Austrian Journal of Statistics
2024, Volume 53, 99–113.
http://www.ajs.or.at/
doi:10.17713/ajs.v53i4.1872

Except-Extremes Ranked Set Sampling for Estimating the Population Variance with Two Applications of Real Data Sets

Mahmoud Zuhier Aldrabseh
Universiti Sains Malaysia
School of Mathematical Science
Penang - Malaysia

Mohd Tahir Ismail
Universiti Sains Malaysia
School of Mathematical Science
Penang - Malaysia

Amer Ibrahim Al-Omari
Al al-Bayt University
Department of Mathematics
Mafraq - Jordan

Abstract

The ranked set sampling (RSS) procedure was initially established by McIntyre (1952) for estimating the mean of forage and pasture yield as more precise than simple random sampling (SRS). Recently, Aldrabseh and Ismail (2023) suggested the except extreme RSS (EERSS) approach as a modification to RSS for estimating the population mean. In this paper, a new estimator of the population variance is proposed using the EERSS method. The mean squared error and bias equations of the new estimator are derived. When the underlying distribution is non-symmetric, a simulation study is conducted to evaluate the suggested estimator relative to SRS and RSS, based on the same number of measured units, in terms of the relative precision and bias values for several sample sizes. For symmetric distributions, the exact values of the bias and relative precision of the EERSS variance estimator are evaluated. Two real datasets are utilized to illustrate the performance of the suggested variance estimator. It is found that the EERSS variance estimator is more efficient than the SRS estimator and more precise than RSS in most cases, especially for small set sizes.

Keywords: ranked set sampling, except extreme ranked set sampling, judgment ordered, variance estimation, efficiency.

1. Introduction

The problem of estimating the population parameters is very important in statistics and various fields of science. While simple random sampling (SRS) is a commonly used method, McIntyre (1952) presented ranked set sampling (RSS) as an alternative for estimating the production mean of pasture and forage yields. The RSS method is useful when the sample units can be ranked visually or by a method other than direct quantification, which is less expensive. Hence, it is widely used in agricultural, forest, and environmental fields. The RSS procedure can be described as follows:

Step 1: Select m random samples each of size m units.

Set 1	$X_1, X_2, \ldots, X_m$
Set 2	$X_1, X_2, \ldots, X_m$
$\vdots$	$\vdots$
Set m	$X_1, X_2, \ldots, X_m$

Step 2: Rank the units within each sample with respect to the variable of interest judgmentally or by any inexpensive method.

Step 3: Choose the i^{th} ranked unit from the i^{th} set for $i = 1, 2, \ldots, m$.

Set	Ranking	Selected units
Set 1	$\mathbf{X_{(1:m)}}, X_{(2:m)}, \ldots, X_{(m:m)}$	$\mathbf{X_{(1:m)}}$
Set 2	$X_{(1:m)}, \mathbf{X_{(2:m)}}, \ldots, X_{(m:m)}$	$\mathbf{X_{(2:m)}}$
$\vdots$	$\vdots$	$\vdots$
Set m	$X_{(1:m)}, X_{(2:m)}, \ldots, \mathbf{X_{(m:m)}}$	$\mathbf{X_{(m:m)}}$

Step 4: The above process can be repeated c times (cycles) to obtain a sample of size $n = cm$.

Now, let $x_{j(i:m)}$ denote the measurements of the produced ranked set sample, where $j = 1, 2, \cdots, c$ represents the cycle number. The RSS estimator of the population mean is defined as $\bar{X}_{RSS} = (1/cm)\sum_{j=1}^{c}\sum_{i=1}^{m} x_{j(i:m)}$. In their study, Takahasi and Wakimoto (1968) defined the theoretical properties of RSS and proved that $\bar{X}_{RSS}$ is an unbiased estimator of the mean.

In a recent study, Aldrabseh and Ismail (2023) provided the except extreme ranked set sampling (EERSS) method as a new approach to estimate the population mean. They showed that the EERSS estimator is more efficient than the RSS, median ranked set sampling (MRSS), moving extreme ranked set sampling (MERSS), and SRS estimators. The EERSS approach is described using the following steps:

Step 1: Select m random samples each of size $m + 2$ units.

Set 1	$X_1, X_2, X_3, \ldots, X_{m+1}, X_{m+2}$
Set 2	$X_1, X_2, X_3, \ldots, X_{m+1}, X_{m+2}$
$\vdots$	$\vdots$
Set m	$X_1, X_2, X_3, \ldots, X_{m+1}, X_{m+2}$

Step 2: Rank the units within each sample based on the variable of interest judgmentally or by any inexpensive method.

Step 3: Choose the $(i+1)^{th}$ ranked unit from the i^{th} set for $i = 1, 2, \ldots, m$.

Set	Ranking	Selected units
Set 1	$X_{(1:m+2)}, \mathbf{X_{(2:m+2)}}, X_{(3:m+2)}, \ldots, X_{(m+1:m+2)}, X_{(m+2:m+2)}$	$\mathbf{X_{(2:m+2)}}$
Set 2	$X_{(1:m+2)}, X_{(2:m+2)}, \mathbf{X_{(3:m+2)}}, \ldots, X_{(m+1:m+2)}, X_{(m+2:m+2)}$	$\mathbf{X_{(3:m+2)}}$
$\vdots$	$\vdots$	$\vdots$
Set m	$X_{(1:m+2)}, X_{(2:m+2)}, X_{(3:m+2)}, \ldots, \mathbf{X_{(m+1:m+2)}}, X_{(m+2:m+2)}$	$\mathbf{X_{(m+1:m+2)}}$

Step 4: The above process can be repeated c times (cycles) to obtain a sample of size $n = cm$.

