

Estimation of the Population Mean Based on Extremes Ranked Set Sampling

B. S. Biradar^{1,*}, Santosha C. D.²

¹Department of Studies in Statistics, University of Mysore, Mysore, India

²All India Institute of Speech and Hearing, Mysore, India

Abstract This paper is concerned with ranked set sampling theory which is useful to estimate the population mean when the order of a sample of small size can be found without measurements or with other methods. In practice ranking a sample of moderate size and observing the i -th ranked unit (ranking of middle ordered units) is a difficult task. Therefore, in this paper we propose two estimators of the population mean based on extremes ranked set sampling methods. The proposed estimators are unbiased for the population mean when the underlying distribution is symmetric. It is shown that the proposed estimators are more efficient than their counter part simple random sampling method for distributions considered in this study.

Keywords Ranked set sampling, Extremes ranked set sampling, Population mean, Relative efficiency, Errors in Ranking

1. Introduction

Ranked set sampling (RSS) was introduced by McIntyre [1] (reprinted in [2]) for estimating the pasture and forage yields. McIntyre indicates that RSS is a more efficient sampling method than simple random sampling (SRS) method for estimating the population mean. It is appropriate for situations where quantification of sampling units is either costly or difficult, but ranking the units in a small set is easy and inexpensive. Ranking can be performed based on expert judgment, visual inspection or any means that does not involve actually quantifying the observations. 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. The 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 repeated r times to obtain a sample of size $n=rm$. These chosen elements are called a ranked set sample.

Takashi and Wakimoto [3] established a very important statistical foundation for the theory of RSS. They proved that the sample mean of the RSS is an unbiased estimator of the population mean with smaller variance than the sample mean of a SRS with the same sample size. Dell and Clutter [4] showed that the mean of the RSS is an unbiased estimator of

the population mean, whether or not there are errors in ranking and more efficient than the mean of SRS. Stokes [5] showed that the estimator of the variance based on RSS is an asymptotically unbiased estimator of the population variance, and it is more efficient than the usual estimator based on SRS data with same sample size, when the sample size is large. Stokes and Sager [6] studied the empirical distribution function based on RSS and showed that it is an unbiased estimate of the underlying distribution function. A recent summaries of RSS literature appear in two survey articles by Wolfe [7] and [8] and a monograph by Chen et al. [9].

Samawi et al. [10] 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 [11] proposed the use of median ranked set sampling (MRSS) method for estimating the population mean. Muttlak [12] investigated quartile ranked set sampling (QRSS) for estimating the population mean. Jemain et al. [13] suggested balanced groups ranked set sampling (BGRSS) for estimating population mean. These procedure are based on ranking the i -th unit (other than the extremes) of the sample which is a difficult task. Recently, Balci et al. [14] introduced another RSS based on extremes for both even and odd sample sizes. They have studied modified maximum likelihood estimator (MMLE) and Best linear unbiased estimator (BLUE) when the underlying distribution is normal.

In this paper another modification RSS is introduced namely independent extreme ranked set sampling (IERSS) in the case of both even and odd sample sizes. The main objective of this paper is to propose nonparametric estimators of the population mean using these two extremes RSS and are compared with estimator based on SRS method.

* Corresponding author:

biradarbs@statistics.uni-mysore.ac.in (B. S. Biradar)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2015 Scientific & Academic Publishing. All Rights Reserved

Since it is not difficult to identify maximum or minimum units, extremes RSS is a very useful modification of RSS. It allows for an increase in set size without introducing too many ranking errors.

2. Estimation of Population Mean Based on Two New RSS

In this study two new estimators of population mean based on modified ranked set sampling methods (i) choosing both extremes of each sample and (ii) choosing extreme form two independent samples have been developed.

2.1. Ranked Set Sampling by Choosing Extremes of Samples (RSS(E))

Balci et al. [14] introduced ranked set sample by choosing extremes of the samples and they have called this sampling scheme as RSS(E). The procedure of RSS(E) is described as follows:

1. Select m random samples each of size m.
2. Each sample is ranked in itself as in ranked set sampling design.
3. Then smallest and largest order statistics from each sample are observed.
4. Repeat above steps r times until the desired sample size n= 2rm is obtained.

We assume that the lowest and largest units of this set can be detected visually, or by any other means easily.

