Article

Not peer-reviewed version

Weighted Ranked Set Sampling for Skew Distributions

Dinesh S Bhoj and Girish Chandra *

Posted Date: 30 May 2024

doi: 10.20944/preprints202405.2029.v1

Keywords: rdered observations; Neyman's allocation; Relative precision; Skewness; Unbiased estimator; Weight

Article

Weighted Ranked Set Sampling for Skew Distributions

Dinesh S. Bhoj [1] and Girish Chandra [2,*]

[1] Department of Mathematical Sciences, Rutgers University, Camden, New Jersey, USA; dbhoj@camden.rutgers.edu
[2] Department of Statistics, University of Delhi, Delhi-110007, India
* Correspondence: gchandra@stats.du.ac.in

Abstract: Ranked Set Sampling (RSS) is a useful technique for improving the estimator of population mean when the sampling units in a study can be easily ranked than the actual measurement. RSS performs better than simple random sampling (SRS) when the mean of units corresponding to each rank is used. The performance of RSS can be increased further by assigning weights to the ranked observations. In this paper, we propose weighted RSS procedures to estimate the population mean of positively skew distributions. It is shown that the gain in the relative precisions of the population mean for chosen distributions are uniformly higher than those based on RSS. The gains in relative precisions are substantially high. Further, the relative precisions of our estimator are slightly higher than the ones based on Neyman's optimal allocation model for small sample sizes. Moreover, it is shown that, the performance of the proposed estimator increases as the skewness increases by using the example of lognormal family of distributions.

Keywords: ordered observations; Neyman's allocation; relative precision; skewness; unbiased estimator; weight

MSC: 62D05; 62G09; 94A20

---

1. Introduction

The Ranked Set Sampling (RSS) procedure has been used advantageously in agriculture, forestry, environmental, ecological and recently in human studies where the exact measurement of units is either difficult or expensive. For example, in forestry, the measurement of stem volume of standing trees is difficult but the ranking of the trees using their height and diameter at breast height is rather easy. For such situations, McIntyre (1952) introduced RSS to estimate the population mean. The RSS is a cost-efficient alternative to simple random sampling (SRS) if observations can be ranked according to the characteristic under investigation by means of visual inspection or other methods not requiring actual measurements. McIntyre (1952) indicated that the RSS procedure is superior to SRS procedure to estimate the population mean. However, Dell and Clutter (1972) and Takahasi and Wakimoto (1968) provided mathematical foundation for RSS. Dell and Clutter (1972) 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 though when there are ranking errors. Bhoj (2001) introduced RSS with unequal samples. Bhoj and Kushary (2016) proposed RSS with unequal samples for positively skew distributions with heavy right tails. RSS is a nonparametric procedure. However, recently, RSS has also been used in the parametric setup (see Bhoj and Ahsanullah (1996); Bhoj (1997a, 1997b); Lam et al. (1994); Stokes (1995).

The selection of ranked set sample of size k involves drawing k random samples with k units in each sample. The units in each sample are ranked by using judgment or other methods not requiring actual measurements. The unit with lowest rank is measured from the first sample, the unit with second lowest rank is measured from the second sample, and the procedure is continued until the

 doi:10.20944/preprints202405.2029.v1

unit with the highest rank is measured from the last sample. The k^2 ordered observations in k samples can be displayed in the matrix form as:

$$y_{(11)}, y_{(12)}, \cdots, y_{(1k)}$$

$$y_{(21)}, y_{(22)}, \cdots, y_{(2k)}$$

$$\cdots$$

$$y_{(k1)}, y_{(k2)}, \cdots, y_{(kk)}$$

We measure only $k(y_{(ii)}, i = 1,2\ldots, k)$ diagonal observations, and they constitute the RSS. We note that these k observations are independently but not identically distributed. In RSS, k is usually small to reduce the ranking errors and therefore, to increase the sample size, the above procedure is repeated $m \geq 2$ times to get the sample of size $n = mk$. In this paper, we assume m=1.

