# Stratified Dual-Rank Ranked Set for Estimating Population Mean

Zahid Khan[1], Amina Shahzadi[2], Naureen Riaz[3], Ammara Nawaz Cheema[4*], Zaid Muhammad[5]

[1]*Department of Statistics, University of Malakand, Chakdara, Dir Lower, Pakistan, Zahid.uom1@gmail.com*
[2]*Department of statistics, GC University Lahore, aminashahzadi@gcu.edu.pk*
[3]*Department of Software Engineering, Lahore Garrison University DHA phase 6 Lahore, noureenriaz@lgu.edu.pk*
[4]*Department of Mathematics, Air University Islamabad Pakistan, ammara.cheema@au.edu.pk*
[5]*M.Phil Scholar, Department of Statistics, University of Malakand, zaidstatsuom@gmail.com*

***Abstract:*** In surveys when measuring units is expensive, ranked set sampling (RSS) is a popular and economical sampling technique. The RSS algorithm selected units using a ranking procedure. Either eye inspection or an auxiliary variable is used for ranking. In the present paper, 'stratified dual-rank ranked set sampling' (SDuRSS) method is suggested to estimate population mean. The proposed design used dual ranking instead of traditional ranking method. The mean and variance of the suggested scheme is derived. The performance of the mean estimator of proposed scheme is investigated by relative efficiency (RE) of the estimator. A simulation study is conducted for computing such relative efficiency which shows that the proposed design is more efficient than stratified ranked set sampling (SRSS) and stratified extreme ranked set sampling (SERSS). The proposed scheme is illustrated with real data set where it also shows superiority on the SRSS and SERSS.

Keywords: dual-rank, suggested, design, efficient, SRSS, SERSS.

## 1. Introduction

Ranked set sampling (RSS) not only depend only on the actual quantification of the sample element, but it uses ranking of elements for improving efficiency of parameter estimation. The ranking is done by any cost-free method usually be eye or by concomitant variables. In many filed like agriculture, environmental studies, and finance the classical random methods are less practical due to high expenses on measurement of elements. McIntyre (1952) introduced RSS method in order to decreases the quantification cost in field of agriculture. Dell and Clutter (1972) were among the first to highlight the efficiency of RSS compared to Simple Random Sampling (SRS).

The RSS method is then not only a method of selection element from population but it uses for estimation of parameters. Due to enrollment of ranking mechanism in RSS procedure the estimates of parameters improved. It was shown that RSS provides more accurate estimates, especially when ranking was easier than direct measurement. In RSS method, initially random sample of size $n^2$ is drawn which then distributing in n sets each of size n elements. Then, ranking is made within each set by cost-free method, and one element is drawn from each set from lowest to largest position.

When ranking is easier than quantification, this approach lowers the variance of estimators in comparison to simple random sampling (SRS). It is especially helpful in domains where efficient and accurate ranking is possible. The ranking process's correctness determines how effective this design is. Biased estimations or decreased precision can result from inaccurate ranking (McIntyre, 1952).

Identifying the extreme elements, that is, lowest ranked and highest ranked, from each ordered set is the procedure of extreme RSS (ERSS) scheme. The design is effective in case of tails distribution. In some field like environmental monitoring and risk assessment, the ERSS method estimate the estimator with higher efficiency (Samawi et al., 1996). The underlying distribution has a major impact on how effective the design is. It might not always perform better than traditional RSS, particularly in cases when the distribution is not heavy-tailed. Every rank position (first, second, third, etc.) is guaranteed to be equally represented throughout the sampled sets thanks to balanced RSS.

Using different ranking techniques or choosing more than one observation from each ranked set are examples of variants on generalized RSS. These designs are flexible enough to meet many kinds of research requirements. In situations where standard approaches might not work well, generalized RSS designs can be tailored for different situations and provide better results (Bai and Chen, 2003).

