STRATIFIED FOLDED RANKED SET SAMPLING FOR ASYMMETRIC DISTRIBUTIONS

Chainarong Peanpailoon[1], Noppakun Thongmual[2,*], and Nipa Nasinprom[3]

Received: February 14, 2024; Revised: April 09, 2024; Accepted: June 20, 2024

Abstract

This study evaluates the comparative efficiency of Stratified Folded ranked set sampling (SFRSS) in estimating the population mean against three other sampling methods: Simple Random Sampling (SRS), Stratified Simple Random Sampling (SSRS), and Random Sampling (RSS). The analysis spans various probability distributions, Exp(1), Geo(0.5), Gamma, Weibull, Log N(0,1), CHI(1). The results show that SFRSS outperforms or remains competitive with the other sampling methods, especially in simpler conditions (r=2), across a wide range of distributions. For instance, in the Exponential and Geometric distributions, SFRSS demonstrates superior efficiency at r=2, indicating its effectiveness in less complex scenarios. Similarly, for the Gamma and Weibull distributions, SFRSS maintains high efficiency at r=2, showcasing its adaptability and robustness in various conditions. Notably, in the Log-Normal distribution at r=5, SFRSS's efficiency dramatically improves against SSRS, suggesting its particular suitability for skewed data under certain conditions. Additionally, SFRSS exhibits exceptional efficiency in the Chi-Square distribution at r=2, underscoring its effectiveness across diverse statistical landscapes. However, at the more complex condition (r=5), although the efficiency of SFRSS generally decreases, it still competes closely with SRSS and outperforms SSRS in certain cases, retaining its relevance and demonstrating a degree of adaptability to complex scenarios. This comprehensive analysis highlights the importance of selecting appropriate sampling strategies based on the study's distribution characteristics and conditions. SFRSS emerges as a highly efficient method for estimating population means, particularly in less complex scenarios, and remains a competitive choice in more demanding conditions. The findings underscore the potential of SFRSS in optimizing estimation accuracy across various statistical distributions and study designs, affirming its value in statistical research and applications.

Keywords: Folded Ranked Set Sampling; Ranked Set Sampling; Simple Random Sampling; Stratified Folded Ranked Set Sampling

---

[1] *Department of Organization Curriculum and Teaching Program (Mathematics). Faculty Education Sakon Nakhon Rajabhat University, Sakon Nakhon 47000, Thailand. E-mail: tankku555@gmail.com*
[2] *Department of Sciences and Mathematics, Faculty of Technology and Health of Science, Kalasin University, Kalasin, Thailand. E-mail: nop_stat@hotmail.com*[*]
[3] *Department of Animal Production, Faculty of Agricultural Technology, Kalasin University, Kalasin, Thailand. E-mail: nipa.na@ksu.ac.th*
[*] *Corresponding author*

DOI: https://doi.org/10.55766/sujst-2024-04-e03604

Introduction

Simple Random Sampling (SRS) is a foundational statistical method used to select a subset of individuals from a larger population in such a way that every individual has an equal chance of being chosen. This method is pivotal in ensuring the generalizability of the study findings to the broader population, thus playing a critical role in both theoretical and applied research across various disciplines. SRS is characterized by the principle of equal probability, where each member of the population has an identical chance of being included in the sample (Levy and Lemeshow, 2008). This is achieved through a process that does not allow any form of bias to influence which members are selected. The process ensures that the sample drawn is representative of the population, thereby enabling researchers to make inferences about the population based on the sample. The significance of SRS lies in its capacity to produce unbiased estimates of population parameters (Cochran, 1977). By ensuring that every individual has an equal chance of selection, SRS minimizes the risk of sample bias, enhancing the reliability and validity of the research findings. This methodological rigor is essential for advancing knowledge and informing practice in fields such as social sciences, medicine, and business. Despite its advantages, SRS has limitations. A significant challenge is the need for a complete and accurate listing of the population, known as the sampling frame. In the absence of a comprehensive sampling frame, the risk of sampling bias increases, potentially compromising the representativeness of the sample (Thompson, 2012). Additionally, SRS may not be the most cost-effective or practical approach for studies involving large or geographically dispersed populations due to the logistics and resources required to randomly select and reach out to potential participants. An advancement derived from SRS is the method known as Range Set Sampling (RSS). RSS addresses limitations encountered in SRS by incorporating auxiliary information to pre-select and rank units before actual measurements are taken. This method significantly improves the representativeness of the sample, thereby increasing the precision of estimates for population parameters (McIntyre, 1952). The methodology of RSS involves selecting sets of units from the population, ranking these units based on an easily obtainable characteristic correlated with the variable of interest, and then measuring a specific unit from each set. This process enhances the efficiency of the sampling technique (Dell and Clutter, 1972). While RSS presents numerous advantages, its implementation requires careful consideration, especially regarding the accurate ranking of units and the correlation between the auxiliary and main variables of interest (Stokes and Sager, 1988). RSS has found utility in environmental studies, agriculture, and public health, among other areas, proving particularly beneficial in scenarios where direct measurements of each unit are impractical or prohibitively expensive (Muttlak and McDonald, 1990). RSS remains a vital tool in statistical sampling, offering enhanced accuracy and efficiency for estimating population parameters. Its adaptability to various research contexts underscores its continued relevance in statistical methodology (Wolfe, 2004). Over time, various authors have further refined and adapted RSS to estimate population parameters. Takahasi and Wakimoto (1968) provided a mathematical proof, establishing that the sample mean derived from RSS serves as an unbiased estimator of the population mean, exhibiting smaller variance compared to the sample mean derived from a SRS with an equivalent sample size. Building on this foundation, Dell and Clutter (1972) illustrated the variance of the sample mean derived from RSS is either less than or equal to that of an SRS, regardless of whether errors in ranking are present. Subsequent developments in RSS approachology include the introduction of Stratified Ranked Set Sampling (SRSS) by Samawi (1996), which further enhanced the efficiency of the sampling approach. Folded Ranked Set Sampling (FRSS) is based on the principle of ranking units in the population according to some measure of size related to the study variable and then selecting samples in a manner that ensures a higher selection probability for units with higher ranks. Unlike traditional RSS, FRSS involves a folding process over the ranked units, which aims to improve representativeness and reduce bias by ensuring that units from both ends of the distribution have a higher likelihood of being sampled (McIntyre, 2005). The primary advantage of FRSS lies in its ability to provide more accurate estimates for skewed distributions than simple random sampling or even stratified sampling methods. This is particularly beneficial when the mean, total, or other aggregate measures of the population are significantly influenced by outliers or extreme values (Stokes, 1980). Moreover, FRSS can lead to efficiency gains by reducing the variance of the estimator, making it a cost-effective option for many practical applications. Mustafa *et al.* (2011) introduced a FRSS tailored for asymmetric distributions, expanding the applicability of RSS to diverse statistical scenarios. Continuing the evolution of RSS, Matthews and Wolfe (2017) proposed a FRSS, presenting an additional variant of the approach. One notable feature of the RSS approach is its efficiency, particularly as the number of sets and cycles increases. This characteristic underscores the versatility and robustness of RSS in accurately estimating population parameters, making it a valuable tool in statistical sampling approachologies.

