Research article Available online www.ijsrr.org ISSN: 2279–0543

International Journal of Scientific Research and Reviews

Partitioned Ranked Set Sampling: Introduction To A New Efficient Estimation Technique

K.B.Panda[1*] and M. Samantaray[2]

Department of Statistics, Utkal University, Bhubaneswar, Odisha, India

ABSTRACT

In this paper we introduce a new Ranked Set Sampling Procedure called Partitioned Ranked Set Sampling (PRSS). The estimator based thereon is found to be unbiased for the population mean. It fares better than the usual estimation than in Ranked Set Sampling (RSS) and Simple Random Sampling (SRS).This may be serves as a generalisation of Quartile Ranked Set Sampling (QRSS). Also supremacy of estimation reveals through numerical illustration.

KEY WORDS: Simple Random Sampling, Ranked Set Sampling, Population Mean, Quartile Ranked Set Sampling, Partitioned Ranked Set Sampling.

***Corresponding author:**

K.B. Panda

Department of Statistics,

Utkal University, Bhubaneswar, Odisha, India

E mail - kunjast@utkaluniversity.ac.in

INTRODUCTION

Contributing the technique of Ranked set Sampling to the sample survey method, McIntyre(1952) proposed an unbiased estimator having lesser variance than SRS for population mean. Muttalak (2003) proposed a estimator using QRSS in order to improve the efficiency of estimator and reduce the error in ranking. It is also applicable for reducing error as compared to RSS.

SAMPLING METHODS

Ranked Set Sampling (RSS)

According to RSS, select m^2 elements from a population and divide these elements randomly into m sets each of size m units. The m units in each sample are ranked with respect to a variable of interest without actually measuring them. Then the smallest rank is measured from first set, the second smallest rank from second set and the procedure is continued till the unit with highest rank is measured from last m set. These ranked m observation are accurately measured to form RSS data. If m is small, then the cycle may be repeated for r times so as to obtained a combined sample of size mr.

Quartile Ranked Set Sampling (QRSS)

QRSS procedure involves selection of m random samples each of size m units from population and ranked them within each sample. If sample size m is odd, then select (q_1(m+1))th ranks from first (m-1)/2samples, the median from (m+1)/2th sample and the (q_3(m+1))th ranks from last (m-1)/2 samples . If sample size m is even, then select (q_1(m+1))th rank from the first m/2 samples and (q_3(m+1))th rank from last m/2 samples. If m is small, then the cycle may be repeated for r times to have a combined sample size of mr. The ranked units are then quantified.

Partitioned Ranked Set Sampling (PRSS)

PRSS procedure comprises the following steps:

Step - 1: Select m^2 units from the target population and then divide them into m sets each of having size m.

Step - 2: If the number sample is odd, then select (p (m + 1))th rank from first (m-1/2) sets and (q (m + 1))th rank from last (m-1/2) sets, with median from (m+1/2) set. If sample size is even, then select (p (m + 1))th rank from first (m/2) sets and (q(m+1))th rank from last (m/2) sets.

Where $0\leq p \leq 1$ and p + q = 1, such that p and q stands for pth and qth partitioned value of given observation.

Step - 3 :The above step gives size of m units out of m^2 units.

Step - 4 :The above procedure may be repeated for r times, to have combined sample of size mr units using PRSS.

ESTIMATION OF POPULATION MEAN:

Let $(Y_{11}, Y_{12}, ...Y_{1m}), (Y_{21}, Y_{22}, ...Y_{2m}), ...(Y_{m1}, y_{m2}, ...Y_{mm})$ be m independent random sets of size m each.

Let us assume that, each variable Y_{ij} has the same distribution function F (x) with mean μ and variance σ^2 .

Again, Let ${Y^{(1)}}_{i(p(m+1))}$ and ${Y^{(1)}}_{i(q(m+1))}$ denote the (p(m + 1))th and (q (m + 1))th order statistic of the i^{th} sample respectively. Where i = 1, 2 . . m.