In the present paper, our main interest is to estimate the population mean for positively skew distributions with longer right tail. We propose estimators based on weighted ranked set sampling (WRSS) and compare their performance with the ones based on the usual RSS procedure and Neyman's optimal allocation model. In section 2, we summarize the estimators of population mean based on RSS procedure and Neyman's optimal solution. In section 3, we propose our WRSS procedure to estimate the population mean of skew distributions. First, we introduce WRSS procedure where we assign one low weight to the highest order statistics and calculated the relative precisions of the estimator based on WRSS, RSS and Neyman's optimal procedure with respect to the estimator based on SRS. The procedures are used to obtain the relative precisions by using the four positively skew distributions. We also computed one set of weights for all four distributions for each k. In section 4, we derived optimal weights for the lowest and highest order statistics for the chosen distributions for each k. We then obtained one set of weights for the lowest and highest order statistics for each k which will maximize the sum of relative precisions of four distributions. In section 5, we generalize the use of all optimal weights for all order statistics for k=4 and k=5 for each distribution. We also obtained one set of weights for each k for the four chosen distributions. In section 6, to see the effect of increasing skewness, the relative precisions of estimators for lognormal family of distributions have been compared. In section 7, we summarize the results with recommendations.

2. Estimation of Mean

We consider first the usual RSS to estimate the population mean. Let $y_{(ii)}, i = 1,2, \ldots, k$ denote the value of characteristic under study of i^{th} order statistic. The mean and variance of the i^{th} rank order statistic for set size k are denoted by $\mu_{(ii)}$ and $\sigma^2_{(ii)}$, respectively. We denote the population mean and variance by μ and σ^2, respectively. Then the unbiased estimator for μ based on RSS is given by

$$\bar{\mu} = \frac{1}{k}\sum_{i=1}^{k} y_{(ii)},$$

with the variance

$$Var(\bar{\mu}) = \frac{1}{k^2}\sum_{i=1}^{k} \sigma^2_{(ii)}.$$

The relative precision of $\bar{\mu}$ compared to the estimator based on SRS with the same number of observations k (Bhoj and Chandra, 2019) is

$$RP_1 = \frac{\sigma^2}{\overline{\sigma^2}}, \tag{2.1}$$

where $\overline{\sigma^2} = \frac{1}{k}\sum_{i=1}^{k} \sigma^2_{(ii)}$ is the average within-rank variance.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2024 doi:10.20944/preprints202405.2029.v1

For the skewed distribution, Neyman's allocation $m_i = \frac{n\sigma_{(ii)}}{\sum_{i=1}^{k}\sigma_{(ii)}}$ provides the optimal allocation and the relative precision of the unbiased estimator of μ based on this model with respect of SRS with the same number of observations n and is given by (Bhoj and Chandra, 2019).

$$RP_3 = \frac{\sigma^2}{\bar{\sigma}^2}, \tag{2.2}$$

where, $\overline{\sigma} = \frac{1}{k}\sum_{i=1}^{k}\sigma_{(ii)}$ is the average within-rank standard deviation.

There are some unequal allocation models for the skew distributions in the literature (see, 't' and 's, t' model (Kaur et al., 1997); Systematic model (Tiwari and Chandra, 2011) and simple model (Chandra et al., 2018 and Bhoj and Chandra, 2019)). The Neyman's allocation does not provide the integer values of m_i which are necessary for any application. The procedure of making them integer is shown in Bhoj and Chandra (2019) and used in this paper. It is noted that the inequality $RP_3 > RP_1$ always holds for the skew distributions.

3. WRSS with One Optimal Weight

In this section, we propose a weighted ranked set sampling (WRSS) with the optimal weight for the largest order statistic since the largest order statistic has the highest variance and higher bias of the estimator for the mean when we deal with the positively skew distributions. We define that the weights w_i $\left(with\ 0 \le w_i \le 1\ and\ \sum_{i=1}^{k} w_i = 1\right)$ as,

$$w_i \propto 1, \text{ for } i = 1, 2, \ldots, k-1$$

$$w_k \propto \frac{1}{C_k},$$

The exact values of weights are proposed as follows:

