y}_r - \bar{Y}_N)^2$$

$$= \frac{1}{N}\sum_{r=1}^{N}\left(\frac{1}{n}\sum_{j=0}^{n-1} y_{r+jk} - \bar{Y}_N\right)^2$$

$$= \frac{\sigma_y^2}{n}\left\{1+(\text{n}-1)\rho_{yw}\right\}$$

$$V(\bar{x})_{css} = \frac{\sigma_x^2}{n}\left\{1+(\text{n}-1)\rho_{xw}\right\}$$

$$COV(\bar{y},\bar{x})_{css} = \frac{\rho_{yx}\sigma_y\sigma_x}{n}\sqrt{1+(n-1)\rho_{yw}}\sqrt{1+(n-1)\rho_{xw}}$$

$$\rho_{yx} = \rho = \frac{E(y_{r+jk}-\bar{Y}_N)\,(x_{r+jk}-\bar{X}_N)}{\sqrt{E(y_{r+jk}-\bar{Y}_N)^2}\sqrt{E(x_{r+jk}-\bar{X}_N)^2}}$$

$$\rho_{yy} = \rho_{yw} = \rho_w = \frac{E(y_{r+jk}-\bar{Y}_N)\,(y_{r+j'k}-\bar{Y}_N)}{E(y_{r+jk}-\bar{Y}_N)^2}$$

$$= \frac{\sum_{r=1}^{N}\sum_{j=0}^{n-1}\sum_{j'>j}(y_{r+jk}-\bar{Y}_N)\,(y_{r\ldots}}{\sum_{r=1}^{N}\sum_{j=0}^{n-1}(y_{r+jk}-\bar{Y}_N)^2}$$

$$= 1-\left(\frac{n}{n-1}\right)\frac{\sigma_{yw}^2}{\sigma_y^2}$$

Here, (σ^2_y, σ^2_{yw}) are they population and within sample variance respectively defined as

$$\sigma^2_{\ y} = \frac{1}{N}\sum_{r=1}^{N}\sum_{j=0}^{n-1}(y_{r+jk}$$

$$\sigma^2_{\ yw} = \sigma^2_{\ w} = \frac{1}{N}\sum_{r=1}^{N}\sigma^2_{rw} = \frac{1}{N}\sum_{r=1}^{N}\frac{1}{n}\sum_{j=0}^{n-1}(y_{r+jk} - \bar{y}_n)^2$$

$$C_Z = \frac{S_Z}{\bar{Z}}, (z = y, x),$$

Here ρ_{xy} is the population correlation coefficient and ρ_w is the interclass correlation coefficient for both the variables (y, x) and have been assumed to be the same (see Murty, 1977, pp.374-375), (c_y, c_x) are the c.v.'s of the variables.

As motivated by Rao (1987), we propose a general class of product type estimators to estimate the population mean $\bar{Y}_N$ which is written as

$$T_{\alpha p} = \propto \bar{y}_p + \{1 - E(f(\propto))\}\bar{y}_{p.} \tag{2.5}$$

where 'α' is a random variable and f(α) is a function of it . Then

$E(T_{\alpha p}) = \bar{y}_N$ if

$$E[\alpha\bar{y}_p - E(f(\alpha))\bar{y}_p] = E(\bar{Y} - \bar{y}_p) \tag{2.6}$$

For which $\alpha = \frac{\bar{X}_N}{\bar{x}}$ and $f(\alpha) = \frac{\bar{x}}{\bar{X}_N}$ is a solution in the sample mean.

Introducing a constant 'q' in the right hand side of (2.6), we can rewrite it as

$$E[\alpha\bar{y}_p - E(f(\alpha))\bar{y}_p] = E(\bar{y} - \bar{y}_p - q'\bar{y}_p + q'\bar{y}_p) \tag{2.7}$$

From equation (2.4), we can have

$$\frac{B(\bar{y}_p)}{B(\bar{y}_{p.})} = \frac{1+(n-1)\rho_w}{g+(n-g)\rho_w} = (1-d) \text{ , (say)}$$

From which, we have

$$d = \frac{(g-1)(1-\rho_w)}{g+(n-g)\rho_w}$$

and $B(\bar{y}_p) = (1-\mathrm{d})B(\bar{y}_{p.})$

$$\mathrm{E}(\bar{y}_p) - \bar{y}_N = (1-d)\{\mathrm{E}(\bar{y}_{p.}) - \bar{y}_N\}$$

$$\mathrm{E}(\bar{y}_p) = \mathrm{E}(\mathrm{d}\bar{y}) + (1-d)\mathrm{E}(\bar{y}_{p.})$$

$$(\bar{y}_p) = (\mathrm{d}\,\bar{y}) + (1-\mathrm{d})(\bar{y}_{p.}) \tag{2.8}$$

Now, the following lemma can be stated.