Hence, the EERSS estimator of the population mean is defined as follows:

$$\bar{X}_{EERSS} = \frac{1}{cm} \sum_{j=1}^{c} \sum_{i=2}^{m+1} X_{j(i:m+2)}, \tag{1}$$

with mean

$$E\left(\bar{X}_{EERSS}\right) = \frac{1}{m} \sum_{i=2}^{m+1} \mu_{(i:m+2)}, \tag{2}$$

and variance

$$V\left(\bar{X}_{EERSS}\right) = \frac{1}{cm^2} \sum_{i=2}^{m+1} \sigma^2_{(i:m+2)}. \tag{3}$$

The RSS is also employed in variance estimation, Stokes (1980) utilized the RSS method to estimate population variance and suggested an asymptotically unbiased estimator given by

$$S^2_{RSS} = \frac{1}{cm-1} \sum_{j=1}^{c} \sum_{i=1}^{m} \left(X_{j(i:m)} - \bar{X}_{RSS}\right)^2, \tag{4}$$

with mean

$$E\left(S^2_{RSS}\right) = \sigma^2 + \frac{1}{m\,(cm-1)} \sum_{i=1}^{m} \tau^2_{(i:m)}, \tag{5}$$

where σ^2 is the population variance, $\tau_{(i:m)} = \mu_{(i:m)} - \mu$, μ is the population mean, $\mu_{(i:m)}$ is the mean of the $(i:m)^{th}$ ordered random variable. The variance of the estimator is given as follows:

$$\begin{aligned} V\left(S^2_{RSS}\right) =& \frac{c}{(cm-1)^2} \left[\frac{(cm-1)^2}{c^2m^2} \sum_{i=1}^{m} \mu^{(4)}_{(i:m)} + 4 \sum_{i=1}^{m} \tau^2_{(i:m)} \sigma^2_{(i:m)} + 4\frac{cm-1}{cm} \sum_{i=1}^{m} \tau_{(i:m)} \mu^{(3)}_{(i:m)} \right. \\ & \left. + \frac{4c}{c^2m^2} \sum\sum\nolimits_{i<j} \sigma^2_{(i:m)} \sigma^2_{(j:m)} + \frac{2(c-1)-(cm-1)^2}{c^2m^2} \sum_{i=1}^{m} \sigma^4_{(i:m)} \right], \end{aligned} \tag{6}$$

In recent decades, many modifications of the RSS have been presented. Muttlak (1997) proposed the MRSS, which considers only the median-ranked units. Subsequently, a multistage version of MRSS was developed (Jemain, Al-Omari, and Ibrahim 2007). Al-Odat and Al-Saleh (2001) established the MERSS, which utilizes only the extreme ranks and varies the set sizes. The MERSS method is employed to derive the Bayesian estimator of variance by Al-Hadhrami and Al-Omari (2009). Later, Zamanzade and Al-Omari (2016) suggested the neoteric RSS procedure, which is used for estimating both the mean and variance. Robust RSS (LRSS) (Al-Nasser 2007), robust extreme RSS (Al-Nasser and Mustafa 2009), double LRSS (Al-Omari and Haq 2019), minimax RSS (Al-Nasser and Al-Omari 2018), modified minimax RSS (Hanandeh and Al-Nasser 2021), and mixed methods combining SRS and RSS (Haq, Brown, Moltchanova, and Al-Omari 2014). Most of these RSS modifications aim to improve the efficiency of estimating the population mean.

Several researchers have addressed the problem of population variance estimation using RSS and its modifications. For example, Yu, Lam, and Sinha (1999) compared the performance of Stokes's estimator for the normal distribution with their new unbiased estimator. MacEachern, Öztürk, Wolfe, and Stark (2002) introduced an RSS-unbiased estimator of variance for the location-scale family. Additionally, Stokes (1995) investigated maximum likelihood (ML) and best linear unbiased estimators for both variance and mean. Similarly, Balci, Akkaya, and Ulgen (2013) introduced a modified ML for these estimators. Chen and Lim (2011) suggested plug-in estimators for the variances and standard errors of strata. Ozturk and Bayramoglu Kavlak (2020) used stratified judgment post-stratified samples to infer population mean and total. Zamanzade and Vock (2015) discussed variance estimation with a concomitant variable. Likewise, Alam, Hanif, Shahbaz, and Shahbaz (2022) suggested two generalized estimators under RSS using information from the auxiliary variable. Mahdizadeh and Zamanzade (2021) proposed a kernel estimator of the cumulative distribution function. The main objective of the current study is to discuss the EERSS procedure for estimating population variance and compare it with SRS and RSS estimators.

The remainder of the paper is structured as follows: Section 2 covers the suggested EERSS variance estimator along with its properties. Section 3 presents the results of various comparisons. In Section 4, we provide applications of the proposed method using two real datasets. Finally, Section 5 offers a summary of the study's conclusions.

2. General setup and suggested variance estimator

In this section, we define some important moments that simplify the calculations for our variance estimation results. Let $X_1, X_2, \ldots, X_{m+2}$ denote a sample of independent and identically distributed (iid) random variables from a population with mean μ, and variance σ^2. The order of the sample units is denoted by $X_{(1:m+2)}$, $X_{(2:m+2)}$, $\ldots$, $X_{(m+2:m+2)}$. Then, following Dell and Clutter (1972), for any constants v and k, we have:

$$\sum_{i=1}^{m+2} \left(X_{(i:m+2)} - v \right)^k = \sum_{i=1}^{m+2} \left(X_i - v \right)^k .$$

Accordingly, by taking the expectations of both sides, we obtain

$$\left[\sum_{i=1}^{m+2} E \left(X_{(i:m+2)} - v \right)^k \right] = (m+2)\, E \left(X_i - v \right)^k .$$

In particular, the first and second moments around the mean μ are expressed as follows:

$$\sum_{i=1}^{m+2} \left(\mu_{(i:m+2)} - \mu \right) = \sum_{i=1}^{m+2} \tau_{(i:m+2)} = 0, \tag{7}$$

and

$$\sum_{i=1}^{m+2} \sigma^2_{(i:m+2)} + \sum_{i=1}^{m+2} \tau^2_{(i:m+2)} = (m+2)\, \sigma^2. \tag{8}$$

In our case, if we exclude the extremes using Equations 7 and 8 and assuming symmetry, we obtain

$$\sum_{i=2}^{m+1} \left(\mu_{(i:m+2)} - \mu \right) = \sum_{i=2}^{m+1} \tau_{(i:m+2)} = 0, \tag{9}$$

and

$$\sum_{i=2}^{m+1} \sigma^2_{(i:m+2)} = (m+2)\, \sigma^2 - \sigma^2_{(1:m+2)} - \sigma^2_{(1:m+2)} - \sum_{i=1}^{m+2} \tau^2_{(i:m+2)}. \tag{10}$$

Based on the above discussion and for later use, we can state the following lemma 1:

Lemma 1. *Assuming that $\bar{X}_{EERSS}$ is an unbiased estimator of the population mean μ under the EERSS method, then*

1. $\sum_{i=2}^{m+1} \mu^2_{(i:m+2)} = \sum_{i=2}^{m+1} \tau^2_{(i:m+2)} + m\mu^2$.
2. $\sum_{i=2}^{m+1} E \left(X^2_{(i:m+2)} \right) = \sum_{i=2}^{m+1} \left[\sigma^2_{(i:m+2)} + \mu^2_{(i:m+2)} \right] = \sum_{i=2}^{m+1} \sigma^2_{(i:m+2)} + \sum_{i=2}^{m+1} \tau^2_{(i:m+2)} + m\mu^2$.
3. $E \left(\bar{X}_{EERSS} \right)^2 = \frac{1}{cm^2} \sum_{i=2}^{m+1} \sigma^2_{(i:m+2)} + \mu^2$.

