Open Journal of Statistics, 2015, 5, 97-103
Published Online April 2015 in SciRes. http://www.scirp.org/journal/ojs
http://dx.doi.org/10.4236/ojs.2015.52012

Estimation of the Population Mean Using Paired Ranked Set Sampling

B. S. Biradar[1], C. D. Santosha[2]

[1]Department of Studies in Statistics, University of Mysore, Mysore, India
[2]All India Institute of Speech and Hearing, Mysore, India
Email: biradarbs@statistics.uni-mysore.ac.in, getsanthoshcd@gmail.com

Received 24 February 2015; accepted 1 April 2015; published 13 April 2015

Abstract

In the situation where the sampling units in a study can be easily ranked than quantified, the ranked set sampling methods are found to be more efficient and cost effective as compared to SRS. In this paper we propose an estimator of the population mean using paired ranked set sampling (RSS) method. The proposed estimator is an unbiased estimator of the population mean when the set size is even. In case of odd set size the estimator is unbiased when the underlying distribution is symmetric. It is shown that the proposed estimator is more efficient than its counterpart SRS method for all distributions considered in this study.

Keywords

Ranked Set Sampling, Paired Ranked Set Sampling, Population Mean, Relative Efficiency, Errors in Ranking

1. Introduction

Ranked set sampling (RSS) enables one to provide more structure for the collected sample items, and use this structure to develop efficient inferential procedures. This approach to data collection was first proposed by McIntyre ([1], reprinted in [2]) for situations where taking the actual measurements for sample observations was difficult (maybe costly, destructive, time-consuming), but mechanisms for either informally or formally ranking a set of sample units was relatively easy and reliable. In RSS one first draws m^2 units at random from the population and partitions them into m sets of m units. The m units in each set are ranked without making actual measurements. From the first set of m units the unit ranked lowest is chosen for actual quantification. From the second set of m units the unit ranked second lowest is measured. This process is continued until the unit ranked largest is measured from the m-th set of m units. If a larger sample size is required then the procedure can be re-

How to cite this paper: Biradar, B.S. and Santosha, C.D. (2015) Estimation of the Population Mean Using Paired Ranked Set Sampling. *Open Journal of Statistics*, **5**, 97-103. http://dx.doi.org/10.4236/ojs.2015.52012

peated r times to obtain a sample of size $n = rm$. These chosen elements are called a ranked set sample.

Dell and Clutter [3] and Takahasi and Wakimoto [4] provided mathematical foundations for RSS. Dell and Clutter [3] also showed that the estimator for population mean based on RSS is at least as efficient as the estimator based on SRS with the same number of measurements even when there were ranking errors. Samawi *et al*. [5] used extreme ranked set sample (ERSS) in case of even sample size which is easier to use than the usual RSS procedure to estimate the population mean. Muttlak [6] proposed the use of median ranked set sampling (MRSS) method for estimating the population mean. Muttlak [7] investigated quartile ranked set sampling (QRSS) for estimating the population mean. Jemain *et al*. [8] suggested balanced groups ranked set sampling (BGRSS) for estimating population mean. Biradar and Santosha [9] studied the use of extremes RSS for estimating population mean. Recent summaries of RSS literature appear in two survey articles by Wolfe [10] [11] and a monograph by Chen *et al*. [12]. These procedures are based on quantification of single unit from each sample. However, more than one order statistics from each sample contain additional information about the unknown parameter. Therefore it is sensible to have more than one quantified observations (order statistics) from each sample to construct an estimator or test of a hypothesis. Recently, Balci *et al*. [13] introduced two modified RSS by choosing two elements from each sample. They have studied modified maximum likelihood estimator (MMLE) and best linear unbiased estimator (BLUE) when the underlying distribution is normal**.** The main objective of this paper is to propose a nonparametric estimator using these paired RSS and to compare with estimators based on SRS and extremes RSS (RSS (E)) recently studied by Biradar and Santosha [9] under both perfect and imperfect ranking (with errors in ranking).