$$w_i = \frac{1}{(k-1)+\frac{1}{C_k}}, \text{ for } i = 1, 2, \ldots, k-1,$$

$$w_k = \frac{1}{C_k\left((k-1)+\frac{1}{C_k}\right)}$$

Our weighted estimator for the population mean μ is

$$\bar{\mu}_{W_1} = \sum_{i=1}^{k} w_i y_{(ii)}, \tag{3.1}$$

The relative precision of our biased estimator $\bar{\mu}_{W_1}$ with respect to the estimator based on SRS is

$$RP_2 = \frac{\sigma^2[1 + C_k(k-1)]^2}{k\left[C_k^2\sum_{i=1}^{k-1}\sigma_{(ii)}^2 + \sigma_{(kk)}^2 + \left\{\mu\left(1 + C_k(k-1)\right) - C_k\sum_{i=1}^{k-1}\mu_{(ii)} - \mu_{(kk)}\right\}^2\right]} \tag{3.2}$$

The value of C_k is to be chosen such that the RP_2 is maximum. To find the optimum value of C_k (for each k), the excel program of RP_2 was developed and using the different iterations on C_k, the values of RP_2 was tested until it gets maximum. All the other values above and below from this optimal C_k, RP_2 starts decreasing.

We computed C_k for all four chosen distributions lognormal (LN(0, 1)), Pareto (P(3.5) and P(4.5)) and Weibull (W(0.5)) and k=2(1)5. The values of RP_2, C_k, RP_3 and RP_1 for these distributions and k=2(1)5 are presented in Table 1. The values of RP_2 are much higher than RP_1, i.e., the relative precisions of the estimator based on RSS procedure. Furthermore, the RP_2 are higher than RP_3, i.e., those based on Neyman's optimal allocation model for all four distributions when $k \le 4$. All relative precisions increase as k increases for LN(0,1), P(3.5) and P(4.5). However, for W(0.5), RP_2 decreases as k increases. This may be because the distribution W(0.5) has extremely large skewness and kurtosis.

Table 1. The RPs (RP_1, RP_2, RP_3) at an individual optimal C_k of each distribution for k = 2(1)5.

	Set size (k)	2	3	4	5
LN(0,1)	C_k	4.3798	3.2859	2.8028	2.5263
	RP_1	1.1872	1.3393	1.4711	1.5891
	RP_2	2.5946	2.7278	2.8083	2.8845
	RP_3	1.5765	2.1182	2.6219	3.1347
P(3.5)	C_k	4.7900	3.5693	3.0417	2.7427
	RP_1	1.1707	1.3073	1.4238	1.5269
	RP_2	2.7528	2.8579	2.9189	2.9805
	RP_3	1.5834	2.1273	2.6370	3.1434
P(4.5)	C_k	3.8151	2.8990	2.4962	2.2678
	RP_1	1.2134	1.3901	1.5451	1.6847
	RP_2	2.3679	2.5338	2.6535	2.7676
	RP_3	1.5544	2.0810	2.5995	3.0878
W(0.5)	C_k	6.7391	4.4803	3.5837	3.0972
	RP_1	1.1268	1.2362	1.3345	1.4250
	RP_2	3.6271	3.3698	3.2166	3.1379
	RP_3	1.6306	2.2105	2.7913	3.3840

Now we attempt to compute one set of values of C_k for four values of sample sizes, which will work well for all chosen four distributions. In these computations, C_k was determined so that the sum of RP_2 for the four distributions is close to the maximum. This optimum value of C_k was found using the same iteration procedure in the developed excel program.

The values of optimum C_k, and RP_2 for the chosen four distributions and four sample sizes are presented in Table 2. The values of RP_2 in Table 2 are slightly smaller than the ones in Table 1 as is expected. However, the pattern of RP_2 remained the same.

Table 2. The RP values (RP_1, RP_2, RP_3 and total RP_2) at combined optimal C_k for k = 2(1)5.