Proof. To prove (1), we utilize formula of $\tau_{(i:m+2)}$, yielding:

$$\mu_{(i:m+2)} = \tau_{(i:m+2)} - \mu.$$

By squaring and taking the sum of both sides and then using Equation (7), we obtain

$$\sum_{i=2}^{m+1} \mu^2_{(i:m+2)} = \sum_{i=2}^{m+1} \tau^2_{(i:m+2)} - 2\mu \sum_{i=2}^{m+1} \tau_{(i:m+2)} + m\mu^2 = \sum_{i=2}^{m+1} \tau^2_{(i:m+2)} + m\mu^2.$$

Now, (2) can be proven using the variance formula and (1) as follows:

$$\sum_{i=2}^{m+1} E\left(X^2_{(i:m+2)}\right) = \sum_{i=2}^{m+1} V\left(X_{(i:m+2)}\right) + \sum_{i=2}^{m+1} \mu^2_{(i:m+2)} = \sum_{i=2}^{m+1} \sigma^2_{(i:m+2)} \sum_{i=2}^{m+1} \tau^2_{(i:m+2)} + m\mu^2.$$

Finally, (3) can be proven directly using the variance formula and Equation (3) as follows:

$$E\left(\bar{X}_{EERSS}\right)^2 = V\left(\bar{X}_{EERSS}\right) + \left[E\left(\bar{X}_{EERSS}\right)\right]^2 = \frac{1}{cm^2} \sum_{i=2}^{m+1} \sigma^2_{(i:m+2)} + \mu^2.$$

□

Let $X_1, X_2, \ldots, X_{mc}$ be a random sample of size $n = mc$ selected from the parent population, with a known population mean μ and an unknown population variance σ^2. Therefore, the most commonly used unbiased estimator of σ^2 under the SRS method is defined as follows:

$$S^2_{SRS} = \frac{1}{mc-1} \sum_{i=1}^{mc} \left(X_i - \bar{X}\right)^2 = \frac{\sum_{i=1}^{mc} X_i^2 - mc\bar{X}^2}{mc-1}, \tag{11}$$

meanwhile, the variance of the estimator in the case of SRS is given as follows:

$$V\left(S^2_{SRS}\right) = \frac{1}{mc} E\left(X-\mu\right)^4 - \frac{mc-3}{mc\left(mc-1\right)}\sigma^4. \tag{12}$$

By referring to the EERSS method with a sample of size $n = mc$, $X_{j(i:m+2)}$ for $i = 2, \ldots, m+1$ and $j = 1, 2, \ldots, c$. Therefore, the proposed EERSS variance estimator is defined by

$$S^2_{EERSS} = \frac{1}{cm-1} \sum_{j=1}^{c} \sum_{i=2}^{m+1} \left(X_{j(i:m+2)} - \bar{X}_{EERSS}\right)^2, \tag{13}$$

with mean

$$E\left(S^2_{EERSS}\right) = \frac{c}{cm-1} \left[\sum_{i=2}^{m+1} E\left(X^2_{(i:m+2)}\right) - mE\left(\bar{X}^2_{EERSS}\right)\right]. \tag{14}$$

According to David and Nagaraja (2004), in the case of a symmetric distribution, employing Lemma 1 and some simplifications, we obtain:

$$E\left(S^2_{EERSS}\right) = \frac{1}{m} \sum_{i=2}^{m+1} \sigma^2_{(i:m+2)} + \frac{c}{cm-1} \sum_{i=2}^{m+1} \tau^2_{(i:m+2)}. \tag{15}$$

Now, utilizing Equation (10), it can be written as:

$$\begin{aligned} E\left(S^2_{EERSS}\right) = \sigma^2 &+ \frac{2\sigma^2 - \sigma^2_{(1:m+2)} - \sigma^2_{(m+2:m+2)} - \tau^2_{(m+2:m+2)} - \tau^2_{(1:m+2)}}{m} \\ &+ \frac{1}{m\left(cm-1\right)} \sum_{i=2}^{m+1} \tau^2_{(i:m+2)}. \end{aligned} \tag{16}$$

Thus, S^2_{EERSS} is a biased estimator for σ^2, with a bias given by:

$$Bias\left(S^2_{EERSS}\right) = \frac{2\sigma^2 - \sigma^2_{(1:m+2)} - \sigma^2_{(m+2:m+2)} - \tau^2_{(m+2:m+2)} - \tau^2_{(1:m+2)}}{m} + \frac{1}{m\,(cm-1)} \sum_{i=2}^{m+1} \tau^2_{(i:m+2)}. \tag{17}$$

If m is sufficiently large such that the bias approaches 0, then S^2_{EERSS} is asymptotically unbiased, with its variance given in the following theorem:

Theorem 1. *Let* $\mu^{(k)}_{(i:m+2)} = E\left[X_{j(i:m+2)} - \mu_{(i:m+2)}\right]^k$ *denote the* k^{th} *moment of the* i^{th} *ordered statistics about its mean* $\mu_{(i:m+2)}$. *If* $\mu^{(4)}_{(i:m+2)}$ *is finite, then the variance of the EERSS estimator of the population variance is given by:*

$$V\left(S^2_{EERSS}\right) = \frac{c}{(mc-1)^2}\left\{\left(\frac{mc-1}{mc}\right)^2 \sum_{i=2}^{m+1} \mu^{(4)}_{(i:m+2)} + 4\sum_{i=2}^{m+1} \tau^2_{(i:m+2)}\sigma^2_{(i:m+2)} + 4\left(\frac{mc-1}{mc}\right)\sum_{i=2}^{m+1} \tau_{(i:m+2)}\mu^{(3)}_{(i:m+2)} + \frac{2\,(c-1)-(mc-1)^2}{m^2c^2}\sum_{i=2}^{m+1}\sigma^4_{(i:m+2)} + \frac{4c}{m^2c^2}\sum_{i<r}\sigma^2_{(i:m+2)}\sigma^2_{(r:m+2)}\right\}.$$

