

Evaluation of Ranked Set Sampling Methods under Skewed and Unskewed Distributions

M. Iqbal Jeelani¹, Ayed AL e'damat², Imran Rashid³, S. N Zafar Geelani¹, Afshan Tabassum³ and Khalid Ul Islam Rather^{3,*}

¹Division of Social and Basic Sciences, Faculty of Forestry, SKUAST-Kashmir, India

²Department of Mathematics, AL Hussain Bin Talal University, Ma'an, Jordan

³Division of Statistics and Computer Science, SKUAST-Jammu, India

Received: 4 May 2022, Revised: 20 Sep. 2022, Accepted: 24 Sep. 2022.

Published online: 1 May 2023.

Abstract: The use of ranked set sampling (RSS), a unique and cost-effective sampling technique, is appropriate when determining units is simple and straightforward. The estimates based on RSS are more effective than the conventional unconstrained sampling approaches, and in recent years, their modified versions have gained popularity because of their remarkably increased effectiveness in research projects on forest management. In light of this, the current study was conducted on skewed and unskewed distributions to assess the efficacy of the various modified versions of RSS. Simulated data was produced in R Studio (version 4.1.2 – 2021) using skewed and unskewed probability distributions like the gamma, exponential, uniform, and normal distributions to fulfil the specified goals. In accordance with this, a ranked set sample of sizes 150, 300, 450, 600, 750, 900, 1050, and 1200 with a set size of 3, 6, 9, 12, 15, 18, and 24 using a constant cycle (r) of 50 was drawn from the simulated data using various modified RSS techniques, including: Extreme ranked set sampling (ERSS), Median ranked set sampling (MRSS), Percentile rank set sampling (PRSS), Balanced grouped ranked set sampling (BGRSS), Balanced grouped ranked set sampling (BGRSS), and Truncation based rank set sampling (TBRSS) by means of library(RS Sampling) of R Studio. Since these samples are based on modified versions of RSS that are more regularly spaced and induce stratification at the sample level, which involves a gain in efficiency, it can be seen from the results that modified versions of RSS performed better in unskewed distributions in comparison to skewed distributions in terms of efficiency. Additionally, it was discovered that the effectiveness of every modified RSS algorithm rises as the sample size does. According to empirical research using simulated data, truncation-based rank set sampling (TBRSS), among the modified RSS methods, outperformed its competitors in terms of efficiency. The value of AIC & BIC reduced as the set size across the modified RSS methods increased, according to the goodness of fit results, showing that less information is lost as set size increases. Finally, it can be said that RSS has practical ramifications, and that R packages greatly aid in the implementation of modified RSS methods, which are quite informative and suitable to sample surveys.

Keywords: Ranked Set, Distributions, Efficiency, Information criteria, Simulation, R Studio

1 Introduction

It is commonly recognised that a well-planned sample survey can yield very accurate information, but because the results of a sample survey are based on a subset of the population rather than the complete population, they are likely to deviate from the actual values of the population. The sample survey has the advantage that this kind of mistake may be measured and managed. In order to increase sampling effectiveness, sampling theory works to provide sample selection and estimation techniques that deliver precise results at the least expensive price. One of the most important aspects of any statistical inference is that the data used was collected in a systematic manner that allows the experimenter to reach reliable conclusions on the relevant question(s). One of the most popular ways to gather such information is by straightforward random sampling. The fact that each individual measurement in the sample is probably representative of the population characteristic of interest, such as mean or median, does not make the resulting statistical inference procedures appropriate when we select a simple random sample X_1, X_2, \dots, X_n from a fixed population of interest. Instead, we should "on average" be able to produce a set of sample observations that are accurately representative of the full population by using the idea of

*Corresponding author e-mail: khalidstat34@gmail.com

sampling distributions of the pertinent statistics. The "on average" concept is not particularly helpful if the specific population items chosen for our sample are in fact not very representative of the entire population, as is the case in practise where we only acquire a single random sample. [1] proposed the use of ranked set sampling (RSS) as an alternative to SRS, which involves drawing m SRS, each of size m , and then choosing just one measurement from each SRS. To do this, choose the smallest observation from the first SRS, the second-smallest from the second SRS, and so on, all the way up to the largest observation from the last SRS. More of a data gathering method than a more representative sample approach, ranked set sampling is. It utilises the fundamental intuitive characteristics of simple random samples while also utilising additional population information to create a "artificially stratified" sample that has more structure. This allows us to concentrate on the actual measurement of more representative units in the population. As a result, there is a set of measurements that is more likely than a random sample to cover the population's range of values. The arithmetic average of RSS measurements, like SRS measurements, provides an unbiased point estimate of the population mean, but the accompanying confidence interval is potentially significantly lower than that obtained with SRS measurements, showing less uncertainty. m random samples of size m should be selected in order to plan an RSS design. This is how the RSS technique is used:

Step 1: choose m^2 units at random from the given population.

Step 2: Next, m^2 units are divided into m sets, each of size m , as randomly as is humanly possible.

Step 3: Without actually measuring them, rank these m units.

Step 4: From the ranked set, keep the smallest judged unit.

Step 5: Randomly choose a second group of m units from a given population, rank them without measuring them, and keep the second-smallest evaluated unit.

Step 6: Continue until m rated units have been measured..

"A cycle" is the term used to describe the first five steps. The process is then repeated r times, yielding a ranked set sample with size $n = mr$. The estimator of population mean according to the RSS technique is defined as:

$$\hat{\mu}_{RSS} = \frac{1}{m} \left(\sum_{i=1}^r X_{i(r+1)} + \sum_{i=r+1}^{m-r} X_{i(i)} + \sum_{i=m-r+1}^m X_{i(m-r)} \right)$$

And its variance is given by:

$$\text{var}(\hat{\mu}_{RSS}) = \frac{1}{m^2} \left(\sum_{i=1}^r \text{var}(X_{i(r+1)}) + \sum_{i=r+1}^{m-r} \text{var}(X_{i(i)}) + \sum_{i=m-r+1}^m \text{var}(X_{i(m-r)}) \right)$$

It is important to note that $\hat{\mu}_{RSS}$ is an unbiased estimator of population mean μ , and has smaller variance than $\hat{\mu}_{SRS}$ i.e. (simple random sample estimator), if the underlying distribution is unskewed. Due to their effectiveness, RSS techniques and their modified forms have recently gained in popularity. The underlying distribution, in particular the mean of skewness, has an impact on the relative efficacy of RSS. Several conjectures are made based on calculations on a number of underlying probability distributions, such as: the relative efficiency of RSS with regard to SRS, in the estimation of the population mean, is between 1 and $(m+1)/2$ where m is the set size; The relative efficiency is not significantly less than $(m+1)/2$ for underlying probability distributions that are not skewed. Nevertheless, as the underlying distribution grows skewed the relative efficiency declines but never falls below 1. When ranking is perfect, it is shown that the variation of RSS, which depends on set size (m), drops as m increases, indicating that the relative efficiency rises as m rises. Through a survey on horticultural data, [2] examined the viability of employing RSS to enhance population mean estimates relative to basic random sampling in horticultural research. The findings showed that, for the specific horticultural survey utilising the same sample size, RSS produces better accurate estimates than ordinary random sampling. [3] provided an RSS sampling package for R studio with respect to various modified versions of RSS like extreme RSS, median RSS, percentile RSS, balanced groups RSS, double versions of RSS, L-RSS, truncation-based RSS, and robust extreme RSS. They also introduced the R package as a free and simple-to-use analysis tool for both sampling processes and statistical inferences based on RSS and its modified versions. Recent studies by [4] and [5] on the characteristics of the maximum likelihood estimator of the population proportion in RSS using extreme rankings found that RSS once again gave more accurate estimates than its rivals. Recently work on Ranked Set sampling was performed by [6, 7] and [8, 9, 10]. With the aforementioned facts in mind, the study's goals are to compare the effectiveness of various ranked set sampling techniques under various skewed and unskewed distributions, as well as to fit various ranked set sampling techniques to simulation data.