	Set size (k)	2	3	4	5
	C_k	5.0655	3.6004	2.9955	2.6618
	Total RP_2	11.2063	11.3956	11.5215	11.7050
	Total Maximum RP_2*	11.3424	11.4893	11.5973	11.7705
LN(0,1)	RP_1	1.1872	1.3393	1.4711	1.5891
	RP_2	2.5809	2.7207	2.8036	2.8811
	RP_3	1.5765	2.1182	2.6219	3.1347
P(3.5)	RP_1	1.1707	1.3073	1.4238	1.5269
	RP_2	2.7507	2.8578	2.9186	2.9794
	RP_3	1.5834	2.1273	2.6370	3.1434
P(4.5)	RP_1	1.2134	1.3901	1.5451	1.6847
	RP_2	2.3205	2.4963	2.6199	2.7363
	RP_3	1.5544	2.0810	2.5995	3.0878
W(0.5)	RP_1	1.1268	1.23617	1.3345	1.4250
	RP_2	3.5542	3.3208	3.1793	3.1082

 doi:10.20944/preprints202405.2029.v1

RP_3	1.6306	2.2105	2.7913	3.3840

*Total Maximum RP_2 is the sum of RP_2 of all the distributions at their respective optimum C_k.

4. WRSS with Two Optimal Weights

In this Section, we propose a WRSS with two optimal weights for the two extreme order statistics. Here the weights w_i $\left(with\ 0 \le w_i \le 1\ and\ \sum_{i=1}^{k} w_i = 1\right)$ for k>2 are defined as

$$w_1 \propto \frac{1}{C_1},$$

$$w_i \propto 1,\ \text{for}\ i = 2, 3, \dots, k-1$$

$$w_k \propto \frac{1}{C_k},$$

The proposed exact weights are as follows:

$$w_1 = \frac{1}{DC_1},$$

$$w_i = \frac{1}{D},\ \text{for i} = 2, 3, \dots, k-1$$

$$w_k = \frac{1}{DC_k},$$

where $D = (k-2) + \frac{1}{C_1} + \frac{1}{C_k}$

Our estimator of population mean is

$$\bar{\mu}_{W_2} = \sum_{i=1}^{k} w_i y_{(ii)} \tag{4.1}$$

The relative precision of $\bar{\mu}_{W_2}$ with respect to the estimator based on SRS is

$$RP_2 = \frac{\sigma^2 D^2}{k\left[\sum_{i=2}^{k-1} \sigma_{(ii)}^2 + \frac{\sigma_{(11)}^2}{C_1^2} + \frac{\sigma_{(kk)}^2}{C_k^2} + \left\{\mu D - \sum_{i=2}^{k-1} \mu_{(ii)} - \frac{\mu_{(11)}}{C_1} - \frac{\mu_{(kk)}}{C_k}\right\}^2\right]} \tag{4.2}$$

We calculate the optimal values of C_k and C_1 using the iteration method. Based on these values, we computed RP_2 along with RP_1 and RP_3 for chosen four distributions and sample sizes k=3, 4 and 5 are presented in Table 3. The gains in precisions of the estimator $\bar{\mu}_{W_2}$ over $\bar{\mu}_{W_1}$ are marginal. The gains of RP_2 based on $\bar{\mu}_{W_2}$ are substantially higher than the estimator based on RSS. $\bar{\mu}_{W_2}$ is superior to the estimator based on Neyman's optimal allocation model for all k for the LN(0,1) and P(3.5) distributions. The values of RP_2 are higher than those of RP_3 for the other two distributions for k=3 and 4. The gains of RP_3 over RP_2 for k=5 for these two distributions are marginal.

Table 3. The RPs (RP_1, RP_2, RP_3) at individual optimal C_k, C_1 of each distribution for k = 3(1)5.

	Set size (k)	3	4	5
LN(0,1)	C_k	3.3322	3.8383	4.3430
	C_1	1.0217	1.9580	6.9907
	RP_1	1.3393	1.4711	1.5891
	RP_2	2.7280	2.8982	3.1621
	RP_3	2.1182	2.6219	3.1347
P(3.5)	C_k	3.6003	4.1931	4.8037
	C_1	1.0134	2.0183	9.4583
	RP_1	1.3073	1.4238	1.5269
	RP_2	2.8579	3.0154	3.2847

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2024 doi:10.20944/preprints202405.2029.v1

	RP_3	2.1273	2.6370	3.1434
P(4.5)	C_k	2.8188	3.2036	3.6176
	C_1	0.9587	1.6485	3.9315
	RP_1	1.3901	1.5451	1.6847
	RP_2	2.5344	2.7072	2.9638
	RP_3	2.0810	2.5995	3.0878
W(0.5)	C_k	3.3947	3.7117	3.9996
	C_1	0.6463	1.0814	2.4127
	RP_1	1.2362	1.3345	1.4250
	RP_2	3.4286	3.2176	3.1936
	RP_3	2.2105	2.7913	3.3840

