NEW MODIFICATIONS OF RANKED SET SAMPLING FOR ESTIMATING SOME POPULATION PARAMETERS

MAHMOUD ZUHIER ABDULRAHEEM ALDRABSEH

UNIVERSITI SAINS MALAYSIA

2025

NEW MODIFICATIONS OF RANKED SET SAMPLING FOR ESTIMATING SOME POPULATION PARAMETERS

by

MAHMOUD ZUHIER ABDULRAHEEM ALDRABSEH

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

March 2025

ACKNOWLEDGEMENT

I would like to express my gratitude to Allah SWT for making my task easy and helping me complete this thesis. I would also like to thank my esteemed supervisor, Dr. Mohd Tahir Bin Ismail, and my co-supervisor, Dr. Amer Ibrahim Al-Omari, for their support, valuable assistance, and suggestions at every stage of the research and thesis writing. Additionally, I would like to express my gratitude to everyone I interacted with during my study period, including the faculty members, staff, and students at the School of Mathematical Sciences at Universiti Sains Malaysia. I also extend my thanks to the IPS members for their efforts in presenting this thesis and for assisting me during the submission stages. Furthermore, I am grateful to the examination committee members for reading and evaluating this work. Finally, I express my heartfelt appreciation to my parents and siblings for their support during my study period; without their kindness, I would not have reached this stage. I also cannot forget the support of my beloved wife for her companionship and dedication during this period of my life, along with our children. Furthermore, I am grateful to my friends and colleagues at work for their continuous support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

m	The sample size achieved in one cycle.
c	The number of cycles.
n	The sample size that achieved using c cycles, where $n = mc$.
μ	The population mean.
σ^2	The population variance.
θ	Any unknown parameter.
$\hat{\theta}$	The estimator of the unknown parameter θ.
X	The variable of interest.
$\bar{X}$	The classical estimator of the population mean.
$f(x)$	The probability density function of the variable *X*.
$F(x)$	The cumulative probability function of the variable *X*.
$E(X)$	The expectation of the random variable *X*.
$Var(X)$	The variance of the random variable *X*.
X_i	The i^{th} random variable.
$X_{j(i)}^k$	The i^{th} ordered variable in the j^{th} set for cycle k.
Y_i^k	The i^{th} double ranked set variable in cycle *k*.
Z_i^k	The i^{th} double except extreme ranked set variable in cycle *k*.
$f_{X_{(i)}}(x)$	The probability density function of the ordered variable $X_{(i)}$.
$F_{X_{(i)}}(x)$	The cumulative probability function of the ordered variable $X_{(i)}$.
$g_{Y_i^k}(x)$	The probability density function of the double ordered variable Y_i^k.
$G_{Y_i^k}(x)$	The cumulative probability function of the double ordered variable Y_i^k.
$h_{Z_i^k}(x)$	The probability density function of the double except extreme variable Z_i^k.
$H_{Z_i^k}(x)$	The cumulative probability function of the double except extreme variable Z_i^k.
$\mu_{(i)}$	The expectation of the i^{th} ordered variable.

$\sigma_{(i)}^{2}$	The variance of the i^{th} ordered variable.
$\mu_{(i)}^{(k)}$	The k^{th} moment a round zero of the i^{th} ordered variable.
$\mu_{(i)}^{*}$	The $mean$ of the i^{th} double ordered variable in a set of size *m*.
$\sigma_{(i)}^{2*}$	The variance of the i^{th} double ordered variable in a set of size *m*.
$\mu_{(i)}^{**}$	The $mean$ of the i^{th} double except extreme variable in a set of size *m+2*.
$\sigma_{(i)}^{2**}$	The variance of the i^{th} double except extreme variable in a set of size *m+2*.
$\hat{\mu}_{SRS}$	The estimator of the population mean using simple random sampling.
$\hat{\mu}_{RSS}$	The estimator of the population mean using ranked set sampling.
$\hat{\mu}_{MRSS}$	The estimator of the population mean using median ranked set sampling.
$\hat{\mu}_{MERSS}$	The estimator of the population mean using moving extreme ranked set sampling.
$\hat{\mu}_{EERSS}$	The estimator of the population mean using except extreme ranked set sampling.
$\hat{\mu}_{DEERSS}$	The estimator of the population mean using double except extreme ranked set sampling.
$\hat{\mu}_{DRSS}$	The estimator of the population mean using double ranked set sampling.
S_{SRS}^{2}	The estimator of the population variance using simple random sampling.
S_{RSS}^{2}	The estimator of the population variance using ranked set sampling
S_{EERSS}^{2}	The estimator of the population variance using except extreme ranked set sampling.
R	The strength stress reliability model $R = P(X > Y)$.
$\hat{R}_{SRS}$	The simple random sampling estimator of the *R* model.
$\hat{R}_{RSS}$	The ranked set sampling estimator of the *R* model.
$\hat{R}_{EERSS}$	The except extreme ranked set sampling estimator of the *R* model.

LIST OF ABBREVIATIONS

SRS	Simple Random Sampling
RSS	Ranked Set Sampling
MRSS	Median Ranked Set Sampling
MERSS	Moving Extreme Ranked Set Sampling
EERSS	Except Extreme Ranked Set Sampling
DRSS	Double Ranked Set Sampling
DEERSS	Double Except Extreme Ranked Set Sampling
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimation
MML	Modified Maximum Likelihood
MMLE	Modified Maximum Likelihood Estimation
PDF	Probability Density Function
CDF	Cumulative Distribution Function
MSE	Mean Square Error
Abias	Absolute Bias
RE	Relative Efficiency
PE	Percent Error
MAE	Mean Absolute Error
SSR	Strength Stress Reliability