2. Ranked Set Sampling by Choosing Diagonals of Samples (RSS (D))

Balci *et al*. [13] introduced modified RSS by choosing paired units from each sample and they have called this sampling scheme as RSS (D).

The procedure of RSS (D) is described as follows:

1) Select m simple random samples each of size m.

2) Each sample is ranked in itself as in ranked set sampling design.

3) Then the i-th smallest and ($m + i - 1$)-th largest order statistics from i-th sample for $i = 1, 2, \cdots, m$ are measured.

4) Repeat above steps r times until the desired sample size $n = 2rm$ is obtained.

We assume that the i-th lowest and ($m + i - 1$)-th largest units of this set can be detected visually, or by any other means easily.

Let $X_1, X_2, \cdots, X_{2m}$ be a random sample of size $2m$ with probability density function $f(x)$ with a finite mean μ and variance σ^2. Let $\bar{X}$ be the mean of the SRS of size $2m$. The mean and variance of $\bar{X}$ are known to be $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \sigma^2/2m$, respectively. Let $\{X_{i1}, X_{i2}, \cdots, X_{im}\}$, $i = 1, 2, \cdots, m$ be m sets of independent random samples each of size m from a population with distribution function $F(x)$ and probability density function $f(x)$ with mean μ and variance σ^2. Let $X_{i(i)}$ and $X_{i(m+1-i)}$ denote the i-th and ($m + 1 - i$)-th order statistics of the i-th sample respectively, $(i = 1, 2, \cdots, m)$. Then

$$\left\{X_{1(1)}, X_{1(m)}, X_{2(2)} X_{2(m-1)}, \cdots, X_{i(i)}, X_{i(m+1-i)}, \cdots, X_{m(m)}, X_{m(1)}\right\}$$

is a RSS (D) of size $2m$. Note that the order statistics within the sample are dependent and between the samples are independent. For all $i = 1, 2, \cdots, m$, let

$$\mu_{(i)} = E\left(X_{i(i)}\right),$$

$$\mu_{(m+1-i)} = E\left(X_{i(m+1-i)}\right),$$

$$\sigma^2_{(i)} = Var\left(X_{i(i)}\right),$$

$$\sigma^2_{(m+1-i)} = Var\left(X_{i(m+1-i)}\right),$$

$$\sigma_{(i,m+1-i)} = Cov\left(X_{i(i)}, X_{i(m+1-i)}\right).$$

The estimator of the population mean based on RSS (D) can be defined in case of even sample size m as

$$\bar{X}_D = \frac{1}{2m}\sum_{i=1}^{m}\left(X_{i(i)} + X_{i(m+1-i)}\right). \tag{1}$$

The mean and variance of $\bar{X}_D$ can be shown to be

$$E\left(\bar{X}_D\right) = \mu$$

and

$$Var\left(\bar{X}_D\right) = \frac{1}{4m^2}\sum_{i=1}^{m}\left(\sigma_{(i)}^2 + \sigma_{(m+1-i)}^2 + 2\sigma_{i,(m+1-i)}^2\right). \tag{2}$$

In case of an odd sample size m, the estimator of the population mean can be defined as

$$\bar{X}_D = \frac{1}{(2m-1)}\left(\sum_{i=1}^{m} X_{i(i)} + \sum_{\substack{i=1 \\ i\neq\frac{m+1}{2}}}^{m} X_{i(m+1-i)}\right). \tag{3}$$

And it follows that

$$Var\left(\bar{X}_D\right) = \frac{1}{(2m-1)^2}\left(\sum_{i=1}^{m}\sigma_{(i)}^2 + \sum_{\substack{i=1 \\ i\neq\frac{m+1}{2}}}^{m}\sigma_{(m+1-i)}^2 + 2\sum_{\substack{i=1 \\ i\neq\frac{m+1}{2}}}^{m}\sigma_{i,(m+1-i)}\right). \tag{4}$$