2.1 Lemma1

If $(\bar{y}_{p.}, \bar{y}_{p.})$ are the estimators based on original samples of size n and g sub-samples each of size m=(n/g), then we have

$$(\bar{y}_p) = \mathrm{d}\,\bar{y} + (1-\mathrm{d})\bar{y}_{p.} \tag{2.9}$$

Using lemma (2.1) in (2.7), we can write

$$E[\alpha\bar{y}_p - E(f(\alpha))\bar{y}_{p.}] = E\left[\left\{1 - q'd\frac{\bar{X}_N}{\bar{x}} + q'\right\}\bar{y}_p - \{1+(1-d)q'\}\bar{y}_{p.}\right] \tag{2.10}$$

From which it follows that a solution is now given as

$$\alpha = (1-q'd)\frac{\bar{X}_N}{\bar{x}} \text{ and } f(\alpha) = \alpha\frac{\bar{x}}{\bar{X}_N}$$

$$\text{with } E(f(\alpha)) = 1 - q'd \tag{2.11}$$

Using (2.11) in (2.6), a general class of exactly unbiased product type estimation in C.S.S. scheme is given as

$$T_{\alpha PU} = q'\bar{y}_p + (1-d\,\mathrm{q}')\bar{y} - (1-d)q'\bar{y}_{p.} \tag{2.12}$$

and we can state the following theorem -:

2.2 Theorem 1

The class of estimators $T_{\alpha p}$ written as

$T_{\alpha p} = \alpha\bar{y}_p + \{1 - E(f(\alpha))\}\bar{y}_{p.}$ would be unbiased if

$\alpha = (1-q'd)\frac{\bar{X}_N}{\bar{x}}$ and $f(\alpha) = \alpha\frac{\bar{x}}{\bar{X}_N}$ for which

$E(f(\alpha)) = (1-d)q'$

3. SAMPLING VARIANCE OF THE CLASS T_{APU}

Form equation (2.12), we write

$$V(T_{\alpha PU}) = q'^2V(\bar{y}_p) + (1-q'd)^2V(\bar{y}) + (1-d)^2q'^2V(\bar{y}_{p.}) + 2q'(1-q'd)Cov(\bar{y}_p,\bar{y}) - 2(1-q'd)(1-d)q'Cov(\bar{y},\bar{y}_{p.}) - 2q'^2(1-d)Cov(\bar{y}_p,\bar{y}_{p.}) \tag{3.1}$$

Following Sukhatme and Sukhatme (1970, pp-162-165),we can write to the order of approximation o(n^{-1}) that

$$V(\bar{y}) = \frac{\bar{Y}_N^2}{n}\{1+(n-1)\rho_w\}C_y^2$$

$$V(\bar{y}_p) = V(\bar{y}_{p.}) = \text{Cov}(\bar{y}_p, \bar{y}_{p.}) \quad (3.2)$$

$$= \frac{\bar{Y}_N^2}{n}\{1+(n-1)\rho_w\}\left[C_y^2 + (1+2k^*)C_x^2\right]$$

$$Cov(\bar{y},\bar{y}_p) = \text{Cov}(\bar{y}_p,\bar{y}_{p.}) = \frac{\bar{Y}_N^2}{n}\{1+(n-1)\rho_w\}(C_y^2 + K^*C_x^2)$$

Substituting the result (3.1) in (3.2) and simplifying it, we get

$$V(T_{\alpha PU}) = d^2q'^2V(\bar{y}_p) + (1-q'd)^2V(\bar{y}) + 2dq'(1-q'd)\text{Cov}(\bar{y}_p,\bar{y})$$

$$= V[dq'\bar{y}_p + (1-q'd)\bar{y}]$$

$$= \frac{\bar{Y}_N^2}{n}\{1+(n-1)\rho_w\}\left[C_y^2 + dq'(\text{dq}'+2k^*)C_x^2\right] \quad (3.4)$$

This is minimum for

$$q' = -\left(\frac{k^*}{d}\right) = q_0' \text{ (say)} \quad (3.4)$$

Using the result(3.4) in (3.3),we get the minimum value of V($T_{\alpha pu}$) as

$$V_0(T_{\alpha PU}) = \frac{\bar{Y}_N^2}{n}\{1+(n-1)\rho_w\}\left(C_y^2 - 2k^{*2}C_x^2\right) \quad (3.5)$$

The $V_0(T_{\alpha PU})$ is equivalent to the approximate variance of usual biased linear regression estimator $\bar{y}_{tr}$ in circular systematic sampling scheme written as

$$\bar{y}_{tr} = \bar{y} + b_{yx}(\bar{X}_N - \bar{x}) \quad (3.6)$$

Here b_{yx} is the sample regression coefficient of y on x in circular systematic sample of size n. Form (3.4) and (2.12), we obtain minimum variance unbiased estimator (M.V.U.E.) in the class $T_{\alpha PU}$ written as

$$T_{\alpha PU0} = -\frac{K}{d}\bar{y}_p + (1+K)\bar{y} + \left(\frac{1-d}{d}\right)K\,\bar{y}_p \quad (3.7)$$

