Available online at http://www.ssstj.sci.ssru.ac.th

Suan Sunandha Science and Technology Journal

Stratified Unified Ranked Set Sampling for Asymmetric Distributions

Chainarong Peanpaylun[1*], Chanankarn Saengprasan[1], Suwiwat Witchakool[1]

[1]Department of Mathematics and Statistics, Faculty of Science and Technology, Sakon Nakhon Rajabhat University, Sakon Nakhon 47000, Thailand

Corresponding author e-mail: *tankku555@gmail.com

Received: *8 Nov, 2019* / **Accepted:** *7 Jan, 2020*

Abstract

The purposes of this study are to propose a modified ranked set sampling, which is called stratified unified ranked set sampling (SURSS), for estimating the population mean and compare the efficiency of the empirical mean estimator based on SURSS with their counterparts in simple random sampling (SRS), stratified simple random sampling (SSRS), and stratified ranked set sampling (SRSS) via a simulation. We compare efficiency using criteria mean square error (MSE) and simulate the data from three asymmetric distributions: Exponential (1), Geometric (0.5), and Gamma (1, 2). It is found that the estimator in SURSS provides more efficient than their counterparts in SRS, SSRS, and SRSS for three parent asymmetric distributions with small sample size. For the larger sample size, the proposed estimator in SURSS still provide more efficient than SRS, but it gives more efficient than SSRS and SRSS in some cases.

Keywords: simple random sampling, ranked set sampling, unified ranked set sampling, stratified unified ranked set sampling

1. Introduction

In 1952, McIntyre (1952) proposed a ranked set sampling (RSS) method to estimate the population mean of average yields. Later, RSS was developed and modified by many authors to estimate the population parameters. In 1968, Takahasi and Wakimoto (1968) provided the mathematical proof for RSS. They proved that the sample mean based on RSS is an unbiased estimator of the population mean which gave smaller variance than the sample mean based on a simple random sample (SRS) with the same sample size. In 1972, Dell and Clutter (1972) demonstrated that the variance of the sample mean based on RSS is less than or equal to that of the SRS, whether or not there are errors in ranking. In 1996, Samawi (1996) suggested a stratified ranked set sampling (SRSS). In 2011, Mustafa, Al-Nasser, and Aslam (2011) introduced a folded ranked set sampling for asymmetric distributions. In 2017, Matthews and Wolfe (2017) suggested a unified ranked sampling (URSS). The RSS method is efficiency increasing the number of set and the number of cycles.

The aim of this study are to propose the stratified unified ranked set sampling (SURSS) for estimating the population mean of asymmetric distributions and to study the efficiency of the empirical mean estimator based on SURSS.

2. Materials and methods

2.1 Simple Random Sampling (SRS)

SRS is a method of selecting n units out of N units such that every one of the ${}_N C_n$ distinct samples has an equal chance of being drawn.

2.2 Stratified Sampling

In stratified sampling method, the population of N units is divided into L non overlapping sub-groups known as strata each stratum has $N_1, N_2, ..., N_L$ units, respectively, such that $N_1 + N_2 + ... + N_L = N$. For full benefit from stratification, the size of the h^{th} strata, denoted by N_h for $h = 1, 2, ..., L$, must be known. Then the samples are drawn independently from each stratum, producing samples sizes denoted by $n_1, n_2, ..., n_L$,

Suan Sunandha Science and Technology Journal

such that the total sample size is $n=\sum_{h=1}^{L} n_h$. If a simple random sample is taken from each stratum, the whole procedure is known as **a stratified simple random sampling (SSRS).**

2.3 Ranked Set Sampling (RSS)

RSS technique can be described as follows:

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

Step 2: Allocate the m^2 selected units randomly into m sets, each of size m.

Step 3: Rank the m units in each set with respect to the variable of interest.

Step 4: Choose a sample by taking the smallest ranked unit in the first set, the second smallest ranked unit in the second set, continue the process until the largest ranked unit is selected from the last set. Then the taken samples are measured the variable of interest.

Step 5: Repeat step 1 through step 4 for r cycles to draw the RSS sample of size $n=mr$.

(Al-Omari & Bouza, 2014, pp. 215-235)

2.4 Unified Ranked Set Sampling (URSS)

URSS technique can be described as follows:

Step 1: Use a SRS method to select m^2 units from the population of interest and rank them with respect to the variable of interest.

