การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิด้วยการเลือกลำดับที่ของชุดตัวอย่างแบบครบวงจรเมื่อจัดลำดับแบบสมบูรณ์
Stratified Unified Ranked Set Sampling with Perfect Ranking

ชัยณรงค์ เพียรภายลุน[*], ชนัญกาญจน์ แสงประสาน และ สุวิวรรธน์ วิชกูล

Chainarong Peanpaylun[*], Chanankarn Saengprasan and Suwiwat Witchakool

สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร

Department of Mathematics and Statistics, Faculty of Science and Technology, Sakon Nakhon Rajabhat University

บทคัดย่อ

ในบทความนี้นำเสนอวิธีการเลือกตัวอย่างสุ่มแบบเลือกลำดับที่ของชุดตัวอย่างแบบใหม่สำหรับประมาณค่าเฉลี่ย ของประชากร คือ การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิด้วยการเลือกลำดับที่ของชุดตัวอย่างแบบครบวงจร (SURSS) และ เปรียบเทียบประสิทธิภาพกับการเลือกตัวอย่างสุ่มแบบง่าย (SRS) การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิด้วยการเลือกตัวอย่าง สุ่มแบบง่าย (SSRS) และการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิด้วยการเลือกตัวอย่างสุ่มแบบเลือกลำดับที่ของชุดตัวอย่าง (SRSS) ผ่านการจำลองข้อมูลภายใต้การแจกแจงแบบสมมาตร 3 การแจกแจง ได้แก่ การแจกแจงแบบปรกติมาตรฐาน การ แจกแจงเอกรูป และการแจกแจงที่ของสติวเด็นท์ พบว่า ตัวประมาณของ SURSS เมื่อจัดลำดับแบบสมบูรณ์เป็นตัวประมาณที่ ไม่เอนเอียงและมีประสิทธิภาพสูงกว่าตัวประมาณที่ได้จาก SRS SSRS และ SRSS สำหรับการแจกแจงแบบสมมาตร

คำสำคัญ : การเลือกตัวอย่างสุ่มแบบง่าย ; การเลือกตัวอย่างสุ่มแบบเลือกลำดับที่ของชุดตัวอย่าง ; การเลือกตัวอย่างสุ่ม แบบเลือกลำดับที่ของชุดตัวอย่างแบบครบวงจร ; การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิด้วยการเลือกลำดับที่ ของชุดตัวอย่างแบบครบวงจร

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

Abstract

In this paper, a new modified ranked set sampling method is suggested, which is called stratified unified ranked set sampling (SURSS), for estimating the population mean. Then, we compare the efficiency of the empirical mean estimator based on the proposed sampling method with simple random sampling (SRS), stratified simple random sampling (SSRS) and stratified ranked set sampling (SRSS) via a simulation under three symmetric distributions: standard normal, uniform, and Student's t. The simulation results indicate that the estimator based on SURSS with perfect ranking is unbiased and more efficient than competitors based on SRS, SSRS, and SRSS for symmetric distributions.

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

Introduction

Sampling Technique is an important statistical process to select the good representatives for target population in order to save time and cost. Probability sampling should be used in statistical inference. There are various probability sampling such as simple random sampling (SRS) which randomly chooses n units from the population of N units that each unit has the same probability of being chosen, and stratified simple random sampling (SSRS) that the population of N units is partitioned into L strata and the samples are drawn independently from each stratum with SRS, etc. To select the sampling technique for study depends on data, target population, budget and time for study. There were many statisticians proposed the modified sampling techniques.

Ranked Set Sampling (RSS) method was first introduced by McIntyre (1952) in which ranking observations without actual measurement of interest variable. The RSS procedure can be executed as follows:

Step 1: Draw a sample of size m^2 units by using SRS from the target population.

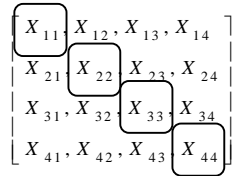**Step 2:** Randomly allocate the m^2 selected units into m sets, each of size m .

