DOI: 10.3195/ejejfs.619271

Eurasian Journal of Forest Science

2020 8(1): 108-122

http://dergipark.gov.tr/ejejfs

Short Communication

A comparative study of Ranked Set Sampling (RSS) and Simple Random Sampling (SRS) in agricultural studies: A case study on the walnut tree

Arzu Altin Yavuz[1]* Derya Öz[2]

[1] *Department of Statistics, Eskişehir Osmangazi University, Eskişehir, Turkey*
[2] *Computer Programming, Nisantasi University, Istanbul, Turkey*
Corresponding author: aaltin@ogu.edu.tr

Abstract
Sampling methods are used extensively in agricultural researches. In the sampling process, taking into consideration the characteristics of the target population and the use of a sampling method which can best represent the population make the estimates more sensitive and lead better estimations for the population. Several sampling methods have been developed for this purpose. In particular, Ranked Set Sampling (RSS) is a method that is used where it is difficult to measure all the units in the population, but it is easy and cheap to sort units in a small-sized population. In this study, we used the data based on the type of walnut tree leafage, Şebin, from a walnut garden established in the province of Giresun. The data was analyzed by using RSS designs and Simple Random Sampling (SRS) method to determine the mean estimators of the population. Finally, estimation performances of the SRS and RSS methods were compared accordingly. The study also emphasizes the advantage the RSS method requires fewer examples than the SRS method.
Keywords: Agriculture, sampling, simple random sampling (SRS), ranked set sampling (RSS), Juglans regia L.

Introduction

As a result of the growing population of the world, the malnutrition and increasing expansion of industrial areas, health problems have increased progressively. Despite great advances in science and technology today, the insensible consumption of natural resources and the economic difficulties have necessitated the multipurpose use of natural resources. On the other side, natural and synthetic antibiotics developed until today to fight against infectious diseases have been underwhelming as a result of resistance of microorganisms. In addition, the presence of various side effects of these antibiotics has led medical science to nature in order to discover new and different antimicrobial substances. For this reason, studies on plants and herbal medicinal raw materials obtained from these plants used in the treatment of various diseases have increased considerably.

Walnut (*Juglans regia* L.), a member of the *Juglandaceae* family, is mostly cultivated in the countries of Northern Hemisphere, such as China, USA, Iran, and Turkey (Zhang et al. 2016). Walnut is species adapted to temperate and continental climates, and it adapts to drought and different soil types. (Salvadó et al. 2005, Figueroa et al. 2017). Walnut is a plant which is used in the production of antimicrobial substances and widely grown throughout the world. Walnut leaves and green bark parts have been used colloquially in our country since ancient times. Green leaves and green bark parts of the plant are collected, dried and stored in mid-June. Then it is boiled and used as a walnut tincture to heal inflamed wounds, as bath admixture against hair loss, fungal infections on hands and feet. Especially, its leaves are brewed to decrease blood sugar level and used because of its regulating effects on digestive disorders, and the leaves boiled to gargle against oral wounds and gingivitis (Yiğit et al. 2009). The green bark and leaf parts of the plant are very important because of their high antioxidant and antimicrobial properties. It is known that bark of the walnut, green rind of fruit and leaves are used in pharmaceutical and cosmetic industries, used as a colourant in the textile industry, and its green bark and leaf parts are used in traditional medicine due to their anti-carcinogenic characteristics (Yiğit et al. 2009). It is known that juglone, which is a substance found in high quantity especially in green leaves, has very strong antioxidant and antimicrobial properties (Yaman 2012). In recent years in Turkey, efforts to establish walnut gardens have been increasing rapidly, and new walnut gardens have been established (Yalçın et al. 2012). Koyuncu and Aşkın (1997) examined the quality change of some walnut types stored in different packaging conditions. Akça (1999) conducted a study on the growth characteristics of Sebin, Bilecik, 32-B-18 and 170B-16 walnut varieties. Kantay et al. (2000) examined the density and resistance properties of walnut wood. Also, Dumanoglu (2000), Şan and Dumanoğlu (2006) and Şan ve Dumanoğlu (2006) studied, the production of walnuts. Sütyemez (2007) investigated the production amounts, flower dust and germination of 32 walnut type under certain conditions. Demir (2018) has developed a new rule (equation) with Prediction and Find Laws algorithms of data mining to calculate colour index (CI) by using an adaptive neuro-fuzzy approach.

Scientific studies on the walnut tree, which has a very significant place for the production of antimicrobial substances, are closely related to the preference of the correct scientific method. Being able to carry out medical research will primarily depend on the yield of the walnut tree. Sampling methods are frequently used in agricultural researches. Studies on obtaining information from samples instead of examining all units about any subject concerned started in the 1940s. The sample is defined as the subgroup containing fewer units than the population, which is created in accordance with various principles from population and represents the population in terms of quality and quantity. Sampling is the process of selection to create the sample representing the population about the research topic (Kılıç 2013). The selection criteria used to create samples are called sampling methods (Çıngı 1994). Thanks to sampling, it is possible to obtain characteristics of the population, such as mean, total, proportion, etc. and interpret this information.