If the underlying distribution is symmetric about zero, then $X_{(i)} \underline{\underline{d}} -X_{(m+1-i)}$ for $i = 1, 2, \cdots, m$. Arnold *et al.* [14] have shown that $\mu_{(i)} = -\mu_{(m+1-i)}$ and $\sigma_{(i)}^2 = \sigma_{(m+1-i)}^2$ for $i = 1, 2, \cdots, m$. This implies that if m is odd, $\mu_{((m+1)/2)} = \mu = 0$

Using the above results for odd sample size $E\left(\bar{X}_D\right) = 0$ and

$$Var\left(\bar{X}_D\right) = \frac{2}{(2m-1)^2}\left[\sum_{\substack{i=1 \\ i\neq\frac{m+1}{2}}}^{m}\left(\sigma_{(i)}^2 + \sigma_{i,(m+1-i)}\right) + \frac{1}{2}\sigma_{\left(\frac{m+1}{2}\right)}^2\right]. \tag{5}$$

3. Efficiency

The efficiency of $\bar{X}_D$ with respect to $\bar{X}$ for estimating the population mean is defined as

$$Eff\left(\bar{X}_D, \bar{X}\right) = \frac{Var\left(\bar{X}\right)}{Var\left(\bar{X}_D\right)}.$$

Similarly, we compare the proposed estimator $\bar{X}_D$ with the estimator based on RSS (E) studied by Biradar and Santosha [9]. Denote

$$X_{i(1)} = \min\left\{X_{i1}, X_{i2}, \cdots, X_{im}\right\} \text{ and } X_{i(m)} = \max\left\{X_{i1}, X_{i2}, \cdots, X_{im}\right\},\ i = 1, 2, \cdots, m.$$

Then

$$\left\{X_{1(1)}, X_{1(m)}, X_{2(1)}X_{2(m)}, \cdots, X_{m(1)}, X_{m(m)}\right\}$$

is a RSS (E) of size $2m$. For all $i = 1, 2, \cdots, m$, let $\mu_{(1)} = E\left(X_{i(1)}\right)$, $\mu_{(m)} = E\left(X_{i(m)}\right)$, $\sigma_{(1)}^2 = Var\left(X_{i(1)}\right)$, $\sigma_{(m)}^2 = Var\left(X_{i(m)}\right)$ and $\sigma_{(1,m)} = Cov\left(X_{i(1)}, X_{i(m)}\right)$. Then the estimator of the population mean based on RSS (E) is defined by

$$\bar{X}_E = \frac{1}{2m}\sum_{i=1}^{m}\left(X_{i(1)} + X_{i(m)}\right) \quad (6)$$

where $\bar{X}_{(1)} = \frac{1}{m}\sum_{i=1}^{m} X_{i(1)}$ and $\bar{X}_{(m)} = \frac{1}{m}\sum_{i=1}^{m} X_{i(m)}$.

Note that if the underlying distribution is symmetric about its mean then $\bar{X}_E$ is an unbiased estimator of the population mean.

The variance of the of $\bar{X}_E$ is given by

$$Var\left(\bar{X}_E\right) = \frac{1}{4m}\left(\sigma_{(1)}^2 + \sigma_{(m)}^2 + 2\sigma_{(1,m)}\right). \quad (7)$$

The efficiency of $\bar{X}_D$ with respect to $\bar{X}_E$ for estimating the population mean is defined as

$$Eff\left(\bar{X}_D, \bar{X}_E\right) = \frac{Var\left(\bar{X}_E\right)}{Var\left(\bar{X}_D\right)}.$$

The relative efficiencies were computed for m = 2(2)10 and are presented in **Table 1**. Considering the results in **Table 1**, a gain in efficiency is obtained by using RSS (D) for different values of m and for all the distributions considered in this study. The estimator $\bar{X}_D$ is more efficient than the $\bar{X}_E$ in the case of exponential, normal and logistic distributions. In the case of uniform distribution $Eff\left(\bar{X}_D, \bar{X}_E\right)$ is 1 for m = 2 and then decreases for $m \geq 4$.

Table 1. The variances and relative efficiencies of estimators of population mean using RSS (E), RSS (D), SRS.