Step 3: Without knowing actual values of interest variable, rank the m sampling units increasingly with respect to interest variable by expert judgment or concomitant variable.

Step 4: Choose the smallest ranked unit in the first set, the second smallest ranked unit in the second set, and so on until the largest ranked unit is selected from the last set. Then measure the actual values for each sample on interest variable.

Step 5: Repeat step 1 through 4, referred as a cycle, for r cycles (times) to yield the RSS of size $n = mr$ (Al-Omari & Bouza, 2014). The RSS procedure for m sets with r cycles is denoted by RSS(m, r).

Example 1: Let $\text{RSS}\,(m = 4, r = 1)$ be

$$\begin{bmatrix} X_{11} & X_{12}, X_{13}, X_{14} \\ X_{21} & X_{22} & X_{23}, X_{24} \\ X_{31}, X_{32} & X_{33} & X_{34} \\ X_{41}, X_{42}, X_{43} & X_{44} \end{bmatrix},$$

where m = number of sets, r = number of cycles (times), $n = mr$ = sample size.

Then measure the actual values of RSS units, which are $X_{11}, X_{22}, X_{33}, X_{44}$. The mean of selected units is used as an estimator of the population mean.

Later, Takahasi & Wakimoto (1968) proved that the sample mean based on RSS is unbiased estimator for the population mean and it provided less variance than its counterpart in simple random sample (SRS) with the same sample size. Dell & Clutter (1972) showed that the variance of the sample mean based on RSS is smaller or equal to that of the SRS, whether there are errors in ranking or not. Then, Samawi (1996) introduced a stratified ranked set sampling (SRSS). Matthews & Wolfe (2017) proposed a unified ranked sampling (URSS). The URSS procedure can be executed as follows:

Step 1: Draw a sample of size m^2 units by using SRS from the target population. Then without knowing actual values of interest variable, rank them increasingly with respect to interest variable by expert judgment or concomitant variable.

Step 2: If m is an odd number, the ranked $\left(\dfrac{m+1}{2} + (i-1)\,m \right)^{th}$ units will be measured for $i = 1, 2, ..., m$.

If m is an even number, the ranked $\left(\dfrac{m}{2} + (i-1)\,m \right)^{th}$ units will be measured for $i = 1, 2, ..., m$.

Step 3: Repeat step 1 and 2, referred as a cycle, for r cycles (times) to yield the URSS of size $n = mr$, for

$j = 1, 2, ..., r$ (Zamanzade, 2014).

Define $X_{[k]j} = X_{[l+(i-1)m]j}$ be the URSS sampled unit of the kth measured from the jth cycle, $j = 1, 2, ..., r$ and $k = 1, 2, ..., m$.

Let m^2 be the simple random units selected from the target population, and let $X_{[1]j}, X_{[2]j}, ..., X_{\left[m^2\right]j}$ be the order statistics of $X_{1j}, X_{2j}, ..., X_{m^2 j}$ for $j = 1, 2, ..., r$.

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

Methods

Stratified Unified Ranked Set Sampling (SURSS)

In stratified sampling if the URSS method is used to select the sample units from each stratum then the whole procedure is called a SURSS. Define $X_{[k]j}^{h} = X_{[l+(i-1)m]j}^{h}$ be the SURSS sampled unit of the kth rank, the jth cycle from the hth stratum, $h = 1, 2, ..., L$, where each stratum URSS is drawn.

To illustrate the SURSS procedure, let us consider the following example.

Example 2: Suppose that there are two strata, i.e. $L = 2$, and $h = 1, 2$. Let $m = 3$ and $r = 2$. Then select a sample of size $mr = 3 \times 2 = 6$ from the first stratum and select a sample of size $mr = 6$ from the second stratum. The process illustrates as follow:

Stratum 1: Draw a simple random sample of size $m^2 = 3^2 = 9$ units for 2 cycles and rank them as follows.