It is pointed out that the class $T_{\alpha PU}$ would be more efficient than the usual sample mean estimator $\bar{y}$ defined under circular systematic sampling scheme according if

$$\text{Either} \quad 0 < q' < -\frac{K}{d}$$

$$\text{Or} \quad -\frac{2K}{d} < q' < 0 \quad (3.8)$$

$$\text{And either} \quad \frac{1}{d} < q' < -\frac{(2K+1)}{d}$$

$$\text{Or} \quad -\frac{(2K+1)}{d} < q' < \frac{1}{d} \quad (3.9)$$

4. EMPIRICAL ILLUSTRATION

To examine the applied usefulness of derived results, we consider the data on 'y' the pound steam used monthly and on 'x' the average atmospheric temperature in degree Fahrenheit from Draper and Smith (1966), pp 615-616. The summarized statistics of the data as

$\bar{Y}_N = 6.328$	$\bar{X}_N = 52.60$
$\sigma_y^2 = 2.64$	$\sigma_x^2 = 286.18$
$C_y^2 = 0.0312$	$C_x^2 = 0.1077$
$S_y^2 = 2.7445$	$S_x^2 = 298.10$
$\rho_w = -0.08$	$\rho_{yx} = -0.845$
$k^* = -0.4547$	d = 0.9643

We have worked out the value of

$$V(.) = \frac{V(T_{\alpha PU})}{\frac{Y_N^2}{n}[1+(n-1)\rho_w]}$$

$$= [C_y^2 + dq'(dq' + 2K^*)C_x^2]$$

and P.R.E. $(.,\bar{y}) = \frac{V(\bar{y})}{V(T_{\alpha PU})} \times 100,$

the percent relative efficiency and are provided in the Table 1.

5. RESULT AND DISCUSSION

Table 1. Value of V(.) and P.R.E. $(.,\bar{y})$

Value of q′	Estimator	V(.)	P.R.E. $(.,\bar{y})$
00.00	$\bar{y}$	31.2*10^{-3}	100.00
$q_0' = 0.4715$	T_{PU0}	8.93*10^{-3}	349.38
$0 < q' < -\frac{2K}{d}$ or $0 < q' < 0.943$	$T_{\alpha PU}$	<31.2*10^{-3}	>100.00
0.4515	$T_{\alpha PU}$	8.97*10^{-3}	347.82
1.4523	$T_{\alpha PU}$	105.27*10^{-3}	29.63
q'=1/d	$\bar{y}_p$ or $\bar{y}_{p.}$	40.93*10^{-3}	76.19
$-\frac{(2K+1)}{d} < q' < \frac{1}{d}$ or $-0.0939 < q' < 1.037$	$T_{\alpha PU}$	<40.95*10^{-3}	>76.19
0.8515	$T_{\alpha PU}$	23.39*10^{-3}	133.39
–0.3515	$T_{\alpha PU}$	76.76*10^{-3}	40.64
–0.0435	$T_{\lambda PU}$	35.5*10^{-3}	87.88

6. CONCLUSION

The above Table revels that the P.R.E. ($T_{\alpha PU0}$)=349.38 which indicates that the estimator $T_{\alpha PU0}$ is the most efficient (optimum) estimator in the class $T_{\alpha PU}$. In practice, one may substitute the estimated values of the variance and covariance in the order to obtain a near optimum value of q'. For the choice of q' in the interval ($0<q'<0.943$), the class $T_{\alpha PU}$ is always more efficient than the sample mean estimator $\bar{y}$. It is also evident that the class $T_{\alpha PU}$ in the interval ($-0.0939<q'<1.037$) is also more efficient than the biased estimator $\bar{y}_p$ as well as it jack-knife version $\bar{y}_p$.

REFERENCES

Draper, N.R. and Smith, H. (1966). *Applied regression analysis*. Wiley series in probability and mathematical statistics, New York.

Kushwaha, K.S. and Singh, H.P. (1989).Class of almost unbiased ratio and product estimator in systematic sampling.*J. Ind. Soc. Agril Statist.*, New Delhi, **vol – 41 No 2**, 193-205.

Murty, M.N. (1977).*Sampling theory and methods*. Statistical publishing society, Kolkatta.

Quenouille, M.H. (1949). Approximate test of correlation in time series, *J. Royal Statist. Soc., London*, **vol 11**, 68-84.

Quenouille, M.H. (1956). Notes on bias in estimation, *Biometrica*, **vol.43**, 353-360.

Rao, T.J. (1971).On certain unbiased product estimators. *Commun. Statist. Theor. Math.* **16(12)**, 3631-3640.

Shukla, N.D. (1971).Systematic sampling and product methods of estimation, Proceeding of all India seminar on demography and statistics, B.H.U., Varanasi, India.

Sudhakar, K. (1978). A note on "circular systematic sampling design". *Sankhya, The Ind. J. of Stats.*, Kolkata, **40(c_1)**, 72-73.

Sukhatme, P.V. and Sukhatme, B.V. (1970).Sampling theory of survey with applications. Iowa Stat. University Press, Iowa, U.S.A.