LIST OF APPENDICES

PENGUBAHSUAIAN BAHARU PENSAMPELAN SET BERPANGKAT BAGI MENGANGGAR BEBERAPA PARAMETER POPULASI

ABSTRAK

Kaedah persampelan tradisional sering kekurangan kecekapan dan menanggung kos pengukuran yang tinggi, yang mencipta keperluan untuk pendekatan yang lebih cekap yang mengurangkan jumlah unit yang diukur sambil memastikan ketepatan. Persampelan set berpangkat (RSS) adalah teknik persampelan yang menjimatkan kos yang meningkatkan penganggaran parameter dengan menggabungkan persampelan rawak mudah (SRS) dengan penilaian pangkat set sampel sebelum pengukuran. Walaupun RSS dan variasinya telah menunjukkan jaminan dalam pengurangan kos, kesan pengecualian pangkat ekstrem terhadap kecekapan RSS masih belum diterokai. Kajian ini memperkenalkan reka bentuk persampelan baru, persampelan set berpangkat pengecualian ekstrem (EERSS), untuk meningkatkan kecekapan penganggar utama. Reka bentuk EERSS yang dicadangkan dinilai dan dibandingkan dengan kaedah persampelan yang ada untuk menganggarkan purata populasi, varians, dan kebolehjadian kekuatan-tekanan $R = P(X > Y)$, dengan menganggap taburan eksponen bebas untuk X (kekuatan) dan Y (tekanan), menggunakan penganggar kebolehjadian maksimum. Selain itu, reka bentuk berganda EERSS (DEERSS) dicadangkan untuk peningkatan kecekapan lebih lanjut. Penemuan menunjukkan prestasi yang lebih baik bagi EERSS berbanding dengan kaedah yang ada. Untuk penganggaran purata populasi, EERSS secara konsisten mengatasi RSS, persampelan set berpangkat median (MRSS), persampelan set berpangkat ekstrem bergerak (MERSS), dan SRS, mencapai kecekapan relatif (RE) sehingga 6.27 dalam taburan Student (4). Dalam penganggaran varians, EERSS menunjukkan kecekapan yang lebih tinggi daripada SRS dan, dalam kebanyakan kes, melebihi RSS,

terutamanya dengan saiz set yang kecil. Penganggaran kebolehjadian menggunakan EERSS memberikan kecekapan relatif yang lebih besar daripada SRS dan RSS, dengan nilai RE melebihi 3.7 dalam konfigurasi tertentu. Selain itu, reka bentuk DEERSS menunjukkan peningkatan yang luar biasa dalam kecekapan untuk estimasi purata, mencapai RE setinggi 23.88 untuk taburan Student (4), sambil juga mengatasi persampelan set berpangkat berganda (DRSS), RSS, dan EERSS dalam semua senario yang diuji. Analisis set data asal mengesahkan hasil ini, menunjukkan bahawa EERSS dan reka bentuk gandanya secara konsisten menghasilkan penganggar yang lebih tepat daripada kaedah pesaing dalam pelbagai aplikasi praktikal. Penemuan ini menekankan potensi EERSS dan DEERSS sebagai alat yang kukuh untuk penganggaran yang kos efektif dan tepat dalam kajian statistik dan kebolehjadian. Kajian akan datang boleh menggunakan reka bentuk yang dicadangkan untuk menganggarkan pelbagai ukuran populasi dan model kebolehpercayaan, termasuk penganggaran Bayesian dan penggunaan DEERSS untuk varians dan kebolehpercayaan.

NEW MODIFICATIONS OF RANKED SET SAMPLING FOR ESTIMATING SOME POPULATION PARAMETERS

ABSTRACT

Traditional sampling methods often lack efficiency and incur high measurement costs, creating a need for more efficient approaches that minimize the number of units measured while ensuring accuracy. Ranked set sampling (RSS) is a cost-effective sampling technique that enhances parameter estimation by combining simple random sampling (SRS) with the judgmental ranking of sample sets before measurement. Although RSS and its variations have shown promise in cost reduction, the impact of excluding extreme ranks on RSS efficiency remains unexplored. This study introduces a novel sampling design, except extreme ranked set sampling (EERSS), to improve the efficiency of key estimators. The proposed EERSS design is evaluated and compared with existing sampling methods for estimating the population mean, variance, and strength-stress reliability $R = P(X > Y)$, assuming independent exponential distributions for X (strength) and Y (stress), using maximum likelihood estimation. Additionally, a double design of EERSS (DEERSS) is proposed for further efficiency gains. The findings demonstrate the superior performance of EERSS compared to existing methods. For population mean estimation, EERSS consistently outperformed RSS, median ranked set sampling (MRSS), moving extreme ranked set sampling (MERSS), and SRS, achieving relative efficiencies (REs) of up to 6.27 in the Student (4) distribution. In variance estimation, EERSS showed higher efficiency than SRS and, in most cases, surpassed RSS, particularly with small set sizes. Reliability estimation using EERSS provided greater relative efficiency than SRS and RSS, with RE values exceeding 3.7 in certain configurations. Furthermore, the DEERSS design demonstrated remarkable improvements in efficiency for mean estimation, achieving REs as high as 23.88 for Student (4) distribution, while also surpassing double-ranked set sampling (DRSS), RSS, and EERSS in all tested scenarios. Analyses of real-life data sets validated these results, showing that EERSS and its double design consistently produce more precise estimators than competing methods across diverse practical applications. These findings underscore the potential of EERSS and DEERSS as robust tools for cost-effective and accurate estimation in statistical and reliability studies. Future studies could apply the proposed designs to estimate various population measures and reliability models, including Bayesian estimation and the use of DEERSS for variance and reliability.

CHAPTER 1

INTRODUCTION

1.1 Introduction

The importance of sampling is shown in its extensive application in various domains, where government organizations, social researchers, and economists are often interested in addressing questions related to critical issues such as unemployment, child labor, and the growth of small businesses. However, these problems cannot be effectively tackled through laboratory experiments or economic, mathematical, and statistical formulations. Instead, survey studies are essential for a comprehensive understanding of these problems (Arnab, 2017). Obtaining a sample from the population and measuring its units can sometimes be costly (Cochran, 1977). Therefore, it is important to acquire a sample that is as representative as possible while minimizing the associated expenses. Consequently, there is an ongoing need for improving sampling designs to enhance flexibility and efficiency.