Al-Saleh and Al-Kadiri (2000) developed double ranked set sampling (DRSS) which is an advanced sampling technique that enhances the efficiency of traditional RSS by incorporating a second round of ranking to improve estimation accuracy. It is particularly useful when measuring the variable of interest is expensive or time-consuming, but ranking the units based on some auxiliary information is relatively easy. Khan, Isamail and Samawi (2020) introduce mixture ranked set sampling (MIRSS) which is a modified version of RSS that combines multiple ranking schemes or set sizes to improve estimation efficiency. It provides better representation and statistical efficiency compared to simple RSS by allowing flexibility in the sampling process. RSS is extended to other designs see (Khan et al, 2022; Khan and Ismail, 2019; Khan, Ismail and Noor-ul-Amin (2022); Khan and Ali, 2022).

## 2. Ranked Set Sampling (RSS)

In RSS first define the population and sampling units. Choose a set size (n) and the number of cycles (r), where the total sample size is n×r. Select $n^2$ random samples from the population without replacement. Using expert judgment, visual assessment, or auxiliary information, rank the n units in each set from smallest to largest (or vice versa). Ranking is done without actual measurement to save cost and resources. From each ranked set, choose one unit based on a predetermined ranking position. For example, in the first set, select the smallest-ranked unit; in the second set, select the second smallest, and so on. Only the selected units from each set are measured precisely. This ensures that each rank position is represented in the final dataset.

Let $X$ be the study variable with probability density function $f_{x(x)}$, cumulative density function $F_{x(x)}$, mean $\mu_X$ and variance $\sigma_X^2$ Let $X_{11j}, X_{12j}, ......, X_{1nj}, X_{21j}, X_{22j}, ......, X_{2nj}, ......, X_{n1j}, X_{n2j}, ......, X_{nnj}$ be the $n$ independent simple random sample each of size $n$ from $j^{th}$ (for $j = 1, 2, ....., r$) cycle drawn from $f_{x(x)}$. The mean and variance of RSS is as follows,

$$\bar{X}_{RSS} = \frac{\sum_{j=1}^{r}\sum_{i=1}^{n} X_{i(i:n)j}}{rn} \quad (1)$$

The variance is as follows,

$$\text{var}\left(\bar{X}_{RSS}\right) = \frac{\sigma_X^2}{rn} - \frac{1}{rn^2}\sum_{i=1}^{n}\left(\mu_{i(i:n)} - \mu\right)^2 \quad (2)$$

Where, $\mu_{i(i:n)}$ is the mean of $i^{th}$ order statistics.

Stratified Ranked Set Sampling (SRSS)

Stratified Ranked Set Sampling is invistaged by Samwi (1996a). The procedure SRSS for getting a sample of size $n$ is to divide the population into H mutually exclusive and exhaustive strata. Then, select an independently ranked set sample of size $j_h n_h$ units from stratum $h$ in $r_h$ cycle. Let $X_{[i:n]j}^h$ be the $i^{th}$ judgment order statistic in the $j^{th}$ cycle of the ranked set sample collected from stratum $h$. The observations $X_{[i:n]j}^h$ $(h = 1, ...., H; i = 1, .., n_h; j = 1, ...r_h)$ are independent, but not identically distributed. For fixed $h$ and $i$, $X_{[i]j}^h$'s $(j = 1, ...r_h)$ are identically distributed, where the common mean and variance are denoted by $\mu_{[i:n]h}$ and $\sigma_{[i:n]h}$ respectively. The population mean estimator under SRSS is given by

$$\bar{X}_{(SRSS)} = \sum_{j=1}^{r} \frac{N_h}{N} \bar{X}_{(SRSS,h)}, \quad (3)$$

where,

$\bar{X}_{(SRSS,h)} = \frac{1}{n_h r_h} \sum_{i=1}^{n_h} \sum_{j=1}^{r_h} X_{[i:n]j}^{h}$ is the RSS estimator of the mean in stratum $h$.