Distribution	m	$Var\left(\bar{X}_D\right)$	$Var\left(\bar{X}_E\right)$	$Var\left(\bar{X}\right)$	$Eff\left(\bar{X}_D, \bar{X}\right)$	$Eff\left(\bar{X}_D, \bar{X}_E\right)$
Uniform	2	0.021	0.021	0.021	1.000	1.000
	4	0.006	0.004	0.010	1.667	0.667
	6	0.003	0.001	0.007	2.333	0.500
	8	0.002	0.001	0.005	3.000	0.400
	10	0.001	0.000	0.004	3.667	0.333
Exponential	2	0.250	0.250	0.250	1.000	1.000
	4	0.080	0.128	0.125	1.565	1.603
	6	0.040	0.161	0.083	2.068	3.987
	8	0.025	0.227	0.063	2.535	9.197
	10	0.017	0.315	0.050	2.977	18.728
Normal	2	0.250	0.250	0.250	1.000	1.000
	4	0.075	0.075	0.125	1.677	1.000
	6	0.036	0.039	0.083	2.331	1.101
	8	0.021	0.026	0.063	2.969	1.217
	10	0.014	0.019	0.050	3.596	1.334
Logistic	2	0.822	0.823	0.822	1.000	1.000
	4	0.250	0.286	0.411	1.645	1.145
	6	0.122	0.170	0.274	2.242	1.390
	8	0.073	0.121	0.206	2.809	1.647
	10	0.049	0.093	0.164	3.354	1.903

4. Paired Ranked Set Sampling with Errors in Ranking

Dell and Clutter [3] considered the case in which there were errors in ranking; that is the quantified observation from the *i*-th sample may not be the *i*-th order statistic rather the *i*-th judgement order statistic. They showed that sample mean of RSS with errors in ranking was an unbiased estimator of the population mean regardless of the errors in ranking, and has smaller variance than the usual estimator based on SRS with same sample size. But the variance of the estimator with errors in ranking will be larger than the variance of the estimator with perfect ranking and less than or equal to the variance of the estimator based on SRS.

Let $X_{i[i]}$ and $X_{i[m+1-i]}$ denote the *i*-th and (*m* + 1 − *i*)-th judgemnet order statistics of the sample for $i = 1, 2, \cdots, m$. Then

$$\left\{X_{1[1]}, X_{1[m]}, X_{2[2]} X_{2[m-1]}, \cdots, X_{i[i]}, X_{i[m+1-i]}, \cdots, X_{m[m]}, X_{m[1]}\right\}$$

denote RSS (D) sample with errors in ranking. The estimators of the population mean using RSS (D) with errors in ranking is defined as

$$\tilde{X}_D = \frac{1}{2m}\sum_{i=1}^{m}\left(X_{i[i]} + X_{i[m+1-i]}\right), \quad \text{when } m \text{ is even,} \tag{8}$$

$$\tilde{X}_D = \frac{1}{(2m-1)}\left(\sum_{i=1}^{m} X_{i[i]} + \sum_{\substack{i=1 \\ i \neq \frac{m+1}{2}}}^{m} X_{i[m+1-i]}\right), \quad \text{when } m \text{ is odd,} \tag{9}$$

with variance

$$Var\left(\bar{X}_D\right) = \frac{1}{4m^2}\sum_{i=1}^{m}\left(\sigma_{[i]}^2 + \sigma_{[m+1-i]}^2 + 2\sigma_{i,[m+1-i]}\right), \quad \text{if } m \text{ is even,} \tag{10}$$

and

$$Var\left(\bar{X}_D\right) = \frac{1}{(2m-1)^2}\left(\sum_{i=1}^{m}\sigma_{[i]}^2 + \sum_{\substack{i=1 \\ i \neq \frac{m+1}{2}}}^{m}\sigma_{[m+1-i]}^2 + 2\sum_{\substack{i=1 \\ i \neq \frac{m+1}{2}}}^{m}\sigma_{([i],[m+1-i])}^2\right), \quad \text{if } m \text{ is odd.} \tag{11}$$