As we did in case of $\bar{\mu}_{W_1}$, we attempt to compute one set of values of C_k and C_1 for three values of sample sizes which will work well for all chosen four distributions. In these computations, C_k and C_1 were determined so that the sum of relative precisions of $\bar{\mu}_{W_2}$ for the four distributions is close to the maximum relative precision. The values of C_k and C_1, RP_1, RP_2 and RP_3 for three sample sizes and four chosen distributions are presented in Table 4. The relative precisions of $\bar{\mu}_{W_2}$ in Table 4 are higher than those of $\bar{\mu}_{W_1}$ for each k in Table 2. The pattern of relative precisions are same as seen in Table 3.

Table 4. The RPs (RP_1, RP_2, RP_3 and total RP_2) at combined optimal C_k, C_1 for k = 3(1)5.

Set size (k)		3	4	5
C_k		3.4245	3.8499	4.2613
	C_1	0.9241	1.7370	5.3284
Total RP_2		11.4036	11.7566	12.5655
Total Maximum RP_2*		11.5489	11.8384	12.6042
LN(0,1)	RP_1	1.3393	1.4711	1.5891
	RP_2	2.7169	2.8937	3.1605
	RP_3	2.1182	2.6219	3.1347
P(3.5)	RP_1	1.3073	1.4238	1.5269
	RP_2	2.8549	3.0112	3.2772
	RP_3	2.1273	2.6370	3.1434
P(4.5)	RP_1	1.3901	1.5451	1.6847
	RP_2	2.4947	2.6841	2.9504
	RP_3	2.0810	2.5995	3.0878
W(0.5)	RP_1	1.2362	1.3345	1.4250
	RP_2	3.3371	3.1677	3.1773
	RP_3	2.2105	2.7913	3.3840

*Total Maximum RP_2 is the sum of RP_2 of all the distributions at their respective C_k and C_1.

5. WRSS with All Optimal Weights

Now, we extend WRSS with optimal weights for all order statistics for k=4 and 5. We take $C = w_1 + w_k$, and determine the optimal values of C and w_k by minimizing MSE of the estimator by using $w_i \propto 1$, for $i = 2, \ldots, k-1$.

In the next step we use

$$w_i = (1-C)f_i,\ i = 2, \dots, k-1 \text{ with } \sum_{i=2}^{k-1} f_i = 1.$$

The values of f_i are chosen so that the value of RP_2 is maximized. Then we repeat the procedure of computing the optimal values of C and w_k with these new $w_i's$. The procedure is repeated until the value of RP_2 achieves the maximum value. We did this by using the developed computer program in Excel.

The values of RP_2 are presented in Table 5. We observe that the values of RP_2 presented in Table 4 are higher than the values of RP_2 based on one or two optimal weights which are given in Tables 1 and 3.

Table 5. The RPs (RP_1, RP_2 and RP_3) at individual optimal C, w_k and $f_i's$ of each distribution for k = 4 and 5.

		LN(0,1)	P(3.5)	P(4.5)	W(0.5)
k=4	C	0.2136	0.2070	0.2693	0.2191
	w_k	0.0937	0.0870	0.1069	0.0830
	f_2	0.5902	0.5827	0.5704	0.6484
	f_3	0.4098	0.4173	0.4296	0.3516
	RP_1	1.4711	1.4238	1.5451	1.3345
	RP_2	2.9401	3.0510	2.7304	3.2862
	RP_3	2.6219	2.7913	2.6370	2.5995
k=5	C	0.0827	0.0767	0.0905	0.0799
	w_k	0.0702	0.0648	0.0785	0.0686
	f_2	0.3346	0.3260	0.3590	0.3491
	f_3	0.4011	0.4013	0.3774	0.4143
	f_4	0.2643	0.2727	0.2636	0.23660
	RP_1	1.5891	1.5269	1.6847	1.4250
	RP_2	3.2583	3.3652	3.0329	3.3215
	RP_3	3.1347	3.1434	3.0878	3.3840

As we did in section 3 and 4, we computed one set of values of C, w_k and different fractions f_i for k=4 and k=5 which work well for all chosen four distributions. In these computations, these values were determined so that the sum of $RP_2's$ for the four distributions is close to the maximum relative precision. These values along with RP_1, RP_2 and RP_3 for k=4 and k=5 and four chosen distributions are presented in Table 6. As we expected the values of RP_2 are smaller in Table 6 when compared to the values of RP_2 in Table 5. However, the pattern of relative precisions remains the same.