Use of a sampling method that can represent the population best by considering the characteristics of the target population during the sampling process and determining the sample size according to this method can make the estimations for the population parameters more sensitive. Various sampling methods have been developed for this purpose. Sampling methods in the literature are divided into random and non-random sampling methods. Study results obtained by using only random sampling methods can be generalized for the population. Random sampling methods are Simple Random Sampling (SRS), Systematic, Stratified and Set Sampling methods. SRS method is the most basic method in sampling. With this method, n units are randomly selected provided that all units from the population of size N are given equal opportunity to be selected. These selected units form the sample. A list of the units that constitute the whole population is required for SRS method. Even if all the units of the population are reached, it causes time loss in some cases. For this purpose, alternative sampling methods are tried to be found.

The ranked set sampling (RSS) is an innovative sampling design and a cost-efficient alternative to simple random sampling (Balci 2013). Especially the RSS is a method used in cases where measuring all the units in the population is difficult, but sorting the units of a smaller sized sample is easy and cost-effective. From this aspect, the RSS can reduce sampling costs by preventing labour and time loss in researches where expensive laboratory analyses or field measurements are required. This method was first used by McIntyre (1952) in order to estimate the average pasture yield. It has been widely used in agriculture, forestry, sociology, ecological and environmental sciences and medical studies. Martin et al. (1980) estimated the value of a heathland located in a forest in Virginia using the RSS method. With the information obtained as a result of the research, they determined that the variance value found by RSS was smaller than the variance value calculated by the SRS. In this study, the use of the RSS method in estimating the amount of the leaves of the walnut tree, which is also remarkable with its use for medical purposes, was discussed. In the study, data about leaves of the walnut tree of Şebin type from a walnut garden established in the province of Giresun was used. Primarily, different RSS methods were introduced in the study. Then with the help of the data about the type of walnut tree , estimators of population mean were obtained using the RSS designs and the SRS method. In the results and suggestions section of the study, the efficiencies of the SRS and RSS methods in estimating the population mean were compared by the evaluation of the findings obtained.

Data Collection

In this section of the research, Şebin variety of walnut from a walnut garden established in Giresun was examined. 125 branches, which were approximate of the same size, were cut from the trees that are randomly selected from Şebin type. Selected branches were separated by an agricultural engineer to 25 sets of 5 in the laboratory and each unit within the sets was ordered in accordance with the leaf numbers on the branches. Data of 125 branches were evaluated as the population units. Sample data of 25 units, which were taken randomly from Şebin walnut variety by means of SRS and RSS sampling methods, were used to estimate the population mean. The data of 125 branches consisting of 25 5-unit sets cut from Şebin walnut variety are as shown in Table 1. Each cell in Table 1 expresses leaf numbers on a branch.

Materials and Methods

In the RSS method, the sample is selected in two steps. Thus, it shows a good spread over the distribution of the variable concerned (Gökpınar et al. 2005). In the first step, k sets with k size from the finite population are selected by the SRS. Then the units are ranked from small to large in terms of the relevant variables visually or by using a method that does not require precise measurement. In the second step, a ranked set sample is generated by measuring the first unit from the first set, the second unit from the second set, and similarly the k^{th} unit from the k^{th} set with the desired precision in terms of the relevant variable. When this operation is repeated r times, ranked set sample of size $n=rk$ is obtained. The obtained ranked set sample is as shown in Table 2.

While k^2 units are ranked by RSS, the only k of these are actually measured. In this way, the cost of sampling is reduced (Deshpande 2013). However, in the RSS method, errors originating from ranking can occur. For this reason, there are various RSS designs that differ in the steps of selecting sample units, and that function to minimize the ranking errors. Among these, commonly used methods are Extreme Ranked Set Sampling (ERSS), Median Ranked Set Sampling (MRSS), Percentile Ranked Set Sampling (PRSS) and L Ranked Set Sampling (LRSS). The most important factor that enables these methods to be different from each other is the way in which samples are created (Akıncı and Ozdemir 2010a).

The estimator of the population mean by RSS method was first suggested by Takahashi and Wakimoto in 1968. Al- Saleh and Samuh (2009) determined that this suggested estimator is more efficient than the estimator in the SRS method.

Table 1. Data relevant to Şebin walnut tree

Repetition	**Set** \ **Unit**	1	2	3	4	5
1	1	71	66	67	58	60
	2	70	68	69	69	67
	3	61	58	67	71	68
	4	65	64	69	72	69
	5	59	66	65	67	65
2	1	65	68	71	58	69
	2	65	69	68	66	66
	3	66	59	65	70	60
	4	65	69	70	65	69
	5	75	62	69	68	69
3	1	69	66	67	69	71
	2	75	66	69	67	69
	3	58	78	70	70	71
	4	79	61	68	70	72
	5	64	69	68	66	71
4	1	70	64	69	69	55
	2	66	64	67	58	59
	3	65	74	70	65	73
	4	67	69	63	66	67
	5	71	62	71	61	69
5	1	68	65	68	82	69
	2	70	63	70	91	69
	3	72	73	64	88	70
	4	63	61	66	57	66
	5	80	71	95	71	68

Table 2. *n=rk* sized ranked set sample

	Repetition			
Set	1	2	…	R
1	$X_{(1)1}$	$X_{(1)2}$	…	$X_{(1)2}$
2	$X_{(2)1}$	$X_{(2)2}$	…	$X_{(2)r}$
…	…	…	…	…
K	$X_{(k)1}$	$X_{(k)2}$	…	$X_{(k)r}$