Step 2: Select the sample units for measurement as follow
If m is an odd number, the ranked $\left(\frac{m+1}{2}+(i-1)m\right)^{th}$ units will be selected for $i=1,2,...,m$. On the other hands, if m is an even number, the ranked $\left(\frac{m}{2}+(i-1)m\right)^{th}$ units will be selected, for $i=1,2,...,m$.

Step 3: Repeat steps 1 and 2 for r cycles (for $j=1,2,...,r$) to draw the URSS of size $n=mr$.

Define $X_{[i]j}$ be the URSS sampled unit of the i^{th} rank from the j^{th} cycle, where $i=1,2,...,m$ and $j=1,2,...,r$.

(Zamanzade, 2014)

2.5 Stratified Unified Ranked Set Sampling (SURSS)

The population of N units is divided into L non overlapping sub-groups known as strata each stratum have $N_1, N_2, ..., N_L$ units, respectively, such that $N_1+N_2+...+N_L=N$. The size of the h^{th} strata denotes by N_h for $h=1,2,...,L$. Then the samples are drawn independently from each stratum, producing samples 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 URSS technique is applied for each stratum then the whole procedure is called a SURSS. Define $X_{[i]j}^{h}$ be the SURSS sampled unit of the i^{th} rank, the j^{th} cycle in the h^{th} stratum, where $i=1,2,...,m$; $j=1,2,...,r$; and $h=1,2,...,L$. The mean of selected units is used as a population mean estimator.

To compare the efficiency of the empirical mean estimator based on SURSS with their counterparts in SRS, SSRS, and SRSS via a simulation in RStudio under the population of 100,000 units divided into two strata each stratum has 50,000 units with the numbers of set in each stratum $m=2,5,10$ and the number of cycles $r=2,5,10$.

3. Results and Discussions

3.1 Estimation of Population Mean

Let $X_1, X_2, ..., X_n$ be n independent random variables from a probability density function $f(x)$, with mean (μ) and variance (σ^2). The empirical mean estimator of URSS is given by

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

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 URSS variance can be estimated by

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

Suan Sunandha Science and Technology Journal

The SURSS estimator of the population mean is given by

$$\bar{X}_{SURSS} = \sum_{h=1}^{L} W_h \left(\bar{X}_{URSS}^{h} \right) \qquad (3)$$

where $W_h = \frac{N_h}{N}$ and $\bar{X}_{URSS}^{h}$ is the URSS mean estimator in the h^{th} stratum.

The variance of $\bar{X}_{SURSS}$ is given by

$$Var\left(\bar{X}_{SURSS}\right) = Var\left[\sum_{h=1}^{L} \frac{W_h}{m_h r} \left(\sum_{i=1}^{m_h} \sum_{j=1}^{r} X_{[l+(i-1)m_h]j} \right) \right]$$

$$= \sum_{h=1}^{L} \frac{W_h^2}{m_h^2 r^2} \left(\sum_{i=1}^{m_h} \sum_{j=1}^{r} Var\left(X_{[l+(i-1)m_h]j} \right) \right)$$

$$= \sum_{h=1}^{L} \frac{W_h^2}{m_h^2 r^2} \left(\sum_{i=1}^{m_h} \sum_{j=1}^{r} \sigma^2_{[l+(i-1)m_h]j,h} \right)$$

$$= \sum_{h=1}^{L} \frac{W_h^2}{m_h r} \sigma^2_{[l+(i-1)m_h]j,h} \qquad (4)$$

3.2 Simulation Study

The simulation study is designed to investigate the performance of SURSS for estimating the population mean compared to their counterparts in SRS, SSRS, and SRSS under asymmetric distributions: Exponential(1), Geometric(0.5), and Gamma(1, 2). The simulations are done based on the population of 100,000 units is divided into two strata each stratum has 50,000 units, which are conducted for the numbers of set in each stratum $m = 2,5,10$ and the number of cycles $r = 2,5,10$ on 5,000 replications. If the underlying distribution is asymmetric, the efficiencies of SURSS relative to SRS, SSRS, and SRSS, respectively are given by

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

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

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

where MSE is the mean square error (MSE).
The simulation results are shown in Tables 1-3.