Table 6. The values of RP_2 and total RP_2 at combined optimal values of C, w_k and $f_i's$ for k = 4 and 5.

Set size	C	w_k	f_2	f_3	f_4	RP_2 LN(0,1)	RP_2 P(3.5)	RP_2 P(4.5)	RP_2 W(0.5)	Total RP_2	Total Max RP_2*
k=4	0.2210	0.0911	0.5967	0.4033		2.9367	3.0467	2.7018	3.2452		
k=5	0.0824	0.0700	0.3397	0.3992	0.2611	3.2577	3.3587	3.0164	3.3078	11.9303	12.0077

*Total Maximum RP_2 is the sum of RP_2 of all the distributions at their respective C , w_k and $f_i's$.

6. WRSS with Increasing Skewness

In this section, we wish to study the performance of the three methods, RSS, WRSS and Neyman's optimum allocation model with increasing values of skewness of a family of distributions. For this purpose, the lognormal distribution, $LN(0,b)$ has been considered. The *pdf* of $LN(a,b)$ is given by

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2024 doi:10.20944/preprints202405.2029.v1

$$f(x) = \frac{1}{xb\sqrt{2\pi}} exp\left[\frac{-1}{2}\left(\frac{log\, x - a}{b}\right)^2\right], for x > 0, a > 0, b > 0, \quad \text{with}$$

population mean=$exp\left(a + \frac{b^2}{2}\right)$ and variance=$exp(2a + 2b^2) - exp(2a + b^2)$

Then skewness (*Sk*) and shape parameter (*p*) are given by

$$Sk = \sqrt{\beta_1} = \sqrt{exp(b^2) - 1}(exp(b^2) + 2) \text{ and } p = Exp(b^2).$$

The performance of these three methods relative to SRS with *k*=4 is presented in Table 7 for lognormal family of distributions for a range of values of population standard deviation. The variances of the order statistics of the family of distributions were computed by using the variances of order statistics for different values of shape parameter (*p*) which are readily available in Balakrishnan and Chen (1999). From Table 7, we observe that as skewness increases the performance of (i) RSS method decreases, and (ii) Neyman's and WRSS methods increases. The values of RP_2 based on all and two optimal weights are higher than RP_3 for all values of shape parameters, However, RP_2 based on one optimal weight is higher than RP_3 for all *p*>1.9. The rate of increase of relative precisions of the proposed estimators based on WRSS are more than that of estimator based on Neyman's method (See Figure 1).

Table 7. The values of RP_1, RP_2 and RP_3 for Lognormal $LN(0, b)$distributions for *k*=4.