The science of collecting, summarizing, and analysing data to make highly accurate inferences about a population is known as statistics. The population determines how the data is collected; if the population is small enough or easy to measure, we utilize the census (measure all parts of the population) to learn about the population's characteristics. However, suppose the population is huge or difficult to measure all elements. In that case, we can carefully select some elements (sample) from the population and measure them to make predictions about population features, inferences about the population, and get the estimators of population parameters. This process is called sampling. In this study, we aim to introduce new highly efficient sampling designs to reduce the number of sampling units to be measured and increase the performance of the current methods.

This chapter will introduce the study by first giving the foundational concepts and background of this study in Section 1.2. The motivations of the study are given in Section 1.3, followed by defining the problem under investigation in Section 1.4, the research main aim, objectives, and questions in Section 1.5, the significance of the study in Section 1.6, the scopes of the study and their limitations will be outlined in Section 1.7, and finally, the structure of this thesis will be detailed in Section 1.8.

1.2 Background of the Study

Sampling is a statistical method that involves selecting a subset of individuals, elements, or units from the population under study to draw statistical inferences about certain population characteristics. The group of all elements is called the population, whereas a subset of elements selected from the population is known as a sample. Sampling plays a critical role in statistical research, with applications spanning numerous fields such as agriculture, healthcare, and industrial quality control (Cochran, 1977). For example, estimating the average yield of a crop in agriculture often requires a balance between cost and accuracy. Traditional methods like simple random sampling (SRS) may be resource-intensive, necessitating the development of more efficient alternatives.

1.2.1 Probability and Non-Probability Sampling Methods

Sampling is broadly categorized into two main types: probability sampling and non-probability sampling. Probability sampling depends on the randomness of their designs, and every member of the population has an equal chance of being included in the sample, i.e. every element has a non-zero probability (Groves et al., 2010). Therefore, the samples produced by probability sampling are more likely to be representative of all members of the population if the sample is sufficient, and thus, it can be used in statistical inference. In contrast, non-probability sampling involves choosing elements from the population non-randomly, based on the researcher's judgment or purpose. In this category, not all elements in the population share an equal chance of being included in the sample, and some may have zero probability. This sample, in this case, might lack representativeness, leading to limitations in statistical inference. The most common probability and non-probability sampling methods, known by researchers in both categories, can be categorized as follows:

Table 1.1 Probability and Non-Probability Sampling Methods

Probability sampling	Non-probability sampling
Simple random sampling	Convenience sampling
Stratified sampling	Purposive sampling
Systematic sampling	Snowball sampling
Cluster sampling	Quota sampling

(Source, Berndt, 2020; Etikan, 2017).

1.2.1(a) Non-probability Sampling Methods

Non-probability sampling is classified into four major methods. The first method, convenience sampling, involves selecting a sample based on the ease of access for the researcher (Stratton, 2023; Kaptein & Heuvel, 2022). The second method, purposive sampling, entails selecting elements based on the researcher's judgment or specific criteria relevant to the study (Stratton, 2023; Kaptein & Heuvel, 2022). The third method, snowball sampling, also known as network sampling, starts with initial participants who then recruit others meeting specific criteria, leading to a 'snowball' effect as the sample size grows. Lastly, quota sampling involves selecting a specific group in predetermined proportions to ensure representation in the sample (Creswell, 2022; Babbie, 2021).

1.2.1(b) Probability Sampling Methods

The most basic and significant sampling method is simple random sampling (SRS). In this case, every sampling unit has an equal chance of being included in the sample. It is considered the main probability sampling method because all other probability methods depend on this method in their design (Groves et al., 2010). The second method is stratified random sampling. In this method, the population is divided into non-overlapping populations called strata, and then a sample is selected from each sub-population (stratum). The third method is systematic random sampling. In this method, the first element is randomly selected from the first k elements, and the next sampling unit is selected for every k element. Finally, cluster random sampling is a sampling method where the population is divided into groups called clusters, and then a group of the clusters is selected to census them. For more details, refer to (Babbie, 2021). In summary, the sampling method is determined based on many standards, such as ease of implementation (flexibility), representativeness, efficiency, low cost, and time-saving measures. For more details about the sampling concepts, refer to (Lohr, 2021).

1.2.2 Ranked Set Sampling Design

Ranked set sampling (RSS) is a method that depends on SRS in selection a group of random samples and on judgment for ordering them. This method was initially proposed by McIntyre (1952) to estimate the mean milk yield in pastures through agricultural sampling. By ensuring that all ranks are represented in the chosen samples, RSS is considered more representative of the population and thus exhibits higher efficiency.

Figure 1.1 presents a flowchart that outlines the step-by-step procedure for implementing RSS. The implementation of RSS involves the following stages:

Step 1. Select m random samples (sets) from the target population, where each sample contains m units. The value of m is a natural number that can be determined based on the requirement to facilitate ranking ease.

Step 2. Rank the units within each set based on the variable of interest using cost-effective methods such as visual assessment, researcher judgment, or an auxiliary variable, without any actual measurements in this step.

Step 3. Select the i^{th} ranked unit from the corresponding i^{th} set for actual measurement, where $i = 1,2, \dots, m$.

Step 4. Repeat Steps 1-3 for the specified number of cycles to obtain a larger sample of size $n = mc$, where m represents the number of measurements produced in each cycle and c denotes the total number of cycles.

The use of RSS improves the representativeness of the sample and enhances the efficiency of population parameter estimation compared to SRS, especially in situations where ranking can be performed accurately and measuring the variable of interest is costly or time-consuming (Dell & Clutter, 1972; Kaur et al., 1995). RSS has been widely applied in various fields, including environmental studies, reliability analysis, and quality control (Chen et al., 2003). Over time, numerous modifications to this method have been introduced, with some offering improved efficiency and others simplifying the ordering process. The approach and its modified designs are primarily used to make statistical inferences about population parameters applicable.