Stratum 1	$(r = 1)$	$X_{[1]1}^{1} \boxed{X_{[2]1}^{1}} X_{[3]1}^{1} \mid X_{[4]1}^{1} \boxed{X_{[5]1}^{1}} X_{[6]1}^{1} \mid X_{[7]1}^{1} \boxed{X_{[8]1}^{1}} X_{[9]1}^{1}$
	$(r = 2)$	$X_{[1]2}^{1} \boxed{X_{[2]2}^{1}} X_{[3]2}^{1} \mid X_{[4]2}^{1} \boxed{X_{[5]2}^{1}} X_{[6]2}^{1} \mid X_{[7]2}^{1} \boxed{X_{[8]2}^{1}} X_{[9]2}^{1}$

For $h = 1$ the sample units from the first stratum are $X_{[2]1}^{1}, X_{[5]1}^{1}, X_{[8]1}^{1}, X_{[2]2}^{1}, X_{[5]2}^{1}, X_{[8]2}^{1}$.

Stratum 2: Draw a simple random sample of size $m^2 = 3^2 = 9$ units for 2 cycles and rank them as follows.

Stratum 2	$(r = 1)$	$X_{[1]1}^{2} \boxed{X_{[2]1}^{2}} X_{[3]1}^{2} \mid X_{[4]1}^{2} \boxed{X_{[5]1}^{2}} X_{[6]1}^{2} \mid X_{[7]1}^{2} \boxed{X_{[8]1}^{2}} X_{[9]1}^{2}$
	$(r = 2)$	$X_{[1]2}^{2} \boxed{X_{[2]2}^{2}} X_{[3]2}^{2} \mid X_{[4]2}^{2} \boxed{X_{[5]2}^{2}} X_{[6]2}^{2} \mid X_{[7]2}^{2} \boxed{X_{[8]2}^{2}} X_{[9]2}^{2}$

For $h = 2$ the sample units from the second stratum are $X_{[2]1}^{2}, X_{[5]1}^{2}, X_{[8]1}^{2}, X_{[2]2}^{2}, X_{[5]2}^{2}, X_{[8]2}^{2}$.

Thus, the measured SURSS units are $X_{[2]1}^{1}, X_{[5]1}^{1}, X_{[8]1}^{1}, X_{[2]2}^{1}, X_{[5]2}^{1}, X_{[8]2}^{1}, X_{[2]1}^{2}, X_{[5]1}^{2}, X_{[8]1}^{2}, X_{[2]2}^{2}, X_{[5]2}^{2}, X_{[8]2}^{2}$.

Simulation Results

The simulation study is conducted via RStudio to investigate the performance of the proposed estimator of the population mean compared to SRS, SSRS, and SRSS estimators under three symmetric distributions, which are standard normal, uniform, and Student's t. Assume that the population of 100,000 units is partitioned into two strata each with 50,000 units. Consider the numbers of set in each stratum $m = 2, 5, 10$ and $r = 2, 5, 10$ cycles for 5,000

replications to compute the means and variances for all estimators. If the distribution is symmetric, the relative efficiency are defined as

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

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

and

$$eff\left(\overline{X}_{SURSS}, \overline{X}_{SRSS}\right) = \frac{Var\left(\overline{X}_{SRSS}\right)}{Var\left(\overline{X}_{SURSS}\right)}.$$

Results

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 $\left(\sigma^2\right)$.

The empirical mean estimators of SRS, SSRS, and SRSS are given $\overline{X}_{SRS} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\overline{X}_{SSRS} = \sum_{h=1}^{L} W_h\left(\overline{X}_{SRS}^{h}\right)$

, and $\overline{X}_{SRSS} = \sum_{h=1}^{L} W_h\left(\overline{X}_{URSS}^{h}\right)$, respectively.