p	Sk	C_k for $\bar{\mu}_{W_1}$	For $\bar{\mu}_{W_2}$		For $\bar{\mu}_{W_3}$				RP_1	RP_2			RP_3
			C_k	C_1	C	w_k	f_2	f_3		$\bar{\mu}_{W_1}$	$\bar{\mu}_{W_2}$	$\bar{\mu}_{W_3}$	
1.8	3.40	2.07	2.715	1.633	0.301	0.124	0.556	0.444	1.702	2.490	2.552	2.568	2.520
1.9	3.70	2.16	2.848	1.674	0.289	0.119	0.561	0.439	1.665	2.521	2.587	2.606	2.535
2.0	4.00	2.24	2.978	1.714	0.279	0.115	0.565	0.435	1.632	2.553	2.623	2.644	2.550
2.1	4.30	2.33	3.105	1.753	0.268	0.112	0.569	0.431	1.603	2.587	2.660	2.684	2.564
2.2	4.60	2.41	3.229	1.790	0.258	0.108	0.573	0.427	1.576	2.621	2.697	2.724	2.577
2.3	4.90	2.48	3.351	1.825	0.249	0.105	0.577	0.424	1.552	2.656	2.735	2.765	2.590
2.4	5.21	2.56	3.471	1.859	0.240	0.102	0.580	0.420	1.530	2.692	2.774	2.806	2.603
2.5	5.51	2.64	3.588	1.891	0.231	0.099	0.583	0.417	1.510	2.728	2.813	2.848	2.615
2.6	5.82	2.71	3.704	1.923	0.223	0.097	0.587	0.413	1.491	2.765	2.852	2.890	2.626
2.7	6.13	2.79	3.818	1.953	0.215	0.094	0.590	0.410	1.474	2.802	2.891	2.932	2.637
2.8	6.44	2.86	3.930	1.981	0.207	0.092	0.593	0.407	1.458	2.839	2.931	2.975	2.648
2.9	6.75	2.94	4.040	2.009	0.200	0.090	0.595	0.405	1.444	2.876	2.970	3.017	2.658
3.0	7.07	3.01	4.149	2.035	0.193	0.088	0.598	0.402	1.430	2.914	3.010	3.060	2.668

Note: Here $\bar{\mu}_{W_3}$represents the proposed estimator based on all optimal weights $RP_2(1), RP_2(2)$ and $RP_2(3)$ are the RP_2s based on one, two and all optimal weights, respectively.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2024 doi:10.20944/preprints202405.2029.v1

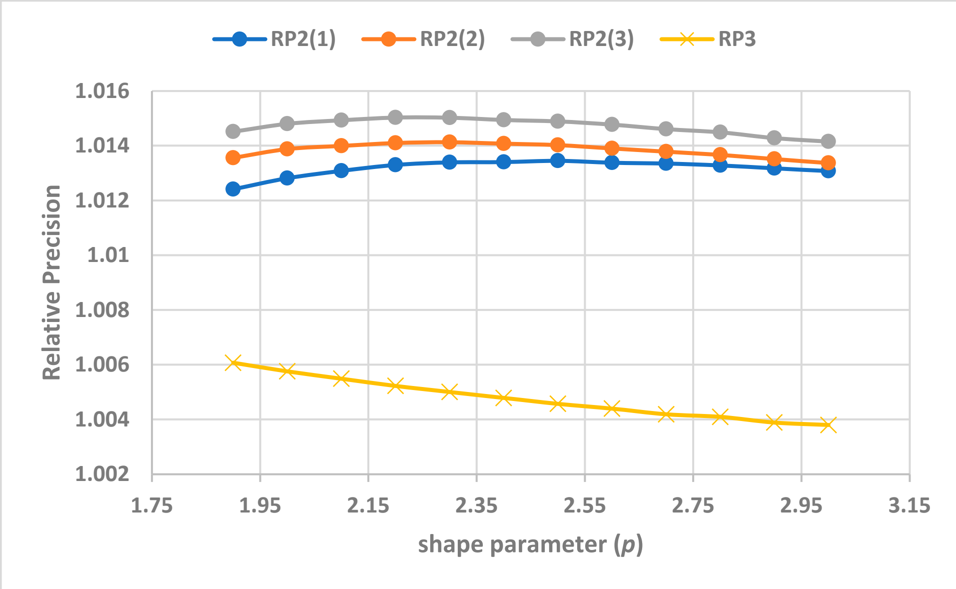

Figure 1. Comparison of rate of relative precisions with increasing skewness.