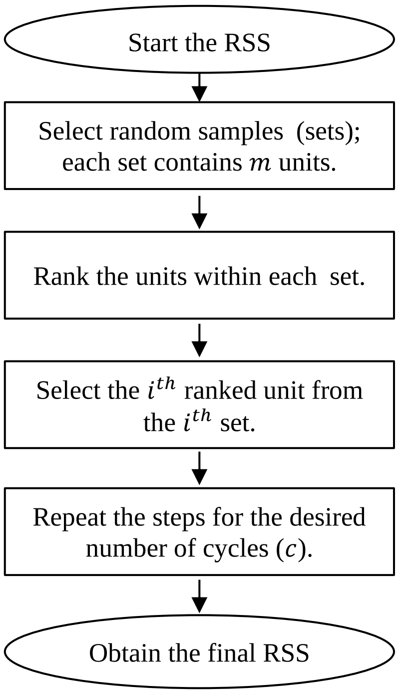

Figure 1.1 Ranked Set Sampling Process Flowchart

1.2.3 Statistical Inference

Statistical inference is the process of drawing conclusions and making predictions about a population using information obtained from a sample. It encompasses the statistical methods used to analyse sample data and make inferences and generalizations about population parameters, such as the mean, variance, proportion, rate, reliability, and correlation. Moreover, statistical inference involves calculations based on probability theory (Sinharay, 2010).

Statistical inference encompasses two primary areas: estimation and hypothesis testing (Lehmann & Casella, 1999). The statistical estimation includes making predictions or inferences about population parameters, which can be either a point estimate or an interval estimate. On the other hand, hypothesis testing is utilized to draw conclusions regarding a population using sample data and involves making a claim about the population's parameters. This process entails formulating both null and alternative hypotheses, computing the test statistics, specifying the level of significance, and finally determining the decision based on the significance level (Chan, 2021; DeGroot & Schervish, 2016). In addition, statistical inference can also be classified into two philosophies, as explained in Table 1.2.

Table 1.2 Differences Between the Statistical Inference Approaches

Bayesian Approach	Frequentist Approach
It uses the term of conditional probability.	Consider the relative frequency for the random variable.
The unknown parameters are considered random variables with a probability distribution. Thus, the beliefs about the parameters can be updated.	The unknown parameters are fixed.
It depends on the likelihood and prior densities.	It only depends on the likelihood.
It is a consistent approach to improve the ideas about the parameters given the evidence that already occurred.	The likelihood is calculated for all potential data sets.

The main difference between the two philosophies lies in their treatment of unknown parameters. In the traditional (frequentist) approach, the unknown parameter under investigation is treated as fixed, whereas in the Bayesian philosophy, the parameter is regarded as a random variable with a probability distribution. Additionally, Bayesian inference incorporates prior information about the parameters into the analysis. To obtain point estimates for the parameters of interest, the Bayesian method of moments (MM) and maximum likelihood (ML) methods can all be employed (Rossi, 2018). The MM and ML are frequentist estimation methods. ML is widely utilized in statistics due to its properties, including its consistency, efficiency, and asymptotic normality (Hastie et al., 2009; Casella & Berger, 2002).

1.3 Motivations for the Study

A significant challenge in statistics and machine learning revolves around drawing samples from arbitrary probability distributions. Sampling approaches are commonly employed during the training, testing, and prediction phases of probabilistic models (Maddison et al., 2014). Consequently, it is evident that a vast majority of studies reliant on data for achieving their objectives implement various sampling methods in accordance with their specific requirements. Furthermore, as noted by Cochran (1977), sampling techniques offer the potential to substantially reduce time, effort, and costs while maintaining high levels of accuracy. This is particularly beneficial for datasets associated with high measurement costs, such as those derived from studies involving large populations or extensive geographical areas.

In many research scenarios, reaching all members of a population is challenging or impractical due to time and cost constraints. It would be beneficial to identify an economical, high-efficiency sampling method (Bhushan et al., 2021). As a result, researchers often draw a representative sample from the population to derive estimators and generalize about its characteristics, known as parameters. RSS methods, as highlighted by Takahasi and Wakimoto (1968), offer enhanced efficiency by requiring fewer samples. Consequently, they contribute to significant time and cost savings in the sampling process.

The drawn sample is primarily utilized to make inferences about the factors and characteristics of the population, aiming for high-accuracy estimations. These estimations encompass aspects such as point estimation, confidence interval estimation, and hypothesis testing regarding these characteristics. Point estimation is a fundamental component of statistical inference, wherein the objective is to ascertain the precise value of population parameters based on information gleaned from the sample (Sahoo, 2013; Lehmann & Casella, 1999). Consequently, this underscores the necessity for ongoing advancements in sampling methods.

In summary, introducing new sampling designs can improve efficiency, reduce variance, and yield estimators of population parameters that are more accurate and stable than traditional methods. Moreover, these designs can be more robust and adaptable to various assumptions in data analysis. Additionally, introducing a new RSS design can help address the limitations and weaknesses of previous RSS sampling designs, including handling outliers, asymmetric distributions, and small sample sizes.

1.4 Problem Statement

Cost-effective sampling methodologies are essential in statistical research for accurately estimating population parameters, as higher efficiency leads to more accurate estimates and reduces the need for larger sample sizes. Traditional approaches, such as SRS, often require substantial sample sizes, making them resource-intensive in terms of cost and time (Cochran, 1977; Eichhorn, 2021; Latpate et al., 2021; Maria et al., 2023). Accordingly, the primary challenge lies in developing advanced sampling methods that minimize resource consumption while preserving the accuracy of estimations

RSS and its variations have been introduced as more cost-effective alternatives to traditional sampling methods, demonstrating the potential to reduce measurement costs (Dell & Clutter, 1972; Hanandeh & Al-Nasser, 2021; Leys et al., 2013; Mahdizadeh & Zamanzade, 2022b; Takahasi & Wakimoto, 1968). RSS is particularly advantageous in scenarios where ordering sample elements is straightforward, but obtaining precise measurements is resource-intensive (Hanandeh et al., 2022a, 2022b; Zamanzade & Mahdizadeh, 2017; Haq et al., 2014a). For example, in studies estimating the average height of trees in a forest, measuring the height of every tree is both time-consuming and expensive. In such cases, RSS offers a viable solution by reducing the number of measured units while maintaining accuracy. Additionally, multi-stage RSS designs (e.g., double-stage RSS) have demonstrated enhanced efficiency in a variety of contexts, further validating the method's utility (Al-Omari & Haq, 2019a; Al-Saleh & Al-Kadiri, 2000; Al-Saleh & Al-Omari, 2002; Samuh et al., 2021). Despite these advantages, there is still room for improving the accuracy and precision of estimators for population parameters, including the mean, variance, and reliability models, indicating a need to refine existing RSS methodologies further to address these limitations.