2. Material and Methods

2.1 Generation of Simulated Data

To achieve the specified goals, simulated data of 2000 observations was created in R Studio (version 4.1.2—2021) using skewed and unskewed probability distributions such as the Gamma, Exponential, Uniform, and Normal distributions. The Gamma, Exponential, Uniform, and Normal distributions were utilised together with other skewed and unskewed distributions to create simulated data using the following R functions.

```
data=runif(2000,0,1)
```

```
data=rnorm(2000,0,1)s
```

```
data=rexp(2000,1)
```

```
data=rgamma(2000,2,1)
```

The following is a quick description of the distributions employed in the current study are gamma distribution, exponential distribution uniform distribution and normal distribution.

Sample Choice

From the simulated data, ranked set samples of sizes 150, 300, 450, 600, 750, 900, 1050, and 1200 were taken using set sizes of 3, 6, 9, 12, 15, 18, 21, and 24 and constant cycle (r)50, respectively. These samples were obtained using various RSS techniques, including: extreme ranked set sampling (ERSS), median ranked set sampling (MRSS), percentile ranked set sampling (PRSS), balanced grouped ranked set sampling (BGRSS), double ranked set sampling (DBRSS), Robust ranked set sampling (L-RSS), Truncation based ranked set sampling (TBRSS), and Robust ranked set sampling (RERSS) using the R studio's following features:

2.2.1 Extreme ranked set sampling (ERSS)

[11] proposed the first modification to RSS, known as ERSS, which only utilises the minimum or maximum rated units from each set to get the population mean. The ERSS process can be summarised as follows: Choose m random sets, each with a size of m units, from the population and rank the units inside each set using a concurrent variable or a human expert. The largest ranked units of each set are selected from the remaining $m/2$ sets if the set size m is even. The lowest ranked units of each set are selected from the first $m/2$ sets. When the set size is odd, the highest ranked units from the other first $(m-1)/2$ sets, the lowest ranked units from the other first $(m-1)/2$ sets, and the median unit from the final set are chosen. If the process is repeated r times, we will get a sample with a size of $n = mr$. The process is illustrated below for $r = 1$ and $m = 4$:

2.2.2 Median ranked set sampling (MRSS)

[12] made the suggestion of MRSS. In this method, the sample for estimating the population mean is exclusively comprised of the median units from the random sets. The $((m + 1)/2)$ th ranked units are selected as each set's median for the odd set sizes. For set sizes that are even, the first $m/2$ ranked units are picked from the first $m/2$ sets, and the last $m/2$ sets are chosen from the $((m + 2)/2)$ ranked units. The technique can be carried out r more times if necessary, giving us a sample size of $n = mr$. Below is an illustration of the process for $r = 1$ and $m = 3$:

2.2.3 Percentile ranked set sampling (PRSS)

Another RSS variant, known as PRSS, was proposed by [13] in which only the upper and lower percentiles of the random sets are chosen as the sample for a given value of p , where $0 \leq p \leq 1$. Consider selecting m random sets of size m from a certain population to sample m units, and then ranking them visually or using a companion variable. The $(p(m + 1))$ th smallest units from the first $m/2$ sets and the $((1 - p)(m + 1))$ th smallest units from the remaining $m/2$ sets are picked if the set size is even. If m is odd, the $(p(m + 1))$ th smallest units are selected from the first $(m - 1)/2$ set, the $((1 - p)(m + 1))$ th smallest units are selected from the second $(m - 1)/2$ sets, and the median unit is selected as the m th unit from the third set. Here is an illustration of the process with $r = 1$, $m = 5$, and $p = 0.3$:

2.2.4 Balanced groups ranked set sampling (BGRSS)

The fusion of ERSS and MRSS is known as BGRSS. [14] Recommended using BGRSS for population mean estimation with a unique sample size of $m = 3k$. According to their study, the BGRSS process is as follows: Every size of m in $m = 3k$ sets (where $k = 1, 2, 3$, etc.) is chosen at random from a certain population. The units in each set are ranked after the sets are divided into three groups at random. The ranking sets are divided into three groups, and the smallest, middle, and largest units from each group are picked. The median unit in the second group is defined as the $((m + 1)/2)$ th

ranked unit in the set when the set size is odd, and as the mean of the $((m/2)$ th and the $((m+2)/2)$ th ranked units when the set size is even. Following is a description of the BGRSS process for one cycle and $k=2$:

2.2.5 Double ranked set sampling (DRSS)

As the first step in a multistage process, [15] proposed double RSS (DRSS), another variation of RSS. A number of scholars improved the DRSS approach, creating the double extreme RSS (DERSS) by [16] the double median RSS (DMRSS) by [17], and the double percentile RSS (DPRSS) by [18,19]. The DRSS technique is explained as follows: m^3 units are chosen from the target population and divided into m groups at random, each of which is made up of m^2 units. The standard RSS process is then applied to each group to produce m samples from each ranked set of size m . In order to obtain a double ranked set sample of size m , the RSS technique is again used to the ranked set samples that were obtained in the previous phase.

2.2.6 Robust ranked set sampling (L-RSS)

[20] created L-RSS, a robust RSS procedure that builds on the concept of the L statistic and serves as a generalisation of several RSS techniques. The selection of m random sets containing m units and ranking of the units in each set constitute the first phase in the L-RSS technique. Assuming $[m\alpha]$ is the greatest integer value less than or equal to m , let k be the L-RSS coefficient, where $k = [m\alpha]$ for 0 to 0.5 . Then, among the remaining sets that are numbered with I , where $I=k+2, \dots, m-k-1$, the $(k+1)$ th ranked units from the first $k+1$ sets, the $(m-k)$ th ranked units from the last $k+1$ sets, and the i th ranked units from the remaining sets that are ranked with I are chosen. The L-RSS method for the scenario where $m=6$ and $k=1$ ($\alpha=0.20$) a cycle can be illustrated as follows:

2.2.7 Truncation based ranked set sampling (TBRSS)

The authors [21] presented the truncation-based RSS (TBRSS). The following succinct description of this process: From the population, choose m sets of m units each at random, and rank the units in each set. Then, as with the L-RSS approach, determine the TBRSS coefficient k and choose the minimum and maximum values for the first k sets. The i th ranking unit of the i th sample ($k+1 \leq i \leq m-k$) should be chosen from the remaining $(m-2k)$ samples. The following diagram illustrates the one cycled TBRSS approach for the instance of $m=8$ and $k=2$ ($\alpha=0.35$).