7. Conclusions and Discussion

In this paper, we proposed weighted ranked set sampling procedure to estimate the population mean of the distributions which are positively skew with heavy right tail. We chose four distributions: lognormal (LN(0, 1)), Pareto (P(3.5) and P(4.5)) and Weibull (W(0.5)). The means and variances of order statistics for these distributions are readily available in Harter and Balakrishnan (1996). We proposed three weighted ranked set sampling procedures. The first procedure is based on one optimal weight for the largest order statistics, the second procedure is to use the two optimal weights for the two extreme order statistics, and the third is the one which is based on k optimal weights. We calculated the relative precisions for each of these four distributions by using the WRSS procedure for each sample size. These relative precisions are much higher than the relative precisions of RSS estimator of mean. Furthermore, relative precisions of our estimators are higher than those which are based on Neyman's optimal procedures for $k \leq 4$. The relative precisions of our estimator are even higher than Neyman's procedure for k=5 for some distributions. Furthermore, we attempted to compute one set of weight(s) for each k for all the distributions and compared the relative precisions of our estimator with those of RSS and Neyman's estimators. Although there is slight loss in the values of relative precisions, they are still higher than those of Neyman's model for $k \leq 4$ for all four distributions and either more than or very close to Neyman's model for k=5. In general, as is expected, the relative precisions of our estimator based on all optimal weights are higher than the relative precisions of our estimator based on two and one optimal weight(s). The gain in relative precisions is however marginal.

We studied the performance of our proposed estimators for increasing skewness of a family of lognormal distributions. The relative precision of our estimator based on one optimal weight is higher than those of Neyman's estimator when the shape parameter exceeds 1.9. The relative precisions of our estimator based on two and k optimal weights is uniformly higher than those of Neyman's estimator for all values of shape parameter considered in Table 7. From Figure 1, we see that with the increasing values of skewness, the rate of increase of relative precisions of our proposed estimators based on WRSS are more than that of estimator based on Neyman's method.

Based on the numerical computations of relative precisions, we recommend our estimator based on WRSS procedures for estimator of population mean of skew distributions with heavy right tail for small values of set sizes.

Conflict of Interest: There is no conflict of interest.

doi:10.20944/preprints202405.2029.v1

References

Balakrishnan, N. and Chen, W. S. (1999), Handbook of Tables of order statistics from lognormal distribution with applications, Springer, U. S. A.

Bhoj, D.S. (1997a). New parametric ranked set sampling. *Journal of Applied Statistical Sciences*, **6,** 275-289.

Bhoj, D.S. (1997b). Estimation of parameters of extreme value distributions with ranked set sampling. *Communication of Statistics: Theory and Methods*, **26(3),** 653-667.

Bhoj, D. S. (2001). Ranked set sampling with unequal samples. *Biometrics*, **57(3)**, 957-962.

Bhoj, D.S. and Ahsanullah, M. (1996). Estimation of the parameters of the generalized geometric distribution using ranked set sampling, *Biometrics*, **52**, 685-694.

Bhoj, D.S. and Chandra, G. (2019). Simple unequal allocation procedure for ranked set sampling with skew distributions. *Journal of Modern Applied Statistical Methods*, **18(2)**, eP2811. doi: 10.22237/jmasm/1604189700.

Bhoj, D.S. and Kushary, D. (2016). Ranked set sampling with unequal samples for skew distributions. *Journal of Statistical computations and simulations*, **86(4),** 676-681.

Chandra, G., Bhoj, D.S. and Pandey, R. (2018). Simple unbalanced ranked set sampling for mean estimation of response variable of developmental programs, *Journal of Modern Applied Statistical Methods*, 17(1). Article 28. DOI: 10.22237/jmasm/1543856083.

Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrics,* **28**,545–555.

Harter, H.L. and Balakrishnan, N. (1996). CRC handbook of tables for the use of order statistics in estimation, CRC Press, Boca Raton, New York.

Kaur, A., Patil, G. P. and Taillie, C. (1997). Unequal allocation models for ranked set sampling with skew distributions. *Biometrics*, **53**, 123-130.

Lam, K., Sinha, B.K. and Wu, Z. (1994). Estimation of parameters of the two-parameter exponential distribution using ranked set sample. *Annals of the Institute Statistical Mathematics*, **46**, 723-736.

McIntyre, G. A. (1952). A method for unbiased selective sampling using ranked sets, *Australian Journal of Agricultural Research,* **3,** 385-390.

Stokes, S. L. (1995). Parametric ranked set sampling. *Annals of the Institute of Statistical Mathematics*, **47,** 465-482.

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.

Tiwari, N. and Chandra, G. (2011). A systematic procedure for unequal allocation for skewed Distributions in Ranked Set Sampling. *Journal of the Society of Agricultural Statistics*, **65(3),** 331-338.