A notable gap in the existing literature is the lack of research investigating the impact of excluding extreme ranks on the efficiency of Ranked Set Sampling (RSS) for estimating population parameters. Previous studies, including those by Muttlak (1997), Al-Odat and Al-Saleh (2001), Al-Nasser and Al-Omari (2018), Hanandeh and Al-Nasser (2021), and Hanandeh et al. (2022b), have not addressed this aspect, leaving an opportunity for further exploration and methodological advancement.

Bridging this gap could offer critical improvements in sampling efficiency by reducing the influence of outliers, which would ultimately enhance the performance of estimators for population parameters. These improvements are particularly valuable for resource-constrained studies in fields such as agriculture, medicine, and environmental science.

To address these challenges, this study proposes a novel sampling design known as except extreme ranked set sampling (EERSS), which excludes the minimum and maximum ranks to evaluate its impact on the efficiency of RSS in estimating population parameters. Furthermore, a double-stage except extreme ranked set sampling (DEERSS) method is introduced to achieve enhanced efficiency. These approaches are expected to achieve higher efficiency in estimating the population parameters. Consequently, this could lead to improved accuracy in estimating population parameters or, alternatively, to a reduction in the number of elements required for measurement, thereby minimizing both the cost and time consumed in data collection.

1.5 Research Aim, Objectives, and Questions

The primary aim of this study is to introduce new efficient and cost-effective sampling designs for estimating population parameters to enhance the efficiency of the RSS method. This study's primary aim aligns with the design-based approach because it focuses on refining and introducing new sampling methods to enhance efficiency in parameter estimation. This overarching aim comprises four specific objectives outlined in subsection 1.5.1. These objectives will be achieved by answering the questions outlined in subsection 1.5.2.

1.5.1 Research Objectives

This main purpose can be implemented by achieving the following objectives:

1. To introduce a new sampling design that involves excluding the smallest and largest ranks, aimed at estimating the population mean parameter in both symmetrical and skewed distributions.
2. To discuss the estimation of population variance using the introduced sampling method in both symmetrical and asymmetrical cases.
3. To apply the introduced method to estimate the strength stress reliability system in an exponential distribution scenario.
4. To propose the double stage of the introduced sampling method for estimating the population mean.

1.5.2 Research Questions

The objectives outlined in the preceding subsection can be attained by addressing each of the corresponding research questions.

1. Is it possible to derive a new sampling design by excluding the minimum and maximum ranks from the ranked set of samples, and how accurate is this method for estimating the population mean?
2. Can the introduced method be used to estimate the population variance?
3. How can the new sampling method facilitate making inferences about the strength stress reliability system?
4. Can the introduced method be implemented in two stages, and what is the effect?

1.6 Significance of the Study

This study introduced two new sampling methods, characterized by their notably high efficiency in estimating the population mean. Additionally, novel estimators for both variance and strength-stress reliability are developed, demonstrating superior accuracy and consistency compared to commonly used estimators employed for comparison in this study. These methods can also be used to infer various population parameters such as the ratio, correlation, median, etc. In summary, these innovative sampling designs play a pivotal role in enhancing relative efficiency and accuracy.

This study explores some applications to real data in various fields, including agriculture, finance, environmental science, and engineering. These applications test the applicability and demonstrate the data analysis procedures using the modified sampling designs. The results show that these methods can be easily applied given the assumption of easy judgment ranking. Consequently, researchers in various domains seeking to utilize these methods can benefit from the guidance provided by this study.

The study directly addresses the issues of cost and time consumption by proposing highly efficient estimators. These methods demonstrate superior relative efficiency compared to existing estimators, as evidenced by the analytical and real data analysis results. Higher relative efficiency implies that fewer samples are required to achieve the same level of accuracy, thereby reducing the resources needed for data collection. This, in turn, minimizes both the financial and temporal costs associated with survey studies, particularly those conducted under resource constraints.

The proposed sampling methods not only enhance the theoretical understanding of ranked set sampling but also hold practical significance in reducing costs through improving efficiency. Applications include precision agriculture, where rapid and cost-effective estimation of crop yields is crucial, and environmental monitoring, where resource constraints often limit the scope of data collection.

1.7 Scopes and Limitations.

This study aims to introduce novel sampling designs for accurately estimating population parameters. Specifically, it focuses on estimating three key parameters: mean, variance, and the strength stress reliability model $P(X > Y)$, where X and Y denote the strength and stress components, respectively. For estimating mean and variance, the study covers both symmetric and asymmetric distributions commonly encountered in statistical analysis. Classical estimator formulas are utilized for estimating the mean and variance. On the other side, the study focuses on using ML estimation for estimating $P(X > Y)$ in the exponential case. The study utilizes a combination of simulation studies and real data applications to demonstrate the effectiveness of the proposed sampling designs and estimators.

The study is limited to estimating the mean and variance, given their importance in descriptive statistics (Wackerly et al., 2008). The study is limited to the estimation of reliability within the strength-stress reliability model for the exponential distribution, a model widely studied by researchers (Akgül & Şenoğlu, 2017; Hussian & Amin, 2017). Additionally, the estimation of $P(X > Y)$ can be effectively achieved using MLE, which is a widely accepted method for estimating parameters in such reliability models. For this reason, Bayesian inference and interval estimation for this reliability model are excluded from the estimation techniques, as the study primarily aims to improve the efficiency and accuracy of estimators for strength-stress reliability using the new RSS design. Including these techniques would fall outside the intended scope and potentially complicate the methodology without providing significant benefits for the specific research objectives. Due to time constraints, other parameters, such as the population ratio of the mean, percentiles, quartiles, proportions, correlations, and rates, are not addressed.