Let X_1, X_2, \dots, X_{2m} be a random sample of size 2m with probability density function $f(x)$ with mean μ and variance σ^2 . Let $\{X_{i1}, X_{i2}, \dots, X_{im}\}$, $i=1,2,\dots,m$ be m sets of independent random samples each of size m from a population with distribution function $F(x)$ and pdf $f(x)$ with mean μ and variance σ^2 . Denote

$$X_{i(1)} = \min\{X_{i1}, X_{i2}, \dots, X_{im}\} \quad \text{and}$$

$$X_{i(m)} = \max\{X_{i1}, X_{i2}, \dots, X_{im}\} \quad i=1,2,\dots,m. \quad \text{Then}$$

$\{X_{1(1)}, X_{1(m)}, X_{2(1)}, X_{2(m)}, \dots, X_{m(1)}, X_{m(m)}\}$ is a

RSS(E) of size 2m. Note that the order statistics within the sample are dependent and between the samples are independent. For all $i= 1,2,\dots,m$ let $\mu = E(X_i)$,

$$\sigma^2 = Var(X_i), \quad \mu_{(1)} = E(X_{i(1)}), \quad \mu_{(m)} = E(X_{i(m)}),$$

$$\sigma_{(1)}^2 = Var(X_{i(1)}) \quad , \quad \sigma_{(m)}^2 = Var(X_{i(m)}) \quad \text{an}$$

$$\sigma_{(1,m)} = Cov(X_{i(1)}, X_{i(m)}).$$

Let \bar{X} be the mean of the SRS of size 2m. The mean and variance of \bar{X} are $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \sigma^2 / 2m$, respectively. The estimator of the population mean based on RSS(E) can be defined as

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

\bar{X}_E can be written as

$$\bar{X}_E = \frac{1}{2}(\bar{X}_{(1)} + \bar{X}_{(m)}),$$

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

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

$$E(\bar{X}_E) = \frac{1}{2}(\mu_{(1)} + \mu_{(m)}) \quad (2)$$

$$Var(\bar{X}_E) = \frac{1}{4m}(\sigma_{(1)}^2 + \sigma_{(m)}^2 + 2\sigma_{(1,m)}) \quad (3)$$

If the underlying distribution is symmetric about zero, then $X_{(1)} \stackrel{d}{=} -X_{(m)}$. Arnold et al. [15] have shown that

$\mu_{(1)} = -\mu_{(m)}$ and $\sigma_{(1)}^2 = \sigma_{(m)}^2$. Using these results

$$E(\bar{X}_E) = 0,$$

and

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

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

2.2. Ranked Set Sampling by Choosing Extremes of Two Independent Samples (IERSS)

Here we introduce another modified RSS called independent Extremes ranked set sampling (IERSS) based on two independent samples of size 2m: first we select 2m random samples of size m each and then identify the maximum within each set of first m samples by visual inspection or by some other cheap method, without actual measurement of the variable of interest. Repeat this for other m simple random samples but for the minima. Repeat above steps for r times until the desired sample size n=2rm is obtained. Let $\{X_{i1}, X_{i2}, \dots, X_{im}\}$ and

$\{Y_{i1}, Y_{i2}, \dots, Y_{im}\}$ $i=1,2,\dots,m$ be 2m sets of random samples each of size m from a population with distribution function

$F(x)$ and pdf $f(x)$ with mean μ and variance σ^2 . Denote

$$X_{i(1)} = \min\{X_{i1}, X_{i2}, \dots, X_{im}\} \quad \text{and}$$

$$Y_{i(m)} = \max\{Y_{i1}, Y_{i2}, \dots, Y_{im}\} \quad , \quad i=1,2,\dots,m. \quad \text{Then}$$

$\{X_{1(1)}, X_{2(1)}, \dots, X_{m(1)}, Y_{1(m)}, Y_{2(m)}, \dots, Y_{m(m)}\}$ be

IERSS of size $2m$. Note that elements of sample are independent of each other. The estimator of the population mean based on IERSS with one cycle can be defined as:

$$\bar{X}_{IE} = \frac{1}{2m} \sum_{i=1}^m (X_{i(1)} + Y_{i(m)}) \quad (5)$$

The mean and variance of \bar{X}_{IE} can be shown to be

$$E(\bar{X}_{IE}) = \frac{1}{2}(\mu_{(1)} + \mu_{(m)}) \quad (6)$$

$$Var(\bar{X}_{IE}) = \frac{1}{4m}(\sigma_{(1)}^2 + \sigma_{(m)}^2) \quad (7)$$

Where $\sigma_{(1)}^2$ is as defined above and $\sigma_{(m)}^2 = Var(Y_{i(m)})$. If the underlying distribution is symmetric about zero then using the above results of Arnold *et al.*(1992), we have $E(\bar{X}_E) = 0$ and

$$Var(\bar{X}_{IE}) = \frac{1}{2m} \sigma_{(1)}^2. \quad (8)$$

We can easily see that if the underlying distribution is symmetric about its mean then \bar{X}_{IE} is an unbiased estimator of the population mean.