The variance is given by

$$\text{var}(\bar{X}_{(SRSS)}) = \sum_{h=1}^{H} \left( \frac{N_h}{N n_h} \right)^2 \frac{1}{r_h} \sum_{i=1}^{n_h} \sigma_{[i:n]h}^2$$

Where

$$\sum_{i=1}^{n_h} \sigma_{[i:n]h}^2 = n_h \sigma_h^2 - \sum_{i=1}^{n_h} (\bar{X}_{(i,n)h} - \mu_h)^2$$

Combining we get,

$$\text{var}(\bar{X}_{(SRSS)}) = \text{var}(\bar{X}_{(SSRS)}) - \sum_{h=1}^{H} \left( \frac{N_h}{N} \right)^2 \frac{1}{n_h^2 r_h} \sum_{i=1}^{n_h} (\bar{X}_{(i,n_h)h} - \mu_h)^2 \quad (4)$$

**3. Dual-rank ranked set sampling (DRRSS)**

Taconeli (2024) developed DRRSS scheme which used double ranking of element within each set. The procedure of DRRSS is to draw $n^2$ elements from the target population, and distributing them into $n$ sets each of size $n$. Each set is ranked visually or by any low-cost method. Again units within each set is ranked row wise. After ranking each set, select median order element from each set for actual quantification. Repeat this procedure r times (cycles) to draw a DRRSS sample of size N = rn. The mean estimator under DRRSS is as under,

$$\bar{X}_{(SDRRSS)} = \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} X_{[\frac{n+1}{2}:n]j}$$

Proposed Sampling Design

The proposed design stratified dual ranked set sampling (SDuRSS) is cost-effective and more effcient compare to its counterpart designs. The procedure of SDuRSS is as following,

First dived the untis in populaiton into H mutually exclusive and exhaustive strata. Then, for h$^{th}$ stratum, select $n^2$ elements randomly from population. Distribute the selcted element into $n_h$ sets each of same size. After ranking each set visually or by any cost free method, pick one element from each set in order, that is, from lowest to higest or vice versa. For actualy qauntification pick median ranked element from each set. Repeat this procedure (cycle) $r_h$ time in order to get stratifed dual-ranked set sample of size $r_h n_h$.

Mean and Variance

The mean estimator under SDuRRSS is given by

$$\bar{X}_{(SDuRSS)} = \sum_{j=1}^{r} \frac{N_h}{N} \bar{X}_{(SDuRSS,h)}, \quad (3)$$

where,

$\bar{X}_{(SDuRSS,h)} = \frac{1}{n_h r_h} \sum_{i=1}^{n_h} \sum_{j=1}^{r_h} X_{[\frac{n+1}{2}:n]j}^{ih*}$ is the SDuRSS estimator of the mean in stratum $h$.

The variance is given by

$$\text{var}(\bar{X}_{(SDuRSS)}) = \sum_{h=1}^{H} \left( \frac{N_h}{N n_h} \right)^2 \frac{1}{r_h} \sum_{i=1}^{n_h} \sigma_{[\frac{n+1}{2}:n]}^{2\ \ ih*}, \quad (4)$$

where,

$$\sum_{i=1}^{n_h} \sigma^{2\ \ ih*}_{[i:\frac{n+1}{2}]} = n_h \sigma_h^2 - \sum_{i=1}^{n_h} (\bar{X}_{(i,\frac{n+1}{2})h} - \mu_h)^2 \,,$$

combining we get,

$$\text{var}(\bar{X}_{(SDuRSS,h)}) = \text{var}(\bar{X}_{(SDuRSS,h)}) - \sum_{h=1}^{H} \left( \frac{N_h}{N} \right)^2 \frac{1}{n_h^2 r_h} \sum_{i=1}^{n_h} (\bar{X}_{(\frac{n+1}{2},n_h)h} - \mu_h)^2$$

Similarly, the mean and variance for even sample sizes is as under,

$$\bar{X}_{(SDuRSS)} = \sum_{j=1}^{r} \frac{N_h}{N} \bar{X}_{(SDuRSS,h)} \,,$$

where,

$$\bar{X}_{(SDuRSS,h)} = \frac{1}{n_h r_h} \sum_{i=1}^{n_h} \sum_{j=1}^{r_h} \left( X^{ih*}_{[\frac{n}{2}:n]j} + X^{ih*}_{[\frac{n+2}{2}:n]j} \right)$$ is the RSS estimator of the mean in stratum $h$ .