2.2. Robust extreme ranked set sampling (RERSS)

By [22], the robust extreme RSS (RERSS) technique was first introduced. The following is a description of this technique: Find m random sets containing m units, and then order the units in each set. Choose the $(k+1)$ th ranked units from the first $m/2$ sets, where $[m]$ is the biggest integer value less than or equal to m and $k = [m\alpha]$ for $0 \leq \alpha < 0.5$. the $(m-k)$ th ranked units from the other $m/2$ sets, and so on. The $((m+1)/2)$ th ranked unit is also chosen from the final set if the set size m is odd. The steps for a single cycle and the scenario with $m=6$ and $k=1$ ($\alpha=0.20$) can be illustrated as follows:

2.3 Informational and Goodness of Fit Standards.

Since it is true that the underlying structure of the probability distribution has an impact on the resilience of RSS in terms of efficiency, A goodness of fit analysis was conducted in light of this to assess how well the distributions under consideration fit the data obtained from various modified RSS algorithms. The Chi-square test and information criteria like the Akaike and Bayesian information criterion (AIC and BIC) were used to achieve this. Below is a brief description of them.

$$\text{Chi-square test: } f(x) = \frac{e^{-x} x^{\frac{v}{2}}}{2^{\frac{v}{2}} \Gamma(\frac{v}{2})}$$

Akaike Information Criterion (AIC): $AIC = n \ln(RMSE) + 2k$

Bayesian Information Criterion (BIC): $BIC = -2 \ln(l) + \ln(n) * k$

3. Results

From Table 1 to Table 8, with regard to the skewed and unskewed distributions employed in this work, the variance of traditional RSS and other modified RSS methods is shown. With regard to both skewed and unskewed distributions, it is shown that the variance of modified approaches was found to be lower than that of traditional RSS. In addition, it was noted that the variance across all distributions of modified RSS methods, including traditional RSS, reduces as the set size grows.

Table1: Modified RSS technique variation across distributions when m=3

Methods Distribution	RSS	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	0.082	0.048	0.045	0.033	0.043	0.032	0.031	0.031	0.032
N (0,1)	0.880	0.544	0.510	0.377	0.501	0.373	0.352	0.352	0.358
Exp. (1)	0.851	0.622	0.580	0.481	0.536	0.476	0.453	0.389	0.475
Gamma (2,1)	1.662	1.088	1.022	0.775	0.994	0.764	0.720	0.717	0.727

Table2: Modified RSS technique variation across distributions when m=6

Methods Distribution	RSS	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	0.813	0.329	0.312	0.237	0.284	0.236	0.183	0.182	0.217
N (0,1)	0.994	0.409	0.389	0.299	0.356	0.294	0.237	0.227	0.270
Exp. (1)	0.999	0.471	0.462	0.309	0.416	0.304	0.288	0.236	0.301
Gamma (2,1)	1.744	0.745	0.706	0.533	0.643	0.520	0.447	0.401	0.498

Table 3: Modified RSS technique variation across distributions when m=9

Methods Distribution	RSS	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	0.079	0.031	0.027	0.017	0.026	0.016	0.014	0.011	0.014
N (0,1)	1.002	0.402	0.372	0.232	0.349	0.215	0.186	0.151	0.186
Exp. (1)	0.871	0.374	0.352	0.212	0.313	0.205	0.166	0.142	0.197
Gamma (2,1)	2.083	0.850	0.823	0.493	0.736	0.454	0.389	0.321	0.412

Table 4: Modified RSS technique variation across distributions when m=12

Methods Distribution	RSS	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	0.080	0.028	0.026	0.013	0.025	0.012	0.012	0.010	0.012

N (0,1)	0.960	0.346	0.333	0.167	0.324	0.150	0.148	0.127	0.150
Exp. (1)	0.934	0.354	0.335	0.181	0.324	0.172	0.152	0.129	0.152
Gamma (2,1)	2.186	0.795	0.772	0.391	0.743	0.362	0.342	0.291	0.343

Table 5: Modified RSS technique variation across distributions when m=15

Methods Distribution	RSS	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	0.081	0.303	0.302	0.169	0.301	0.166	0.127	0.114	0.158
N (0,1)	0.976	0.275	0.273	0.152	0.272	0.151	0.114	0.103	0.141
Exp. (1)	1.046	0.317	0.316	0.197	0.315	0.170	0.127	0.114	0.155
Gamma (2,1)	1.922	0.558	0.555	0.312	0.552	0.307	0.227	0.205	0.282

Table 6: Modified RSS technique variation across distributions when m=18

Methods Distribution	RSS	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	0.083	0.021	0.019	0.012	0.019	0.010	0.007	0.007	0.010
N (0,1)	0.993	0.255	0.236	0.149	0.233	0.133	0.096	0.093	0.126
Exp. (1)	0.851	0.238	0.229	0.152	0.220	0.118	0.087	0.083	0.110
Gamma (2,1)	1.773	0.466	0.432	0.278	0.424	0.241	0.173	0.168	0.226

Table 7: Modified RSS technique variation across distributions when m=21

Methods Distribution	RSS	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	0.086	0.021	0.019	0.010	0.018	0.009	0.006	0.006	0.009
N (0,1)	0.977	0.239	0.218	0.122	0.215	0.114	0.077	0.072	0.111
Exp. (1)	0.989	0.254	0.223	0.145	0.226	0.119	0.080	0.075	0.117
Gamma (2,1)	2.064	0.512	0.468	0.271	0.463	0.244	0.164	0.154	0.239

Table 8: Modified RSS technique variation across distributions when m=24

Methods	RSS	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
----------------	------------	-------------	-------------	-------------	--------------	-------------	--------------	--------------	--------------

Distribution									
U (0,1)	0.085	0.019	0.017	0.010	0.017	0.008	0.005	0.005	0.008
N (0,1)	0.996	0.230	0.210	0.121	0.211	0.103	0.068	0.064	0.095
Exp. (1)	1.065	0.258	0.233	0.151	0.229	0.115	0.075	0.070	0.105
Gamma (2,1)	2.027	0.475	0.435	0.258	0.430	0.214	0.141	0.132	0.196

Tables 9 through 16 show the relative effectiveness of RSS modifications. Exponential, Gamma, Uniform, and Normal distributions were among the four probability distribution functions that were taken into consideration. Based on the findings, it was determined that Truncation Based RSS (TBRSS) was more effective than other modified RSS approaches for both skewed and unskewed distributions. In addition, it was discovered that although the efficiency did decline under a skewed distribution, it never fell below 1. When the underlying distribution was unskewed, the relative efficiency significantly improved in comparison to a skewed distribution.