If the size of sample is odd, Then partitioned samples are $\left[Y^{(1)}_{1(p(m+1))k}, Y^{(1)}_{2(p(m+1))k}, ...Y^{(1)}_{h(p(m+1))k,}\right], \left[Y^{(1)}_{(h+1,(m+1)/2)k,}\right], \left[Y^{(1)}_{h+2(q(m+1))k}, Y^{(1)}_{h+3(p(m+1))k}, ..., Y^{(1)}_{m(q(m+1))k,}\right]$

$$\Rightarrow \overline{Y}^{(1)}_{PRSSO} = \frac{1}{m}\left[\sum_{i=1}^{h} Y^{(1)}_{i(p(m+1));m} + Y^{(1)}_{i(h+2);m} + \sum_{i=h+2}^{m} Y^{(1)}_{i(q(m+1));m}\right], \text{ where } h = \frac{m-1}{2}$$

$$\Rightarrow Var(\overline{Y}^{(1)}_{PRSSO}) = \frac{1}{m^2}\left[\sum_{i=1}^{h} Var(Y^{(1)}_{i(p(m+1));m}) + Var(Y^{(1)}_{i(h+2);m}) + \sum_{i=h+2}^{m} Var(Y^{(1)}_{i(q(m+1));m})\right]$$

If the size of sample is even, Then partitioned samples are $\left[Y^{(1)}_{1(p(m+1))k}, Y^{(1)}_{2(p(m+1))k}, ...Y^{(1)}_{l(p(m+1))k,}\right], \left[Y^{(1)}_{l+1(q(m+1))k}, Y^{(1)}_{h+3(p(m+1))k}, ..., Y^{(1)}_{m(q(m+1))k,}\right]$

$$\Rightarrow \overline{Y}^{(1)}_{PRSSE} = \frac{1}{m}\left[\sum_{i=1}^{l} Y^{(1)}_{i(p(m+1));m} + \sum_{i=l+1}^{m} Y^{(1)}_{i(q(m+1));m}\right], \text{ where } l = \frac{m}{2}$$

$$\Rightarrow Var(\overline{Y}^{(1)}_{PRSSE}) = \frac{1}{m^2}\left[\sum_{i=1}^{l} Var(Y^{(1)}_{i(p(m+1));m}) + \sum_{i=l+2}^{m} Var(Y^{(1)}_{i(q(m+1));m})\right]$$

Let $\overline{Y}_{SRS}$ denotes the sample mean of size n. The properties of $\overline{Y}_{PRSS}$ are

Property-1 :

Let Y_{ij} be a random variable having pdf f (y) and cdf F(y). A random sample of size m was selected and ranked. $\overline{Y}_{PRSS}$ is an unbiased estimator of population mean μ , if the given distribution is symmetric about mean .

Proof:

Let us assume that, $Y^{(1)}{}_{(s;m)} = s^{th}$ smallest value of the sample.

Where S = 1, 2 ... m.

The pdf and cdf for $Y^{(1)}{}_{(s;m)}$ are

$$\left.\begin{aligned} f_{s;m}(y) &= \frac{1}{B(s,m-s+1)} F^{s-1}(y)(1-F(y))^{m-s} f(y) \\ F_{s;m}(y) &= FB(F(y);s,m-s+1) \text{ respectively} \end{aligned}\right\} \quad (3.1)$$

where FB= (F(y);S,m-S+1) is a beta distribution function with parameters (S, m – S + 1)

Then, let us denote mean and variance as

$$\left.\begin{aligned} & E(y_{s;m}) = \mu_{s;m} \\ \text{and} \quad & V(y_{s;m}) = \sigma^2{}_{s;m} \text{ respectively.} \end{aligned}\right\} \quad (3.2)$$

Using Taylor series, as given in David and Nagaranjah (2003), we have

$$E(y_{s;m}) = \mu_{s;m} = y)dx; F_{s;m}^{-1}(p_s)$$

$$\text{Let } F_{s;m}(y) = FB(F(y);s,m-s+1) = p_s \quad (3.3)$$

Using above, we have

$$\Rightarrow \mu_{s;m} = F_{s;m}^{-1}(p_s) = F^{-1}[\alpha(s)] \quad (3.4)$$

Where α (S)= PB (P_s;S,m- S +1)

Let $F_{m-s+1;m}^{-1}(y) = FB(F(Y);m-s+1,s) = q_s$

Where $p_s + q_s = 1$

$$\Rightarrow \mu_{m-s+1;m} = F_{m-s+1;m}^{-1}(q_s)$$

$$= F^{-1}[1-\alpha(s)] \quad (3.5)$$

If f (x) follows symmetric about mean, for any 0≤ α(S) ≤1,

Then

$$\Rightarrow F^{-1}[1-\alpha(s)] - \mu = \mu - F^{-1}[\alpha(s)]$$

$$\Rightarrow F^{-1}[1-\alpha(s)] + F^{-1}[\alpha(s)] = \mu_{m-s+1} + \mu_{s;m} = 2\mu \qquad (3.6)$$