The variance is given by,

$$\text{var}(\bar{X}_{(SDuRSS)}) = \sum_{h=1}^{H} \left( \frac{N_h}{N n_h} \right)^2 \frac{1}{r_h} \sum_{i=1}^{n_h} \left( \sigma^2_{[\frac{n}{2}:n]^h} + \sigma^2_{[\frac{n+2}{2}:n]^h} \right),$$

After simplification we get,

$$\text{var}(\bar{X}_{(SDuRSS,h)}) = \text{var}(\bar{X}_{(SDSRS,h)}) - \sum_{h=1}^{H} \left( \frac{N_h}{N} \right)^2 \frac{1}{n_h^2 r_h} \sum_{i=1}^{n_h} (\bar{X}_{(i,n_h)h} - \mu_h)^2 \,.$$

## 4. Simulation Study

Simulation study is conducted with 50,000 iterations. The equation for relative efficiency (RE) of DRRSS, SRSS with respect to SSRS for symmetric distribution is define as,

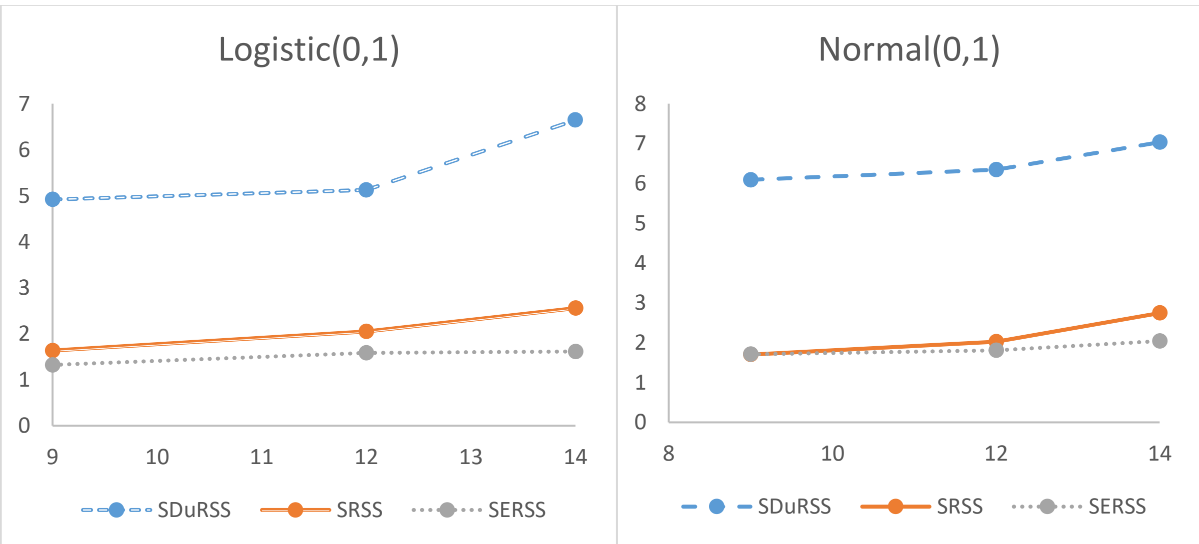
$$RE\left(\bar{X}_t, \bar{X}_{SRS}\right) = \frac{\text{var}\left(\bar{X}_{SRS}\right)}{\text{var}\left(\bar{X}_t\right)}$$

Where, t=SDuRSS, SRSS

For asymmetric distribution efficiency is define as,

$$RE\left(\bar{X}_t, \bar{X}_{SRS}\right) = \frac{\text{var}\left(\bar{X}_{SRS}\right)}{MSE\left(\bar{X}_t\right)}$$

Four distributions, Normal (0,1), Lognormal (0,1), Weibull (5,6), Gamma (3,1) are considered for simulation study.