Table 9: The performance of modified RSS techniques for various distributions when m=3

Methods Distribution	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	1.701	1.811	2.505	1.899	2.541	2.638	2.647	2.548
N (0,1)	1.618	1.726	2.329	1.757	2.354	2.495	2.497	2.457
Exp. (1)	1.368	1.468	1.768	1.589	1.788	1.878	2.189	1.798
Gamma (2,1)	1.527	1.626	2.145	1.672	2.174	2.307	2.317	2.286

Table 10: The performance of modified RSS techniques for various distributions when m=6

Methods Distribution	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	2.47	2.60	3.43	2.86	3.44	4.44	4.47	3.74
N (0,1)	2.43	2.55	3.32	2.79	3.38	4.18	4.38	3.68
Exp. (1)	2.12	2.16	3.23	2.40	3.28	3.47	4.23	3.32
Gamma (2,1)	2.34	2.47	3.27	2.71	3.35	3.90	4.35	3.50

Table 11: The performance of modified RSS techniques for various distributions when m=9

Methods Distribution	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	2.51	2.95	4.45	2.99	4.71	5.46	6.82	5.443
N (0,1)	2.49	2.69	4.31	2.87	4.65	5.39	6.62	5.371
Exp. (1)	2.33	2.47	4.11	2.78	4.24	5.226	6.11	4.41

Gamma (2,1)	2.45	2.53	4.22	2.83	4.58	5.354	6.47	5.05
--------------------	------	------	------	------	------	-------	------	------

Table 12: The performance of modified RSS techniques for various distributions when m=12

Methods Distribution	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	2.80	3.00	5.92	3.11	6.41	6.62	7.72	6.45
N (0,1)	2.77	2.88	5.72	2.96	6.38	6.47	7.52	6.40
Exp. (1)	2.64	2.79	5.15	2.88	5.42	6.15	7.22	6.12
Gamma (2,1)	2.75	2.83	5.59	2.94	6.04	6.39	7.49	6.36

Table 13: The performance of modified RSS techniques for various distributions when m=15

Methods Distribution	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	3.63	3.64	6.49	3.65	6.60	8.65	9.61	6.96
N (0,1)	3.54	3.57	6.41	3.59	6.45	8.52	9.44	6.88
Exp. (1)	3.30	3.31	5.29	3.32	6.14	8.24	9.14	6.75
Gamma (2,1)	3.44	3.46	6.16	3.48	6.26	8.46	9.36	6.80

Table 14: The performance of modified RSS techniques for various distributions when m=18

Methods Distribution	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	3.93	4.32	6.90	4.37	7.60	10.65	10.82	7.87
N (0,1)	3.89	4.196	6.64	4.25	7.47	10.34	10.62	7.84
Exp. (1)	3.58	3.721	5.58	3.86	7.18	9.795	10.27	7.69
Gamma (2,1)	3.80	4.104	6.36	4.18	7.36	10.22	10.52	7.82

Table 15: The performance of modified RSS techniques for various distributions when m=21

Methods Distribution	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	4.12	4.55	8.00	4.59	8.77	12.72	13.764	8.80
N (0,1)	4.09	4.47	7.95	4.53	8.57	12.61	13.553	8.75
Exp. (1)	3.89	4.42	6.80	4.37	8.27	12.27	13.181	8.45

Gamma (2,1)	4.03	4.41	7.61	4.45	8.46	12.54	13.331	8.61
--------------------	------	------	------	------	------	-------	--------	------

Table 16: The performance of modified RSS techniques for various distributions when m=24

Methods Distribution	ERSS	MRSS	PRSS	BGRSS	DRSS	L-RSS	TBRSS	RERSS
U (0,1)	4.363	4.773	8.236	4.798	9.709	14.729	15.649	10.585
N (0,1)	4.326	4.731	8.187	4.718	9.608	14.518	15.468	10.485
Exp. (1)	4.128	4.567	7.041	4.636	9.206	14.116	15.226	10.142
Gamma (2,1)	4.267	4.659	7.847	4.714	9.476	14.346	15.356	10.346

Additionally, a goodness of fit analysis was performed by computing the Chi-square, AIC, and BIC for all fitted distributions in order to compare the empirical distribution for each RSS method which results are listed in tables 17 to 20. The following observations were drawn from the findings: According to the goodness of fit study, all RSS approaches had a non-significant p-value. Additionally, it was discovered that across all of the adjusted RSS approaches, the AIC and BIC values drastically decline as set size increases.

Table 17: Fit and information standards under uniform (0,1)

Design	M	Chi-square	p-value	AIC	BIC
ERSS	3	7.35	Ns	114.05	121.40
	6	5.22		109.03	118.75
	9	4.02		112.75	116.52
	12	3.18		111.55	113.24
	15	2.17		107.97	109.62
	18	2.15		104.50	106.10
	21	2.14		93.28	94.710
	24	1.13		92.27	92.03
MRSS	3	7.41	ns	108.75	114.95
	6	6.28		103.19	112.43
	9	4.22		106.73	110.30
	12	3.02		105.58	107.20
	15	3.16		102.18	103.75
	18	2.55		98.89	100.40
	21	1.14		88.23	89.59
	24	1.08		86.16	87.46
PRSS	3	7.24	Ns	104.90	110.90
	6	6.21		99.53	108.46

	9	5.15		102.95	106.41
	12	4.17		101.84	103.4
	15	3.15		98.56	100.07
	18	2.13		95.37	96.84
	21	2.11		85.06	86.38
	24	1.10		83.30	84.94
BGRSS	3	7.36	Ns	93.64	99.04
	6	6.17		88.81	96.84
	9	4.16		91.88	94.99
	12	4.15		90.89	92.29
	15	3.14		87.93	89.29
	18	2.13		85.06	86.38
	21	1.11		75.78	76.97
	24	1.09		73.28	74.84
DRSS	3	7.19	Ns	92.62	98.02
	6	6.15		87.79	95.82
	9	6.14		90.86	93.97
	12	5.13		89.87	91.27
	15	4.12		86.91	88.27
	18	3.11		84.04	85.36
	21	2.10		74.76	75.95
	24	1.08		73.74	74.61
L-RSS	3	7.16	Ns	91.59	96.99
	6	6.12		90.42	94.79
	9	5.11		89.83	92.94
	12	4.01		88.84	90.24
	15	3.09		86.88	87.24
	18	2.08		83.01	84.33
	21	1.07		80.73	74.92
	24	1.05		77.49	73.96
TBRSS	3	7.14	Ns	88.00	95.95
	6	6.10		83.12	93.75
	9	5.09		80.43	91.90
	12	4.18		77.32	89.20
	15	4.07		72.22	86.20
	18	3.06		68.12	83.29

	21	2.05		64.54	73.88
	24	2.03		57.17	72.05
RERSS	3	7.11	Ns	90.55	94.94
	6	7.07		89.72	92.74
	9	6.06		88.79	90.89
	12	5.05		87.80	88.19
	15	4.04		84.84	85.19
	18	3.33		81.97	82.28
	21	3.02		72.69	71.04
	24	2.01		71.65	70.93

Table 18: Fit and information standards under Normal (0,1)

Design	M	Chi-square	p-value	AIC	BIC
ERSS	3	7.11	Ns	117.30	124.25
	6	6.24		111.88	121.60
	9	6.22		115.60	119.37
	12	5.20		114.40	116.09
	15	4.19		110.82	112.47
	18	3.17		107.35	108.95
	21	2.16		96.13	97.56
	24	1.15		95.12	94.88
MRSS	3	7.27	Ns	111.6	117.80
	6	6.22		106.04	115.28
	9	6.20		109.58	113.15
	12	5.19		108.43	110.05
	15	4.18		105.03	106.60
	18	3.17		101.74	103.25
	21	3.06		91.08	92.44
	24	1.14		89.01	90.31
PRSS	3	7.26	Ns	107.75	113.75
	6	7.22		102.38	111.31
	9	6.20		105.8	109.26
	12	5.19		104.69	106.25
	15	5.17		101.41	102.92
	18	4.15		98.22	99.69
	21	4.13		87.91	89.23
	24	2.12		86.15	87.79