Proof. To simplify the proof, let us deal with the variance of the numerator. Now, let

$$\begin{aligned}
T &= \sum_{j=1}^{c}\sum_{i=2}^{m+1}\left(X_{j(i:m+2)} - \bar{X}_{EERSS}\right)^2 \\
&= \sum_{j=1}^{c}\sum_{i=2}^{m+1}\left[\left(X_{j(i:m+2)} - \mu_{(i:m+2)}\right) - \left(\bar{X}_{EERSS} - \mu\right) + \left(\mu_{(i:m+2)} - \mu\right)\right]^2 \\
&= \sum_{j=1}^{c}\sum_{i=2}^{m+1}\left(X_{j(i:m+2)} - \mu_{(i:m+2)}\right)^2 + \sum_{j=1}^{c}\sum_{i=2}^{m+1}\left(\bar{X}_{EERSS} - \mu\right)^2 + \sum_{j=1}^{c}\sum_{i=2}^{m+1}\left(\mu_{(i:m+2)} - \mu\right)^2 \\
&- 2\sum_{j=1}^{c}\sum_{i=2}^{m+1}\left(X_{j(i:m+2)} - \mu_{(i:m+2)}\right)\left(\bar{X}_{EERSS} - \mu\right) - 2\sum_{j=1}^{c}\sum_{i=2}^{m+1}\left(\bar{X}_{EERSS} - \mu\right)\left(\mu_{(i:m+2)} - \mu\right) \\
&+ 2\sum_{j=1}^{c}\sum_{i=2}^{m+1}\left(X_{j(i:m+2)} - \mu_{(i:m+2)}\right)\left(\mu_{(i:m+2)} - \mu\right).
\end{aligned}$$

Now, let $X_{j(i:m+2)} - \mu_{(i:m+2)} = Y_{j(i)}$. Then, $E\left(Y_{j(i)}\right) = 0$, $E\left(Y^2_{j(i)}\right) = \sigma^2_{(i:m+2)}$, $E\left(Y^k_{j(i)}\right) = \mu^{(k)}_{(i:m+2)}$, and $Cov\left(Y_{j(i)}, Y_{s(r)}\right) = 0$, for any $j \neq s$ or $i \neq r$. After omitting the constant and zero terms and simplifying further, it can be expressed as

$$T^* = \left(\frac{mc-1}{mc}\right)\sum_{j=1}^{c}\sum_{i=2}^{m+1} Y^2_{j(i)} - \frac{2}{mc}\sum_{i<r}\sum_{j=1}^{c} Y_{j(i)}Y_{j(r)} - \frac{2}{mc}\sum_{j<s}\sum_{i=2}^{m+1} Y_{j(i)}Y_{s(i)} + 2\sum_{j=1}^{c}\sum_{i=2}^{m+1}\tau_{(i:m+2)}Y_{j(i)},$$

with

$$E\left(T^*\right) = \left(\frac{mc-1}{mc}\right)\sum_{j=1}^{c}\sum_{i=2}^{m+1} E\left(Y^2_{j(i)}\right) = \left(\frac{mc-1}{m}\right)\sum_{i=2}^{m+1}\sigma^2_{(i:m+2)},$$

and

$$E\left(T^{*2}\right)=E\left\{\left(\frac{mc-1}{mc}\right)^2\left[\sum_{j=1}^{c}\sum_{i=2}^{m+1}Y_{j(i)}^4+2\sum_{i<r}\sum_{j=1}^{c}\sum_{s=1}^{c}Y_{j(i)}^2Y_{s(r)}^2+2\sum_{j<s}\sum_{i=2}^{m+1}Y_{j(i)}^2Y_{s(i)}^2\right]\right.$$
$$+\frac{4}{m^2c^2}\sum_{i<r}\sum_{j=1}^{c}\sum_{s=1}^{c}Y_{j(i)}^2Y_{s(r)}^2+\frac{4}{m^2c^2}\sum_{j<s}\sum_{i=2}^{m+1}Y_{j(i)}^2Y_{s(i)}^2+4\sum_{j=1}^{c}\sum_{i=2}^{m+1}\tau_{(i:m+2)}^2Y_{j(i)}^2$$
$$\left.+4\left(\frac{mc-1}{mc}\right)\sum_{j=1}^{c}\sum_{i=2}^{m+1}\tau_{(i:m+2)}Y_{j(i)}^3\right\}.$$

After further simplifications and applying the variance and covariance formulas, we obtain

$$E\left(T^{*2}\right)=c\left\{\left(\frac{mc-1}{mc}\right)^2\sum_{i=2}^{m+1}\mu_{(i:m+2)}^{(4)}+\frac{(mc-1)^2+2}{m^2c^2}\left[2c\sum_{i<r}\sigma_{(i:m+2)}^2\sigma_{(r:m+2)}^2+(c-1)\right.\right.$$
$$\left.\left.\cdot\sum_{i=2}^{m+1}\sigma_{(i:m+2)}^4\right]+4\sum_{i=2}^{m+1}\tau_{(i:m+2)}^2\sigma_{(i:m+2)}^2+4\left(\frac{mc-1}{mc}\right)\sum_{i=2}^{m+1}\tau_{(i:m+2)}\mu_{(i:m+2)}^{(3)}\right\}.$$

Finally, the variance of the estimator is expressed as

$$V\left(S^2{}_{EERSS}\right)=\frac{1}{(mc-1)^2}\left\{E\left(T^{*2}\right)-\left[E\left(T^*\right)\right]\right\}=\frac{c}{(mc-1)^2}\left\{\left(\frac{mc-1}{mc}\right)^2\sum_{i=2}^{m+1}\mu_{(i:m+2)}^{(4)}\right.$$
$$+4\sum_{i=2}^{m+1}\tau_{(i:m+2)}^2\sigma_{(i:m+2)}^2+4\left(\frac{mc-1}{mc}\right)\sum_{i=2}^{m+1}\tau_{(i:m+2)}\mu_{(i:m+2)}^{(3)}$$
$$\left.+\frac{2(c-1)-(mc-1)^2}{m^2c^2}\sum_{i=2}^{m+1}\sigma_{(i:m+2)}^4+\frac{4c}{m^2c^2}\sum_{i<r}\sigma_{(i:m+2)}^2\sigma_{(r:m+2)}^2\right\}.$$

The proof has been completed. □

3. Results and discussion

Since our estimator is biased, we need to calculate the mean squared error (MSE) using the following formula:

$$MSE\left(S_{EERSS}^2\right)=V\left(S_{EERSS}^2\right)+\left[Bias\left(S_{EERSS}^2\right)\right]^2. \tag{18}$$

For comparisons between $EERSS$ and RSS relative to SRS for variance estimation, the relative precision (RP) of S_{EERSS}^2 with respect to the usual sample variance S_{SRS}^2 and S_{RSS}^2 is given by

$$RP\left(S_l^2,S_{SRS}^2\right)=\frac{V\left(S_{SRS}^2\right)}{MSE\left(S_l^2\right)},l=EERSS,RSS. \tag{19}$$

This section discusses the bias and the RP of the variance estimator. All calculations for symmetric distributions were performed exactly using Wolfram Mathematica 13.3. For non-symmetric distributions, all results were obtained through simulation using MATLAB R2023a, with $L=100,000$ repetitions.