Balanced Ranked Set Sampling

If the sizes of sets constituted in the RSS method are equal, this method is called Balanced Ranked Set Sampling (BRSS). In $X_{(i:k)j}$ j^{th} repetition, the estimator of the population mean to denote i^{th} ranked sample unit of size k is,

$$\bar{X}_{RSS} = \frac{\sum_{j=1}^{r}\sum_{i=1}^{k} X_{(i:k)j}}{kr} \quad (1)$$

and its variance is

$$Var(\bar{X}_{RSS}) = \frac{1}{kr}\{\sigma^2 - \frac{1}{k}\sum_{i=1}^{k}(\mu_{(i:k)} - \mu)^2\} \quad (2)$$

(Takahasi and Wakimoto 1968, Yıldız 2007). σ^2 given in Equation 2 denotes the population variance. When the population variance is not known, variance estimation by means of RSS method is calculated as;

$$\hat{\sigma}^2 = \frac{1}{(kr-1)}\sum_{j=1}^{r}\sum_{i=1}^{k}[(X_{(i:k)j} - \bar{X}_{RSS})]^2 \quad (3)$$

$$Var(\bar{X}_{RSS}) = \frac{1}{kr} - \left[\hat{\sigma}^2 - \frac{1}{k}\sum_{i=1}^{k}(\hat{\mu}_{(i:k)} - \bar{X}_{RSS})^2\right] \quad (4)$$

When this variance value is compared to the variance value in the SRS method,

$$Var(\bar{X}_{RSS}) = Var(\bar{X}_{SRS}) - \frac{1}{k^2 r}\sum_{i=1}^{k}(\mu_{(i:k)} - \mu)^2 \quad (5)$$

it is seen that the variance of the population mean in the RSS method is smaller (Yıldız 2007).

Extreme Ranked Set Sampling

The ranking should be carried out precisely in order to obtain a better estimator of the population mean of RSS method in comparison with the SRS method. In 1968, Takahashi and Wakimoto proposed ERSS method, which is a practical method for the cases where the sample size is larger than 4, and the ranking cannot be carried out easily. In this method, it is not necessary to rank all units. Only the largest and smallest units are selected (Samawi et al. 1996). In the ERSS method, sampling is done by following the steps below:

- *k* samples of size *k* are selected from the population.
- If the sample size is even, the smallest units from *k/2* sample and the largest units from the other *k/2* sample are selected for the actual measurement. If sample size *k* is odd, the smallest units from *(k-1)/2* samples that are ranked separately, the largest units from *(k-1)/2* sample and median value from *(k+1)/2* sample are selected for the actual measurement. This cycle is repeated *r* times for *kr* units. As a result, *kr* units denote the units of ERSS.

$X_{i(1)}$ and $X_{i(k)}$ are the smallest and largest values of ith sample, respectively (*i*=1,2,3,…,*k*); if this cycle is repeated once, the estimator of the population mean based on ERSS method is calculated as,

$$\bar{X}_{ERSS1} = \frac{1}{k}\left(\sum_{i=1}^{L} X_{i(1)} + \sum_{i=L+1}^{n} X_{i(k)}\right) \quad (6)$$

if the sample size is even. Here, $L = k/2$. If sample size *k* is odd, the estimator of the population mean is calculated as in Equation 7.

$$\bar{X}_{ERSS2} = \frac{1}{k}\left(\sum_{i=1}^{L_1} X_{i(1)} + \sum_{i=L_1+2}^{k} X_{i(k)} + X_{i(\frac{k+1}{2})}\right) \quad (7)$$

Here; $L_1 = (k-1)/2$ and $X_{i(\frac{k+1}{2})}$, denote the median of the sample (k+1)/2. In order to simplify this notation, if *k* is even; the smallest unit of $X_{(i:e)}$: *i*th sample denotes the largest unit (*i=L+1,L+2,…,k*) of (*i*=1,2,…,*L*) and *i*th sample. The smallest of the *i*th sample (*i*=1,2,…,L_1) denoting the largest of the *i*th sample (*i=k+1)/2*), the estimator of the population mean is written as follows: (Muttlak and Abu-Dayyeh 2004).

$\bar{X}_{ERSS} = \frac{1}{k}\sum_{i=1}^{k} X_{(i:e)}$ (8)

Variance is defined as,

$Var(\bar{X}_{ERSS}) = \frac{1}{k^2}\sum_{i=1}^{k} \sigma^2_{(i:e)}$ (9)

Here, it is expressed as,

$\sigma^2_{(i:k)} = E[X_{(i:e)} - E\big(X_{(i:e)}\big)]^2$ (10)

Median Ranked Set Sampling

In 1997, Muttlak suggested MRSS method in order to find population mean estimator in unimodal symmetric distributions such as normal distribution. This method is based on the median values selected from each set (Akıncı and Ozdemir 2010b). Sample selection steps in MRSS method are as follows:

- k samples of size k are selected from the population.
- Units of each set are ranked.
- If the k sample size is odd, (th smallest ranked units are selected from each sample for measurement. But if the sample size is even, $(k/2)^{th}$ units from the first k/2th sample and $((k+2)/2)^{th}$ units from the remainder $k/2$ the sample is selected. This cycle can be repeated *r* times to obtain *kr* units.

When the abovementioned cycle is repeated once, if the sample size is odd; $X_{i((k+1)/2)}$ denotes the median of i^{th} (i=1,2,3,…,k) sample. If sample size k is even, $X_{i(k/2)}$ is denoted as $(k/2)^{th}$ order statistics of i^{th} (i=1,2,3,…, $L=k/2$) sample, and $X_{i((k+2)/2)}$ is denoted as $((k+2)/2)^{th}$ order statistics of i^{th} ($i=L+1,L+2,…,k$) sample, for cases where k is even and odd, the estimators of MRSS method based population mean are given in Equation 11 and 12, respectively.