BGRSS	3	6.23	Ns	96.49	101.89
	6	5.19		91.66	99.69
	9	4.18		94.73	97.84
	12	3.17		93.74	95.14
	15	2.46		90.78	92.14
	18	2.15		87.91	89.23
	21	1.33		78.63	79.82
	24	1.11		76.13	77.69
DRSS	3	6.21	Ns	95.47	100.87
	6	6.17		90.64	98.67
	9	5.16		93.71	96.82
	12	4.45		92.72	94.12
	15	4.14		89.76	91.12
	18	3.13		86.89	88.21
	21	2.12		77.61	78.80
	24	1.11		76.59	77.46
L-RSS	3	7.18	Ns	90.44	99.84
	6	6.14		89.61	97.64
	9	5.33		87.68	95.79
	12	5.12		85.69	93.09
	15	4.11		83.73	90.09
	18	3.10		81.86	87.18
	21	2.09		78.58	77.77
	24	2.07		75.34	76.81
TBRSS	3	6.56	Ns	87.98	98.80
	6	6.12		82.00	96.60
	9	5.11		81.23	94.75
	12	4.10		78.33	92.05
	15	3.09		75.99	89.05
	18	2.08		72.11	86.14
	21	1.07		70.54	76.73
	24	1.05		62.19	74.90
RERSS	3	7.13	Ns	93.4	97.79
	6	6.09		90.57	95.59
	9	5.58		91.64	93.74
	12	5.07		90.65	91.04

	15	3.06		87.69	88.04
	18	2.65		84.82	85.13
	21	2.02		75.54	73.89
	24	1.01		74.50	76.93

Table 19: Fit and information standards under Gamma (2,1)

Design	M	Chi-square	p-value	AIC	BIC
ERSS	3	7.25	ns	121.33	127.99
	6	6.21		115.62	125.34
	9	6.19		119.34	123.11
	12	5.17		118.14	119.83
	15	4.16		114.56	116.21
	18	3.14		111.09	112.69
	21	2.13		99.87	101.30
	24	1.12		98.86	98.62
MRSS	3	7.24	Ns	115.34	121.54
	6	7.19		109.78	119.02
	9	6.17		113.32	116.89
	12	5.16		112.17	113.79
	15	5.05		108.77	110.34
	18	4.14		105.48	106.99
	21	3.13		94.82	96.18
	24	1.11		92.75	94.05
PRSS	3	6.23	Ns	111.49	117.49
	6	6.19		106.12	115.05
	9	5.17		109.54	113.00
	12	4.16		108.43	109.99
	15	4.14		105.15	106.66
	18	2.12		101.96	103.43
	21	1.10		91.65	92.97
	24	1.09		89.79	91.53
BGRSS	3	7.20	Ns	100.23	105.63
	6	6.16		95.40	103.43
	9	5.15		98.47	101.58
	12	5.14		97.48	98.88
	15	4.13		94.52	95.88
	18	3.12		91.65	92.97

	21	3.01		82.37	83.56
	24	2.08		79.87	81.43
DRSS	3	6.18	Ns	99.21	104.61
	6	6.14		94.38	102.41
	9	5.13		97.45	100.56
	12	4.12		96.46	97.86
	15	3.11		93.50	94.86
	18	2.10		90.63	91.95
	21	1.09		81.35	82.54
	24	1.07		80.33	81.20
L-RSS	3	7.15	ns	95.18	103.58
	6	6.11		93.35	101.38
	9	5.10		91.42	99.53
	12	4.09		89.43	96.83
	15	4.08		87.45	93.83
	18	3.07		85.43	90.92
	21	2.06		82.63	81.51
	24	1.04		80.31	80.55
TBRSS	3	7.13	Ns	89.54	102.54
	6	6.09		84.71	100.34
	9	5.08		87.78	98.49
	12	4.07		86.79	95.79
	15	3.06		83.83	92.79
	18	3.05		80.96	89.88
	21	3.04		70.64	80.47
	24	1.02		60.00	78.64
RERSS	3	6.10	Ns	97.14	101.53
	6	5.06		92.31	99.33
	9	5.05		95.38	97.48
	12	4.04		94.39	94.78
	15	4.03		91.43	91.78
	18	3.02		88.56	88.87
	21	2.01		79.28	77.63
	24	1.01		78.24	78.65

Table 20: Goodness of Fit and information criteria under Exponential (1)

Design	<i>m</i>	Chi-square	p-value	AIC	BIC
ERSS	3	7.03	Ns	132.3	129.13
	6	6.25		127.00	126.48
	9	6.23		120.48	124.25
	12	4.21		119.28	120.97
	15	4.20		115.70	117.35
	18	3.18		112.23	113.83
	21	2.17		101.01	102.44
	24	1.16		100.00	99.76
MRSS	3	6.28	Ns	116.48	122.68
	6	6.23		110.92	120.16
	9	5.21		114.46	118.03
	12	4.20		113.31	114.93
	15	4.19		109.91	111.48
	18	3.18		106.62	108.13
	21	2.17		95.96	97.32
	24	1.15		93.89	95.19
PRSS	3	7.27	Ns	112.63	118.63
	6	6.23		107.26	116.19
	9	5.21		110.68	114.14
	12	4.20		109.57	111.13
	15	3.18		106.29	107.80
	18	2.16		103.10	104.57
	21	1.14		92.79	94.11
	24	1.13		91.03	92.67
BGRSS	3	7.24	Ns	101.37	106.77
	6	6.20		96.54	104.57
	9	6.19		99.61	102.72
	12	4.18		98.62	100.02
	15	3.17		95.66	97.02
	18	2.16		92.79	94.11
	21	2.14		83.51	84.70
	24	1.12		81.01	82.57
DRSS	3	6.22	Ns	100.35	105.75
	6	6.18		95.52	103.55

	9	5.17		98.59	101.70
	12	4.16		97.60	99.00
	15	4.15		94.64	96.00
	18	3.14		91.77	93.09
	21	2.13		82.49	83.68
	24	1.11		81.47	82.34
L-RSS	3	7.19	Ns	99.32	104.72
	6	7.15		97.23	102.52
	9	6.14		95.43	100.67
	12	6.13		93.44	97.97
	15	5.12		91.87	94.97
	18	5.11		88.22	92.06
	21	4.01		86.35	82.65
	24	2.08		85.12	81.69
TBRSS	3	7.17	ns	96.13	103.68
	6	6.33		91.30	101.48
	9	5.12		94.37	99.63
	12	4.11		93.38	96.93
	15	4.01		90.42	93.93
	18	3.09		87.55	91.02
	21	2.38		77.23	81.61
	24	2.06		65.12	79.78
RERSS	3	6.14	Ns	95.12	102.67
	6	6.01		90.29	100.47
	9	5.09		93.36	98.62
	12	4.08		92.37	95.92
	15	3.47		89.41	92.92
	18	3.06		86.54	90.01
	21	2.03		83.76	78.77
	24	1.02		75.45	77.13

Utilizing a library, R Studio was used to evaluate changed RSS techniques graphically in terms of relative effectiveness and goodness of fit (ggplot2). With regard to relative effectiveness and goodness of fit, numerous ggplots were drawn using various R Studio programmes. Figures 1 through 3 show a perusal of data visualisation using ggplots. Appendix 1 provides descriptions of numerous R codes from graphical evaluation.