3. Efficiency

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

$$Eff(\bar{X}_E, \bar{X}) = \frac{Var(\bar{X})}{MSE(\bar{X}_E)}. \quad (9)$$

If the distribution is symmetric then $MSE(\bar{X}_E) = Var(\bar{X}_E)$.

Similarly,

$$Eff(\bar{X}_{IE}, \bar{X}) = \frac{Var(\bar{X})}{MSE(\bar{X}_{IE})}. \quad (10)$$

If the distribution is symmetric then $MSE(\bar{X}_{IE}) = Var(\bar{X}_{IE})$.

And finally

$$Eff(\bar{X}_{IE}, \bar{X}_E) = \frac{MSE(\bar{X}_E)}{MSE(\bar{X}_{IE})}. \quad (11)$$

For uniform distribution over (0,1) the efficiency of \bar{X}_E with respect to \bar{X} is given by

$$Eff(\bar{X}_E, \bar{X}) = \frac{(m+1)(m+2)}{12} > 1, \text{ for } m > 2. \quad (12)$$

Similarly the efficiency of \bar{X}_{IE} with respect to \bar{X} is given by

Table 1. The Relative Efficiency of estimators of population mean using RSS(E) and IERSS

Distribution	N	$Biase(\bar{X}_E)$	$Biase(\bar{X}_{IE})$	$Eff(\bar{X}_E, \bar{X})$	$Eff(\bar{X}_{IE}, \bar{X})$	$Eff(\bar{X}_{IE}, \bar{X}_E)$
Uniform	2	0.000	0.000	1.000	1.500	1.500
	4	0.000	0.000	2.500	3.120	1.250
	6	0.000	0.000	4.667	5.440	1.167
	8	0.000	0.000	7.500	8.437	1.125
	10	0.000	0.000	11.000	12.100	1.110
Exponential	2	0.00	0.00	1.000	1.330	1.333
	4	0.167	0.167	0.977	1.039	1.024
	6	0.308	0.308	0.518	0.526	1.014
	8	0.421	0.421	0.276	0.277	1.004
	10	0.514	0.514	0.159	0.165	1.036
Normal	2	0.000	0.000	1.000	1.467	1.4671
	4	0.000	0.000	1.677	2.030	1.2142
	6	0.000	0.000	2.117	2.404	1.3548
	8	0.000	0.000	2.441	2.710	1.0987
	10	0.000	0.000	2.695	2.904	1.0774
Logistic	2	0.000	0.000	1.000	1.437	1.4368
	4	0.000	0.000	1.435	1.706	1.1872
	6	0.000	0.000	1.613	1.801	1.1170
	8	0.000	0.000	1.706	1.850	1.0831
	10	0.000	0.000	1.763	1.880	1.0662

$$Eff(\bar{X}_{IE}, \bar{X}_E) = \frac{m+1}{m} > 1, \text{ for } m > 1 \text{ and tends to 1 as } m \text{ goes to infinity.}$$

This indicates that as the set size m increases $\sigma_{(1,m)} = Cov(X_{i(1)}, X_{i(m)})$ tends to zero and both estimators \bar{X}_{IE} and \bar{X}_E are equally efficient. The results of efficiency and bias of the estimators are presented in Table 1 for uniform, exponential, normal and logistic distributions using SRS, RSSE and IERSS sampling schemes. From Table 1 we observed that in the case of uniform, normal and logistic distributions the estimators based on \bar{X}_{IE} and \bar{X}_E are both more efficient than \bar{X} and \bar{X}_{IE} is more efficient than \bar{X}_E . For exponential distribution $Eff(\bar{X}_E, \bar{X})$ and $Eff(\bar{X}_{IE}, \bar{X})$ decrease as m increases. This is because \bar{X}_{IE} and \bar{X}_E are bias estimators and bias diverges as m goes to ∞ .

We can easily see that the estimator of population mean based on \bar{X}_{IE} is equivalent to the one of the estimators proposed by Samawi et al. [10] based on extreme ranked set sample with number of cycles equal two and when set size m is even.

4. Extreme RSS with Errors in Ranking

Dell and Clutter [4] considered the case in which there are errors in ranking; that is, the quantified observation from the i -th sample in the j -th cycle may no be the i -th order statistic but rather the i -th judgement order statistic. They showed that the sample mean of RSS with errors in ranking is unbiased estimator of the population mean regardless of the errors in ranking and has smaller variance than the usual estimator based on SRS with the same sample size.