$\bar{X}_{MRSS1=} \frac{1}{k}\sum_{i=1}^{k} X_{i((k+1)/2)}$ (11)

and

$\bar{X}_{MRSS2=} \frac{1}{k}\left(\sum_{i=1}^{L} X_{i(k/2)} + \sum_{i=L+1}^{k} X_{i((k+2)/2)}\right)$ (12)

Here; $L = k/2$. In order to simplify the notation above; when $X_{(i:m)}$ the sample size is odd, it is the median of the i^{th} sample when the sample size is even, it denotes $(k/2)^{th}$ order statistics of i^{th} ($i=1,2,…,L=k/2$) sample and $((k+2)/2)^{th}$ order statistics of i^{th} ($i=L+1, L+2,…, k$) sample.

Accordingly, the estimator of MRSS based population mean is determined as,

$\bar{X}_{MRSS} = \frac{1}{k}\sum_{i=1}^{k} X_{(i:m)}$ (13)

Variance relevant to population mean is given in Equation 14.

$Var(\bar{X}_{MRSS}) = \frac{1}{k^2}\sum_{i=1}^{k} \sigma^2_{(i:m)}$ (14)

Here, it is calculated as,

$\sigma^2_{(i:m)} = E\big[X_{(i:m)} - E\big(X_{(i:m)}\big)\big]^2$ (15)

(Muttlak and Dayyeh 2004).

In symmetric distributions, $\bar{X}_{MRSS1}$ and $\bar{X}_{MRSS2}$ estimators are unbiased. But in asymmetric distributions, there are biased estimators. For this reason, MSE value is used instead of variance.

$$MSE(\bar{X}_{MRSS}) = Var(\bar{X}_{MRSS}) + (bias)^2 \quad (16)$$

Percentile Ranked Set Sampling

PRSS was suggested by Muttlak (2003) as an alternative method for RSS. With this method, which is based on the percentage value in the first step of the research, it is aimed to select sample units from different areas of the population (Akıncı and Ozdemir 2010b).

Sampling in the PRSS method is carried out by following the steps below:

- *k* samples of size *k* are selected from the population.
- Units within each set are ranked.
- If sample size *k* is even; $p(k+1)^{th}$ units from the first *k/2* sample and $(q(k+1))^{th}$ units from the second *k/2* sample are selected. Here, $0 \leq p \leq 1$ and $q = (1-p)$. When selecting sample units, $[(p(k+1))]$ and $[(q(k+1))]$ values are rounded to the closest integer value. If sample size is odd, $(p(k+1))^{th}$ ranked units from the first $((k-1)/2)$ sample and $(q(k+1))$ ranked units from the remainder $((k-1)/2)$ sample and the median of $\left(\frac{k+1}{2}\right)^{th}$ sample is selected for the actual measurement. This cycle can be repeated *r* times to obtain *kr* units.

$X_{i(p(k+1))}$ and $X_{i(q(k+1))}$ denotes $(p(k+1))^{th}$ and $(q(n+1))^{th}$ order statistics of i^{th} $(i=1,2,3,\dots,k)$ sample. In practice, $(p(k+1))$ and $(q(k+1))$ values are rounded to integer value. When *k* is even for the cycle which repeats once using PRSS method, the estimator of population mean is defined as,

$$\bar{X}_{PRSS1} = \frac{1}{k}\left(\sum_{i=1}^{L_1} X_{i(p(k+1))} + \sum_{i=L_1+1}^{k} X_{i(q(k+1))}\right) \quad (17)$$

Here, $L_1 = k/2$. When *k* is odd, the estimator of the population mean is given in Equation 18.

$$\bar{X}_{PRSS2} = \frac{1}{k}\left(\sum_{i=1}^{L_2} X_{i(p(k+1))} + \sum_{i=L_2+2}^{k} X_{i(q(k+1))} + X_{i(k+1)/2}\right) \quad (18)$$

Here, $L_2 = (k-1)/2$ and $X_{i((k+1)/2)}$ are the median of $i=(k+1)/2^{th}$ sample. When the sample size is even and odd, variance values are as specified in Equation 19 and 20, respectively.

$$Var(\bar{X}_{PRSS1}) = \frac{1}{n^2}\left[\sum_{i=1}^{L_2} Var\left(X_{i[p(n+1),n]}\right) + \sum_{i=L_2+2}^{n} Var\left(X_{i[q(n+1),n]}\right) + Var\left(X_{\frac{n+1}{2}\left[\frac{n+1}{2},n\right]}\right)\right] \quad (19)$$

and

$$Var(\bar{X}_{PRSS2}) = \frac{1}{n^2}\left[\sum_{i=1}^{L_1} Var\left(X_{i[p(n+1),n]}\right) + \sum_{i=L_1+1}^{n} Var\left(X_{i[q(n+1),n]}\right)\right] \quad (20)$$