The empirical mean estimator of URSS is given by

$$\overline{X}_{URSS} = \frac{1}{mr}\sum_{j=1}^{r}\sum_{k=1}^{m} X_{[k]j},$$

where $X_{[k]j}$ be the URSS sampled unit of the kth measured from the jth cycle for $k = 1, 2, ..., m$ and $j = 1, 2, ..., r$, with estimated variance

$$Var(\overline{X}_{URSS}) = \frac{1}{mr - 1}\left\{\sum_{j=1}^{r}\sum_{k=1}^{m}\left(X_{[k]j} - \overline{X}_{URSS}\right)^2\right\}.$$

Then the SURSS estimator of the population mean is given by

$$\overline{X}_{SURSS} = \sum_{h=1}^{L} W_h\left(\overline{X}_{URSS}^{h}\right),$$

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

Lemma : If the distribution is symmetric about μ , then $E\left(\overline{X}_{SURSS}\right) = \mu$. Therefore, $\overline{X}_{SURSS}$ is an unbiased estimator of population mean.

Proof : For the sample size $\left(m_h r = n_h\right)$ in the hth stratum, we have

$$E\left(\overline{X}_{SURSS}\right) = E\left[\sum_{h=1}^{L} W_h \left(\overline{X}_{URSS}^{h}\right)\right],$$

$$= E\left[\sum_{h=1}^{L} \frac{W_h}{m_h r}\left(\sum_{j=1}^{r}\sum_{k=1}^{m_h} X_{[k]j}^{h}\right)\right],$$

$$= \sum_{h=1}^{L} \frac{W_h}{m_h r}\left[\sum_{j=1}^{r}\sum_{k=1}^{m_h} E\left(X_{[k]j}^{h}\right)\right],$$

$$= \sum_{h=1}^{L} \frac{W_h}{m_h r}\left[\sum_{j=1}^{r}\sum_{k=1}^{m_h} \mu_{[k]j}^{h}\right].$$

Since the distribution is symmetric about μ , then $\mu_{[k]j}^{h} = \mu_{h}$. Hence, we have

$$E\left(\overline{X}_{SURSS}\right) = E\left[\sum_{h=1}^{L} \frac{W_h}{m_h r}\left(\sum_{j=1}^{r}\sum_{k=1}^{m_h} \mu_{h}\right)\right],$$

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

$$= \sum_{h=1}^{L} W_h \cdot \mu_{h} = \mu,$$

where $W_h = \dfrac{N_h}{N}$, N_h is the stratum size and N is the population size. The variance of SURSS is given by

$$Var\left(\overline{X}_{SURSS}\right) = Var\left[\sum_{h=1}^{L} W_h \left(\overline{X}_{URSS}^{h}\right)\right],$$

$$= Var\left[\sum_{h=1}^{L} \frac{W_h}{m_h r}\left(\sum_{j=1}^{r}\sum_{k=1}^{m_h} X_{[k]j}^{h}\right)\right],$$

$$Var\left(\overline{X}_{SURSS}\right) = \sum_{h=1}^{L} \frac{W_h^2}{m_h^2 r^2}\left(\sum_{j=1}^{r}\sum_{k=1}^{m_h} Var\left(X_{[k]j}^h\right)\right),$$

$$= \sum_{h=1}^{L} \frac{W_h^2}{m_h^2 r^2}\left(\sum_{j=1}^{r}\sum_{k=1}^{m_h} \sigma_{[k]j,h}^2\right),$$

$$= \sum_{h=1}^{L} \frac{W_h^2}{m_h r}\sigma_{[k]j,h}^2 .$$

The simulation results are summarized in Tables 1 – 3.

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

Distribution	r	$eff\left(\overline{X}_{SURSS}, \overline{X}_{SRS}\right)$	$eff\left(\overline{X}_{SURSS}, \overline{X}_{SSRS}\right)$	$eff\left(\overline{X}_{SURSS}, \overline{X}_{SRSS}\right)$
Uniform	2	8.3143	2.0376	0.8374
	5	16.2174	4.0425	0.7480
	10	29.8389	7.2286	0.7097
Normal	2	7.3557	2.6888	1.2392
	5	6.8353	6.2725	1.3988
	10	33.4775	12.0541	1.5171
Student's t	2	9.7045	5.5415	3.6062
	5	29.0793	16.6271	7.0207
	10	59.8125	37.4124	12.5095

Table 2 The efficiency of SURSS relative to SRS, SSRS, and SRSS for $m = 5$, $r = 2, 5, 10$.