This study is undertaken with the primary goals of introducing the Stratified Folded Ranked Set Sampling (SFRSS) approach and evaluating its effectiveness in estimating the population mean in the context of asymmetric distributions. Additionally, a key focus is placed on scrutinizing the efficiency of the empirical mean estimator derived from the application of SFRSS.

Materials and Approaches

Simple Random Sampling (SRS)

SRS is an approach used to select n units from a N population, ensuring that each of the ${}_N C_n$ distinct samples, comprised of a subset of units from the total population.

Ranked Set Sampling (RSS)

The RSS technique can be outlined as follows:

Step 1: Utilize a SRS approach to select m^2 units from the population of interest.

Step 2: Randomly allocate the m^2 selected units into sets, each consisting of a predefined size.

Step 3: Rank the m units within each set based on the variable of interest.

Step 4: Construct a sample by sequentially selecting the smallest ranked unit from the first set, followed by the second smallest ranked unit from the second set, and continue this process until the largest ranked unit is chosen from the last set. Subsequently, measure the variable of interest for each unit in the sample.

Step 5: Repeat steps 1 through 4 for multiple cycles to draw the Resampling with RSS sample of the desired size $n = mr$. This approach enables the creation of an RSS sample, facilitating robust statistical analyses and inference through resampling from the original population data (Al- Omari and Bouza, 2014).

Folded Ranked Set Sampling (FRSS)

The FRSS technique can be outlined as follows:

Step 1: Draw a simple random sample of size m^2 units at random from the target population. Then, divide the m^2 units into m sets.

Step 2: Select $\frac{m+1}{2}$ samples for an odd number, and $\frac{m}{2}$ samples for an even number.

Step 3: Rank the units within each sample according to a variable of interest through visual inspection.

Step 4: Choose the 1^{st} and m^{th} units from the first sample for actual measurement.

Step 5: Select the 2^{nd} and $(m-1)^{th}$ units from the second sample for actual measurement.

Step 6: Continue this process until the m^{th} unit have been selected from the sample.

Step 7: Repeat steps 1 through 5 for r cycles to draw the FRSS sample of the desired size $n = mr$. This approach facilitates systematic and efficient sampling by systematically selecting units based on their ranks within each sample, ensuring representative samples for robust statistical analysis (Bani-Mustafa *et al.*, 2011).

Example 1 Consider the case of $(m = 6, r = 1)$. Draw a simple random sample of size $m^2 = 6^2 = 36$ units as:

$$\begin{bmatrix} X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16} \\ X_{21}, X_{22}, X_{23}, X_{24}, X_{25}, X_{26} \\ X_{31}, X_{32}, X_{33}, X_{34}, X_{35}, X_{36} \\ X_{41}, X_{42}, X_{43}, X_{44}, X_{45}, X_{46} \\ X_{51}, X_{52}, X_{53}, X_{54}, X_{55}, X_{56} \\ X_{61}, X_{62}, X_{63}, X_{64}, X_{65}, X_{66} \end{bmatrix}$$

Select $\frac{6}{2} = 3$ samples.

$$\begin{bmatrix} X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16} \\ X_{21}, X_{22}, X_{23}, X_{24}, X_{25}, X_{26} \\ X_{31}, X_{32}, X_{33}, X_{34}, X_{35}, X_{36} \end{bmatrix}$$

We select samples each of size

$$\begin{bmatrix} \boxed{X_{11}}, X_{12}, X_{13}, X_{14}, X_{15}, \boxed{X_{16}} \\ X_{21}, \boxed{X_{22}}, X_{23}, X_{24}, \boxed{X_{25}}, X_{26} \\ X_{31}, X_{32}, \boxed{X_{33}}, \boxed{X_{34}}, X_{35}, X_{36} \end{bmatrix}$$

Let X_{11}, X_{16}, X_{22}, X_{25}, X_{33}, X_{34} is FRSS of size 6

Example 2 Consider the case of $(m=5, r=1)$. Draw a simple random sample of size $m^2 = 5^2 = 25$ units as;

$$\begin{bmatrix} X_{11}, X_{12}, X_{13}, X_{14}, X_{15} \\ X_{21}, X_{22}, X_{23}, X_{24}, X_{25} \\ X_{31}, X_{32}, X_{33}, X_{34}, X_{35} \\ X_{41}, X_{42}, X_{43}, X_{44}, X_{45} \\ X_{51}, X_{52}, X_{53}, X_{54}, X_{55} \end{bmatrix}$$

Select $\frac{5+1}{2} = 3$ samples.

$$\begin{bmatrix} X_{11}, X_{12}, X_{13}, X_{14}, X_{15} \\ X_{21}, X_{22}, X_{23}, X_{24}, X_{25} \\ X_{31}, X_{32}, X_{33}, X_{34}, X_{35} \end{bmatrix}$$

We select samples each of size

$$\begin{bmatrix} \boxed{X_{11}}, X_{12}, X_{13}, X_{14}, \boxed{X_{15}} \\ X_{21}, \boxed{X_{22}}, X_{23}, \boxed{X_{24}}, X_{25} \\ X_{31}, X_{32}, \boxed{X_{33}}, X_{34}, X_{35} \end{bmatrix}$$

Let $X_{11}, X_{15}, X_{22}, X_{24}, X_{33}$ is FRSS of size 5

The purpose of this research is to suggest the modified RSS, namely the stratified Folded ranked set sampling (SFRSS) with imperfect ranking to estimate the population mean. This study also illustrates the efficiency of the mean estimator based on SFRSS via a simulation under Asymmetric distributions.