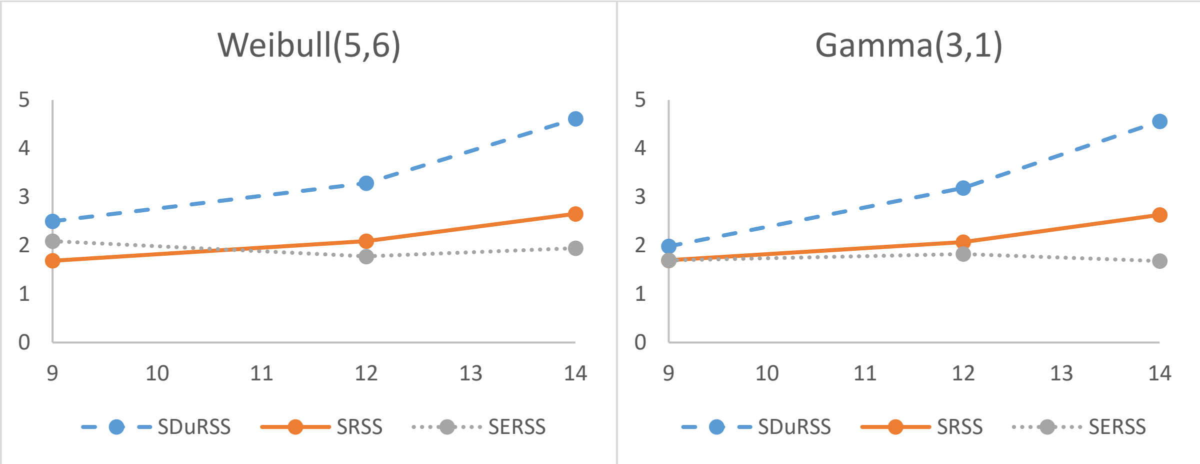



Figure 1 Relative Efficiency of SDuRSS, SRSS, SERSS relative of SRS

The relative efficiencies of SDuRSS, SRSS, and SERSS is presented in Figure 1. The Normal, Logistic, Gamma, and Weibull distribution are considered for generating samples. The selected distribution consists of both symmetric and asymmetric. The Figure shows higher relative efficiencies for SDuRSS than SRSS and SERSS. Thus, in selected distributions the suggested distribution performs well.
The SDuRSS has very high relative efficiency than its counterpart scheme SRSS. This shows that the dual ranking technique is very effective. Thus, experimenter may use the SDuRSS instead of SRSS.

## 5. Practical application

The DuRSS is illustrated with practical data of high of individuals (Statistics online computation resource data; 1993). Initially, sample of size 16 are randomly drawn from the given data. Then, it is distributed into 4 sets each of size 4. Then, the data in each set is ranked. These four sets are presented in Table 1. The Table 2 shows the sets with obtained after ranking of 1st, 2nd, 3rd and 4th element of each set. Thus, new sets formed, according to the procedure of DuRSS. Finally, lowest ranked unit from first set and second set, highest from third and fourth set is drawn. Thus, sample of size 4 DuRSS are selected. The RSS procedure is applied in Table 1, and four samples are selected. Similarly, from the given data of Hight of individual, 4 simple random sample are drawn. The DuRSS, RSS, and SRS samples are shown in Table 3. Finally, the mean and variance of each sample calculated and presented in Table 4.

