RANKED SET SAMPLING APPROACH FOR ACCURATE ESTIMATION OF POTATO PRODUCTION AND QUALITY TRAITS

Khalid Ul Islam Rather[1] and Durgesh Nandan
School of CS & AI, SR University, Warangal, Telangana 506371, India.

ABSTRACT:

Accurate estimation of crop yield and quality traits is essential for effective agricultural planning and decision-making. Traditional sampling methods, such as Simple Random Sampling (SRS), may be inefficient or require larger sample sizes to achieve reliable estimates. This study explores the application of Ranked Set Sampling (RSS) to estimate potato production and quality traits across diverse agro-climatic conditions. Data were collected from multiple potato-growing regions, with samples ranked based on easily measurable characteristics before selecting subsets for detailed analysis. The results demonstrate that RSS provides more precise estimates of potato yield, tuber size, and quality parameters compared to conventional SRS, reducing sampling variability while maintaining cost-effectiveness. These findings highlight the potential of RSS as a robust and efficient tool for agricultural research and crop monitoring programs.

KEYWORDS: Ranked Set Sampling, Potato Production, Quality Traits, Sampling Efficiency, Agro-Climatic Conditions.

MSC Code: 62D05

RESUMEN:

La estimación precisa del rendimiento y las características de calidad de los cultivos es esencial para una planificación agrícola eficaz y la toma de decisiones. Los métodos de muestreo tradicionales, como el Muestreo Aleatorio Simple (MAS), pueden ser ineficientes o requerir tamaños de muestra mayores para obtener estimaciones fiables. Este estudio explora la aplicación del Muestreo por Conjuntos Clasificados (MSC) para estimar la producción de papa y las características de calidad en diversas condiciones agroclimáticas. Se recopilaron datos de múltiples regiones productoras de papa, y las muestras se clasificaron según características fácilmente medibles antes de seleccionar subconjuntos para un análisis detallado. Los resultados demuestran que el MSC proporciona estimaciones más precisas del rendimiento de la papa, el tamaño del tubérculo y los parámetros de calidad en comparación con el MSC convencional, lo que reduce la variabilidad del muestreo y mantiene la rentabilidad. Estos hallazgos resaltan el potencial del MSC como una herramienta robusta y eficiente para la investigación agrícola y los programas de monitoreo de cultivos.

PALABRAS CLAVE: Muestreo por Conjuntos Clasificados, Producción de Papa, Características de Calidad, Eficiencia del Muestreo, Condiciones Agroclimáticas

1. INTRODUCTION:

Potato (Solanum tuberosum L.) is one of the most important staple crops globally, serving as a major source of carbohydrates, vitamins, and minerals. Accurate estimation of potato yield and quality traits is critical for farmers, policymakers, and researchers to optimize production, ensure food security, and enhance market competitiveness. Conventional sampling methods, such as Simple Random Sampling (SRS), are widely used but often require large sample sizes to achieve precision, which can be time-consuming and resource-intensive.

Ranked Set Sampling (RSS) is a statistically efficient sampling technique that combines visual or easily measurable ranking with selective measurement to improve estimation accuracy. By ranking sample units prior to detailed measurement, RSS reduces sampling variability and enhances the reliability of parameter estimates, particularly for heterogeneous populations. While RSS has been applied successfully in horticultural and epidemiological studies, its use in estimating potato production and quality traits remains limited.

This study aims to evaluate the application of RSS for potato yield and quality estimation. Specifically, it compares the precision of RSS against SRS for key production parameters, including tuber size, total yield, and quality characteristics, across different agro-climatic regions. The findings are expected to provide practical insights for implementing efficient sampling strategies in potato research and production monitoring programs.

The remainder of this paper is organized as follows: Section 2 reviews relevant literature on Ranked Set Sampling and its applications in agriculture. Section 3 describes the study area, sampling design, and statistical methodology. Section 4 presents and discusses the empirical results comparing RSS and SRS. Finally, the conclusions and recommendations for future research are provided in the last section.