Stratified Folded Ranked Set Sampling (SFRSS)

The population consisting of N units are partitioned into L non-overlapping sub-groups known as strata, with each stratum containing $N_1, N_2, \cdots, N_L$ units, respectively, such that $N_1 + N_2 + \cdots + N_L = N$. The size of the h^{th} stratum is denoted by N_h for $h = 1, 2, \cdots, L$. Subsequently, independent samples are drawn from each stratum, resulting in sample sizes denoted by $n_1, n_2, ..., n_L$, such that the total sample size is $n = \sum_{h=1}^{L} n_h$. If the FRSS technique is applied within each stratum, the entire procedure is referred to as SFRSS. Let $X_{[i]j}^{h}$ represent the i^{th} sampled unit of rank in the h^{th} stratum during the j^{th} th cycle, where $i = 1, 2, ..., m$; $j = 1, 2, ..., r$; and $h = 1, 2, ..., L$. The mean of the selected units is utilized as an estimator of the population mean.

Example 3 Suppose that we have two strata, i.e. $L = 2$ and $h = 1,2$. Let (m, r) Assume that from the first stratum we select a sample of size $m \times r = 4 \times 2 = 8$ and from the second stratum we want a sample of size $m \times r = 4 \times 2 = 8$ Then the process as illustrates as follows :

Stratum 1: Now, select 8 samples as follows:

Draw a simple random sample of size $m^2 = 4^2 = 16$ units, 2 times and select of sample units.

Stratum 1	$(r=1)$	$\begin{bmatrix} X^1_{11[1]}, X^1_{12[1]}, X^1_{13[1]}, X^1_{14[1]} \\ X^1_{21[1]}, X^1_{22[1]}, X^1_{23[1]}, X^1_{24[1]} \end{bmatrix}$
	$(r=2)$	$\begin{bmatrix} X^1_{11[2]}, X^1_{12[2]}, X^1_{13[2]}, X^1_{14[2]} \\ X^1_{21[2]}, X^1_{22[2]}, X^1_{23[2]}, X^1_{24[2]} \end{bmatrix}$

For $h=1$ we have:

$X^1_{11[1]}, X^1_{14[1]}, X^1_{22[1]}, X^1_{23[1]}, X^1_{11[2]}, X^1_{14[2]}, X^1_{22[2]}, X^1_{23[2]}$

Stratum 2 Now, select 8 samples as follows:

Draw a simple random sample of size $m^2 = 4^2 = 16$ units, 2 times and select of sample units.

Stratum 2	$(r=1)$	$\begin{bmatrix} X^2_{11[1]}, X^2_{12[1]}, X^2_{13[1]}, X^2_{14[1]} \\ X^2_{21[1]}, X^2_{22[1]}, X^2_{23[1]}, X^2_{24[1]} \end{bmatrix}$
	$(r=2)$	$\begin{bmatrix} X^2_{11[2]}, X^2_{12[2]}, X^2_{13[2]}, X^2_{14[2]} \\ X^2_{21[2]}, X^2_{22[2]}, X^2_{23[2]}, X^2_{24[2]} \end{bmatrix}$

For $h=2$ we have:

$X^2_{11[1]}, X^2_{14[1]}, X^2_{22[1]}, X^2_{23[1]}, X^2_{11[2]}, X^2_{14[2]}, X^2_{22[2]}, X^2_{23[2]}$

(Define: $X^{h}_{[i]j}$, $h =$ stratum size)

Therefore, the measured SFRSS units are;

$$X^1_{11[1]}, X^1_{14[1]}, X^1_{22[1]}, X^1_{23[1]}, X^1_{11[2]}, X^1_{14[2]}, X^1_{22[2]}, X^1_{23[2]},$$
$$X^2_{11[1]}, X^2_{14[1]}, X^2_{22[1]}, X^2_{23[1]}, X^2_{11[2]}, X^2_{14[2]}, X^2_{22[2]}, X^2_{23[2]}$$

where their mean of these units is used as an estimator of the population mean.

To evaluate the efficiency of the empirical mean estimator using SFRSS compared to its counterparts in SRS, SSRS, and SRSS, a simulation was conducted in R (Version 4.1.0). The population consisted of 100,000 units, divided into two strata, with each stratum containing 50,000 units. The simulation involved specifying the number of sets in each stratum $m = 2,4,6$ and the number of cycles $r = 2,5$ to be performed.

Results and Discussion

Estimation of Population Mean

Let $X_1, X_2, ..., X_n$ be n independent random variables with mean (μ) and variance (σ^2). The mean estimator of FRSS is

$$\bar{X}_{FRSS} = \frac{1}{mr}\sum_{i=1}^{m}\sum_{j=1}^{r} X_{[l+(i-1)m]j}\ , \quad (1)$$

where $l = \frac{m}{2}$ if i is an even number and $l = \frac{m+1}{2}$ if i is an odd number (for $i = 1,2,...,m$).