Table 1 contains the exact results of the RP and bias for the EERSS and RSS estimators applied to symmetrical distributions: $N(0,1)$, $Uquadratic\,(0,1)$, $Uniform\,(0,1)$, $Lablace\,(0,1)$, and $Beta\,(3,3)$, at various sample sizes ($m=2,4,6,8,12,16,20$). Similarly, Table 2 presents

Table 1: Relative precision comparison of EERSS and RSS estimators vs. SRS estimator for population variance, and bias values for some symmetric distributions

Distribution	m	*EERSS* *RP*	*EERSS* $\|bias\|$	*RSS* *RP*	*RSS* $\|bias\|$
	2	2.6723	0.4631	0.6768	0.3183
	4	1.8653	0.4355	0.9234	0.1913
	6	1.5579	0.4016	1.1419	0.1372
$N(0,1)$	8	1.4111	0.3713	1.3437	0.1071
	12	1.2860	0.3228	1.7156	0.0746
	16	1.2447	0.2868	2.0595	0.0573
	20	1.2358	0.2590	2.3844	0.0465
	2	2.0019	0.0187	1.2636	0.0459
	4	2.1860	0.0049	1.4386	0.0276
	6	2.3475	0.0118	1.8026	0.0199
$Uquadratic(0,1)$	8	2.3584	0.0139	2.1774	0.0155
	12	2.0699	0.0150	3.0055	0.0109
	16	1.6928	0.0151	3.9842	0.0084
	20	1.3975	0.0149	5.1650	0.0068
	2	1.7626	0.0233	0.7241	0.0278
	4	1.5052	0.0213	1.0637	0.0167
	6	1.4312	0.0187	1.3483	0.0119
$Uniform(0,1)$	8	1.4416	0.0165	1.6248	0.0093
	12	1.5634	0.0132	2.1744	0.0064
	16	1.7393	0.0110	2.7240	0.0049
	20	1.9374	0.0094	3.2742	0.0040
	2	4.3703	1.2430	0.6935	0.5625
	4	2.7464	1.1875	0.8783	0.3396
	6	2.1517	1.1196	1.0171	0.2463
$Laplace(0,1)$	8	1.8496	1.0573	1.1338	0.1943
	12	1.5495	0.9539	1.3311	0.1374
	16	1.4039	0.8733	1.5000	0.1066
	20	1.3208	0.8086	1.6512	0.0873
	2	2.1882	0.0140	0.6877	0.0117
	4	1.6179	0.0131	0.9726	0.0070
	6	1.4087	0.0118	1.2364	0.0050
$Beta(3,3)$	8	1.3230	0.0108	1.4927	0.0039
	12	1.2823	0.0091	1.9930	0.0027
	16	1.3060	0.0079	2.4828	0.0021
	20	1.3539	0.0070	2.9656	0.0017

the simulation results for the RP and bias of the EERSS and RSS estimators applied to asymmetric distributions: $Beta(5,2)$, $Rayleigh(1)$, $HalfNormal(2)$, $Weibull(1,1)$, $Exp(1)$, $Gamma(2,3)$, and $Chisquare(5)$, also at various sample sizes $(m = 2, 4, 6, 8, 12, 16, 20)$.

From Tables 1 - 2, we can conclude that:

- The RP varies across different distributions with fixed sampling methods at the same value of m. For example, at $m = 2$, the RP of the EERSS estimator compared to the SRS estimator is highest for the $Laplace(0,1)$ distribution at 4.3703 and lowest for the $Uniform(0,1)$ distribution at 1.7636.

- The RP of the EERSS estimator compared to the SRS estimator decreases as m decreases for all parent distributions. This decrease is pronounced at small values of m, gradual at medium values of m, and in some cases, there is an increase at large values of m, as shown in the RP for beta and uniform distributions. In contrast, the RP of the RSS estimator relative to the SRS estimator increases as m increases for all parent

Table 2: Relative Precision comparison of EERSS and RSS estimators vs. SRS estimator for population variance, and bias values for some asymmetric distributions

		$EERSS$		RSS	
$Distribution$	m	RP	$\lvert bias \rvert$	RP	$\lvert bias \rvert$
	2	2.5419	0.0108	0.7222	0.0082
	4	1.8514	0.0100	0.9450	0.0049
	6	1.6179	0.0091	1.1708	0.0035
$Beta\,(5,2)$	8	1.5146	0.0084	1.4452	0.0028
	12	1.4299	0.0072	1.7999	0.0019
	16	1.4486	0.0063	2.2211	0.0015
	20	1.4568	0.0056	2.5658	0.0012
	2	2.7685	0.1911	0.7005	0.1376
	4	2.0106	0.1786	0.9075	0.0840
	6	1.7473	0.1650	1.1641	0.0576
$Rayleigh\,(1)$	8	1.6233	0.1516	1.3409	0.0448
	12	1.4980	0.1318	1.6802	0.0315
	16	1.4677	0.1162	1.9927	0.0246
	20	1.4694	0.1045	2.2979	0.0197
	2	3.0087	0.6530	0.7449	0.4325
	4	2.2258	0.6095	0.8927	0.2653
	6	1.9786	0.5640	1.1182	0.1954
$HalfNormal\,(2)$	8	1.9111	0.5171	1.3362	0.1408
	12	1.7006	0.4468	1.5416	0.1026
	16	1.7078	0.3992	1.8697	0.0858
	20	1.6798	0.3630	2.1455	0.0640
	2	4.9230	0.5775	0.7074	0.2555
	4	3.8017	0.5445	0.8950	0.1513
	6	3.2180	0.5165	0.9836	0.1113
$Weibull\,(1,1)$	8	2.8319	0.4884	1.1049	0.0962
	12	2.5355	0.4382	1.3350	0.0675
	16	2.3144	0.3997	1.5066	0.0515
	20	2.2765	0.3702	1.5626	0.0490
	2	5.5139	0.5742	0.7749	0.2652
	4	3.8812	0.5469	0.8645	0.1696
	6	3.2610	0.5148	1.0267	0.1137
$Exp\,(1)$	8	2.9214	0.4851	1.1763	0.0926
	12	2.5659	0.4371	1.3076	0.0685
	16	2.3784	0.3977	1.4671	0.0538
	20	2.2122	0.3691	1.6448	0.0416
	2	4.5578	9.5543	0.8032	5.0975
	4	3.0155	8.9068	0.7867	3.3173
	6	2.6044	8.2803	1.0708	2.2824
$Gamma\,(2,3)$	8	2.3110	7.7544	1.1589	1.7873
	12	2.0662	6.8951	1.3882	1.1985
	16	2.0662	6.8951	1.3882	1.1985
	20	1.9387	5.6778	1.7726	0.8291
	2	3.8760	5.1454	0.7111	2.9644
	4	2.8830	4.8335	0.9326	1.7575
	6	2.5215	4.506	1.1148	1.2378
$Chisquare\,(5)$	8	2.2363	4.2021	1.2072	1.0322
	12	1.9810	3.7208	1.4521	0.7028
	16	1.8548	3.3413	1.6227	0.5551
	20	1.7888	3.0548	1.7807	0.4318

distributions. This increase is significant at small values of m and smooth at large values of m.