Sub case :–

(1) For m is even :

$$\bar{Y}_{PRSSE} = \frac{1}{m}\left(\sum_{i=1}^{l} Y^{(1)}{}_{i(s;m)} + \sum_{i=l+1}^{m} Y^{(1)}{}_{i(m-s+1;m)}\right)$$

where $l = \frac{m}{2}$

$$\Rightarrow E(\bar{Y}_{PRSSE}) = \frac{1}{m}\left(\sum_{i=1}^{l} E(Y^{(1)}{}_{i(s;m)}) + \sum_{i=l+1}^{m} E(Y^{(1)}{}_{i(m-s+1;m)})\right)$$

$$= \frac{1}{m}\left[\left(\frac{m}{2} * \mu_{s;m}\right) + \left(\frac{m}{2} * \mu_{m-s+1;m}\right)\right]$$

$$= \frac{1}{m}\left[\left(\frac{m}{2} * (\mu_{s;m} + \mu_{m-s+1;m}\right)\right]$$

$$= \frac{1}{m}\left[\left(\frac{m}{2} * (2\mu\right)\right] = \mu$$

$$\Rightarrow E(\bar{Y}_{PRSSE}) = \mu \qquad (3.7)$$

Hence, $\bar{Y}_{PRSSE}$ is unbiased estimator of population mean μ .

(2) For m is odd :

$$\bar{Y}_{PRSSE} = \frac{1}{m}\left(\sum_{i=1}^{h} Y^{(1)}{}_{i(s;m)} + \sum_{i=h+2}^{m} Y^{(1)}{}_{i(m-s+1;m)} + Y_{i(M;m)}\right)$$

where $h = \frac{m-1}{2}$ and M=median

$$\Rightarrow E(\bar{Y}_{PRSSE}) = \frac{1}{m}\left(\sum_{i=1}^{h} E(Y_{i(s;m)}) + \sum_{k=h+2}^{m} E(Y_{i(m-s+1;m)}) + E(Y_{i(M;m)}\right)$$

$$= \frac{1}{m}\left[\left(\frac{m-1}{2} * \mu_{s;m}\right) + \left(\frac{m-1}{2} * \mu_{m-s+1;m}\right) + \mu\right]$$

$$= \frac{1}{m}\left[\left(\frac{m-1}{2} * (\mu_{s;m} + \mu_{m-s+1;m}\right) + \mu\right]$$

$$= \frac{1}{m}\left[\left(\frac{m-1}{2} * (2\mu\right) + \mu\right] = \mu$$

$$\Rightarrow E(\overline{Y}_{PRSSE}) = \mu \qquad (3.8)$$

Again, $\overline{Y}_{PRSSE}$ is also unbiased estimators of population mean μ .

Property-2:

The variance of $\overline{Y}_{PRSS}$ is

$$Var(\overline{Y}_{PRSS}) = \frac{1}{m^2}\sum_{i=1}^{m}\sigma^2_{i(s;m)}$$

where $\sigma^2_{i(s;m)} = E\left[Y_{i(s;m)} - E(Y_{i(s;m)})\right]$

and $X_{i(s;m)} = p(m+1)th$ order of kth sample

Then $\sigma^2_{(s;m)} = \int \left(Y - \mu_{(s;m)}\right)^2 f_{s;m}(y)dy$

$$= \int (Y-\mu)^2 f_{s;m}(y)dy - (\mu_{(s;m)} - \mu)^2$$

$$\Rightarrow \sigma^2_{(s;m)} + (\mu_{(s;m)} - \mu)^2 = \int (Y-\mu)^2 f_{s;m}(y)dy$$

$$= \int (Y-\mu)^2 \left(\frac{1}{B(s;m-s+1)} F^{s-1}(y)(1-F(y))^{m-s}\right) f\ (y)dy$$

$$< \int (Y-\mu)^2 f\ (y)dy = \sigma^2$$

where $\frac{1}{B(s;m-s+1)} F^{s-1}(y)(1-F(y))^{m-s} < 1$

$$\Rightarrow \sigma^2_{(s;m)} + (\mu_{s;m} - \mu)^2 < \sigma^2 \qquad (3.9)$$

Taking summation and dividing $\frac{1}{m^2}$ we have

$$\Rightarrow \frac{1}{m^2}\sigma^2_{(s;m)} < \frac{1}{m^2}\sigma^2$$

$$\Rightarrow Var(\overline{Y}_{PRSS}) < Var(\overline{Y}_{SRS}) \qquad (3..10)$$