The FRSS variance can be estimated by

$$S^2_{FRSS} = \frac{1}{mr-1}\left\{\sum_{i=1}^{m}\sum_{j=1}^{r}\left(X_{[l+(i-1)m]j} - \bar{X}_{FRSS}\right)^2\right\}. \quad (2)$$

The SFRSS estimator of the population mean is

Lemma 1 If the distribution is symmetric about μ, then $E\left(\bar{X}_{SFRSS}\right) = \mu$, $\left(E\left(\bar{X}_{SFRSS}\right)\right.$ is U.E. of $\left.\mu\right)$

Proof the sample size $(m_h r = n_h)$ whining the strata, we have

$$E\left(\bar{X}_{SFRSS}\right) = E\left[\sum_{h=1}^{L} W_h \left(\bar{X}_{FRSS}(m,r)\right)h\right]$$

$$= E\left[\sum_{h=1}^{L}\frac{W_h}{m_h r}\left(\sum_{i=1}^{m_h}\sum_{j=1}^{r} x_{(l+(i-1)m_h)j}\right)\right]$$

$$= \sum_{h=1}^{L}\frac{W_h}{m_h r}\left[\sum_{i=1}^{m_h}\sum_{j=1}^{r} E\left(x_{(l+(i-1)m_h)j}\right)\right]$$

$$= \sum_{h=1}^{L}\frac{W_h}{m_h r}\left[\sum_{i=1}^{m_h}\sum_{j=1}^{r} \mu_{(l+(i-1)m_h)j}\right]$$

Since the distribution is symmetric about μ, then $\mu_{(l+(i-1)m_h)j} = \mu_h$ Therefore, we have

$$E\left(\bar{X}_{SFRSS}\right) = E\left[\sum_{h=1}^{L}\frac{W_h}{m_h r}\left(\sum_{i=1}^{m_h}\sum_{j=1}^{r}\mu_h\right)\right]$$

$$= \sum_{h=1}^{L}\frac{W_h}{n_h}\left(n_h \mu_h\right)$$

$$= \sum_{h=1}^{L} W_h.\mu_h = \mu. \quad (3)$$

Where $W_h = \frac{N_h}{N}, N_h$ is the stratum size. The variance of SFRSS is given by

$$Var\left(\bar{X}_{SFRSS}\right) = Var\left[\sum_{h=1}^{L}\frac{W_h}{m_h r}\left(\sum_{i=1}^{m_h}\sum_{j=1}^{r} x_{(i:m_h)j}\right)\right]$$

$$= \sum_{h=1}^{L}\frac{W_h^2}{m_h^2 r^2}\left(\sum_{i=1}^{m_h}\sum_{j=1}^{r} Var\left(x_{(i:m_h)j}\right)\right)$$

$$= \sum_{h=1}^{L}\frac{W_h^2}{m_h^2 r^2}\left(\sum_{i=1}^{m_h}\sum_{j=1}^{r}\sigma^2_{\left[x_{(i:m)j}\right]h}\right)$$

$$= \sum_{h=1}^{L}\frac{W_h^2}{m_h^2 r^2}\sigma^2_{\left[x_{(i:m)j}\right]h} \quad (4)$$

Simulation Study

The simulation study is crafted to assess the efficacy of SFRSS in estimating population means, comparing its performance with SRS, SSRS, and SRSS across asymmetric distributions. The selected distributions for the investigation include Exponential(1), Geometric(0.5), Gamma(0.5,1), Gamma(1,2), Weibull(0.5,1), Weibull(1,2), Log N(0,1), and CHI(1). The simulations are executed with a population of 100,000 units, stratified into two groups, each containing 50,000 units. The study systematically varies the number of sets in each stratum and the number of cycles across 5,000 replications. The primary focus is on assessing the performance of SFRSS in comparison to SRS, SSRS, and SRSS under conditions of asymmetry in the underlying distributions, shown in Table 1.

Table 1. The mean squared error of estimators by various sampling methods when the population is distributed as Exp(1) with (*m* = 2,4,6) and (*r* = 2,5).

m	*r*	*MSE*			
		SRS	SSRS	SRSS	SFRSS
2	2	1.2307	0.6230	1.1612	1.2338
	5	1.1072	0.5499	17.1243	16.3668
4	2	1.1238	0.5796	0.5469	0.5175
	5	1.0459	0.5211	8.0728	8.1073
6	2	1.0843	0.5487	0.3069	0.3031
	5	1.0482	0.5199	5.2464	5.2215

For m = 2 and r = 2, the minimum MSE is achieved by SSRS with a value of 0. 6230. For m = 2 and r = 5, the minimum MSE is found with SSRS again, presenting a value of 0. 5499. Notably, this is before the dramatic increase in MSE values for SRSS and SFRSS methods.

For m = 4 and r = 2, the lowest MSE is observed with SFRSS, yielding a value of 0.5175, indicating a significant improvement in estimator accuracy through this method. For m = 4 and r = 5, again, SFRSS provides the minimum MSE, but this time the value is 0. 5211. It's noteworthy that the MSE for SRSS and SFRSS significantly increases beyond this point for larger stratum sizes.

For m = 6 and r = 2, SRSS and SFRSS both exhibit the lowest MSE, with SFRSS having a slight edge at 0. 3031. This represents the most efficient estimation accuracy among the configurations presented. For m = 6 and r = 5, similar to the previous case, the lowest MSE is found with SRSS and SFRSS, with SFRSS slightly leading at 0.3031, highlighting the effectiveness of systematic approaches in larger strata configurations, shown in Table 2.

Table 2. The mean squared error of estimators by various sampling methods when the population is distributed as Geo(0.5) with (*m* = 2,4,6) and (*r* = 2,5)

m	*r*	*MSE*			
		SRS	**SSRS**	**SRSS**	**SFRSS**
2	2	1.2307	0.6230	1.1612	1.2338
	5	1.1072	0.5499	17.1243	16.3668
4	2	1.1238	0.5796	0.5469	0.5175
	5	1.0459	0.5211	8.0728	8.1073
6	2	1.0843	0.5487	0.3069	0.3031
	5	1.0482	0.5199	5.2464	5.2215

For m = 2 and r = 2, the lowest MSE is 0.6230, achieved by SSRS. For m = 2 and r = 5, the minimum MSE is 0.5499, also achieved by SSRS.

For m = 4 and r = 2, the lowest MSE is 0.5175, found with SFRSS. For m = 4 and r = 5, the minimum MSE is 0.5211, observed with SFRSS as well.

For m = 6 and r = 2, the lowest MSE is 0.3069, achieved by SRSS. For m = 6 and r = 5, the minimum MSE is 0. 3031, found with SFRSS, shown in Table 3.

