*Karadeniz Fen Bilimleri Dergisi*, 15(1), 303-318, 2025. DOI: 10.31466/kfbd.1547405

KFBD

**Karadeniz Fen Bilimleri Dergisi**
The Black Sea Journal of Sciences
ISSN (Online): 2564-7377 https://dergipark.org.tr/tr/pub/kfbd

Araştırma Makalesi / Research Article

# RSStest: An R Package for Hypothesis Test for Two Population Means with Ranked Set Sampling and Median Ranked Set Sampling

Fikri GÖKPINAR[1], Yaprak Arzu ÖZDEMİR[2*], Meral EBEGİL[3], Esra GÖKPINAR[4]

## Abstract

The hypothesis test for the equality of two population means of independent groups is the most commonly used statistical analysis with applications in medical sciences, agriculture, biology, engineering, social sciences and so on. This paper presents the RSStest package created to perform hypothesis testing for equality of two population means using ranked set sampling and median ranked set sampling. The package gives the test results such as test statistics, $p$ value, critical values and basic comments. In addition, using this package, it is possible to select a sample from ranked set sampling and median ranked set sampling by ranking according to auxiliary variable from a given finite population. In addition, to demonstrate the use of these tests, a real data application on Copoeta trutta fish from Elazığ Keban Dam Lake was performed.

**Keywords:** Hypothesis test, Ranked set sampling, Median ranked set sampling, R-Package.

# RSStest: Sıralı Küme Örneklemesi ve Medyan Sıralı Küme Örneklemesi ile İki Yığın Hipotez Testi için R Paketi

## Öz

İki bağımsız grubun yığın ortalamalarının eşitliği için hipotez testi, tıp bilimleri, tarım, biyoloji, mühendislik, sosyal bilimler gibi birçok alanda yaygın olarak kullanılan bir istatistiksel analizdir. Bu makalede, sıralı küme örneklemesi ve medyan sıralı küme örneklemesi kullanılarak iki yığın ortalamasının eşitliği hipotez testini gerçekleştirmek için oluşturulan RSStest paketi sunulmuştur. Paket, test istatistikleri, $p$ değeri, kritik değerler ve temel yorumlar gibi test sonuçlarını vermektedir. Ayrıca, bu paket kullanılarak, verilen sonlu bir yığından yardımcı değişkene göre sıralama yapılarak, sıralı küme örneklemesi ve medyan sıralı küme örneklemesi ile örnek seçmek de mümkündür. Ek olarak, bu testlerin kullanımını göstermek amacıyla Elazığ Keban Baraj Gölü Copoeta trutta türü balıklara ilişkin gerçek veri uygulaması yapılmıştır.

**Anahtar Kelimeler:** Hipotez testi, Sıralı küme örneklemesi, Medyan sıralı küme örneklemesi, R-Paketi.

[1,2,3,4] Gazi University, Faculty of Science, Department of Statistics, Ankara, Turkey,
fikri@gazi.edu.tr yaprak@gazi.edu.tr mdemirel@gazi.edu.tr eyigit@gazi.edu.tr

[*] Sorumlu Yazar/Corresponding Author

**Geliş/Received:** 10.09.2024 **Kabul/Accepted:** 03.02.2025 **Yayın/Published:** 15.03.2025

Gökpınar, F., Özdemir, Y.A., Ebegil, M., Gökpınar, E. (2025). RSStest: An R Package for Hypothesis Test for Two Population Means with Ranked Set Sampling and Median Ranked Set Sampling. *Karadeniz Fen Bilimleri Dergisi*, 15(1), 303-318.

## 1. Introduction

Ranked Set Sampling (RSS) is a sampling method introduced by McIntyre in 1952 and applied in various fields such as health, environment, ecology and agriculture. In these fields, measuring sampling units for the variable of interest can often be difficult or costly. RSS is particularly effective in such scenarios when compared to Simple Random Sampling (SRS) for estimating population parameters. The mathematical theory underlying RSS was first proposed by Takahasi and Wakimoto in 1968. The errors in ranking were considered by Dell and Clutter (1972). Many modifications of RSS have been proposed to increase the efficiency of parameter estimation. One of these is Median Ranked Set Sampling (MRSS). MRSS was suggested by Muttlak (1997) to reduce the errors in ranking, and to increase the efficiency of estimator of the population mean for unimodal symmetrical distributions.