1.8 Structural Outlines

The subsequent chapters of the thesis are organized as follows: Chapter 2 comprehensively reviews previous studies relevant to the study's objectives, while Chapter 3 outlines the procedural steps for implementing the study and achieving its objectives. In Chapter 4, the novel sampling design is introduced, along with a discussion of the population mean estimator. Chapter 5 focuses on the estimation of variance, followed by an exploration of strength stress reliability estimation in Chapter 6. The two-stage design of our novel approach is elaborated in Chapter 7. Finally, Chapter 8 presents a summary of the study's key findings and provides recommendations for future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This study focuses on statistical sampling methods aimed at enhancing the efficiency of population parameter estimation, a critical task across various scientific and practical fields, including environmental studies, agriculture, and industrial quality control. Ranked Set Sampling (RSS) and its modifications have demonstrated significant potential in improving estimation efficiency, making them highly relevant for advancing statistical methodologies while minimizing cost and time.

In this chapter, we will examine key concepts related to RSS, including its advanced modifications and their applications in estimating the mean, variance, reliability models, and other parameters. The purpose of this literature review is to critically evaluate existing research on RSS and its modifications, identifying gaps in current methodologies. By synthesizing advancements and highlighting Issues under study, this review provides a foundation for the proposed research, which aims to introduce a novel RSS design that addresses these shortcomings.

The chapter is structured as follows: Section 2.2 presents an overview of modifications to RSS, including recent advancements and various design approaches. Section 2.3 discusses multi-stage modifications of RSS, demonstrating its potential to improve sampling efficiency. Section 2.4 focuses on methods for estimating population variance using RSS and its variations. Section 2.5 examines the use of RSS in estimating the stress-strength system, with a particular focus on its application in reliability analysis. Section 2.6 reviews diverse applications of RSS in estimation theory, particularly its use in parameter estimation across different domains. Finally, Section

2.7 summarizes the key findings from the literature and identifies research gaps and areas for further investigation.

2.2 Overview of Modifications to Ranked Set Sampling

In 1952, McIntyre introduced a new sampling method called ranked set sampling (RSS). He used this method in pasture sampling to estimate the mean production in pastures. It has also been proven to be more effective and accurate than simple random sampling (SRS) methods. If visual inspection is feasible and inexpensive, while actual measuring sampling units is costly, then this method would be advantageous to use. This method is implemented by selecting m^2 units from the population of interest and subsequently divide them into m sets, ranking the units within each set based on the variable under study by visual inspection or any costless method, measuring the i^{th} element within its corresponding i^{th} set, and then repeating this process c times to obtain a sample of size $n = cm$.

The RSS method has attracted the attention of many researchers for its important role in reducing cost; a large number of researchers have suggested many designs for the ranked set sampling method to improve their efficiency and suit various applications, including statistical inference and various fields such as forestry, management, ecology, environmental, sociology, and agriculture (Haq et al., 2016).

Muttlak (1996) introduced the pair RSS (PRSS) design for estimating the population means. This design can be implemented as follows: for an even set size, $(m/2)$ random samples (sets) are drawn, each of size m, from the target population. Then, the $(i)^{th}$ and $(m-i+1)^{th}$ units are chosen from the $(i)^{th}$ ranked set, for $i = 1,2,\dots,m/2$. In the case of an odd set size, the same steps are followed, but with $(m/2)$

replaced by $(m+1)/2$. In the same way, Samawi et al. (1996) introduced the extreme RSS (ERSS) method. The ERSS entails selecting m ranked set samples, each consisting of m units. In cases where m is even, the smallest elements are chosen from the first $m/2$ ranked sets, while the maximum is selected from the remaining $m/2$ sets. However, if m is odd, the smallest element is selected from the first $(m-1)/2$ ranked sets, the largest element from the second $(m-1)/2$ ranked sets, and the median from the remaining last set.

The median RSS (MRSS) design proposed by Muttlak (1997) as a modification of the RSS method reduces the number of considered ranks by selecting only the median ranks. The moving extreme ranked set sampling (MERSS) method is a sampling design proposed by Al-Odat and Al-Saleh (2001) as a modification of the RSS method, wherein the set size is changed from 1 to m. This method involves two iterations, one using the maximum and another using the minimum.

In later years, Muttlak (2003a) also introduced a new sampling design known as quartile RSS (QRSS). The QRSS is implemented differently depending on whether m is even or odd. If m is even, the method involves measuring the $[0.25(m+1)]^{th}$ element in the first $m/2$ ranked sets and the $[0.75(m+1)]^{th}$ element in the remaining $m/2$ ranked sets. On the other hand, if m is odd, measurements include the $[0.25(m+1)]^{th}$ element in the first $(m-1)/2$ ranked sets, the $[0.75(m+1)]^{th}$ element in the next $(m-1)/2$ ranked sets, and the $[0.5(m+1)]^{th}$ element (median) in the remaining set. Similarly, Muttlak (2003b) suggested the percentile RSS (PRSS). The steps of QRSS implement the PRSS but instead of measuring $[0.25(m+1)]^{th}$ and $[0.75(m+1)]^{th}$, measure the $[P(m+1)]^{th}$, and $[(1-P)(m+1)]^{th}$respectively, where $0 < P < 1$.

Chen and Shen (2003) introduced a new sampling approach called two-layer RSS (TRSS), which utilizes two concomitant variables. This method is employed for estimating the population mean and the coefficients of regression models for the concomitant variables. The study assessed performance using MSE. Later, Gulay and Demirel (2019) described a new sampling approach called two-layer MRSS (TMRSS), which combines MRSS and TRSS using two concomitant variables. The estimator of the population mean is compared with the MRSS estimator. Performance comparison is based on RE, MSE, and bias of the regression coefficients and mean estimate.