- In the same distribution and for a fixed m, the results of RP indicate that the EERSS estimator is more efficient than the RSS estimator in most cases, especially for small values of m.
- The RP of the EERSS estimator is greater than one for all cases and all the considered values of m. This indicates that the EERSS estimator outperforms SRS in all scenarios. Moreover, it consistently surpasses the RP of RSS in most cases.
- The bias results indicate that the bias of the variance estimator decreases as m decreases for both EERSS and RSS methods. However, at a fixed m and sampling method, the bias value varies across different distributions. Across all distributions and at a fixed m, the bias of the RSS estimator is lower than that of the EERSS estimator.
- In summary, the results of the RP strongly indicate that EERSS is the most efficient sampling method for estimating population variance, especially at small set sizes.

4. Applications to real data

In this section, two real datasets are considered to illustrate the efficiency of the suggested estimator. The first dataset represents the yearly crude birth rate in Jordan from 1960 to 2021, which is an important measure of the annual live births per 1,000 people estimated at midyear. The natural increase rate is obtained by subtracting the crude death rate from the crude birth rate. This natural increase rate reflects the rate of population change without considering the impact of migration (Division 2022a). The second dataset represents the yearly percentage of Jordanian people under the age of 14, from 1960 to 2022, from all populations (Division 2022b).

Table 3 presents the raw data for the two datasets. Table 4 displays the results of descriptive statistics for both datasets. Specifically, it shows that the mean and variance of the crude birth rate are 37.27 and 91.315, respectively. Additionally, it indicates that the mean and variance of the percentage of Jordanian people under the age of 14 are 43.258 and 36.778, respectively. Figures 1 and 2 depict histograms and time series plots of the two datasets. It is evident that the crude birth rate is slightly left-skewed and decreasing, while the percentage of Jordanian people under 14 is symmetric and remains approximately constant.

Table 3: Raw datasets for two populations

Dataset 1										
21.950	22.265	22.600	22.999	23.377	23.708	24.441	25.398	26.207	27.018	27.840
28.605	28.605	28.729	28.785	29.024	29.044	29.068	29.085	29.153	29.399	29.659
30.551	31.479	32.784	33.971	34.843	35.439	35.883	36.184	36.572	36.921	37.915
38.654	39.329	40.094	40.509	41.023	41.685	42.181	42.865	43.577	44.030	44.543
45.418	45.877	46.445	46.787	47.285	47.806	48.184	48.443	48.886	49.847	50.308
50.566	50.803	51.033	51.217	51.231	51.307	51.318				
Dataset 2										
45.7576	46.3203	46.8786	47.4050	47.9155	48.4247	48.8960	49.3088	49.6683	49.9636	50.1977
50.3821	50.5170	50.5977	50.6221	50.5926	50.5035	50.3610	50.1678	49.9304	49.6622	49.3675
49.0531	48.7268	48.4016	48.0751	47.7348	47.3721	46.9891	46.5898	45.7967	44.6471	43.5221
42.5032	41.7014	41.1739	40.9007	40.7824	40.6850	40.5379	40.3323	40.0622	39.7497	39.4117
39.0290	38.6318	38.2539	37.8850	37.4908	37.0804	36.6475	36.2102	35.7832	35.4660	35.3182
35.0718	34.7015	34.3305	33.9283	33.4993	33.0585	32.6091	32.0942			

To illustrate the applicability of the EERSS estimator to the two considered populations, Table 5 outlines the procedure for calculating the EERSS variance estimator. This estimator is evaluated using Equation 13 with $m = 6$. It can be observed that for the first and second

Table 4: Descriptive statistics of the two datasets

	N	Min	Max	Mean	Q1	Median	Q3	Var	Skewness
Dataset 1	62	21.95	51.318	37.27	29.024	36.746	46.445	91.315	0.0222
Dataset 2	63	32.094	50.622	43.258	37.977	44.647	49.014	36.778	-0.2970

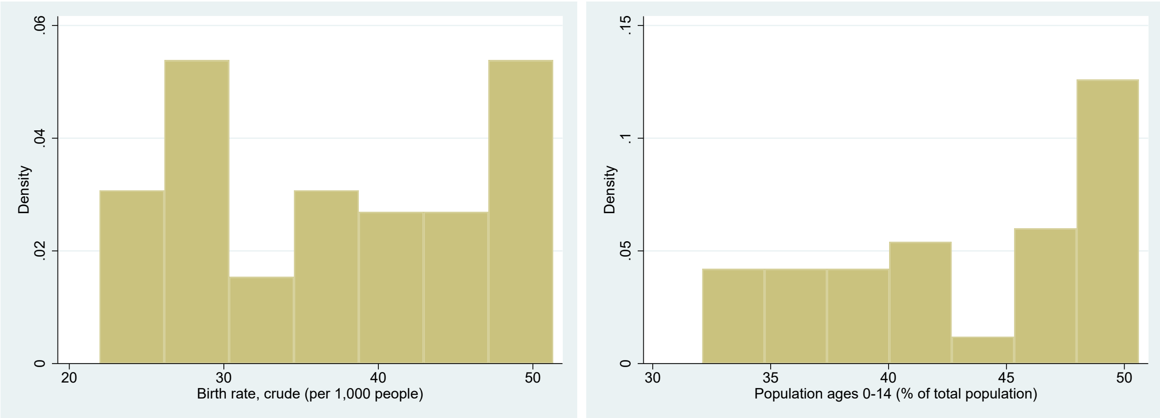

Figure 1: Histogram plots of dataset 1 (left) and dataset 2 (right)