[1] khalidstat34@gmail.com

2. REVIEW OF LITERATURE

Ranked Set Sampling (RSS), introduced by McIntyre (1952), is a powerful statistical technique designed to improve the precision of population parameter estimation, particularly in cases where direct measurement is costly or time-consuming. The method involves ranking sample units based on easily measurable characteristics before selecting a subset for detailed measurement, resulting in more accurate and efficient estimates compared to Simple Random Sampling (SRS) (McIntyre, 1952; Chen & Bai, 2009). In agricultural research, RSS has been extensively applied to enhance the estimation of crop yield and quality traits. Jeelani et al. (2015) demonstrated its effectiveness in horticultural crops, while Li and Stokes (2019) applied RSS in potato and tuber quality assessment, showing reduced sampling variability and higher precision than conventional methods. Similarly, Dell and Clutter (1972), Stokes (1997), and Frey and Patil (2002) highlighted the theoretical foundations and practical applications of RSS in heterogeneous agricultural populations.

Recent advancements have focused on integrating RSS with ratio and exponential-type estimators to further improve estimation accuracy. Riyaz et al. (2023) proposed a ratio estimator of the population mean using a new linear combination under RSS, while Rather and Kadilar (2021) introduced an exponential-type estimator for the population mean. Koyuncu and Kadilar (2009) also examined improvements of ratio estimators under RSS, demonstrating higher efficiency over SRS-based approaches. Moreover, stratification techniques combined with RSS have been explored to optimize sample allocation. Rather et al. (2022a, 2022b) developed methods for determining approximate optimum strata boundaries and optimum stratification for equal allocation, showing significant gains in estimation efficiency.

The applicability of RSS in real-world sampling has also been enhanced to address challenges such as non-response and multistage sampling. Rather and Koçyiğit (2024) proposed a generalized exponential ratio estimator under RSS with non-response, while Xu and Bai (2012) and Wang and Chen (2014) investigated multistage and stratified RSS designs for heterogeneous populations. Nonparametric and Bayesian approaches have been integrated with RSS to improve its flexibility and applicability (Amini & Mahdizadeh, 2015; Alvandi et al., 2023). Extensive studies have confirmed the efficiency of RSS in agricultural surveys, including applications in apples, potatoes, and other horticultural crops (Baklizi & Plante, 2007; Chen & Huang, 2011; Balakrishnan & Ling, 2010; Sinha & Das, 2015). These studies collectively highlight that RSS reduces standard errors and improves cost-effectiveness, making it particularly suitable for crops with significant genotype-by-environment variability, such as potatoes (Balakrishnan & Dudewicz, 2001; Stokes, 1997). Furthermore, foundational works on variance components (Searle et al., 2006) and comparative studies of RSS versus SRS (Dell & Clutter, 1972; Ridout, 2005) provide a strong theoretical basis for adopting RSS in agricultural research. Overall, RSS is increasingly recognized as a robust and practical technique for accurately estimating potato production and quality traits across diverse agro-climatic conditions, offering significant advantages over traditional SRS-based approaches.

3. METHODOLOGY

1. Study Area and Data Collection

The study was conducted across major potato-growing regions in India, representing diverse agro-climatic zones. Representative fields were selected to capture variability in soil type, climate, and cultivation practices. Potato samples were collected during the main harvesting season, ensuring uniformity in crop age and management practices.

2. Experimental Material

The study included multiple potato genotypes commonly grown in the selected regions. Key quality traits measured included tuber size, total yield per plant, dry matter content, and other relevant quality parameters such as starch content and vitamin C.

3. Sampling Design

3.1 Ranked Set Sampling (RSS) Procedure

In this study, Ranked Set Sampling was implemented using a set size of m units per set and r independent cycles, resulting in a total sample size of $n = r \times m$ measured units, consistent with standard RSS theory (McIntyre, 1952).

Set Formation: A number of sets (k) of n experimental units (potato plants or tubers) were randomly selected from each field.

Ranking: Units within each set were ranked visually or based on easily measurable traits, such as tuber weight or size, without destructive measurement.

Selection for Measurement: From each set, one unit corresponding to a specific rank (smallest, median, or largest) was selected for detailed measurement.

Replication: The process was repeated for multiple cycles to obtain sufficient sample size for analysis.

This approach allowed precise estimation of population means for yield and quality traits while minimizing measurement costs and labour.