RELATIVE EFFICIENCY:

The relative efficiency of unbiased estimators using PRSS procedures for estimating population mean with SRS is defined as

$$eff\left(\overline{Y}_{SRS}, \overline{Y}_{PRSS}\right) = \frac{Var\left(\overline{Y}_{SRS}\right)}{Var\left(\overline{Y}_{PRSS}\right)}$$

$$= \frac{A+B}{A} > 1 \qquad (4.1)$$

where $A = \frac{1}{m^2}\sum \sigma^2_{s;m}$ and $B = \frac{1}{m^2}\sum \left(\mu_{s;m} - \mu\right)^2$

Comparison of two estimator for μ based on RSS and PRSS procedures. For this purpose, we define the following Relative Precision (RP).

I. For RSS:

$$RP = \frac{Var\left(\overline{Y}_{SRS}\right)}{Var\left(\overline{Y}_{S;m}\right)}, \text{if } \overline{Y}_{s;m} \text{ is an unbiased estimator}$$

$$= \frac{Var\left(\overline{Y}_{SRS}\right)}{MSE\left(\overline{Y}_{S;m}\right)}, \text{if } \overline{Y}_{s;m} \text{ is a biased estimator}$$

II. For PRSS:

$$RP' = \frac{Var\left(\overline{Y}_{SRS}\right)}{Var\left(\overline{Y}^{(1)}_{s;m}\right)}, \text{if } \overline{Y}^{(1)}_{s;m} \text{ is an unbiased estimator}$$

$$= \frac{Var\left(\overline{Y}_{SRS}\right)}{MSE\left(\overline{Y}^{(1)}_{s;m}\right)}, \text{if } \overline{Y}^{(1)}_{s;m} \text{ is a biased estimator}$$

Let us assume that , the cycle is repeated once. The relative efficiency of RSS , PRSS estimators with sample size m=6, 7, 11 and 12 for each 50,000 iteration were performed for 10 symmetric and asymmetric distribution for each simulation.

Table-1: RP for RSS and PRSS at p= 25% w.r.t. SRS with sample size 6,7, 11 and 12

Distribution	m	RP		RP'	
			Bias		Bias
Uniform(0,1)	6	3.400	0.000	3.114	0.000
	7	3.815	0.000	3.706	0.000
	11	6.213	0.000	5.617	0.000
	12	6.500	0.000	6.649	0.000
Uniform(0,2)	6	3.400	0.000	3.132	0.000
	7	3.815	0.000	3.671	0.000
	11	6.6213	0.000	5.632	0.000
	12	6.503	0.000	6.651	0.000
Normal(0,1)	6	3.191	0.000	3.593	0.000
	7	3.585	0.000	3.927	0.000
	11	5.112	0.000	5.980	0.000
	12	5.237	0.000	6.127	0.000
Normal(1,2)	6	3.110	0.000	3.445	0.000
	7	3.535	0.000	4.251	0.000
	11	5.195	0.000	6.240	0.000
	12	5.652	0.000	6.412	0.000
Logistic(-1,1)	6	2.668	0.000	3.592	0.000
	7	3.243	0.000	4.112	0.000
	11	4.599	0.000	6.755	0.000
	12	4.911	0.000	6.728	0.000
Exponential(1)	6	2.135	0.000	2.995	0.219
	7	2.564	0.000	3.213	0.049
	11	3.671	0.000	3.542	0.105
	12	3.922	0.000	4.693	0.061
Exponential(2)	6	2.207	0.000	3.122	0.168
	7	2.476	0.000	2.751	0.013
	11	3.659	0.000	3.521	0.053
	12	3.962	0.000	4.735	0.031
Gamma(1,2)	6	2.218	0.000	3.022	0.183
	7	2.537	0.000	3.111	0.012
	11	3.638	0.000	3.539	0.314
	12	3.990	0.000	4.711	0.184
Gamma(1,3)	6	2.416	0.000	3.025	0.0279
	7	2.669	0.000	3.282	0.023
	11	3.728	0.000	3.594	0.210
	12	3.918	0.000	4.697	0.123
Weibull(1,3)	6	2.459	0.000	3.029	0.274
	7	2.755	0.000	3.334	0.227
	11	3.699	0.000	3.576	0.313
	12	3.960	0.000	4.751	0.158

From above, we get following information,

1. A gain in efficiency attainted using PRSS for estimation population mean for symmetric distribution. As example for N(1,2) with m=12, the relative efficiency of the PRSS is 6.412 comparing it, with RSS is 5.652.