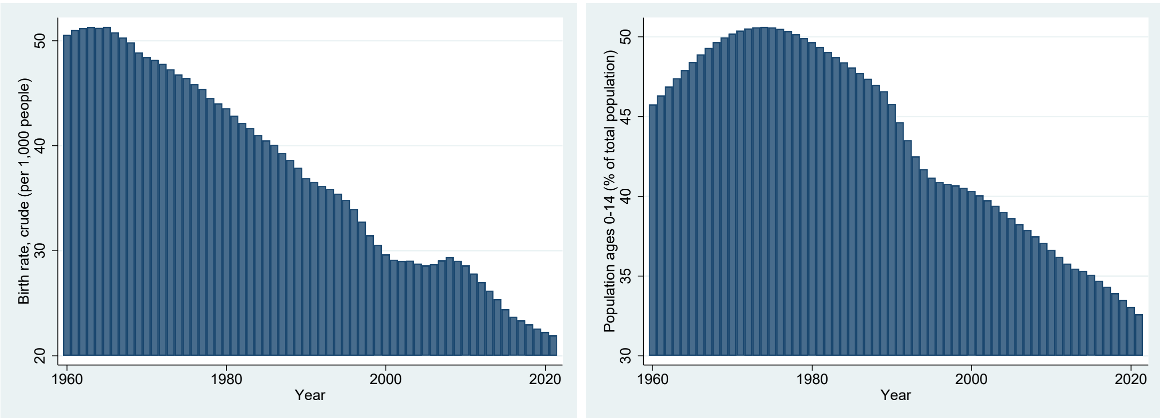

Figure 2: Time series plots of dataset 1 (left) and dataset 2 (right)

datasets, S^2_{EERSS} values are 75.0252 and 31.8805, with bias values of -16.2898 and -4.8975, respectively.

Furthermore, to assess the performance of this estimator and compare it with both RSS and SRS, we calculate the EERSS estimator using the same procedure for $m = 2, 3, 4, 5$, and 6. This involves computing the RP, absolute bias, and percentage error (PE). The PE serves as an indicator of the absolute difference between the estimated and actual variance values relative to the actual value, and it can be calculated using the formula (20).

$$PE = \frac{|\hat{\sigma}_l^2 - \sigma^2|}{\sigma^2} \times 100\%, \tag{20}$$

where l is the sampling method, either EERSS or RSS. All calculations are performed using MATLAB R2023b with $L = 100,000$ repetitions. Table 6 presents the results of RP, bias, and percent error (PE) for the two real datasets at $m = 2, 3, 4, 5, 6$. These results demonstrate the quality and applicability of EERSS in estimating the population variance, compared to both SRS and RSS procedures. The RP and absolute bias are evaluated using Equations (19) and (17), respectively.

The results of RP in Table 6 for both EERSS and RSS estimators of population variance

Table 5: EERSS procedure for a sample size of $m = 6$ from two datasets

EERSS sample of size $m = 6$ from dataset 1									
Select m SRS and ranking them								EERSS	S^2_{EERSS}
22.265	**28.605**	29.659	31.479	32.784	34.843	46.787	51.217	**28.605**	**75.0252**
22.600	24.441	**28.729**	35.439	41.685	42.181	48.184	51.231	**28.729**	
28.605	29.044	29.068	**29.085**	46.445	48.886	49.847	50.803	**29.085**	
23.708	26.207	27.840	30.551	**33.971**	40.509	42.865	44.030	**33.971**	
21.950	22.999	25.398	35.883	36.184	**36.572**	48.443	50.566	**36.572**	
23.377	28.785	36.921	41.023	47.806	50.308	**51.033**	51.318	**51.033**	
EERSS sample of size $m = 6$ from dataset 2									
Select m SRS and ranking them								EERSS	S^2_{EERSS}
33.9283	**35.0718**	40.5379	43.5221	47.7348	48.7268	49.9304	50.5926	**35.0718**	**31.8805**
33.0585	37.8850	**40.3323**	40.7824	46.9891	48.0751	50.1977	50.6221	**40.3323**	
37.0804	38.6318	40.0622	**41.1739**	42.5032	46.8786	48.4016	48.4247	**41.1739**	
34.3305	35.4660	36.6475	37.4908	**44.6471**	46.5898	48.8960	50.1678	**44.6471**	
35.3182	41.7014	45.7576	47.4050	49.3675	**49.6622**	49.6683	49.9636	**49.6622**	
32.0942	35.7832	39.0290	39.7497	46.3203	47.3721	**49.3088**	50.3610	**49.3088**	

Table 6: Relative precision, bias, and percent error of the sample variance of the two datasets at $m = 2, 3, 4, 5$ and 6

		EERSS			*RSS*		
Dataset	m	RP	$\|bias\|$	PE	RP	$\|bias\|$	PE
Dataset 1	2	1.7025	18.506	20.5979	0.7387	32.5325	36.2108
	3	1.7038	17.656	19.6525	0.9561	24.1859	26.9204
	4	1.7116	16.549	18.4205	1.1136	19.6234	21.8421
	5	1.7177	15.394	17.1347	1.2585	16.6946	18.5821
	6	1.7731	14.326	15.9457	1.3910	14.5467	16.1913
Dataset 2	2	1.7259	6.4581	17.8431	0.7609	12.7800	35.3087
	3	1.7997	6.2408	17.2426	0.9726	9.5919	26.5013
	4	1.8351	5.9055	16.3161	1.1320	7.8293	21.6315
	5	1.9046	5.4655	15.1004	1.2776	6.6002	18.2356
	6	1.9720	5.1148	14.1315	1.4267	5.8119	16.0577

increase with m, while the absolute bias and PE decrease. Furthermore, at the same values of m, the RP of the EERSS estimator exceeds that of the RSS estimator. However, the absolute bias and PE of the EERSS estimator are lower than those of the RSS estimator.

These indicators provide some insight into the consistency of the EERSS variance estimator based on real applications, aligning somewhat with the simulation results presented in the previous section.

5. Conclusions

This paper introduced a new estimator for population variance applicable to both symmetric and non-symmetric distributions. Two real datasets are utilized for illustration purposes. The study demonstrates that the EERSS estimator is asymptotically unbiased and more efficient than the SRS estimator when using the same number of measured units. Furthermore, the EERSS estimator outperforms the RSS estimator in most cases, especially for small set sizes.

In future work, further exploration of population variance estimation under alternative RSS modifications and investigation of additional applications can be discussed.

Acknowledgements

The authors would like to extend our heartfelt appreciation to the editors and reviewers for their insightful comments and constructive suggestions, which greatly contributed to enhancing the quality of this research.

References

Al-Hadhrami SA, Al-Omari AI (2009). "Bayesian Inference on the Variance of Normal Distribution Using Moving Extremes Ranked Set Sampling." *Journal of Modern Applied Statistical Methods*, **8**(1), 25. doi:10.22237/jmasm/1241137440.