Table 3. The mean squared error of estimators by various sampling methods when the population is distributed as Gamma(0.5,1) with (*m* = 2,4,6) and (*r* = 2,5)

m	*r*	*MSE*			
		SRS	**SSRS**	**SRSS**	**SFRSS**
2	2	0.3248	0.3248	0.3248	0.3008
	5	0.2587	0.2587	0.2587	4.0612
4	2	0.3044	0.1538	0.1209	0.1332
	5	0.2676	0.1375	2.1424	2.0891
6	2	0.2292	0.2292	0.2292	0.0634
	5	0.2860	0.2860	0.2860	1.4132

For m = 2 and r = 2, the minimum MSE is 0. 3008, observed with SFRSS, indicating its superiority in this specific configuration. For m = 2 and r = 5, the lowest MSE among the methods is 0.2587, shared by SRS, SSRS, and SRSS. Notably, SFRSS's MSE dramatically increases to 4. 0612 in this setting.

For m = 4 and r = 2, SRSS exhibits the lowest MSE at 0. 1209, showcasing the efficiency of systematic sampling in this particular scenario. For m = 4 and r = 5, SSRS achieves the minimum MSE of 0. 1375, whereas SRSS and SFRSS show a significant increase in MSE, indicating potential inefficiencies or issues with these methods in this configuration.

For m = 6 and r = 2, SFRSS provides the lowest MSE value at 0. 0634, demonstrating a significant advantage over the other sampling methods in this setup. For m = 6 and r = 5, the minimum MSE is 0. 2860, consistently seen across SRS, SSRS, and SRSS, with SFRSS again showing a notable increase to 1.4132, shown in Table 4.

Table 4. The mean squared error of estimators by various sampling methods when the population is distributed as Gamma(1,2) with (*m* = 2,4,6) and (*r* = 2,5)

m	*r*	*MSE*			
		SRS	**SSRS**	**SRSS**	**SFRSS**
2	2	0.2999	0.1545	0.2972	0.2908
	5	0.2766	0.1402	4.1809	4.2252
4	2	0.2877	0.1404	0.1213	0.1205
	5	0.2640	0.1315	2.0453	2.0259
6	2	0.2678	0.1341	0.0760	0.0751
	5	0.2549	0.1303	1.3080	1.3064

For m = 2 and r = 2, SSRS shows the minimum MSE of 0. 1545, indicating superior performance in this scenario. For m = 2 and r = 5, again, SSRS has the lowest MSE, at 0. 1402, suggesting its effectiveness in stratified sampling even with a larger r = 5. SRSS and SFRSS have significantly higher MSE values in this case.

For m = 4 and r = 2, the lowest MSE is observed with SFRSS at 0.1205, closely followed by SRSS at 0. 1213, highlighting the efficiency of these methods in more complex stratification with smaller r = 2. For m = 4 and r = 5, SSRS presents the minimum MSE of 0. 1315, while SRSS and SFRSS exhibit a substantial increase, indicating potential challenges with these methods in handling larger r = 5 within this number of strata.

For m = 6 and r = 2, SFRSS has the slightest edge with an MSE of 0. 0751, with SRSS close behind at 0. 0760, showing the effectiveness of systematic strategies in dense stratification settings. For m = 6 and r = 5, the minimum MSE is 0.1303 with SSRS, while SRSS and SFRSS again show higher MSE values, albeit less dramatically than in some other configurations, shown in Table 5.

Table 5. The mean squared error of estimators by various sampling methods when the population is distributed as Weibull(0.5,1) with (*m* = 2,4,6) and (*r* = 2,5)

m	*r*	*MSE*			
		SRS	**SSRS**	**SRSS**	**SFRSS**
2	2	0.4635	0.3089	0.5634	0.5209
	5	0.4502	0.1990	5.4408	5.1812
4	2	0.3662	0.1877	0.1926	0.1602
	5	0.4865	0.1982	3.0307	3.0731
6	2	0.2566	0.1358	0.0883	0.0872
	5	0.2527	0.1211	1.3749	1.2545

For m = 2 and r = 2, The lowest MSE is 0.3089 with SSRS, suggesting it provides the most accurate estimation in this scenario. For m = 2 and r = 5, SSRS again shows the minimum MSE, at 0.1990, indicating its effectiveness for r = 5 in this configuration, whereas SRSS and SFRSS exhibit significantly higher MSE values.

For m = 4 and r = 2, the lowest MSE is observed with SFRSS at 0.1602, demonstrating the strength of stratified systematic sampling in more complex stratification settings with r = 2. For m = 4 and r = 5, SSRS offers the minimum MSE of 0.1982, while SRSS and SFRSS show higher MSEs, suggesting potential inefficiencies with these methods for r = 5 in this stratification level.

For m = 6 and r = 2, the lowest MSE is with SFRSS at 0.0872, closely followed by SRSS at 0.0883, indicating both systematic strategies perform exceptionally well in densely stratified settings. For m = 6 and r = 5, the minimum MSE is 0.1211 with SSRS, whereas SRSS and SFRSS again have higher MSE values, though the gap narrows compared to some other configurations, shown in Table 6.

Table 6. The mean squared error of estimators by various sampling methods when the population is distributed as Weibull (1,2) with (*m* = 2,4,6) and (*r* = 2,5)

m	*r*	*MSE*			
		SRS	**SSRS**	**SRSS**	**SFRSS**
2	2	0.5966	0.3112	0.5706	0.6569
	5	0.7542	0.3788	11.6874	11.3415
4	2	0.8287	0.4020	0.3762	0.3825
	5	1.3011	0.6674	9.4856	9.631
6	2	0.4188	0.2014	0.1207	0.1153
	5	1.0525	0.5311	5.3846	5.3701

For m = 2 and r = 2, SSRS presents the minimum MSE at 0.3112, indicating its superior accuracy in this scenario over the other methods.

For m = 2 and r = 5, SSRS again demonstrates the lowest MSE, with a value of 0.3788, suggesting it is the most effective method in this configuration as well, despite a general increase in MSE values across methods.