Table 1. The efficiency of SURSS relative to SRS, SSRS, and SRSS for estimating the population mean with $m = 2$ and $r = 2,5,10$

Distribution	r	Efficiency		
		SURSS vs. SRS	**SURSS vs. SSRS**	**SURSS vs. SRSS**
Exp(1)	2	**4.5738**	**2.3453**	**1.1372**
	5	**2.4891**	**1.2585**	0.3571
	10	**1.0906**	0.5499	0.0990
Geo(0.5)	2	**4.6105**	**2.3540**	**1.1868**
	5	**2.4206**	**1.2129**	0.3565
	10	**1.0614**	0.5784	0.0998
Gamma (1,2)	2	**4.5327**	**2.2916**	**1.2159**
	5	**2.2539**	**1.2350**	0.3530
	10	**1.0749**	0.5506	0.1024

Based on Table 1, the numbers of set in each stratum $m = 2$, it indicates that the SURSS estimator is more efficient than SRS estimator for all numbers of cycle $r = 2,5,10$ based on all three asymmetric distributions. In addition, the SURSS estimator is more efficient compared to SSRS for $r = 2,5$ based on all three asymmetric distributions. Moreover, the SURSS estimator is more efficient than SRSS estimator for $r = 2$ based on all three asymmetric distributions.

Table 2: The efficiency of SURSS relative to SRS, SSRS, and SRSS for estimating the population mean with $m = 5$ and $r = 2,5,10$

Distribution	r	Efficiency		
		SURSS vs. SRS	**SURSS vs. SSRS**	**SURSS vs. SRSS**
Exp(1)	2	**2.3324**	**1.2083**	0.3523
	5	0.8265	0.4251	0.0641
	10	0.3998	0.2073	0.0185
Geo(0.5)	2	**2.2595**	**1.2333**	0.3529
	5	0.8600	0.4372	0.0658
	10	0.3966	0.2112	0.0186
Gamma (1,2)	2	**2.4353**	**1.2216**	0.3608
	5	0.8246	0.4408	0.0661
	10	0.3973	0.2066	0.0185

Based on Table 2, the numbers of set in each stratum $m = 5$, we can conclude that the

Suan Sunandha Science and Technology Journal

SURSS estimator is more efficient compared to SRS and SSRS estimators for the numbers of cycle $r=2$ based on all three asymmetric distributions.

Table 3. The efficiency of SURSS relative to SRS, SSRS, and SRSS for estimating the population mean with $m=10$ and $r=2,5,10$

Distribution	r	Efficiency		
		SURSS vs. SRS	**SURSS vs. SSRS**	**SURSS vs. SRSS**
Exp(1)	2	**1.0782**	0.5517	0.1029
	5	0.3957	0.2094	0.0187
	10	0.1987	0.0976	0.0049
Geo(0.5)	2	**1.0620**	0.5499	0.1002
	5	0.3882	0.2150	0.0188
	10	0.1927	0.1036	0.0047
Gamma (1,2)	2	**1.0815**	0.5694	0.0953
	5	0.3985	0.2110	0.0178
	10	0.2034	0.1031	0.0049

Based on Table 3, the numbers of set in each stratum $m=10$, it implies that the SURSS estimator is more efficient than SRS estimators for the numbers of cycle $r=2$ based on all three asymmetric distributions.

4. Conclusion

In conclusion, the proposed estimator in SURSS provide more efficient than their counterparts in SRS, SSRS, and SRSS for three parent asymmetric distributions in the case of a small sample size. For the larger sample size, the proposed estimator in SURSS still provide more efficient than SRS, but it gives more efficient than SSRS and SRSS in some cases.

5. Acknowledgement

This work was partially funded by the National Research Council of Thailand (NRCT). The authors would like to express our thanks to Research and Development Institute, Sakon Nakhon Rajabhat University and the Department of Mathematics and Statistics, the Faculty of Science and Technology, Sakon Nakhon Rajabhat University. We also thank the referees and the associate editor for their useful comments and suggestions on the earlier draft that led to this improved version.

6. Publication Ethic

Submitted manuscripts must not have been previously published by or be under review by another print or online journal or source.

7. References

Al-Omari, A.I., & Bouza, C.N. (2014). Review of Ranked Set Sampling: Modifications and Applications. *Revista Investigacion Operacional*, *35*(3), 215-235.

Dell, T.R., & Clutter, J.L. (1972). Ranked Set Sampling Theory with Order Statistics Background. *Biometrics*, 28, 545–555. doi: 10.2307/2556166

Matthews, M.J. & Wolfe, D.A. (2017). Unified Ranked Sampling. *Statistics & Probability Letters*, *112*(C), 173–178. doi: 10.1016/j.spl.2016.11.015

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

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

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

Takahasi, K., & 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, 1-31.

Zamanzade, E. (2014). Unified Ranked Set Sampling. Retrieved from https://arxiv.org/vc/arxiv/papers/1411/1411.1351v1.pdf