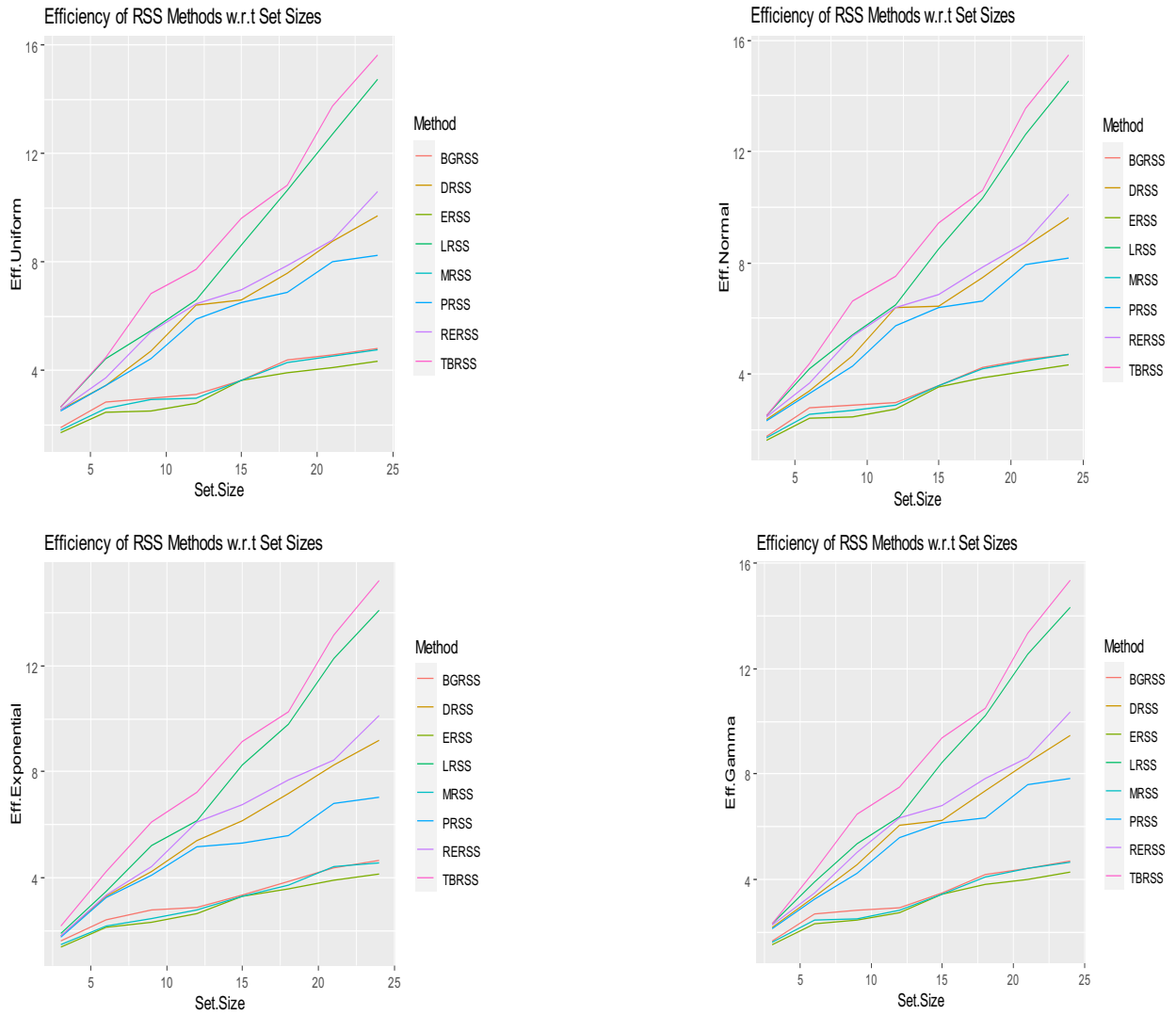
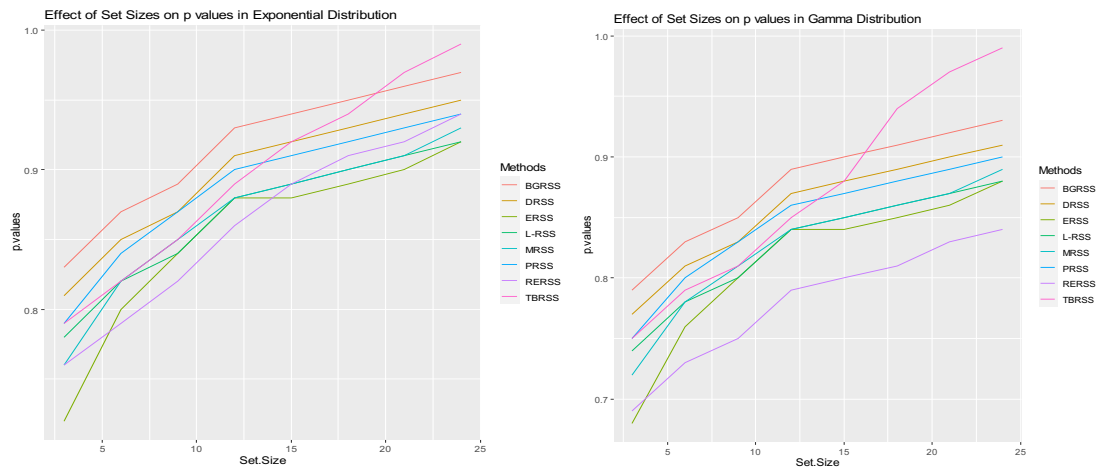