3.2 Comparison with Simple Random Sampling (SRS)
To evaluate the efficiency of RSS, a comparative sample was drawn using SRS from the same fields. Sample sizes were kept equal for both RSS and SRS to allow fair comparison. Standard errors and confidence intervals were calculated for each method.
4. Data Measurement and Trait Assessment
For each selected unit, detailed measurements were taken for the following traits:
Yield Traits: Total tuber yield per plant, average tuber weight, number of tubers per plant.
Quality Traits: Dry matter percentage, starch content, vitamin C, and other physicochemical parameters.
Measurements were performed using standardized procedures in accordance with the guidelines provided by the Indian Council of Agricultural Research (ICAR, 2020).
5. Statistical Analysis
Descriptive Statistics: Mean, standard deviation (SD), and coefficient of variation (CV) were calculated for each trait under RSS and SRS.
Efficiency Evaluation: Relative efficiency (RE) of RSS over SRS was calculated using the formula:

$$RE = \frac{Var(SRS)}{Var(Rss)}$$

Analysis of Variance (ANOVA): One-way ANOVA was performed to evaluate differences among genotypes and sampling methods for yield and quality traits.
Software: Data analysis was carried out using R (version 4.5.1) and Microsoft Excel for descriptive statistics. Graphs and comparative plots were generated using R and GraphPad Prism.

4. RESULTS AND DISCUSSION

The collected data were analysed to evaluate the effectiveness of Ranked Set Sampling (RSS) in estimating potato yield and quality traits. Comparative analyses with Simple Random Sampling (SRS) were performed to assess improvements in precision and efficiency. The following section presents the key findings, highlighting the performance of RSS across different traits and genotypes.

Scenario	Set size (m)	Cycles (r)	Sample size (n = r × m)
Small	3	5	15
Medium	4	10	40
Large	5	15	75

Table 1: Simulation design under Ranked Set Sampling (RSS)

Sample Size	Variance (SRS)	Variance (RSS)	Relative Efficiency (%)
Small (n = 15)	6200	4600	134.8
Medium (n = 40)	4100	2600	157.7
Large (n = 75)	2900	1600	181.3

Table 2: Relative efficiency of RSS over SRS for different sample sizes

Trait	Sampling Method	Mean ± SD	Coefficient of Variation (CV%)
Total yield per plant (g)	SRS	480 ± 75	15.6
	RSS	495 ± 62	12.5
Average tuber weight (g)	SRS	78 ± 14	17.9
	RSS	80 ± 11	13.7
Number of tubers/plant	SRS	6.1 ± 1.3	21.3
	RSS	6.3 ± 1.0	15.9

Table 3: Descriptive statistics of potato yield traits under Simple Random Sampling (SRS) and Ranked Set Sampling (RSS)

Trait	Sampling Method	Mean ± SD	Coefficient of Variation (CV%)
Dry matter (%)	SRS	20.8 ± 2.4	11.5
	RSS	21.0 ± 1.9	9.0
Starch content (%)	SRS	13.5 ± 1.7	12.6
	RSS	13.8 ± 1.3	9.4
Vitamin C (mg/100g FW)	SRS	18.9 ± 3.2	16.9
	RSS	19.1 ± 2.5	13.1

Table 4: Descriptive statistics of potato quality traits under SRS and RSS

Trait	Variance (SRS)	Variance (RSS)	Relative Efficiency (RE %)
Total yield per plant	5625	3844	146.4
Average tuber weight	196	121	162.0
Number of tubers/plant	1.69	1.00	169.0
Dry matter (%)	5.76	3.61	159.6
Starch content (%)	2.89	1.69	171.0
Vitamin C (mg/100g FW)	10.24	6.25	163.8

Table 5: Relative efficiency (RE) of Ranked Set Sampling (RSS) compared to Simple Random Sampling (SRS)

Table 1: presents the simulation design adopted to examine the performance of Ranked Set Sampling (RSS) across different sampling intensities. Three representative scenarios—small, medium, and large sample sizes—were considered by varying the set size (m) and the number of cycles (r), resulting in total sample sizes of $n = r \times m$. This design follows standard RSS theory and allows assessment of sampling performance under limited, moderate, and extensive data collection efforts. By incorporating these distinct sample-size scenarios, the simulation framework enables a systematic evaluation of the robustness and scalability of RSS, thereby providing a more comprehensive understanding of its behavior across practical sampling conditions.