In statistical inference, such as hypothesis testing and confidence intervals for population parameters, RSS and MRSS can be employed to achieve more efficient estimations. The use of RSS in hypothesis testing was first explored by Muttlak and Abu-Dayyeh (1998) and Shen (1994). Abu-Dayyeh and Muttlak (1996) investigated hypothesis tests on the scale parameter of exponential and uniform distributions using ranked set samples. Tseng and Wu (2007) developed a formula for critical values based on relative efficiency values for the means of normal and exponential distributions in RSS and MRSS. Ozturk and Balakrishnan (2009) proposed an exact two-sample nonparametric test procedure for detecting quantile shifts between two populations using RSS. Özdemir et al. (2017) studied hypothesis testing for the difference in means between two populations under RSS for normal distributions with unknown variances, comparing Type I errors and test powers for RSS and SRS, and also evaluated RSS performance in cases of imperfect ranking. In 2019, Özdemir et al. proposed two different MRSS-based approaches for testing the equality of means between two normal populations, deriving critical values under the assumption of homogeneous variances using Monte Carlo simulation, and also presented an alternative method based on t-distribution approximation for cases with heterogeneous variances. The first R package for RSS and its modified versions was proposed by Sevinç et al. (2019). This package also gives applications of some nonparametric methods. The package is available as RSSampling on The Comprehensive R Archive Network (CRAN).

In this study, we present the RSStest package (Gokpinar et al. (2023)) to investigate the difference between two independent ranked set samples and median ranked set samples using the test statistics proposed by Özdemir et al. (2017) and Özdemir et al. (2019). The package gives the test results such as test statistics, $p$ value and critical values for all set and cycle sizes under perfect and imperfect ranking cases. Also, we give a real data example for the usage of these tests.

## 2. Hypothesis Tests for Two Independent Population Means Using RSS and MRSS

In this section, the statistical tests that are used to test the equality of two independent population means are explained. These tests were proposed for RSS by Özdemir et al. (2017) and for MRSS by Özdemir et al. (2019). In these studies, test statistics and critical values are obtained and it was observed that these tests are more powerful than the tests those of SRS for all considered cases. They also propose two distinct approaches based on MRSS for testing the equality of means between two normal populations. Given the theoretical distribution of any statistic based on MRSS is highly complex, critical values for the test statistics were determined using Monte Carlo simulations. The remainder of this section provides details on tests for comparing two population means using both RSS and MRSS.

### 2.1. Tests for two independent population means in RSS

In this section, we give the testing procedure for the equality of two normal means by using RSS. First, we should give the sample selection procedure in RSS. In RSS, the sample selection procedure for one group is given as follows:

1. $m$ simple random samples, each of size $m$, are selected from the population, and each sample is referred to as a set.
2. Within each set, observations are ranked from smallest to largest according to the variable of interest, denoted as $X$. Ranking can be performed using a concomitant variable or through visual ranking, which minimizes measurement costs.
3. The first unit from the first set, the second unit from the second set, and so on up to the $m$-th unit from the $m$-th set are then measured according to variable $X$.
4. This process can be repeated $r$ times (cycle size) to obtain a ranked set sample with a total size of $n = mr$.

This selection procedure can be applied for each group separately.

In biological studies, determining fish age using otolith bone is both time consuming and costly. In addition, age determination by otolith bone causes the death of the fish. Therefore, RSS can be applied using other variables that are highly correlated with fish age, such as length or weight of the fish (Wang et al., 2009; Hatefi et al., 2015). For example, two fish species in a lake are required to be compare in terms of average ages. To obtain the exact age of a fish, otolith bone needs to be extracted from the fish in the laboratories and this results in death of the fish. Alternatively, the length of a fish can be used to determine the approximate fish age. Assume that, we want to take a ranked set sample of fishes with sample size 3 for each species. For this reason, first we should select 9 fish randomly for each species. Let assume that the fishes are given in Figure 1.