To gain some insight of the effect of ranking errors on the efficiencies of the estimators various simulation trails were conducted. We use the simulation method considered by Dell and Clutter [3] and David and Lavine [15]. In the first stage we generate *m* sets of simple random samples $\{X_{i1}, X_{i2}, \cdots, X_{im}\}$, $i = 1, 2, \cdots, m$ from uniform, normal, exponential and logistic distributions. The corresponding *m* sets of random error variables $\{e_{i1}, e_{i2}, \cdots, e_{im}\}$, $i = 1, 2, \cdots, m$ are generated from normal distribution with mean zero and variance σ^2. Define

$$Z_{ij} = X_{ij} + e_{ij}, \quad i = 1, 2, \cdots, m \quad j = 1, 2, \cdots, m,$$

where X_{ij} and e_{ij} are independent. The sets of $\left(Z_{ij}, X_{ij}\right)$ $i = 1, 2, \cdots, m$, $j = 1, 2, \cdots, m$ are ranked with respect to the first components of $\left(Z_{(i):i}, X_{[i]:i}\right)$. The second components are taken as judgement ranked order statistics.

Now the RSS (D) and RSS (E) procedures were used to get the values of the estimators for population mean. Based on 10,000 simulated samples estimates of means and varainces or mean squared error (MSE) of estimators were computed. These trails were run with standard deviation set at 0.05, 0.25, 0.5 and 0.75. The results are presented in **Table 2** and **Table 3**.

The efficieny values in **Table 2** suggest that for all the cases (for allvalues of *m* and distributions considered here) RSS (D) estimator is more efficient than the SRS estimator in the presence of erros in ranking. **Table 2** also shows that efficiency values increase with *m* and decrease with errors in ranking. This indicates that lesser the extent of errors in ranking better the performance of RSS (D) estimator.

Table 2. The relative efficiencies of estimators of population mean based on RSS (D) w.r.t. SRS.

Distribution	m	$\sigma^2=0$	$\sigma^2=0.05$	$\sigma^2=0.25$	$\sigma^2=0.50$	$\sigma^2=0.75$
Uniform	4	1.688	1.327	1.111	1.064	1.039
	6	2.291	1.608	1.189	1.126	1.102
	8	3.044	1.757	1.196	1.106	1.051
	10	3.698	1.889	1.223	1.115	1.058
Exponential	4	1.585	1.504	1.386	1.303	1.264
	6	2.004	1.919	1.692	1.521	1.416
	8	2.555	2.321	1.940	1.800	1.583
	10	2.979	2.811	2.223	1.893	1.719
Normal	4	1.670	1.593	1.433	1.340	1.268
	6	2.303	2.223	1.893	1.656	1.521
	8	2.855	2.660	2.128	1.790	1.607
	10	3.481	3.117	2.324	1.869	1.658
Logistic	4	1.683	1.652	1.586	1.538	1.482
	6	2.195	2.160	2.023	1.916	1.824
	8	2.796	2.715	2.460	2.460	2.248
	10	3.397	3.207	2.852	2.498	2.299

Table 3. The relative efficiencies of estimators of population mean based on RSS (D) w.r.t. RSS (E).

Distribution	m	$\sigma^2=0$	$\sigma^2=0.05$	$\sigma^2=0.25$	$\sigma^2=0.50$	$\sigma^2=0.75$
Uniform	4	0.662	0.733	0.911	0.951	0.962
	6	0.501	0.617	0.838	0.926	0.963
	8	0.409	0.553	0.823	0.913	0.933
	10	0.331	0.425	0.799	0.891	0.921
Exponential	4	1.603	1.624	1.604	1.549	1.512
	6	3.999	4.111	3.796	3.329	2.976
	8	9.423	9.412	8.036	6.666	5.757
	10	18.273	18.249	14.577	11.606	9.675
Normal	4	0.995	0.992	0.982	0.992	0.992
	6	1.109	1.100	1.092	1.062	1.037
	8	1.196	1.161	1.121	1.082	1.063
	10	1.318	1.427	1.169	1.097	1.074
Logistic	4	1.136	1.137	1.131	1.140	1.130
	6	1.381	1.374	1.355	1.356	1.355
	8	1.686	1.672	1.619	1.570	1.542
	10	1.913	1.885	1.801	1.693	1.664

From **Table 3** we can observe that except for uniform distribution RSS (D) estimator performs better than RSS(E) estimator in the presence of errors in ranking. In the case of exponential, normal and logistic distributions the efficency values increase with *m* and decrease with errors in ranking. For uniform distribution the opposite trend can be observed, *i.e.*, efficency values increase with σ^2 and decrease with set size *m*. This indicates that RSS (D) estimator for uniform dustribution improves with samller set size *m* and larger extent of errors in ranking.