Al-Nasser (2007) proposed a new robust RSS method called L-ranked set sampling (LRSS). The LRSS involves defining a coefficient $\beta = [\alpha m], 0 \leq \alpha < 0.5$, and then selecting the $(\beta + 1)^{th}$ element in the first $\beta + 1$ ranked sets and the $(m - \beta)^{th}$ element in the last $\beta + 1$ ranked sets, with the i^{th} element selected for the sets from $k + 2, \cdots, n - \beta + 1.$ Beside this, Al-Nasser and Mustafa (2009) proposed the robust ERSS (RERSS) for estimating the population mean. The method involves defining a coefficient $\beta = [\alpha m]$, where $0 < \alpha < 0.5$. In the case of an even m, the $(\beta + 1)^{th}$ element is selected from the first $m/2$ sets and the $(m - \beta)^{th}$ element from the last $m/2$ sets. If m is odd, the cycle is completed by selecting the median rank from the remaining set. They suggested two estimators for estimating the mean parameter.

Similar to the LRSS method, Al-Omari and Almanjahie (2021) introduced two new RSS (RSS) methods: simple Z RSS (SZRSS) and generalized Z RSS (GZRSS). In SZRSS, the $(i + 1)^{th}$ element is selected from the first $m/2$ ranked sets, the $(i - 1)^{th}$ element from the last $m/2$ ranked sets, and the median from the remaining sets if the set size is odd. In GZRSS, a coefficient $\beta = [\alpha m], 0 \leq \alpha < 0.5,$ is defined. The $(i + 1)^{th}$ element is then chosen from the first β ranked sets, the $(i - 1)^{th}$ element from the last β ranked sets, and the i^{th} element from the remaining sets. These two methods were utilized for estimating both the population mean and median.

Al-Omari and Raqab (2013) suggested the truncated-based RSS (TBRSS) method for estimating mean and median parameters. This method involves defining the coefficient $\beta = [\alpha m]$, where $0 \leq \alpha < 0.5$. The smallest elements are selected from the first β ranked set, the largest elements from the last β ranked sets, and the i^{th} element from the remaining ranked sets. Additionally, Haq et al. (2013) described the partial RSS approach as a method for estimating the population mean, median, and variance by combining SRS and RSS techniques. The partial RSS methodology consists of the following steps: Firstly, a coefficient $\beta = [\alpha m]$ is defined, where α is a real number such that $0 \leq \alpha < 0.5$, and β is the largest integer less than αm. Secondly, 2β random samples are selected, each containing a single unit. In the third step, $m - 2\beta$ random samples of size m are chosen, and after ranking the units into each sample, the $(i + \beta)^{th}$ unit is selected from $(i)^{th}$ ranked set for $i = 1,2, \dots, m - 2\beta$. Similarly, Maria et al. (2023) proposed a method known as partial stratified RSS, wherein the selection of $m - 2\beta$ elements are conducted within each stratum (h).

Zamanzade and Al-Omari (2016) proposed a novel approach called neoteric RSS (NRSS) for estimating population mean and variance. The method involves selecting a sample (set) of size m^2 from the target population. If m is odd, the $((m + 1)/2 + (i - 1)m)^{th}$ ranked element is measured for $i = 1,2, \dots m$. However, measurements are taken for the $(m/2 + (i - 1)m)^{th}$ ranked elements when i is even, and $((m + 2)/2 + (i - 1)m)^{th}$ ranked elements when i is odd, for $i = 1,2, \dots m$. Also, Tayyab et al. (2019) suggested a novel RSS strategy known as even order RSS

(EORSS), incorporating an auxiliary variable for ratio estimation. They compared this approach with SRS, RSS, MRSS, and QRSS.

Biradar and Santosha (2017) introduced a novel sampling method focusing only on maximum ranks and varying set sizes, termed maximum RSS with unequal samples (MaxRSSU). This design is already included within the MERSS design but without a minimum cycle. The study also discusses the maximum likelihood estimation (MLE) and modified ML estimation (MMLE), Fisher information number, and the error in ranking for the exponential parameter. Its modification was introduced by Basikhasteh et al. (2020).

Al-Nasser and Al-Omari (2018) introduced the minimax RSS (MMRSS) approach as a modification of the MERSS method. This approach involves drawing m random samples (sets) from 1 to m. After ranking the elements within each set, the minimum is measured for all odd-numbered sets, and the maximum is measured for all even-numbered sets. Additionally, Hanandeh and Al-Nasser (2021) suggested a modification to the MMRSS (MMMRSS) when m is even. The MMMRSS involves drawing $m + 1$ random samples (sets) of sizes $1,2, \ldots, m$. After ranking the elements within each set, the $(m/2)^{th}$ element is measured in the last two sets, the minimum is measured for all remaining odd-numbered sets, and the maximum is measured for all remaining even-numbered sets. Haq et al. (2014) considered the hybrid RSS (HRSS), and Robertson et al. (2021) proposed the quasi-random RSS (QRRSS). More designs are described by Cui (2018), Sevinç et al. (2018), Noor et al. (2018), and Gulay and Demirel (2019) .

In recent advancements, Kumari et al. (2024) introduced Neutrosophic MRSS (NMRSS) as an extension of the traditional MRSS for estimating population means in bivariate data scenarios, particularly those involving interval-type or neutrosophic data. The performance of this design and its estimators was assessed using metrics such as bias, mean square error (MSE), and percentage relative efficiency (PRE). The results highlighted the enhanced efficiency and accuracy of the proposed estimators. Similarly, Bhoj and Chandra (2024) proposed the weighted ranked set sampling (WRSS) method for estimating population means in skewed distributions. Their findings demonstrated that WRSS consistently achieves higher relative precision (RP) than the traditional RSS, with notable improvements observed in positively skewed distributions. Lastly, Peanpailoon et al. (2024) introduced and evaluated the stratified folded ranked set sampling (SFRSS) method for estimating population means, demonstrating its superior efficiency and adaptability across various probability distributions, including Exponential, Geometric, Gamma, Weibull, Log-Normal, and Chi-Square. The findings highlight SFRSS as a robust and competitive sampling strategy, particularly in less complex scenarios, with significant potential to optimize estimation accuracy in diverse statistical applications. A summary of the existing RSS designs is presented in Table 2.1.