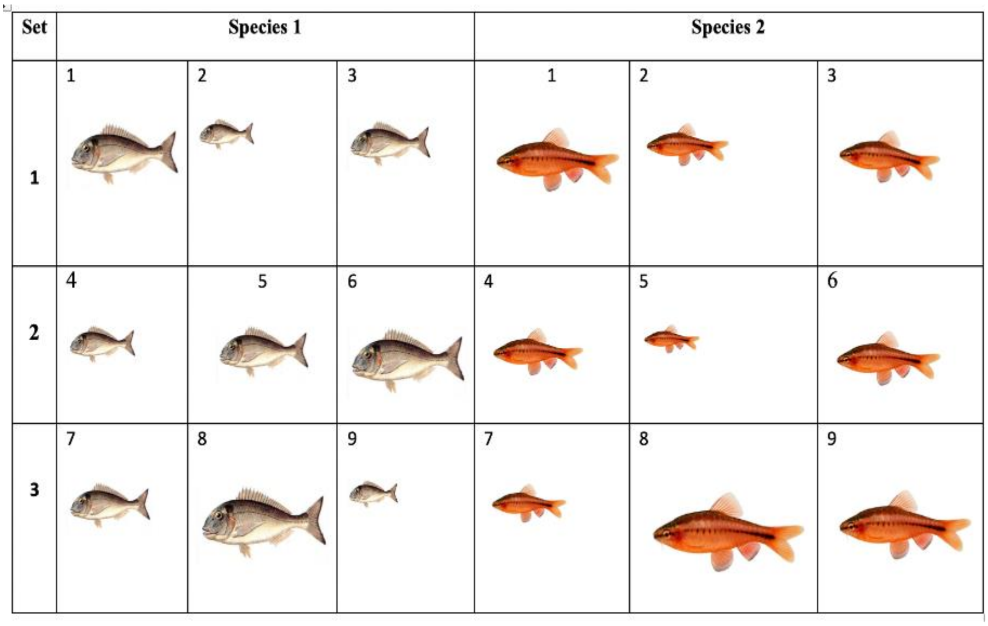

**Figure 1**. Randomly selected fishes for species 1 and species 2.

In Figure 2, ranked and selected fishes in ascending order in each set according to their lengths.