For m = 4 and r = 2, SRSS yields the lowest MSE at 0.3762, closely followed by SFRSS at 0.3825, highlighting the competitive accuracy of systematic sampling methods in moderately complex stratification settings. For m = 4 and r = 5, SSRS achieves the minimum MSE of 0.6674, with SRSS and SFRSS showing significantly higher MSEs, indicating a decrease in their efficiency for r = 5 within this number of strata.

For m = 6 and r = 2, SFRSS shows the best performance with an MSE of 0.1153, slightly better than SRSS at 0.1207, suggesting that highly stratified configurations can benefit from systematic sampling methods. For m = 6 and r = 5, SSRS has the lowest MSE at 0.5311, while SRSS and SFRSS exhibit much higher MSE values, though the difference is less pronounced than in some other configurations, shown in Table 7.

Table 7. The mean squared error of estimators by various sampling methods when the population is distributed as LogN (0,1) with (*m* = 2,4,6) and (*r* = 2,5)

m	*r*	*MSE*			
		SRS	**SSRS**	**SRSS**	**SFRSS**
2	2	13.0452	12.8110	7.5529	16.3802
	5	6.6940	58.6498	2.0199	55.4923
4	2	27.1714	10.0030	12.0343	12.8805
	5	5.7896	2.9618	44.9974	49.6113
6	2	36.0108	18.6484	9.5673	8.7301
	5	3.7584	2.1815	22.1990	20.0878

For 2 m = 2 and r = 2, SRSS demonstrates the lowest MSE at 7.5529, indicating the highest accuracy among the methods in this scenario. For m = 2 and r = 5, SRSS again provides the minimum MSE, significantly lower at 2.0199, suggesting its effectiveness in this particular configuration with r = 5.

For m = 4 and r = 2, SSRS shows the best performance with an MSE of 10.0030, highlighting the value of stratification in reducing estimation errors for a moderate number of strata. For m = 4 and r = 5, SSRS achieves the lowest MSE at 2.9618, far surpassing other methods, which exhibit much higher MSE values in this setting.

For m = 6 and r = 2, SFRSS records the minimum MSE at 8.7301, indicating its efficiency in densely stratified settings with r = 2. For m = 6 and r = 5, SSRS presents the lowest MSE at 2.1815, demonstrating its effectiveness even in complex stratification with r = 5, while SRSS and SFRSS show increased MSE values, shown in Table 8.

Table 8. The mean squared error of estimators by various sampling methods when the population is distributed as CHI(1) with (*m* = 2,4,6) and (*r*= 2,5)

m	*r*	*MSE*			
		SRS	**SSRS**	**SRSS**	**SFRSS**
2	2	36.076	15.107	30.026	25.386
	5	8.218	4.674	127.877	128.403
4	2	31.889	16.388	15.825	15.431
	5	8.338	4.0300	64.369	57.902
6	2	34.484	17.315	10.125	9.727
	5	4.229	2.082	20.987	21.641

For m = 2 and r = 2, SSRS shows the lowest MSE at 15.107, indicating superior performance in this setup compared to the other methods. For m = 2 and r = 5, SSRS again provides the minimum MSE, significantly lower at 4.674, demonstrating its effectiveness in configurations with r = 5.

For m = 4 and r = 2, SFRSS yields the best performance with an MSE of 15.431, slightly better than the other methods, highlighting the benefits of stratified systematic sampling in moderately complex stratification settings. For m = 4 and r = 5, SSRS achieves the lowest MSE at 4.0300, indicating its efficiency in reducing estimation errors for r = 5 within this level of stratification.

For m = 6 and r = 2, SFRSS records the minimum MSE at 9.727, suggesting that in densely stratified settings with smaller sample sizes, combining stratification with systematic sampling can optimize estimation accuracy. For m = 6 and r = 5, SSRS presents the lowest MSE at 2.082, demonstrating its effectiveness in complex stratifications with r = 5, while SRSS and SFRSS show increased MSE values but not as dramatically as in some previous configurations, shown in Table 9.

Efficiency comparisons are made by evaluating the relative efficiencies of SFRSS concerning SRS, SSRS, and SRSS, respectively are given by;

$$eff\left(\bar{X}_{SFRSS}, \bar{X}_{SRS}\right) = \frac{MSE\left(\bar{X}_{SRS}\right)}{MSE\left(\bar{X}_{SFRSS}\right)},$$

$$eff\left(\bar{X}_{SFRSS}, \bar{X}_{SSRS}\right) = \frac{MSE\left(\bar{X}_{SSRS}\right)}{MSE\left(\bar{X}_{SFRSS}\right)},$$

$$eff\left(\bar{X}_{SFRSS}, \bar{X}_{SRSS}\right) = \frac{MSE\left(\bar{X}_{SRSS}\right)}{MSE\left(\bar{X}_{SFRSS}\right)},$$

Table 9. The mean squared error of estimators by various sampling methods when the population is distributed as CHI(1) with (*m* = 2,4,6) and (*r* = 2,5)

Distribution	**r**	**Efficiency**		
		SFRSS vs. SRS	**SFRSS vs. SSRS**	**SFRSS vs. SRSS**
Exp(1)	2	0.9975	0.5049	0.9412
	5	0.0676	0.0336	1.0463
Geo(0.5)	2	1.0798	0.5062	0.9734
	5	0.0637	0.0315	0.9743
Gamma(0.5,1)	2	1.2323	0.5698	1.0488
	5	0.0637	0.0320	1.0227
Gamma(1,2)	2	1.0313	0.5313	1.0220
	5	0.0655	0.0332	0.9895
Weibull(0.5,1)	2	0.8899	0.5931	1.0816
	5	0.0869	0.0384	1.0501
Weibull(1,2)	2	0.9082	0.4738	0.8687
	5	0.0665	0.0334	1.0305
Log N(0,1)	2	0.7964	0.7821	0.4611
	5	0.0701	1.0569	0.0364
CHI(1)	2	1.4211	0.5951	1.1828
	5	0.0640	0.0364	0.9959

In cases where SFRSS has higher efficiency compared to SRS, SSRS, and SRSS, it signifies that SFRSS is more effective and preferable for estimating the population mean under those specific conditions and across the various distributions examined.