Despite the development of various RSS designs that incorporate extreme ranks (e.g., ERSS, MERSS, MMRSS, and MMMRSS) or percentile ranks (e.g., MRSS, QRSS, and PRSS), the effect of excluding extreme ranks on the efficiency of RSS remains insufficiently explored in the literature. To address this gap, this study proposes a novel design that excludes extreme ranks while utilizing all remaining ranks, with the objective of enhancing the efficiency of population mean estimation. This approach has the potential to achieve greater efficiency in estimating population parameters, leading to either improved accuracy or a reduction in the sample size required for measurement, thereby minimizing both the cost and time associated with data collection.

Table 2.1 Summary of Modifications of Ranked Set Sampling

Author(s)	Sampling design	Abbreviation
Muttlak (1996)	pair-ranked set sampling	Pair RSS
Samawi et al. (1996)	extreme ranked set sampling	ERSS
Muttlak (1997)	Median ranked set sampling	MRSS
Al-Odat and Al-saleh (2001)	moving extreme ranked set sampling	MERSS
Muttlak (2003a)	quartile-ranked set sampling	QRSS
Muttlak (2003b)	percentile ranked set sampling	PRSS
Chen and Shen (2003)	two-layer ranked set sampling	TRSS
Gulay and Demirel (2019)	two-layer median ranked set sampling	TMRSS
Al-Nasser (2007)	L-ranked set sampling	LRSS
Al-nasser and Mustafa (2009)	robust extreme ranked set sampling	RERSS
Al-Omari and Raqab (2013)	truncated-based ranked set sampling	TBRSS
Haq et al. (2013)	partial ranked set sampling	Partial RSS
Haq et al. (2014)	hybrid ranked set sampling	HRSS
Zamanzade and Al-Omari (2016)	neoteric ranked set sampling	NRSS
Biradar and Santosha (2017)	maximum ranked set sampling with unequal samples	MaxRSSU
Al-Nasser and Al-Omari (2018)	minimax ranked set sampling	MMRSS
Tayyab et al. (2019)	even order ranked set sampling	EORSS
Gulay and Demirel (2019)	Two-Layer Median Ranked Set Sampling	TLMRSS
Hanandeh and Al-Nasser (2021)	modified minimax ranked set sampling	MMMRSS
Al-Omari and Almanjahie (2021)	simple Z and generalized simple Z-ranked set sampling	SZRSS, GSZRSS
Robertson et al. (2021a)	quasi-random RSS	Q-RRSS
Maria et al. (2023)	partial stratified ranked set sampling	PStrRSS
Kumari et al. (2024)	Neutrosophic median ranked set sampling	NMRSS
Bhoj and Chandra (2024)	weighted ranked set sampling	WRSS
Peanpailoon et al. (2024)	Stratified folded ranked set sampling	SFRSS

2.3 Overview of Ranked Set Sampling in Multiple Stages

Al-Saleh and Al-Kadiri (2000) suggested implementing RSS in two stages as the initial step in multistage designs to enhance efficiency, reduce sampling costs, and provide greater flexibility in estimating the population mean. This design is called double RSS (DRSS). The study demonstrated that ranking in the second stage is easier than in the first stage. Subsequently, Al-Saleh and Al-Omari (2002) introduced multistage RSS (MSRSS), which involves generating samples by applying RSS across multiple stages. The MSRSS includes drawing m^r samples (sets), each of size *m*, and ranking the elements within each set. Then, apply the RSS to the first *m* sets and the second *m* sets up to the last. The process continues through *r* stages; the actual measurements are applied only in the last stage, while in stages 1 to $r-1$, insight or judgment ranking is utilized.

Similarly, Samawi and Tawalbeh (2002) considered the double stage of MRSS, Jemain, and Al-Omari (2006a, 2006b) described the double-stage designs for both QRSS and PRSS, double-pair RSS (DPRSS) by Haq et al. (2016), and LRSS by Al-Omari and Haq (2019). Likewise, multistage designs were outlined for MRSS, ERSS, QRSS, and PRSS in Jemain et al. (2007a, 2007b) and Jemain and Al-Omari (2007a, 2007b).

In recent years, several authors have proposed combining different RSS designs to achieve greater flexibility. Al-Mawan et al. (2018) introduced median double-RSS (MDRSS) by implementing RSS in the first stage and MRSS in the second stage. Similarly, Samuh et al. (2021) suggested the mixed DRSS, while Hanandeh et al. (2022a) described RSS-ERSS and MRSS-ERSS, which combine RSS and MRSS with extreme RSS. Additionally, Hanandeh et al. (2022b) proposed RSS-MMRSS and

MRSS-MMRSS by integrating minimax ranked set sampling with RSS and MRSS, respectively.

Shehzad et al. (2024) proposed a novel double-stage ranked set sampling modification that combines MRSS in the first stage and QRSS in the second stage, referred to as the modified median-quartile DRSS (MMQDRSS). To evaluate the performance of this new design, the RE of the estimator for the population mean was compared to that of other estimators, including DRSS, extreme DRSS, median DRSS, and quartile DRSS. Simulation results from various distributions and real data examples demonstrated improved efficiency for the proposed estimators in most cases. Also, Zubair et al. (2024) suggested the double extreme-cum-MRSS (DEMRSS) by combining the DRSS and EMRSS designs to provide a more representative sample and improve the efficiency of the population mean estimation, making it a competitive alternative to existing sampling strategies. The DEMRSS framework involves selecting an initial sample of size $(2m)^3$ from a population of size N, which is then randomly divided into $2m$ ranked sets, each containing $(2m)^2$ units. The RSS technique is applied to each set to generate $2m$ ranked sets and the EMRSS technique is used on these sets to produce a DEMRSS sample of size $2m$. The process is repeated c times to obtain a sample of size $n = 2cm$. This methodology is particularly effective in situations where the population is heterogeneous and includes outliers.

Additionally, Peanpailoon (2024) introduced stratified double unified ranked set sampling with perfect ranking (SDURSS-PR) as a novel sampling design for accurate population mean estimation. By comparing its performance against stratified simple random sampling (SrSRS), stratified ranked set sampling (SrRSS), and stratified median ranked set sampling (SrMRSS) under standard normal, Student’s t, and uniform distributions, the study showed that SDURSS-PR provides unbiased estimators with