Table 1. Ranked Sets of Hight of individuals.

| set 1 | set2 | set3 | set4 |
|---|---|---|---|
| 65.38356 | 66.36418 | 67.01255 | 63.42577 |
| 66.54376 | 66.50418 | 68.30355 | 66.28644 |
| 70.18447 | 67.537 | 70.06306 | 66.76711 |
| 70.40617 | 68.99958 | 70.55703 | 68.88741 |

Table 2 Ranking of elements row wise in Table 1, new sets are formed

| set 1 | set2 | set3 | set4 |
|---|---|---|---|
| 63.42577 | 65.38356 | 66.36418 | 67.01255 |
| 66.28644 | 66.50418 | 66.54376 | 68.30355 |
| 66.76711 | 67.537 | 70.06306 | 70.18447 |
| 68.88741 | 68.99958 | 70.40617 | 70.55703 |

Table 3. Sample of size 4 under DuRSS, RSS and SRS Designs

| DuRSS | SERSS | SRS |
|---|---|---|
| 63.42577 | 60.42577 | 129.3137 |
| 66.50418 | 66.28644 | 70.84278 |
| 70.06306 | 68.18447 | 69.92442 |
| 70.40617 | 70.55703 | 64.28508 |

Table 4. Mean and variance of DuERSS, RSS, SRS method

| | DuRSS | SERSS | SRS |
|---|---|---|---|
| mean | 67.63343 | 67.965 | 83.5915 |
| variance | 10.8523 | 18.7232 | 30.61904 |

## 6. Conclusion

In the present study a new sampling scheme, stratified dual ranked set sampling (SDuRSS) is suggested. The mean and variance of SDuRSS is developed. The efficiency of suggested scheme is assessed by conducting simulation study which shows higher efficiency of SDuRSS. The suggested design performs well compare to some of its counterpart designs. Ratio and product estimator of the suggested design can be developed in future study.

## References

1. Taconeli, C. A. (2024). Dual-rank ranked set sampling. Journal of Statistical Computation and Simulation, 94(1), 29-49.
2. McIntyre GA. A. (1952). method for unbiased selective sampling using ranked sets. Aust J Agric Res. 3, 385–390.
3. Dell TR, Clutter JL. (1972). Ranked set sampling theory with order statistics background. Biometrics, 28, 545–555.
4. Khan, Z., Muhammad, A., Aftaf, M. H. and Ismail, M. (2022). Modified median ranked set sampling scheme for estimating population mean. International Journal of Computational Intelligence in Control, 14(9), 121-127.
5. Khan, Z., & Ismail, M. (2019). Modified Ratio Estimators of population mean based on ranked set sampling. Pakistan Journal of Statistics and Operation Research,15 (2), 445-449.
6. Khan, Z., Ismail, M., & Noor-ul-Amin, M. (2022). Stratified mixture ranked set sampling for estimation of population mean and median. Pak. J. Statist, 38(2), 197-210.
7. Khan, Z., Ismail, M., & Samawi, H. (2020). Mixture ranked set sampling for estimation of population mean and median. Journal of Statistical Computation and Simulation, 90(4), 573-585.
8. Khan, Z., & Ali, A. (2022). Anemia and its Causes among Children in District Dir Lower Applying Ranked Set Sampling Method for Data Collection. Research Journal of Social Sciences and Economics Review, 3(4), 34-37.
9. Al-Saleh, M. F., & Al-Kadiri, M. A. (2000). Double-ranked set sampling. Statistics & Probability Letters, 48(2), 205-212.
10. Samawi, H. M., Ahmed, M. S., & Abu-Dayyeh, W. (1996). Estimating the population mean using extreme ranked set sampling. Biometrical Journal, 38(5), 577-586.
11. Samawi, H. M. (1996a). Stratified ranked set sampling. Pakistan Journal of statistics. 12(1), 9-16.
12. Bai, Z., & Chen, Z. (2003). On the theory of ranked-set sampling and its ramifications. Journal of Statistical Planning and Inference, 109(1-2), 81-99.
13. Samawi, H. M. (1996). Stratified ranked set sampling. Pakistan Journal of Statistics, 12(1), 9-16.