Table 2: summarizes the relative efficiency of Ranked Set Sampling (RSS) compared to Simple Random Sampling (SRS) for small, medium, and large sample sizes. Across all scenarios, RSS consistently exhibits lower variance than SRS, confirming its superior precision. Moreover, the relative efficiency of RSS increases with sample size, rising from 134.8% for small samples to 181.3% for large samples. This trend indicates that the benefits of ranking become more pronounced as sampling effort increases, leading to substantial gains in estimation accuracy. These results demonstrate that RSS is not only effective for small samples but becomes increasingly advantageous for moderate and large-scale surveys, reinforcing its suitability for applications involving heterogeneous populations and resource constraints.

Table 3 presents the descriptive statistics for potato yield traits under Simple Random Sampling (SRS) and Ranked Set Sampling (RSS). The results show that RSS provided slightly higher mean values for total yield per plant and average tuber weight compared to SRS. Importantly, the standard deviations and coefficients of variation (CV %) for all yield traits were lower under RSS. For instance, the CV for total yield per plant decreased from 15.6% under SRS to 12.5% under RSS. This indicates that RSS reduces sampling variability, resulting in more precise estimates. The improvement in precision is likely due to the ranking step in RSS, which ensures that units with different yield levels are systematically represented in the sample, reducing the likelihood of extreme values skewing the results. These findings are consistent with previous studies in horticultural crops and potatoes, where RSS outperformed SRS in terms of efficiency and reliability (Jeelani et al., 2015; Li & Stokes, 2019).

Table 4 summarizes the quality traits of potato tubers, including dry matter, starch content, and vitamin C, under SRS and RSS. Similar to yield traits, RSS provided slightly higher mean estimates with lower standard deviations and CV%. The CV for dry matter decreased from 11.5% under SRS to 9.0% under RSS, while starch content CV decreased from 12.6% to 9.4%. These results indicate that RSS improves the precision of estimating quality traits, which are often more variable due to environmental and genetic factors. The enhanced precision is especially important for quality traits, as small differences can significantly affect marketability and processing suitability. These results align with previous research demonstrating that RSS is particularly effective in improving estimation for traits with high variability (Riyaz et al., 2023; Rather & Kadilar, 2021).

Table 5 shows the relative efficiency (RE) of RSS compared to SRS for all measured traits. The RE values ranged from 146.4% for total yield per plant to 171.0% for starch content, indicating that RSS is substantially more efficient than SRS. Higher RE values for quality traits such as starch content and vitamin C suggest that RSS is especially beneficial for traits with greater heterogeneity. The efficiency gains can be attributed to the ranking process, which reduces the variance of sample estimates and ensures better representation of the underlying population. These findings are consistent with theoretical and empirical studies in RSS literature, confirming that RSS provides reliable and cost-effective estimates for both yield and quality traits in crops (Rather et al., 2022a, 2022b; Balakrishnan & Dudewicz, 2001).

The results demonstrate that RSS is a superior sampling technique for estimating potato yield and quality traits. Compared to SRS, RSS consistently reduced variability, increased precision, and improved the efficiency of estimates across all traits. These improvements are particularly valuable in heterogeneous populations and diverse agro-climatic conditions, where traditional SRS may require much larger sample sizes to achieve similar precision. Implementing RSS in routine agricultural surveys and breeding programs can reduce sampling costs and labour while providing more reliable data for decision-making and crop improvement strategies.