In order to simplify the notation above, if *k* is even, $X_{(i:p)}$, is $(p(k+1))^{th}$ order statistics of i^{th} $(i=1,2,\dots,L_1)$ sample and $(q(k+1))^{th}$ order statistics of i^{th} $(i=L_1+1, L_1+2, \dots, k)$ sample. Additionally, if *k* is odd, it denotes $(p(k+1))^{th}$ order statistics of i^{th} $(i=1,2,\dots,L_2)$ sample, the median of i^{th} ($i=(k+1)/2$) sample and $(q(n+1))^{th}$ order statistics of i^{th} ($i=L_2+2, L_2+3, \dots, n$) sample. Accordingly, the estimator of PRSS based population mean is determined as,

$$\bar{X}_{PRSS} = \frac{1}{n}\sum_{i=1}^{n} X_{(i:p)} \tag{21}$$

Variance based on $\bar{X}_{PRSS}$ is given in Equation 20.

$$Var(\bar{X}_{PRSS}) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2_{(i:p)} \tag{22}$$

Here, it is calculated as,

$$\sigma^2_{(i:p)} = E\left[X_{(i:p)} - E\left(X_{(i:p)}\right)\right]^2 \tag{23}$$

(Muttlak and Abu-Dayyeh, 2004).

In symmetric distributions $\bar{X}_{PRSS1}$ and $\bar{X}_{PRSS2}$ estimators are unbiased. In asymmetric distributions, MSE values of $\bar{X}_{PRSS1}$ and $\bar{X}_{PRSS2}$ estimators are used.

L Ranked Set Sampling

This method, which was suggested by Al-Nasser in 2007 as an alternative to the RSS method, is based on robust L estimators used in the presence of outliers in the population. LRSS method is used to obtain minimum-variance unbiased estimator for population mean that is not affected by outliers in symmetric distributions.

The selection process of the units in the LRSS method is as follows:

- k sets with the size k are selected randomly.
- The units within the sets are ordered based on the relevant variance visually or by a ranking technique practised without measurement.
- On condition that $0 \le \alpha \le 0.5$, $l = [k.\alpha]$ LRSS coefficient is determined. Here l coefficient is defined as the largest of integer values which equal to $k.\alpha$ or are smaller than $k.\alpha$.
- From the first ($l+1$) set obtained, $(l+1)^{th}$ ordered units, from the last ($l+1$) set obtained $(k\text{-}l)^{th}$ ordered units and for ($j=l+2, \dots, k\text{-}l\text{-}1$) set, j^{th} are selected.
- Selected units are measured in terms of relevant variable, and L ranked set sample is with the size k is obtained (Akıncı and Özdemir 2010a).

When $l = 0$ for any sample size, LRSS method transforms into BRSS method. The estimator of the population mean based on LRSS method is defined as,

$$\bar{X}_{LRSS} = \frac{1}{k}\left(\sum_{i=1}^{l} X_{i[l+1:k]} + \sum_{i=l+1}^{k-l} X_{i[i:k]} + \sum_{i=k-l+1}^{k} X_{i[k-l:k]}\right) \tag{24}$$

$\bar{X}_{LRSS}$ the estimator is an unbiased estimator of the population mean for symmetric distributions. The variance of the estimator of $\bar{X}_{LRSS}$ is calculated as

$$Var(\bar{X}_{LRSS}) = \frac{1}{k^2}\left[\sum_{i=1}^{l} Var\left(X_{i[l+1:k]}\right) + \sum_{i=l+1}^{k-l} Var\left(X_{i[i:k]}\right) + \sum_{i=k-l+1}^{k} Var\left(X_{i[k-l:k]}\right)\right] \tag{25}$$

Quartile Ranked Set Sampling

QRSS is a method suggested by Muttlak in 2003. q coefficients are used in creating sample units. In the QRSS method,

- k random samples of size k are selected from the population.
- The units within the sets are ordered based on the relevant variance visually or by a ranking technique practised without measurement.
- If the sample size is even, $(q_l\ (k+1))^{\text{th}}$ smallest ranked unit from the first $\frac{k}{2}$ sample unit and $(q_u\ (k+1))^{\text{th}}$ smallest ranked unit from the second $\frac{k}{2}$ unit is selected for measurement. Here, q_l=0,25 and $q_u = 0{,}75$. Rounded values of $q_1\ (k+1)$ and $q_u\ (k+1)$ are used. If the sample size is odd, $(q_1(k+1))^{\text{th}}$ smallest ranked unit from the first $\frac{(k-1)}{2}$ sample, $\left(q_u(k+1)\right)^{\text{th}}$ smallest unit from the other $\frac{(k-1)}{2}$ sample and the median value of $\left(\frac{n+1}{2}\right)^{\text{th}}$ sample is selected for the actual measurement.
- This cycle is repeated r times to obtain kr units.

When the sample size is even in a single cycle, the estimator of the population mean based on QRSS method is,

$$\bar{X}_{QRSS} = \frac{1}{k}\left(\sum_{i=1}^{L_1} X_{i(q_{1(k+1)})} \sum_{i=L_1+1}^{k} X_{i(q_{u(k+1)})}\right) \tag{26}$$

Here, $L_1 = \frac{k}{2}$. When the sample size is odd, the estimator of the population mean is defined as,

$$\bar{X}_{QRSS} = \frac{1}{k}\left(\sum_{i=1}^{L_2} X_{i(q_{1(k+1)})} \sum_{i=L_2+2}^{k} X_{i(q_{u(k+1)})} + X_{i((k+1)/2)}\right) \tag{27}$$

Here, $L_2 = \frac{(k-1)}{2}$ and $X_{i((k+1)/2)}$ is the median of $(i = (k+1)/2)^{\text{th}}$ sample.