In case Exp(1), at r = 2, SFRSS is nearly as efficient as SRS and significantly more efficient than SSRS and slightly less so compared to SRSS. At r=5, its efficiency surpasses SRSS notably, indicating SFRSS's superiority in more complex scenarios.

In case Geo(0.5), SFRSS shows superior efficiency compared to SRS and is comparably or more efficient than SSRS and SRSS at r = 2. This suggests that SFRSS excels in conditions that are not overly complex. At r = 5, despite a general decrease in efficiency, SFRSS still outperforms SRSS in the Geo(0.5) distribution, showing its strength in specific conditions.

In case Gamma(0.5,1) and Gamma(1,2), SFRSS again demonstrates higher efficiency at r = 2 across both parameters of the Gamma distribution, indicating its robustness in various conditions.

In case Weibull(0.5,1) and Weibull(1,2), at r = 2, SFRSS shows a mix of efficiency but generally performs well, particularly against SSRS and SRSS in Weibull(0.5,1), highlighting its adaptability.

In case Log N(0,1), Interestingly, at r = 5, SFRSS's efficiency dramatically improves against SSRS, suggesting it might be particularly suited for estimating means in log-normal distributions under certain conditions.

In case CHI(1), SFRSS exhibits the highest efficiency at r = 2 compared to all methods, showcasing its effectiveness in Chi-Square distributions for simpler conditions, shown in Table 10.

Table 10. The comparative efficiency of SFRSS in estimating the population mean, with reference to SRS, SSRS, and SRSS, under the given conditions of *m* = 4 and *r* = 2, 5

Distribution	r	Efficiency		
		SFRSS vs. SRS	SFRSS vs. SSRS	SFRSS vs. SRSS
Exp(1)	2	2.1716	1.1200	1.0568
	5	0.1290	0.0643	0.9957
Geo(0.5)	2	2.4180	1.1603	1.0574
	5	0.1300	0.0657	0.9768
Gamma(0.5,1)	2	2.2853	1.1550	0.9078
	5	0.1281	0.0658	1.0255
Gamma(1,2)	2	2.3876	1.1651	1.0066
	5	0.1303	0.0649	1.0096
Weibull(0.5,1)	2	2.2860	1.1716	1.2025
	5	0.1583	0.0645	0.9862
Weibull(1,2)	2	2.1665	1.0510	0.9836
	5	0.1351	0.0693	0.9849
Log N(0,1)	2	2.1095	0.7766	0.9343
	5	0.1167	0.0597	0.9070
CHI(1)	2	2.0665	1.0620	1.0255
	5	0.1440	0.0696	1.1117

Table 10 provides a comparative analysis of the efficiency of SFRSS in estimating the population mean, compared to SRS, SSRS, and SRSS across various distributions.

In case Exp(1), at r = 2, SFRSS significantly outperforms SRS, SSRS, and SRSS, showcasing its superior efficiency in estimating the population mean. At r = 5, although efficiency drops, SFRSS's performance relative to SRSS remains nearly equivalent.

In case Geo(0.5), SFRSS exhibits even higher efficiency at r = 2 compared to the Exponential distribution, further solidifying its effectiveness. At r = 5, the efficiency slightly decreases but SFRSS remains highly competitive, especially against SRSS.

For both parameters of the Gamma distribution Gamma(0.5,1) and Gamma(1,2), SFRSS shows exceptional efficiency at r = 2. At r = 5, efficiency decreases but remains competitive, especially against SRSS, indicating SFRSS's robustness across Gamma distributions.

SFRSS continues to demonstrate high efficiency at r = 2 in both parameters of the Weibull distribution, with a notable peak in efficiency against SRSS in the Weibull(0.5,1) distribution. At r = 5, efficiency varies, but SFRSS still shows commendable performance, particularly against SRSS.

In case Log N(0,1), at r = 2, SFRSS's efficiency is high, though slightly lower than in some other distributions, indicating it's still an effective method. At r=5, efficiency decreases but SFRSS remains a viable method, especially compared to SRSS.

In case CHI(1), SFRSS shows good efficiency at r = 2, outperforming all compared methods. At r = 5, SFRSS's efficiency increases against SRSS, suggesting its suitability for Chi-Square distributed data under more complex conditions. Shown in Table 11.

Table 11. The comparative efficiency of SFRSS in estimating the population mean, with reference to SRS, SSRS, and SRSS, under the given conditions of *m* = 6 and *r* = 2, 5

Distribution	r	Efficiency		
		SFRSS vs. SRS	SFRSS vs. SSRS	SFRSS vs. SRSS
Exp(1)	2	3.5774	1.8103	1.0125
	5	0.2007	0.0996	1.0048
Geo(0.5)	2	3.6152	1.7204	1.0014
	5	0.2024	0.0975	1.0074
Gamma(0.5,1)	2	3.1213	1.6420	1.0170
	5	0.1844	0.1000	0.8939
Gamma(1,2)	2	3.5659	1.7856	1.0120
	5	0.1951	0.0997	1.0012
Weibull(0.5,1)	2	2.9425	1.5570	1.0123
	5	0.2014	0.0965	1.0960
Weibull(1,2)	2	3.6322	1.7467	1.0470
	5	0.1960	0.0989	1.0027
Log N(0,1)	2	4.1249	2.1361	1.0959
	5	0.1871	0.1086	1.1051
CHI(1)	2	3.5450	1.7800	1.0409
	5	0.1954	0.0962	0.9698

Table 11 provides a detailed comparison of the efficiency of SFRSS in estimating the population mean relative to SRS, SSRS, and SRSS.

In case Exp(1), at r = 2, SFRSS significantly outperforms SRS, SSRS, and is almost equivalent to SRSS, showcasing its exceptional efficiency. At r = 5, its efficiency greatly diminishes but remains close to SRSS, indicating it retains its relevance even in less ideal conditions.