Al-Nasser AD (2007). "L Ranked Set Sampling: A Generalization Procedure for Robust Visual Sampling." *Communications in Statistics—Simulation and Computation®*, **36**(1), 33–43. doi:10.1080/03610910601096510.

Al-Nasser AD, Al-Omari AI (2018). "Minimax Ranked Set Sampling." *Investigación Operacional*, **39**(4), 560–571. URL https://www.researchgate.net/publication/328651243_Minimax_ranked_set_sampling.

Al-Nasser AD, Mustafa AB (2009). "Robust Extreme Ranked Set Sampling." *Journal of Statistical Computation and Simulation*, **79**(7), 859–867. doi:10.1080/00949650701683084.

Al-Odat MT, Al-Saleh MF (2001). "A Variation of Ranked Set Sampling." *Journal of Applied Statistical Science*, **10**(2), 137–146. URL https://www.researchgate.net/publication/267662221_A_variation_of_ranked_set_sampling.

Al-Omari AI, Haq A (2019). "A New Sampling Method for Estimating the Population Mean." *Journal of Statistical Computation and Simulation*, **89**(11), 1973–1985. doi:10.1080/00949655.2019.1604710.

Alam R, Hanif M, Shahbaz SH, Shahbaz MQ (2022). "Estimation of Population Variance under Ranked Set Sampling Method by Using the Ratio of Supplementary Information with Study Variable." *Scientific Reports*, **12**(1), 21203. doi:10.1038/s41598-022-24296-1.

Aldrabseh MZ, Ismail MT (2023). "New Modification of Ranked Set Sampling for Estimating Population Mean." *Journal of Statistical Computation and Simulation*, **93**(16), 2843–2855. doi:10.1080/00949655.2023.2212312.

Balci S, Akkaya AD, Ulgen BE (2013). "Modified Maximum Likelihood Estimators Using Ranked Set Sampling." *Journal of Computational and Applied Mathematics*, **238**, 171–179. doi:10.1016/j.cam.2012.08.030.

Chen M, Lim J (2011). "Estimating Variances of Strata in Ranked Set Sampling." *Journal of Statistical Planning and Inference*, **141**(8), 2513–2518. doi:10.1016/j.jspi.2010.11.043.

David HA, Nagaraja HN (2004). *Order Statistics*. John Wiley & Sons. doi:10.1002/0471722162.

Dell TR, Clutter JL (1972). "Ranked Set Sampling Theory with Order Statistics Background." *Biometrics*, pp. 545–555. doi:10.2307/2556166.

Division WBDUNP (2022a). "Birth Rate, Crude (Per 1,000 People) - Jordan." https://data.worldbank.org/indicator/SP.DYN.CBRT.IN?locations=JO.

Division WBDUNP (2022b). "Population Ages 0-14 (Percent of Total Population)-Jordan." https://data.worldbank.org/indicator/SP.POP.0014.TO.ZS?locations=JO.

Hanandeh AA, Al-Nasser AD (2021). "Modified Minimax Ranked Set Sampling." *Pakistan Journal of Statistics*, **37**(2). URL https://www.researchgate.net/publication/350189391_MODIFIED_MINIMAX_RANKED_SET_SAMPLING.

Haq A, Brown J, Moltchanova E, Al-Omari AI (2014). "Mixed Ranked Set Sampling Design." *Journal of Applied Statistics*, **41**(10), 2141–2156. doi:10.1080/02664763.2014.909781.

Jemain AA, Al-Omari AI, Ibrahim K (2007). "Multistage Median Ranked Set Sampling for Estimating the Population Median." *Journal of Mathematics and Statistics*, **3**(2), 58–64. URL http://localhost:8080/xmlui/handle/123456789/8115.

MacEachern SN, Öztürk Ö, Wolfe DA, Stark GV (2002). "A New Ranked Set Sample Estimator of Variance." *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **64**(2), 177–188. doi:10.1111/1467-9868.00331.

Mahdizadeh M, Zamanzade E (2021). "New Estimator for the Variances of Strata in Ranked Set Sampling." *Soft Computing*, **25**(13), 8007–8013. doi:10.1007/s00500-021-05787-1.

McIntyre GA (1952). "A Method for Unbiased Selective Sampling, Using Ranked Sets." *Australian Journal of Agricultural Research*, **3**(4), 385–390. doi:10.1071/AR9520385.

Muttlak HA (1997). "Median Ranked Set Sampling." *Journal of Applied Statistical Science*, **6**, 245–255. URL https://cir.nii.ac.jp/crid/1573105975290057472.

Ozturk O, Bayramoglu Kavlak K (2020). "Statistical Inference Using Stratified Judgment Post-Stratified Samples from Finite Populations." *Environmental and Ecological Statistics*, **27**(1), 73–94. doi:10.1007/s10651-019-00435-2.

Stokes L (1995). "Parametric Ranked Set Sampling." *Annals of the Institute of Statistical Mathematics*, **47**(3), 465–482. doi:10.1007/BF00773396.

Stokes SL (1980). "Estimation of Variance Using Judgment Ordered Ranked Set Samples." *Biometrics*, pp. 35–42. doi:10.2307/2530493.

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), 1–31. doi:10.1007/BF02911622.

Yu PLH, Lam K, Sinha BMK (1999). "Estimation of Normal Variance Based on Balanced and Unbalanced Ranked Set Samples." *Environmental and Ecological Statistics*, **6**, 23–46. doi:10.1023/A:1009687332668.

Zamanzade E, Al-Omari AI (2016). "New Ranked Set Sampling for Estimating the Population Mean and Variance." *Hacettepe Journal of Mathematics and Statistics*, **45**(6), 1891–1905. doi:10.15672/HJMS.20159213166.

Zamanzade E, Vock M (2015). "Variance Estimation in Ranked Set Sampling Using a Concomitant Variable." *Statistics & Probability Letters*, **105**, 1–5. doi:10.1016/j.spl.2015.04.034.

Affiliation:

Mahmoud Zuhier Aldrabseh
School of Mathematical Science
Universiti Sains Malaysia
11800 Penang - Malaysia
E-mail: mah.darabseh@gmail.com

Mohd Tahir Ismail
School of Mathematical Science
Universiti Sains Malaysia
11800 Penang - Malaysia
E-mail: m.tahir@usm.my

Amer Ibrahim Al-Omari
Department of Mathematics
Al al-Bayt University
Mafraq - Jordan
E-mail: alomari_amer@yahoo.com

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 53 *Submitted:* 2024-01-26
2024 *Accepted:* 2024-05-14