In order to make the formulations more understandable, when sample size k is even; $X_{(i:q)}$ is defined as $(q_1(k+1))^{\text{th}}$ order statistics of i^{th} $(i = 1, 2,\ \dots\ ,\ L_1)$ sample and $(q_u\ (k+1))^{\text{th}}$ order statistics of i^{th} $(i = L_1 + 1,\ L_1,\ \dots\ , k)$ sample. If sample size k is even when it is $\left(q_1(k+1)\right)^{\text{th}}$ order statistics of i^{th} $(i = 1, 2,\ \dots\ , L_2)$ sample, the median of the i^{th} $(i = (k+1)/2)$ sample and $(q_u(k+1))^{\text{th}}$ order statistics of i^{th} +2, $L_2 + 3,\ \dots\ , k)$ sample, the estimator of the population mean based on QRSS is defined as,

$$\bar{X}_{QRSS} = \frac{1}{k}\sum_{i=1}^{k} X_{(i:q)} \tag{28}$$

Variance of $\bar{X}_{QRSS}$ is as given in Equation 29.

$$Var\left(\bar{X}_{QRSS}\right) = \frac{1}{k^2}\sum_{i=1}^{n} \sigma_{(i:q)}^2 \tag{29}$$

Here, $\sigma_{(i:q)}^2 = E\left[X_{(i:q)} - E\left(X_{(i:q)}\right)\right]^2$.

Moving Extreme Ranked Set Sampling

In cases where the sample size is large, the ranking process becomes difficult. In larger sample sizes, the possibility of ranking errors is higher than in smaller samples. For this purpose, MERSS method was discussed by Al-Odat and Al-Saleh in 2001. The steps of MERSS method are as follows:

- *k* SRS sets of size *1,2,3,…,k* are generated.
- Each set is ordered from smallest to largest visually or by other methods without actual measurement.
- Actual measurements are carried out by selecting the largest units from each set ranked from the first set to k^{th} set.
- 3^{rd} set is repeated, but this time, the smallest units from each set are selected and measured.
- Operations above are repeated *r* times until the desired sample size *n=2rk* is obtained.

In MERSS method, only two extreme values are used. These are the sets of largest and smallest units in various sizes. However, in the BRSS method, all units in all sets are required to be ranked. Because determining the largest and smallest units in each set, MERSS method is a much more useful design than BRSS. With this method, the sample size can be increased without too many ranking errors (Chen et al. 2013).

Results

The descriptive statistics for the data in Table 1 and the results of the analysis of goodness of fit carried out to determine the distribution of the population are given in Table 3. Here, $k=5$ and $r=5$.

When Table 3 is examined, it is seen that the distribution of the data of Şebin walnut variety at the significance level of 0.05 conforms to the normal distribution (because it has a p-value greater than 0.05 significance level). Since the distribution of the population is symmetric, the variance values will be considered in the comparison of the estimators from now on. First, a sample of 25 units from the data set was obtained by the SRS method. Table 4 shows the data related to the selected units.

Table 3. Descriptive statistics and results of the analysis of goodness of fit relevant to Şebin data set

Population Characteristics	**Minimum**	**Maximum**	**Mean**	**Standard Deviation**	**Variance**
	55	95	67,792	5,916	35,005
Distribution	**Exponential**	**Gamma**	**Normal**	**Uniform**	**Weibull**
DPLUS	0,2753	0,1608	0,1738	0,4800	0,2314
DMINUS	0,5606	0,1123	0,1105	0,0590	0,1630
DN	0,5606	0,1608	0,1738	0,4800	0,2314
P-Value	0,0000	0,0031	**0,1125**	0,0000	0,0000

Table 4. Units selected by SRS method

Observations				
67	61	64	69	65
65	66	65	70	65
67	71	70	68	69
64	70	73	67	61
70	70	73	66	95

Estimate value of the population mean based on the SRS method using the data in Table 4 is determined as,

$$\bar{X}_{SRS} = \frac{1}{5*5}\sum\sum X^*_{(i)j} = \frac{1}{25}*(67+61+\cdots+66+95) = 68{,}44$$

and variance is determined as

$$Var(\bar{X}_{SRS}) = \frac{\hat{\sigma}^2}{n} = \frac{40{,}923}{25} = 1{,}64$$

In the next step, the estimator of the mean of the population for Şebin walnut variety was tried to be obtained by BRSS method. Locations of the units determined for this are given for the first repetition in Table 5.

Table 5. Units selected for Şebin walnut variety by BRSS method

Repetition	Set \ Unit	1	2	3	4	5
1	**1**	71				
	2		68			
	3			67		
	4				72	
	5					65

Estimate value of the population mean using Equation (1) by means of BRSS is calculated as,

$$\bar{X}_{RSS} = \frac{1}{5*5}\sum\sum X^*_{(i)j} = \frac{1}{25}*(71+68+\cdots+57+68) = 67{,}24$$

and variance is calculated as,

$$Var(\bar{X}_{RSS}) = \frac{1}{kr}\{\sigma^2 - \frac{1}{k}\sum_{i=1}^{k}\left(\mu_{(i:k)} - \mu\right)^2\} = \frac{1}{5*5}\Big[35{,}005 - \frac{1}{5}((68{,}6-67{,}24)^2+\cdots+(64-67{,}24)^2)\Big] = 1{,}228$$

ERSS, MRSS, and PRSS, QRSS when *p=0,30*, and the ranking positions and values of the units selected in the first repetition, the procedure for selecting the units in the first two repetitions in MERSS method and selected units are given in Table 6. For other repetitions, the same procedures were followed, and samples of 25 units were created.