In case Geo(0.5), SFRSS exhibits its superiority at r = 2, with even higher efficiency than in the Exponential distribution, reinforcing its effectiveness. At r = 5, efficiency slightly decreases yet is still competitive, showing SFRSS's robustness.

In cases Gamma(0.5,1) and Gamma(1,2), across both parameters, SFRSS displays high efficiency at r = 2, suggesting it's well-suited for Gamma distributions. At r = 5, there's a general decrease in efficiency; however, it performs notably well against SRSS, especially in the Gamma(0.5,1) distribution.

In cases Weibull(0.5,1) and Weibull(1,2), SFRSS continues to show high efficiency at r = 2, indicating its adaptability to different distribution shapes. At r = 5, while efficiency dips, SFRSS maintains a competitive stance, especially noteworthy against SRSS in the Weibull(0.5,1) scenario.

In case Log N(0,1), At r = 2, SFRSS reaches its peak efficiency compared to the other distributions, highlighting its effectiveness in handling skewed data. At r = 5, despite a decrease, it showcases impressive efficiency, especially against SRSS.

In case CHI(1), SFRSS performs excellently at r = 2, asserting its capability across a wide range of distributions. At r = 5, its efficiency slightly decreases but still competes well, particularly against SRSS.

Conclusions

The comparative analysis of the efficiency of SFRSS in estimating the population mean, relative to SRS, SSRS, and SRSS, across a variety of distributions. Through the examination of Exp(1), Geometric (Geo(0.5)), Gamma, Weibull, Log N(0,1), and CHI(1). SFRSS consistently demonstrates superior efficiency in estimating the population mean at the simpler condition (r = 2) across all examined distributions. This indicates SFRSS's strong capability in environments that are less complex, making it an optimal choice for studies where conditions are controlled or can be simplified. Even as conditions become more complex (r = 5), SFRSS remains a competitive method, particularly against SRSS. Its ability to still deliver relatively high efficiency in such scenarios suggests that SFRSS is adaptable and can be effectively utilized in a wide range of research settings. The efficiency of SFRSS varies across different distributions, yet it generally maintains a level of performance that is either superior to or competitive with the traditional sampling methods. This variability highlights the necessity of considering the specific characteristics of the data distribution when selecting a sampling method. The findings emphasize the importance of careful method selection in statistical study designs. For researchers, choosing SFRSS could significantly enhance the accuracy of population mean estimates, particularly in studies that align with the conditions where SFRSS has shown to be highly efficient. Particularly in the case of Log-Normal distributions, SFRSS's improved efficiency at r=5 against SSRS suggests that it is well-suited for handling skewed data under specific conditions. This adaptability makes SFRSS a valuable tool in a broader spectrum of statistical analyses, including those involving non-normal distributions. The study underscores the strategic importance of selecting SFRSS in scenarios that match its strengths, as evidenced by its performance across various distributions and conditions. Such strategic selection can lead to more accurate and reliable estimation of population means, thereby improving the quality of research outcome.

Acknowledgement

The authors express their gratitude to Kalasin University and Sakon Nakhon Rajabhat University for their support and collaboration.

References

Al-Omari, A.I. and Bouza, C.N. (2014). Review of ranked set sampling: Modifications and applications. Revista Investigacion Operacional, 35(3):215-235.

Bani-Mustafa, A.S., AL-Nasser, A., and Aslam M. (2011). Folded Ranked Set Sampling for Asymmetric Distributions. Communications of the Korean Statistical Society, 18(1):147-153. https://doi.org/10.5351/CKSS.2011.18.1.147

Cochran, W.G. (1977). Sampling Techniques (3rd ed.). New York: John Wiley & Sons.

Dell, T.R. and Clutter, J.L. (1972). Ranked Set Sampling Theory with Order Statistics Background. International Biometric Society, 28(2):545-555. https://doi.org/10.2307/2556166

Levy, P.S. and Lemeshow, S. (2008). Sampling of Populations: Methods and Applications (4th ed.). Wiley-Interscience. https://doi.org/10.1002/9780470374597

Matthews, M.J. and Wolfe, D.A. (2017). Folded Ranked Sampling. Statistics & Probability Letters, 112(C):173-178. https://doi.org/10.1016/j.spl.2016.11.015

McIntyre, G.A. (1952). A Method for Unbiased Selective Sampling Using Ranked Sets. Australian Journal of Agricultural Research, 3(4):385-390. https://doi.org/10.1071/AR9520385

McIntyre, G.A. (2005). A Method for Unbiased Selective Sampling Using Ranked Set. The American Statistician, 59(3):230-232. https://doi.org/10.1198/000313005X54180

Mustafa, A.B., Al-Nasser, A.D., and Aslam, M. (2011). Folded Ranked Set Sampling for Asymmetric Distributions. Communications of the Korean Statistical Society, 18(1):147-153. https://doi.org/10.5351/CKSS.2011.18.1.147

Muttlak, H.A. and McDonald, L.L. (1990). Ranked set sampling with size-biased probability of selection. International Biometric Society, 46(2):435-445. https://doi.org/10.2307/2531448

Samawi, H.M. (1996). Stratified ranked set sample, Pakistan Journal of Statistics, 12(1):9-16.

Stokes, S.L. (1980). Estimation of variance using judgement post-stratification. Journal of the American Statistical Association, 75(370):320-330. https://doi.org/10.1080/01621459.1980.10477469

Stokes, S.L. and Sager, T.W. (1988). Characterization of a ranked-set sample with application to estimating distribution functions. Journal of the American Statistical Association, 83(402):374-381. https://doi.org/10.1080/01621459.1988.10478607

Takahasi, K. and Wakimoto, K. (1968). On Unbiased Estimates of the Population Mean Based on the Sample Stratified by Means of Ordering. Annals of the Institute of Statistical Mathematics, 20(1):1-31. https://doi.org/10.1007/BF02911622

Thompson, S.K. (2012). Sampling (3rd ed.). Wiley Series in Probability and Statistics. https://doi.org/10.1002/9781118162934

Wolfe, D.A. (2004). Ranked set sampling: An approach to more efficient data collection. Statistical Science, 19(4):636-643. https://doi.org/10.1214/088342304000000369