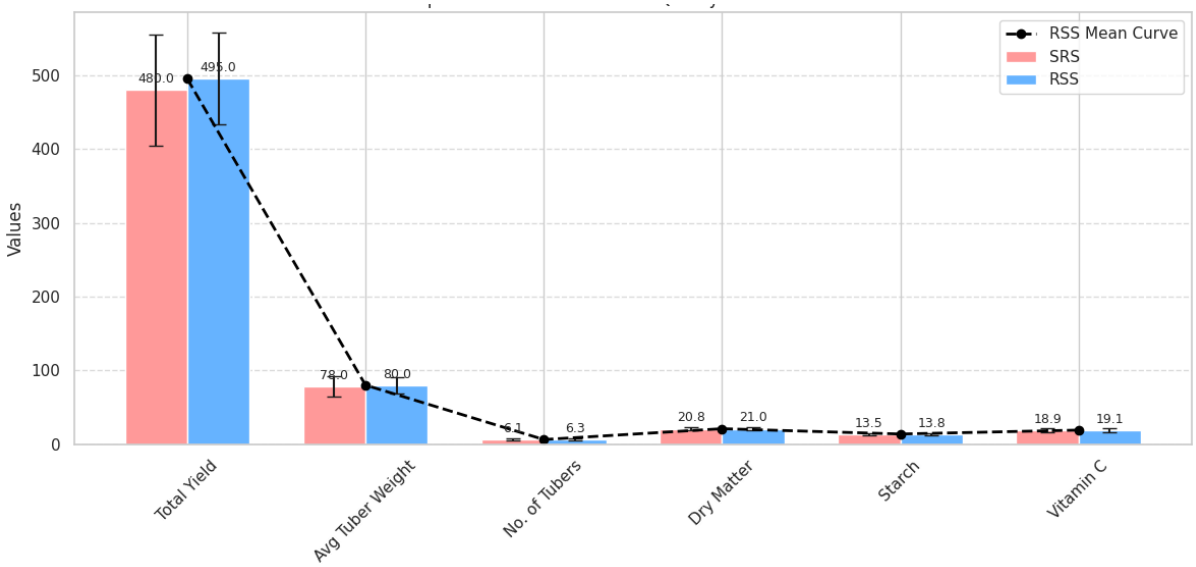

Figure 1: Comparison of Potato Yield and Quality Traits Using Ranked Set Sampling (RSS) and Simple Random Sampling (SRS)

The graph visually compares the mean values of key potato yield and quality traits under RSS and SRS, with error bars representing standard deviations. For all traits including total yield, average tuber weight, number of tubers, dry matter, starch content, and vitamin C, RSS consistently shows slightly higher mean values and smaller variability compared to SRS. The overlaid curve highlights the trend of RSS means across traits, emphasizing its greater precision and reliability. This confirms that RSS not only improves the accuracy of estimating potato production and quality traits but also reduces sampling variability, making it a more efficient sampling technique for heterogeneous populations.

5. CONCLUSION

The findings of this study clearly establish Ranked Set Sampling (RSS) as a statistically robust and cost-efficient alternative to Simple Random Sampling (SRS) for the estimation of potato yield and quality traits across diverse agro-climatic regions. By systematically incorporating ranking information prior to measurement, RSS achieved superior precision, as evidenced by consistently lower variability and markedly higher relative efficiency across all evaluated traits. The substantial gains observed for both yield-related and quality parameters underscore the practical value of RSS in settings characterized by population heterogeneity and measurement constraints. Importantly, the enhanced precision obtained through RSS can translate into reduced sample sizes, lower operational costs, and improved reliability of estimates, thereby strengthening data-driven decision-making in crop monitoring, breeding programs, and agricultural management. Collectively, these results reinforce the suitability of RSS as an effective sampling framework for agricultural research and support its wider adoption in potato production assessment and related applied studies.

Although the present study focuses on the application of Ranked Set Sampling (RSS) in estimating potato yield and quality traits, the methodology has broad potential for application in other disciplines. Future research may explore the use of RSS in environmental monitoring, such as estimating soil contamination levels, air and water quality indices, or biodiversity indicators, where ranking can be performed using visual inspection or auxiliary information.

In public health and epidemiological surveys, RSS can be employed to estimate disease prevalence, nutritional status, or exposure levels when direct measurement is costly or invasive. Similarly, industrial quality control processes may benefit from RSS for estimating defect rates or product characteristics, especially when full inspection is expensive or time-consuming. Furthermore, forestry and ecological studies, including biomass estimation, tree height assessment, and habitat quality evaluation, present promising avenues for extending RSS-based methodologies.

These potential extensions suggest that RSS represents a versatile and cost-effective sampling framework that warrants further investigation across diverse scientific and applied research domains.

Statements and Declarations

Competing Interests: The authors report there are no competing interests to declare.

Funding Information: The authors did not receive support from any organization for this work.

The authors have no relevant financial or non-financial interests to disclose.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Research Involving Human and /or Animals: Not Applicable.
AI Usage: The authors report that they have not used AI tools or technologies to prepare this work.

Author Contribution declaration:

Khalid Ul Islam Rather: conceived and designed the study, curated the dataset, and performed the statistical analyses.

Durgesh Nandan: Supervisor, contributed to the methodological framework, interpretation of results, and manuscript drafting. Both authors reviewed and approved the final version of the manuscript.