Estimate value and variance of the population mean for Şebin walnut variety were obtained by using all other RSS methods. The results obtained are given in Table 7.

According to Table 7, it is seen that RSS methods give better results than the SRS method for Şebin walnut variety. It was determined that the distribution of Şebin walnut variety population is normal. It is seen that the RSS method that gives the most efficient estimate value for this data set is MERSS. Methods that give efficient results after MERSS are ERSS, QRSS, BRSS, MRSS and PRSS, respectively.

Table 6. Sample selection steps related to other RSS methods

Method	**Repetition**	**Sets**	**1. Unit**	**2. Unit**	**3. Unit**	**4. Unit**	**5. Unit**
ERSS	1	**1**	71	-	-	-	-
		2	70	-	-	-	-
		3	-	-	67	-	-
		4	-	-	-	-	69
		5	-	-	-	-	65
MRSS	1	**1**	-	-	67	-	-
		2	-	-	69	-	-
		3	-	-	67	-	-
		4	-	-	69	-	-
		5	-	-	65	-	-
PRSS(0.30)	1	**1**	-	66	-	-	-
		2	-	68	-	-	-
		3	-	-	67	-	-
		4	-	-	-	72	-
		5	-	-	-	67	-
QRSS	1	**1**	-	66	-	-	-
		2	-	68	-	-	-
		3	-	-	67	-	-
		4	-	-	-	-	69
		5	-	-	-	-	65
MERSS	1	**1**	-	-	-	-	60
		2	-	-	-	-	67
		3	-	-	-	-	68
		4	-	-	-	-	69
		5	-	-	-	-	65
	2	**1**	69	-	-	-	-
		2	66	-	-	-	-
		3	60	-	-	-	-
		4	69	-	-	-	-
		5	69	-	-	-	-

Table 7. Relative efficiencies for Şebin walnut variety

	Şebin		
Design	**Mean**	**Var**	**RE**
SRS	68,440	1,640	100,000
BRSS	67,240	1,228	74,878
ERSS	68,400	0,780	47,561
MRSS	69,000	1,480	90,244
PRSS	66,320	1,590	96,951
QRSS	67,200	0,800	48,780
MERSS	67,240	0,760	**46,341**

In addition, RE values obtained for RSS methods according to SRS method for Şebin walnut variety was found to be smaller than 1. Therefore, RSS methods give more efficient results than the SRS method. This indicates the number of units to be considered in order to obtain the related estimate value. When Table 7 is examined, it is seen that MERSS method is more efficient than the classical SRS method in estimation of leaf mean of Şebin walnut variety. The estimate to be obtained with 25 units in the SRS method can be obtained with 12 units in MERSS method.

Discussion

In this study, the parameter estimators of population obtained by RSS method and designs suggested as an alternative to SRS method with reference to leaf data of Şebin walnut variety was compared in terms of efficiency. The application of RSS method, which is a new alternative for sampling methods commonly used in agricultural studies, was presented. Regardless of how large the population is, RSS methods are advantageous because of providing efficient result in small sample sizes. In addition to this advantage, they are widely used especially in ecological studies since they require less cost and labour force. As in the literature, the RSS method allows us to obtain more efficient estimates than the SRS method in our study, as well.

In the application section, data belonging to walnut tree leaf research carried out in Giresun province were used. RE values of RSS designs applied for data set of Şebin walnut variety are calculated in comparison with SRS. The reason for the difference in the RSS method, which gives the most efficient estimate value of population mean of Şebin walnut variety, is a distribution type of population. The efficiency of RSS methods varies depending on ranking errors for symmetric and asymmetric distributions. However, the generalization of this result obtained from a single application is not appropriate. Studies about efficiencies of different RSS designs on estimating the population mean under different distributions are present in the literature. As the RSS methods are more efficient than the SRS method for both distribution types, the preference of the RSS method in studies where the SRS method is applicable will be useful to obtain more efficient results.

References

Akça, Y. (2000). Cevizlerde Meyve Büyümesi ve Meyve Gelişiminin Saptanması Üzerinde bir Araştırma. Turk. J. Agric. For. 24 (2000), 349-354.

Akıncı, N., Özdemir, Y. A., (2010a). L- Sıralı Küme Örneklemesi Kullanılarak Yığın Ortalamasının Tahmin Edilmesi. VII. İstatistik Günleri Sempozyumu, VII. İstatistik Günleri Sempozyumu, 28–30 Haziran 2010, Ankara, Turkey.

Akıncı, N., Özdemir, Y. A., (2010b). A Study on The Efficiencies of Multistage Ranked Set Sampling Designs. Journal of Natural & Applied Sciences. Vol. 15 Issue 2, pp.127-134.

Al-Nasser, A. D. (2007). L ranked set sampling: A generalization procedure for robust visual sampling. Communications in Statistics-Simulation and Computation. 36: 33-43.

Al-Odat,M.T., Al-Saleh,M.F.(2001). A variation of ranked set sampling. Journal of Applied Statistical Science. Vol.10,137–146.

Al-Saleh, M. F., Samuh, M. H., (2009). On multistage ranked set sampling for distribution and median estimation. Computational Statistics & Data Analysis. 52, 2066-2078.