References

[1] McIntyre, G.A. (1952) A Method for Unbiased Selective Sampling Using Ranked Sets. *Australian Journal of Agricultural Research*, **3**,385-390. http://dx.doi.org/10.1071/AR9520385

[2] McIntyre, G.A. (2005) A Method for Unbiased Selective Sampling, Using Ranked Sets. *The American Statistician*, **59**, 230-232. *Originally Appeared in Australian Journal of Agricultural Research*, **3**, 385-390. http://dx.doi.org/10.1198/000313005X54180 http://dx.doi.org/10.1071/AR9520385

[3] Dell, T.R. and Clutter, J.L. (1972) Ranked Set Sampling Theory with Order Statistics Background. *Biometrics*, **28**, 545-555. http://dx.doi.org/10.2307/2556166

[4] Takahasi, K. and Wakimoto, K. (1968) On Unbiased Estimates of the Population Mean Based on the Sample Stratified by Means of Ordering. *Annals of the Institute of Statistical Mathematics*, **20**, 1-31. http://dx.doi.org/10.1007/BF02911622

[5] Samwi, H., Ahmad, M. and Abu-Dayyeh, W. (1996) Estimating the Population Mean Using Extreme Ranked Set Sampling. *Biometrical Journal*, **38**, 577-586. http://dx.doi.org/10.1002/bimj.4710380506

[6] Muttlak, H.A. (1997) Median Ranked Set Sampling. *Journal of Applied Statistical Sciences*, **6**, 245-255.

[7] Muttlak, H.A. (2003) Investigating the Use of Quartile Ranked Set Samples for Estimating the Population Mean. *Journal of Applied Mathematics and Computation*, **146**, 437-443. http://dx.doi.org/10.1016/S0096-3003(02)00595-7

[8] Jemain, A.A., Al-Omari, A. and Ibrahim, K. (2008) Some Variations of Ranked Set Sampling. *Electronic Journal of Applied Statistical Analysis*, **1**, 1-15.

[9] Biradar, B.S. and Santosha, C.D. (2015) Estimation of the Population Mean Based on Extrmemes Ranked Set Sampling. *American Journal of Mathematics and Statistics*, **5**, 32-36.

[10] Wolfe, D.A. (2004) Ranked Set Sampling: An Approach to More Efficient Data Collection. *Statistical Science*, **19**, 636-643. http://dx.doi.org/10.1214/088342304000000369

[11] Wolfe, D.A. (2010) Ranked Set Sampling. *Wiley Interdisciplinary Reviews*: *Computational Statistics*, 2, 460-466. http://dx.doi.org/10.1002/wics.92

[12] Chen, Z.H., Bai, Z.D. and Sinha, B.K. (2004) Ranked Set Sampling: Theory and Applications. Springer, New York. http://dx.doi.org/10.1007/978-0-387-21664-5

[13] Balci, S., Akkaya, A.D. and Ulgen, B.E. (2013) Modified Maximum Likelihood Estimators Using Ranked Set Sampling. *Journal of Computational and Applied Mathematics*, **238**,171-179. http://dx.doi.org/10.1016/j.cam.2012.08.030

[14] Arnold, B.C., Balkrishna, N. and Nagaraj, H.N. (1992) A First Course in Order Statistics. John Wiley and Sons, New York.

[15] David, H.A. and Levine, D.N. (1972) Ranked Set Sampling in the Presence of Judgment Ranking Error. *Biometrics*, **28**, 553-555.