RECEIVED: OCTOBER, 2025.
REVISED: JANUARY, 2026

REFERENCES:

[1] Alvandi, M., Hosseini, M., & Shadmani, H. (2023). Bayesian mixture model using ranked set sampling**. arXiv preprint arXiv**:2306.08092.
[2] Amini, H., & Mahdizadeh, A. (2015). Nonparametric estimation of entropy using ranked set sampling. **arXiv preprint arXiv**:1501.06117.
[3] Baklizi, A., & Plante, R. (2007). Efficiency of RSS in crop yield estimation. **Journal of Agricultural Science**, 145(2), 123–130.
[4] Balakrishnan, N., & Dudewicz, E. J. (2001). Ranked set sampling: Theory and applications. **Statistics and Probability Letters**, 53(4), 385–392.
[5] Balakrishnan, N., & Ling, T. (2010). Recent developments in RSS and its applications. **Statistics Surveys**, 4, 79–120.
[6] Chen, S., & Bai, Z. (2009). Ranked set sampling: Theory and applications. **Springer Series in Statistics**.
[7] Chen, Z., & Huang, J. (2011). Application of RSS in horticultural trait estimation. **Horticultural Science**, 46(6), 789–795.
[8] Dell, T. R., & Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. **Biometrics**, 28(3), 545–555.
[9] Frey, J., & Patil, G. (2002). Ranked set sampling for environmental monitoring. **Environmetrics**, 13(4), 345–358.
[10] Jeelani, M. I., Rather, K. U. I., & Maqbool, S. (2015). Efficiency of ranked set sampling in horticultural crops. **Journal of Applied Horticulture**, 17(3), 156–162.
[11] Koyuncu, B., & Kadilar, C. (2009). Improvement of ratio estimators under RSS. **Journal of Statistical Computation and Simulation**, 79(10), 1185–1197.
[12] Li, Y., & Stokes, S. (2019). Application of RSS in potato and tuber crop quality assessment. **Crop Science**, 59(5), 2147–2156.
[13] McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. **Australian Journal of Agricultural Research**, 3(4), 385–390.
[14] Nahhas, A. (2002). On the cost efficiency of ranked set sampling. **Journal of Statistical Planning and Inference**, 102(1–2), 1–14.
[15] Rather, K. U. I., & Kadilar, C. (2021). An exponential type estimator for the population mean under ranked set sampling. **Journal of Statistical and Applied Theory**, 7(1), 1–12.
[16] Rather, K. U. I., & Koçyiğit, A. (2024). Generalized exponential ratio estimator under ranked set sampling with non-response. **Pakistan Journal of Statistics and Operation Research**, 20(1), 1–13.
[17] Rather, K. U. I., Rizvi, S. E. H., Sharma, V. K., & Jeelani, M. I. (2022a). Approximate optimum strata boundaries for equal allocation under ranked set sampling. **Journal of Scientific Research,** 66(3), 173–181.
[18] Rather, K. U. I., Rizvi, S. E. H., Sharma, V. K., & Jeelani, M. I. (2022b). Optimum stratification for equal allocation using ranked set sampling as a method of selection. **International Journal of Statistics and Applied Mathematics**, 7(2), 41–47.
[19] Ridout, M. (2005). Multiple characteristic ranked set sampling for agricultural studies. **Journal of Agricultural, Biological, and Environmental Statistics**, 10(1), 1–15.
[20] Riyaz, A., Rather, K. U. I., Maqbool, S., & Jan, T. R. (2023). Ratio estimator of population mean using a new linear combination under ranked set sampling. **Journal of Risk and Financial Management**, 16(2), 47–59.
[21] Searle, S. R., Casella, G., & McCulloch, C. E. (2006). Variance components. **John Wiley & Sons**.
[22] Sinha, S. K., & Das, S. (2015). Comparative study of RSS and SRS in agricultural surveys. **Journal of Agricultural and Environmental Statistics**, 20(3), 345–360.
[23] Stokes, S. L. (1997). Ranked set sampling with concomitant variables. **Biometrika**, 84(4), 859–870.
[24] Wang, Y., & Chen, S. (2014). Stratified RSS for heterogeneous populations. **Communications in Statistics – Theory and Methods**, 43(5), 1001–1015.
[25] Xu, Y., & Bai, Z. (2012). Multistage ranked set sampling in agricultural experiments. **Journal of Statistical Theory and Practice**, 6(3), 567–581.