Balci, S., Akkaya, A., Ulgen, B.E. (2013). Modified maximum likelihood estimators using ranked set sampling. Journal of Computational and Applied Mathematics. 238, 171–179.

Chen, W., Xie, M., Wu, M. (2013). Parametric estimation for the scale parameter for scale distributions using moving extremes ranked set sampling. Statistics and Probability Letters. 82, 2060-2066.

Çıngı, H. (1994). Örnekleme Kuramı. Hacettepe Üniversitesi Fen Fakültesi Basımevi, p. 2, 11.

Demir, B. (2018). Application of data mining and adaptive neuro-fuzzy structure to predict colour parameters of walnuts (Juglans regia L.). Turk. J. Agric. For. 42, 216-225.

Deshpande, J. V. (2013). Ranked set sampling for environmental studies. Indian Institute of Science Education and Research. pp. 16.

Dumanoğlu, H. (2000). Dessiccation Using saturated Salt Solutions and Improvement Germination Rate of Walnut (Juglans regia L.) Somatic Embryos. Turk. J. Agric. For. 24, 491-498.

Figueroa, F., Marhuenda, J., Gironés-Vilaplana, A., Villaño, D., Martínez-Cachá, A., Mulero, J., Cerdá, B., Zafrilla P. (2017). Soil and climate determine the antioxidant capacity of walnuts. Emirates Journal of Food and Agriculture. 29(7): 557-561.

Gökpınar, F., Özdemir, Y. A., Esin, A. A. (2005). Comparing the efficiency of the estimators for the population mean under different designs of ranked set sampling. Gazi Üniversitesi Fen Bilimleri Dergisi. 18, 3, 329-339.

Kantay, R., As. Nusret, Ünsal, Ö. (2000). Ceviz (Juglans regia L.) odununun Yoğunluğu ve Bazı Mekanik Özellikleri. Turk. J. Agric. For. 24 (2000), 751-756.

Kılıç, S. (2013). Sampling Methods. Journal of Mood Disorders. 3(1): 44-6.

Koyuncu, M.A., Aşkın, M.A. (1999), Van Gölü çevresinde yetiştiriciliği yapılan bazı ceviz tiplerinin depolanması üzerine çalışmalar. Turk. J. Agric. For. 23 (1999), 785-796.

Martin, W.L., Sharik, T.L., Oderwald, R.G., Smith, D.W. (1980). Evaluation of ranked set sampling for estimating shrub phytomass in Appalachian oak forests, Publication Number FWS-4-80. School of Forestry and Wildlife Resources. Virginia Polytechnic Institute and State University, Blacksburg.

McInytre, G. A. (1952). A method for unbiased selective sampling using ranked sets. Australian Journal of Agricultural Research. 3, 385-390.

Muttlak H. A. (1997). Median ranked set sampling. J. App. Stat. Scien. 6 (4), 245-255.

Muttlak, H. A. (2003). Investigating the use of quartile ranked set samples for estimating the population mean. Applied Mathematics and Computation. 146, 437.

Muttlak, H. A., Abu- Dayyeh, W. (2004). Weighted modified ranked set sampling methods. Applied Mathematics and Computation. 151, 645-657.

Salvadó, J. S., E. R. Rahola, J. S. Casellas (2005). Frutos Secos. Salud y Culturas Mediterráneas, Editorial Glosa. SL, Barcelona, Julio.

Samawi, H. M., Ahmed, M. S., Abu-Dayyeh, W. (1996). Estimation of the population mean using extreme ranked set sampling. Biometric Journal. 38, 5, 577-586.

Sütyemez, M. (2007). Determination of Pollen Production and Quality of Some Local and Foreign Walnut Genotypes in Turkey. Turk. J. Agric. For. 31 (2007), 109-114.

Şan, B., Dumanoğlu, H. (2006). Somatic Embryogenesis from Immature Cotyledons of Apomictic and Non-Apomictic Seeds in Walnut (Juglans regia L.). Turk. J. Agric. For. 30, 111-117.

Şan, B., Dumanoğlu, H. (2006). Determination of Apomictic Fruit Set Ratio in Some Turkish Walnut (Juglans regia L.) Genotypes. Turk. J. Agric. For. 30 (2006), 189-193.

Takahasi, K., Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. Ann. Inst. Statist. Math. 20, 1-31.

Yalçın, M. et al. (2012). Ceviz meyvesinin hasadında kullanılan farklı yöntemlerin karşılaştırılması. Türk Mikrobiyoloji Cem Derg. 41 (2) : 13-21.

Yaman, K. (2012). Bitkisel atıkların değerlendirilmesi ve ekonomik önemi. Kastamonu Üni. Orman Fakültesi Dergisi. 12 (2): 339-348.

Yıldız, H. (2007). Sıralı Küme Örneklemesi Yöntemi İle Yatırımcı Profili Analizi. Marmara Üniversitesi. Unpublished Thesis, pp. 27-66.

Yiğit, D., Yigit, N., Aktas, E., Ozgen, U. (2009). Cevizin antimikrobiyal aktivitesi. Türk Mikrobiyol Cem Derg. 41 (2): 13-21.

Zhang, B., Zheng, A., Zhou,L., Huang, Z., Wang, S. (2016). Developing hot air-assisted radio frequency drying protocols for in-shell walnuts. Emirates Journal of Food and Agriculture. 28(7): 459-467.

Submitted: 12.09.2019 Accepted: 06.01.2020