2. For asymmetric asymmetric distribution, gain in efficiency is attainted with smaller bias using DPRSS. for example, for Weibull with m=12, the relative efficiency of PRSS is 4.751 with bias 0.185.for estimating population mean having parameter1 and 3, comparing with RSS and MRSS is 3.960.

PRSS with errors in ranking:

Dell and clutter (1973) showed that the sample mean using RSS is unbiased estimator of the population mean regardless of whatever the ranking is perfect or not and has a smaller variance comparing with SRS with same sample size.

Muttalak (2003), showed QRSS with error in ranking is unbiased estimator of population mean when the distribution follows symmetric about its mean. Hence, applying the above with PRSS method in ranking with error may be defined as follows

Let $Y^*_{i(p(m+1))}th$ and $Y^*_{i(q(m+1))}th$ are the first and last judgement partitioned unit from the ith sample (i = 1, 2...M) respectively with errors in ranking.

The estimator of population mean with error in ranking may be defined as

1. If size of m is even,

$$\text{then } \hat{Y}^*_{PRSSE_e} = \frac{1}{m}\left(\sum_{i=1}^{l} Y^*_{i(p(m+1))} + \sum_{i=l+1}^{m} Y^*_{i(q(m+1))}\right), \qquad \text{where } l = \frac{m}{2}$$

2. If size of m is odd,

$$\text{then } \hat{Y}^*_{PRSSO_e} = \frac{1}{m}\left(\sum_{i=1}^{h} Y^*_{i(p(m+1))} + \sum_{i=h+1}^{m} Y^*_{i(q(m+1))} + Y^*_{h+1(\frac{m+1}{2})}\right), \text{ where } h = \frac{m-1}{2} \qquad (5.1)$$

The estimator of population mean μ with error in ranking has the following properties

(i) $\hat{Y}_{PRSS_e}$ is unbiased estimator of the population mean, if the population is symmetric about mean

(ii) $Var(\hat{Y}_{PRSS_e}) \leq Var(\overline{Y}_{SRS})$

(iii) For asymmetric distribution about its mean $MSE(\hat{Y}_{PRSS_e}) \leq Var(\overline{Y}_{SRS})$

CONCLUSION

In this paper, it is observed that, the PRSS estimator is unbiased of population mean if the given distribution is symmetric, and is also more efficient than the SRS, RSS, and gives almost equal result with QRSS when p=25%. In case of asymmetric distribution, although there is small bias but give better result than RSS.

REFERENCES

1. Al- Saieh M.F.; and Al-Kadiri, M.; Double ranked Set Sampling. statist. Probab. Lett., 2000; 48: 205-212
2. Dell, T.R.; and Clutter, J.L.; Ranked Set Sampling theory with order statistics background. Biometrics, 1972; 28: 545 -555.
3. McIntyre, G.A.; A method of unbiased selective sampling, using Ranked sets. Australian Journal of Agriculture Research, 1952; 3: 385-390
4. Muttalak, H.A.; Median Ranked set Sampling. Journal of Applied Statistical Science, 1997; 6: 245-255.
5. Muttlak, H.A.; Investigating the use of quintiles ranked set samples foe estimating the population mean. J. Appl. Math. Comp., 2003;146: 437-443.
6. Samawi, H.M.; Varied Set size ranked Set sampling with applications to mean and ratio estimation. International Journal of modelling and simulation, 2011; 31: 6-13.
7. Strokes, S.L.; Parametric ranked set sampling. Annals of the Institute of Statistical Mathematics , 1995; 47: 465- 482
8. Takahasi, K.; and Wakimoto, K. ; On unbiased estimates of the population mean based on the sample stratified by means of ordering. Annals of the institute of Statistical Mathematics, 1968; 20 : 1-31.