Let $X_{i[1]}$ and $X_{i[m]}$ denote the smallest and largest judgment order statistic of the sample respectively ($i=1,2,\dots,m$). Then $\{X_{1[1]}, X_{1[m]}, X_{2[1]}, X_{2[m]}, \dots, X_{m[1]}, X_{m[m]}\}$ and $\{X_{1[1]}, X_{2[1]}, \dots, X_{m[1]}, Y_{1[m]}, Y_{2[m]}, \dots, Y_{m[m]}\}$ denote RSS(E) and IERSS samples with errors in ranking. The estimators of the population mean using RSS(E) and IERSS with errors in ranking can be defined as

$$\tilde{X}_E = \frac{1}{2m} \sum_{i=1}^m (X_{i[1]} + X_{i[m]}), \tag{13}$$

and

$$\tilde{X}_{IE} = \frac{1}{2m} \sum_{i=1}^m (X_{i[1]} + Y_{i[m]}) \tag{14}$$

The variance of \tilde{X}_E and \tilde{X}_{IE} can be defined as

$$Var(\tilde{X}_E) = \frac{1}{4m} (\sigma_{[1]}^2 + \sigma_{[m]}^2 + 2\sigma_{[1,m]}) \tag{15}$$

and

$$Var(\tilde{X}_{IE}) = \frac{1}{4m} (\sigma_{[1]}^2 + \sigma_{[m]}^2) \tag{16}$$

where $\sigma_{[1]}^2 = Var(X_{i[1]})$, $\sigma_{[m]}^2 = Var(X_{i[m]})$ and $\sigma_{[1,m]} = Cov(X_{i[1]}, X_{i[m]})$ for $i=1,2,\dots,m$.

It can easily seen that \tilde{X}_E and \tilde{X}_{IE} are unbiased estimators of the population mean if the underlying distribution is symmetric about its mean.

5. Conclusions

We proposed two new estimators of the population mean using two modified ranked set sampling methods. The proposed estimators are unbiased of the population mean when the underlying distribution is symmetric about its mean. We showed that both estimators have smaller variances than the estimator using SRS and provide more efficient estimators. The estimators using extremes RSS (RSS (E) and IERSS) will reduce errors in ranking compared to RSS, MRSS, QRSS and BGRSS, since we have to identify and measure the smallest and largest of the i th sample.

ACKNOWLEDGEMENTS

The Authors are very grateful to the referee and editor for their thoughtful comments and suggestions.

REFERENCES

- [1] McIntyre, G.A. (1952). A method for unbiased selective sampling using ranked sets. *Aust.J Agri.Res.* 3:385-390.
- [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.
- [3] Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Stat. Math.*, 20, 1-31.
- [4] Dell, T.R. and Clutter, J.L. (1972). Ranked Set Sampling Theory with Order Statistics Background. *Biometrics*, 28, 545-555.
- [5] Stokes, S.L. (1980) Estimation of variance using judgment ordered ranked set samples, *Biometrics*, 36, 35-42.
- [6] Stokes, S.L and Sager, T. (1988). Characterization of a ranked set sample with application to estimating distribution

- functions, *J. Amer. Statist. Assoc.*, 83, 374-381.
- [7] Wolfe, D.A. (2004). Ranked set sampling: an approach to more efficient data collection. *Statistical Science* 19, 636-643.
- [8] Wolfe, D.A., (2010). Ranked set sampling. *Wiley Interdisciplinary Reviews: Computational Statistics* 2, 460-466.
- [9] Chen, Z., Bai, Z., Sinha, B.K., (2004). *Ranked Set Sampling: Theory and Applications*. Springer, New York.
- [10] Samwi, H., Ahmad, M. and Abu-Dayyeh, W. (1996). Estimating the population mean using extreme ranked set sampling. *Biom. Journal*, 38, 577-586.
- [11] Muttlak, H.A (1997) Median Ranked set sampling, *Journal of applied statistical Sciences*, 6, 245-255.
- [12] 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.
- [13] Jemain A.A., Al-omari A, Ibrahim, K (2008). Some variations of Ranked set sampling. *Electronic J.App. Stat. Anal.*, 1, 1-15.
- [14] Balci, S, Akkaya A.D., Ulgen, B.E (2013) Modified Maximum Likelihood estimators using ranked set sampling. *Journal of Computational and Applied Mathematics* 238, 171-179.
- [15] Arnold, B.C. Balkrishna, N and Nagaraj, H.N. (1992). *A First course in order statistics*. John Wiley and Sons, New York.