Fig 1: GGplots of the effectiveness of RSS techniques



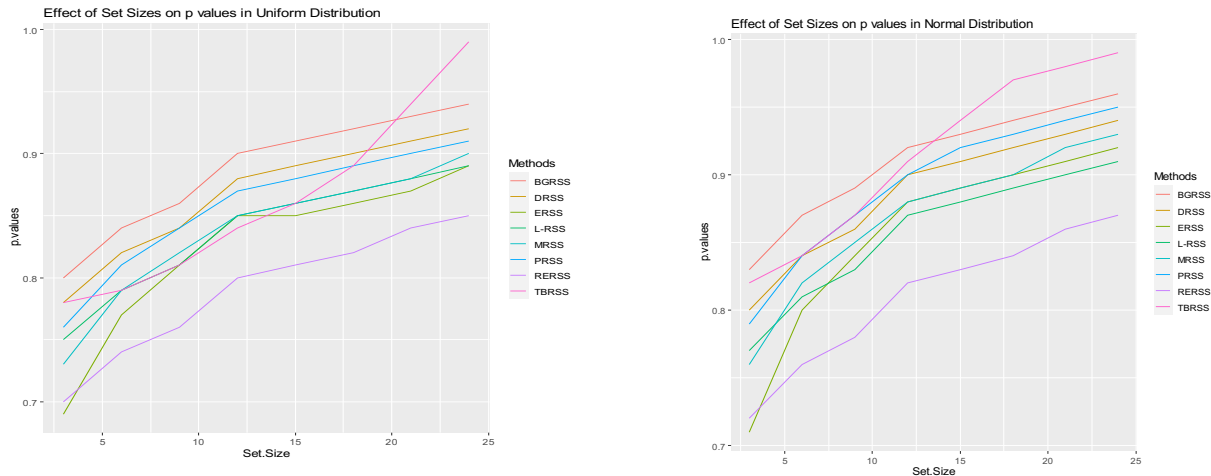


Fig 2: GGplots of p-values in relation to set sizes

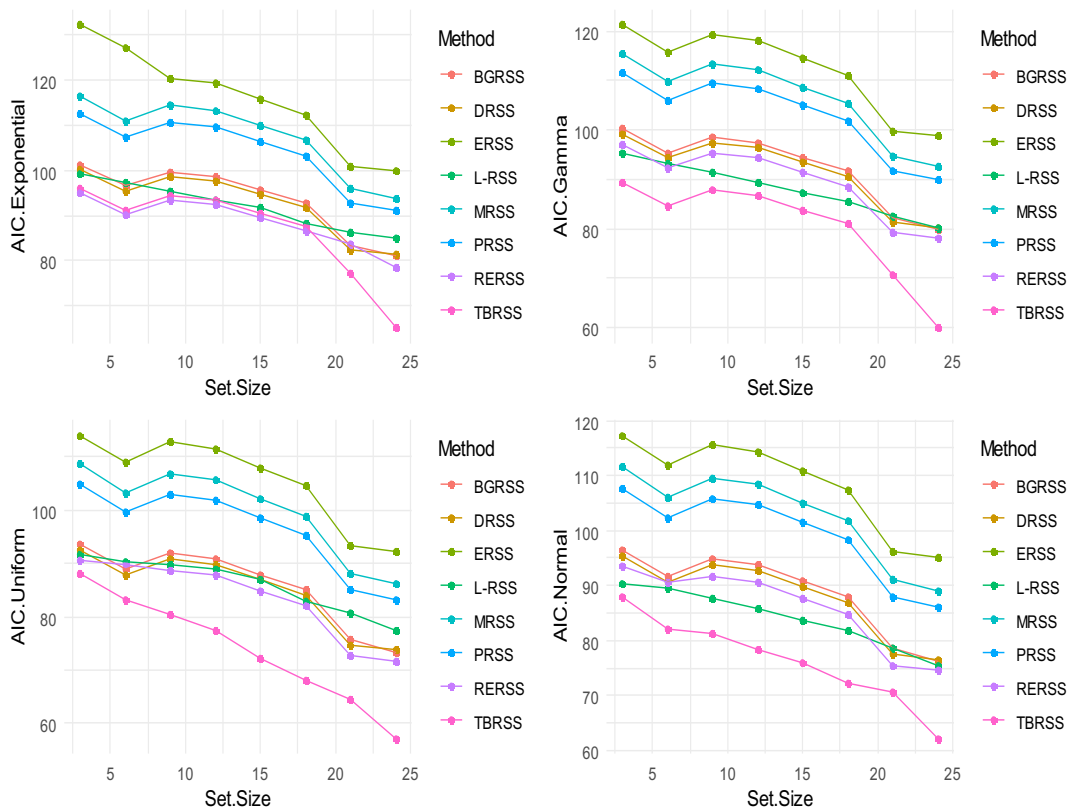


Fig 3: GGplots of Information standards

4 .Discussions

In R Studio, simulated data for 2000 observations was created using both skewed and unskewed probability distributions. All of these modified techniques were also employed to collect samples under various set sizes and constant cycle combinations. The purpose of the study was to determine whether RSS was practical in a framework with different changes and whether it was appropriate for both skewed and unskewed distributions. To create the simulated data, distributions like exponential, gamma, uniform, and normal were used. Utilizing the library (RSSampling) package, modified methods of RSS such as Extreme ranked set sampling (ERSS), Median ranked set sampling (MRSS), Percentile ranked set sampling (PRSS), Balanced groups ranked set sampling (BGRSS), Double ranked set sampling (DRSS), Robust ranked set sampling (L-RSS), Truncation based ranked set sampling (TBRSS), and Robust ranked set sampling (RERSS) were used to

draw samples with respect to different combination of set sizes. For graphical analysis, library (ggplot2) was used in addition to the library. We evaluated the performances of RSS-modified algorithms for both skewed and unskewed distributions. In compared to other modified RSS approaches, including traditional RSS, TBRSS was proven to be superior in terms of variance under skewed distribution. TBRSS worked better with an unskewed distribution this time, resulting in a smaller variance value compared to the alternatives. The relative efficacy of the RSS modified methods' performances was also evaluated. Under set size $m=3$, relative efficiency values for TBRSS were determined to be 2.647, 2.497, 2.189, and 2.317, respectively, spanning uniform, normal, exponential, and gamma distributions. These values were much higher than those for other modified methods of RSS.

The relative effectiveness increased when the set size was set to $m=6$. With relative efficiency values of 4.47, 4.38, 3.32, and 3.50 over uniform, normal, exponential, and gamma distributions, the results were once again in favour of TBRSS. For set sizes $m=9$ and $m=12$ value of relative efficiency was found significantly higher than those of other modified methods of RSS and once again in favour of TBRSS with the relative efficiency values of 6.82, 6.62, 6.11, 6.47 and 7.72, 7.52, 7.22, 7.49 across uniform, normal, exponential and gamma distributions, respectively relative efficiency values of 9.61, 9.44, 9.14 and 9.36 were found in case of TBRSS method across uniform, normal, exponential and gamma distributions, which were higher in comparison to its counterparts under set size $m=15$. According to uniform, normal, exponential, and gamma distributions, relative efficiencies of 10.82, 10.62, 10.27, 10.52, and 13.76, 13.55, 13.18, and 13.33 were observed, respectively, for set sizes $m=18$ and $m=21$. This indicates that TBRSS has greater relative efficiencies than other approaches. TBRSS once more surpassed other RSS algorithms with a value of 15.64, 15.46, 15.22, and 15.35 across uniform, normal, exponential, and gamma distributions when the set size was $m=24$. All of the adjusted RSS techniques under uniform distribution had AIC and BIC values that varied from 57.17 to 114.05 and 70.93 to 121.40, respectively, for goodness of fit analysis. The values of AIC and BIC for normal distribution across all RSS adjusted techniques ranged from 62.19 to 117.30 and 73.89 to 124.25, respectively. AIC and BIC values for exponential distributions across all RSS adjusted techniques ranged from 65.12 to 132.30 and 78.77 to 129.13, respectively. The AIC and BIC values for the gamma distribution for all RSS adjusted techniques ranged from 60.01 to 121.33 and 77.63 to 127.99, respectively. From the aforementioned data, it was discovered that TBRSS performed better across all distributions with lower and higher AIC and BIC values.

5. Conclusions

Compared to SRS, RSS is a more effective way of data collection, particularly when ranking is more cost-effective but measuring a unit is more expensive. In this study, TBRSS produced low variance values for skewed distributions when compared to other RSS methods and classical RSS, and the lowest variance values for unskewed distributions when compared to both classical RSS and other modified versions of RSS. TBRSS produced lower variance values for both skewed and unskewed distributions under various set size variations, with the lowest values for skewed distributions being 0.0700 and 0.1320 for exponential and gamma distributions under set size $m=24$, and the lowest values for unskewed distributions being 0.0055 and 0.0644 for uniform and normal distributions under set size $m=24$. Regardless of the kind of distribution, TBRSS surpassed all modified RSS methods in terms of relative efficiency, with a relative efficiency score ranging from 2.189 to 15.649. A more structured sample is drawn using the truncation notion in TBRSS, which is a modified version of different modified RSS methods. This result in an increase in the efficiency of the procedure. A noteworthy finding from this study was that, according to the McIntyre principle, the relative efficiency of all approaches grew as set size increased and reduced when the underlying distribution was asymmetric in character. However, it never fell below 1.