Distribution	r	$eff\left(\overline{X}_{SURSS}, \overline{X}_{SRS}\right)$	$eff\left(\overline{X}_{SURSS}, \overline{X}_{SSRS}\right)$	$eff\left(\overline{X}_{SURSS}, \overline{X}_{SRSS}\right)$
Uniform	2	16.5812	3.9995	0.7414
	5	36.3914	8.9309	0.6569
	10	70.6524	18.2277	0.6762
Normal	2	17.0967	5.9315	1.3409
	5	44.7578	15.8524	1.6239
	10	84.3977	31.1180	1.7262
Student's t	2	30.2390	16.5906	7.7660
	5	75.8580	46.4221	12.3991
	10	164.7786	90.2784	26.3307

Based on Table 2, we can conclude the following:

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

Distribution	r	$eff\left(\overline{X}_{SURSS}, \overline{X}_{SRS}\right)$	$eff\left(\overline{X}_{SURSS}, \overline{X}_{SSRS}\right)$	$eff\left(\overline{X}_{SURSS}, \overline{X}_{SRSS}\right)$
Uniform	2	28.9198	7.3918	0.7042
	5	69.3193	17.6941	0.6478
	10	158.0451	39.5009	0.7089
Normal	2	34.1389	12.3953	1.5045
	5	81.0558	29.8732	1.6407
	10	191.5909	68.2321	1.8963
Student's t	2	61.7976	35.4796	12.7364
	5	153.4307	91.5412	22.1702
	10	277.4122	177.4810	74.7536

Discussion

Based on Table 1 – 3, we can conclude the following:

1. The SURSS estimator is more efficient than SRS, SSRS, and SRSS estimators based on the same number of sample size under standard normal and Student's t distribution. For example, when $m = 5$ and $r = 2$ the relative efficiency of SURSS with respect to SRS, SSRS, and SRSS are 17.0967, 5.9315, and 1.3409 respectively for estimating the population mean of the standard normal distribution.

2. The SURSS estimator is more efficient than SRS and SSRS estimators but slightly less efficient than SRSS estimator based on the same number of sample size under uniform distribution. For instance, when $m = 2$ and $r = 2$ the relative efficiency of SURSS with respect to SRS and SSRS are 8.3143, 2.0376, and 0.8374 respectively for estimating the population mean of the uniform distribution.

3. The efficiency of SURSS is increasing as the number of sample size is increasing. For example, when $m = 2$, the relative efficient of SURSS with respect to SRS are 9.7045, 29.0793, and 59.8125 with 2, 5, and 10 cycles, respectively for estimating the population mean of the Student's t distribution. When $r = 2$, the relative efficient of SURSS with respect to SRS are 9.7045, 30.2390, and 61.7976 with the numbers of set in each stratum are 2, 5, and 10, respectively for estimating the population mean of the Student's t distribution.

Conclusions

In this paper, the SURSS is suggested for estimating the population mean in the case of perfect ranking. It is found that the estimator based on SURSS is unbiased and more efficient than their counterparts in SRS, SSRS, and SRSS under standard normal, and Student's t for all cases of simulation. In addition, the relative efficiency increases

as the number of sample size increases. In the simulation under uniform distribution, the estimator based on SRSS is more efficient than SURSS.

Acknowledgements

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.

Matthews, M.J. & Wolfe, D.A. (2017). Unified Ranked Sampling. *Statistics & Probability Letters*, *112*(C), 173–178.

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

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   August 1,2019,   from https://arxiv.org/vc/arxiv/papers/1411/1411.1351v1.pdf