References

- [1] G.A. McIntyre., A method for unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research*, 3:385-390, 1952.
- [2] M.I. Jeelani, and C.N. Bouza., Efficiency of Ranked Set Sampling in Horticultural Surveys. *Cadernos do IME – Série Estatística*, 38:37-48, 2015b.
- [3] B., Sevinc, B., Cetintav, M. Esemem, and S. Gurler., RSSampling: A Pioneering Package for Ranked Set Sampling. *The R Journal*, 11(1): 401-405, 2019.
- [4] B., Robertson, J. Brown, and T.McDonald., The Use of Ranked Set Sampling with Balanced Acceptance Sampling. 34th International Workshop on Statistical Modelling (IWSM), Guimaraes, Portugal. pp 7-12, 2019.
- [5] E. Zamanzade, and M. Mahdizadeh., Using ranked set sampling with extreme ranks in estimating the population proportion. *Statistical Methods in Medical Research*, 29(1):165–177, 2020.

- [6] K.U.I. Rather, and C. Kadilar,. Exponential Type Estimator for the population Mean under Ranked Set Sampling. *Journal of Statistics: Advances in Theory and Applications*, 25(1), 1–12, (2021).
- [7] K.U.I., Rather, S.E.H., Rizvi, M. Sharma, and , M. I. Jeelani. Optimum Stratification for Equal Allocation using Ranked Set Sampling as Method of Selection. *International Journal of Statistics and applied mathematics*, 7(2): 63-67, (2022).
- [8] K.U.I., Rather, A., Asad, M. Iqbal Jeelani an improved difference cum – exponential ratio type estimator in ranked set sampling. *Reliability: Theory & Applications.*, 4(71): 469-476, 2022.
- [9] R., Khalid, E. G., Koçyiğit, & C. Ünal,. New Exponential Ratio Estimator in Ranked Set Sampling. *Pakistan Journal of Statistics and Operation Research*, 18(2), 403-409, (2022). <https://doi.org/10.18187/pjsor.v18i2.3921>
- [10] K.U.I. Rather, and S.E.H. Rizvi,. Approximate Optimum Strata Boundaries for Equal Allocation under Ranked Set Sampling. *Journal of Scientific Research*, 66(3): 302-307, (2022).
- [11] H.M. Samawi, and H.A. Muttalak,. Estimation of ratio using rank set sampling. *Biometrical Journal*, 38: 753-764, 1996.
- [12] H.A. Muttalak,. Median Ranked Set Sampling. *Journal of Applied Statistical Sciences*, 6:245-255, 1997.
- [13] H.A. Muttalak,. Modified ranked set sampling methods. *Pakistan Journal of Statistics*, 19(3):315-323, 2003a.
- [14] A.A., Jemain, A.I. Al-Omari, and K. Ibrahim,. Some variations of ranked set sampling. *Electronic Journal of Applied Statistical Analysis*, 1:1–15, 2008a.
- [15] M.F. Al-Saleh, and M. Al-Kadiri,. Double ranked set sampling. *Statistics & Probability Letters*, 48:205–212, 2000.
- [17] H.M. Samawi,. On double extreme ranked set sample with application to regression estimator. *Metron*, 1(2):53-66, 2002.
- [18] A.A., Jemain, A.I., Al-Omari, and K. Ibrahim,. Balanced groups ranked set sampling for estimating the population median. *Journal of Applied Statistical Sciences*, 17(1): 39-46, 2008b.
- [19] H.M. Samawi, and E.M. Tawalbeh,. Double median ranked set sampling: Comparison to other double ranked set samples for mean and ratio estimators. *Journal of Modern Applied Statistical Methods*, 1(2):428-442, 2002.
- [20] H.M, Samawi, M.S. Ahmed, and W.A. Abu-Dayyeh,. Estimating the population mean using extreme ranked set sampling. *Biometrical Journal*, 38:577-586, 1996.
- [21] H.A. Muttalak,. Investigating the use of quartile ranked set samples for estimating the population mean. *Applied Mathematics and Computation*, 146(2):437–443, 2003b.
- [22] A.A. Jemain, and A.I. Al-Omari,. Double percentile ranked set samples for estimating the population mean. *Advances and Applications in Statistics*, 6(3):261-276, 2006.

Appendix 1

(a) R Studio Codes for ggplots in term of efficiency:

```
library(ggplot2)
data <- read.table("clipboard", header = TRUE)
names(data)
p1 <- ggplot(data, aes(x=Set.Size, y=Eff.Uniform, colour=Method ,group=Method)) +
geom_line() +
ggtitle("Efficiency of RSS Methods w.r.t Set Sizes ")
p1
p2 <- ggplot(data, aes(x=Set.Size, y=Eff.Normal, colour=Method ,group=Method)) +
geom_line() +
```

```

ggtitle("Efficiency of RSS Methods w.r.t Set Sizes ")
p2
p3 <- ggplot(data, aes(x=Set.Size, y=Eff.Exponential , colour=Method ,group=Method)) +
geom_line() +
ggtitle("Efficiency of RSS Methods w.r.t Set Sizes ")
p3
p4 <- ggplot(data, aes(x=Set.Size, y=Eff.Gamma , colour=Method ,group=Method)) +
geom_line() +
ggtitle("Efficiency of RSS Methods w.r.t Set Sizes ")
p4

```

(b) R Studio codes of ggplots for p-value:

```

library(ggplot2)
p1<- ggplot(data, aes(x=Set.Size, y=p.values, colour=Methods ,group=Methods)) +
geom_line() +
ggtitle("Effect of Set Sizes on p values in Exponential Distribution ")
p1
p2 <- ggplot(data, aes(x=Set.Size, y=p.values, colour=Methods ,group=Methods)) +
geom_line() +
ggtitle("Effect of Set Sizes on p values in Gamma Distribution ")
p2
p3 <- ggplot(data, aes(x=Set.Size, y=p.values, colour=Methods ,group=Methods)) +
geom_line() +
ggtitle("Effect of Set Sizes on p values in Uniform Distribution ")
p3
p4 <- ggplot(data, aes(x=Set.Size, y=p.values, colour=Methods ,group=Methods)) +
geom_line() +
ggtitle("Effect of Set Sizes on p values in Normal Distribution ")
p4

```

(c) R Studio codes of ggplots for Information criteria:

```

rss=read.table("clipboard",header=T)
library(ggpubr)
names(rss)
E1=ggplot(data=rss,aes(x=Set.Size, y=AIC.Exponential,color=Method)) + geom_point() +geom_line() + theme_minimal()
E2=ggplot(data=rss,aes(x=Set.Size, y=AIC.Gamma,color=Method)) + geom_point() +geom_line() + theme_minimal()
E3=ggplot(data=rss,aes(x=Set.Size, y=AIC.Uniform,color=Method)) + geom_point() +geom_line() + theme_minimal()
E4=ggplot(data=rss,aes(x=Set.Size, y=AIC.Normal,color=Method)) + geom_point() +geom_line() + theme_minimal()
p=ggarrange(E1,E2,E3,E4)
p

```