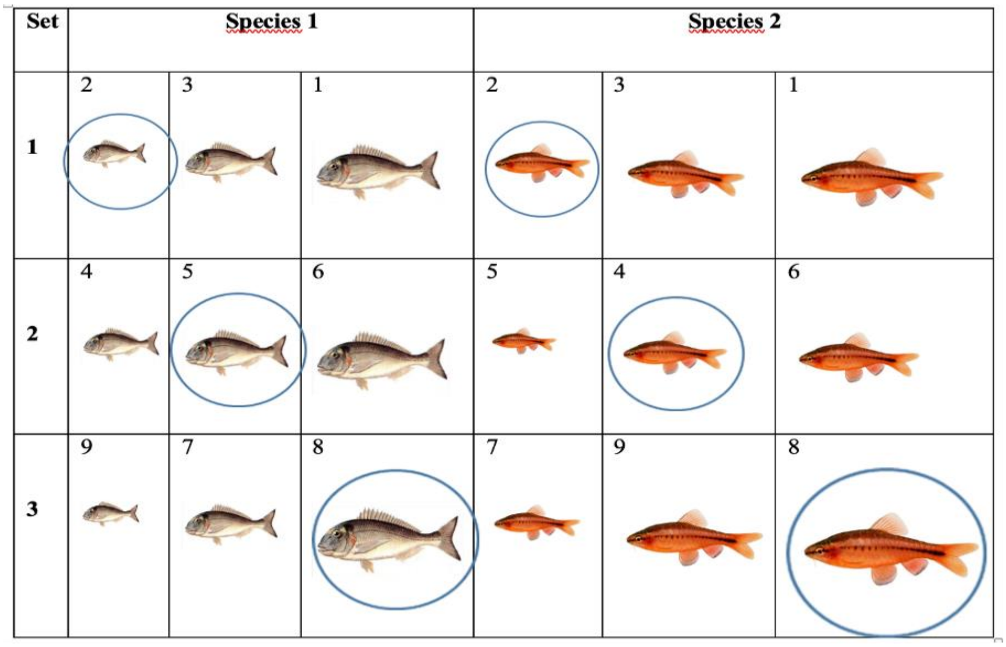

**Figure 2**. Sorted fishes in ascending order for species 1 and 2.

As a result, the fishes numbered as 2, 5 and 8 consist of the ranked sample for species 1. Also, the fishes numbered as 2, 4 and 8 consist of the ranked sample for species 2. These samples are given in Figure 3.

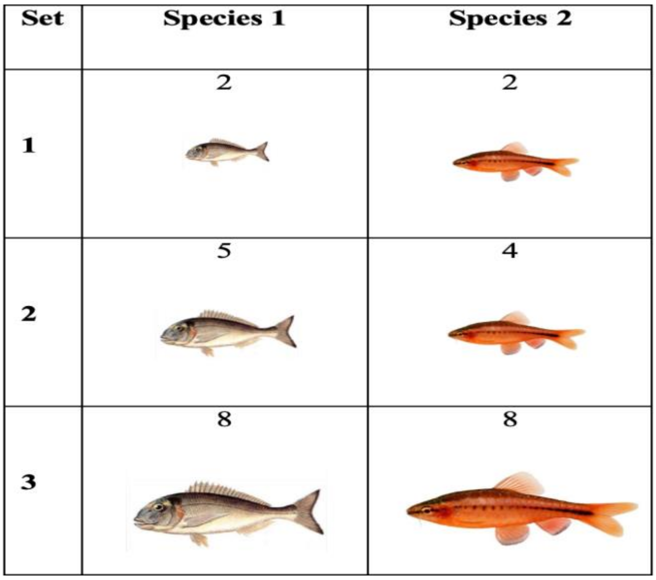

**Figure 3.** Final ranked set samples with size 3 for each species.

Using the selection procedure given above the data for the first and second group is shown as follows, respectively;

$$\begin{bmatrix} X_{111}^{(1)} & X_{112}^{(1)} & \cdots & X_{11r_1}^{(1)} \\ X_{121}^{(2)} & X_{122}^{(2)} & & X_{12r_1}^{(2)} \\ \vdots & & \ddots & \vdots \\ X_{1m_11}^{(m_1)} & X_{1m_12}^{(m_1)} & \cdots & X_{1m_1r_1}^{(m_1)} \end{bmatrix}$$

and

$$\begin{bmatrix} X_{211}^{(1)} & X_{212}^{(1)} & \cdots & X_{11r_2}^{(1)} \\ X_{221}^{(2)} & X_{222}^{(2)} & & X_{22r_2}^{(2)} \\ \vdots & & \ddots & \vdots \\ X_{2m_21}^{(m_2)} & X_{2m_22}^{(m_2)} & \cdots & X_{2m_2r_2}^{(m_2)} \end{bmatrix}$$

where $X_{kij}^{(s)}$ is the $s$th order statistic in the $i$th set and $j$th cycle for the $k$th population $(s = 1,2,..,m_k, i = 1,2,\ldots,m_k, j = 1,2,\ldots,r_k, k = 1,2)$, $m_k$ and $r_k$ are the set size and the cycle size of the $k$th population, respectively.

Hypothesis for the means of two normal populations with unknown variances for one tailed and two-tailed hypotheses can be given as follows, respectively:

$$H_0: \mu_1 = \mu_2 \ or \ \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 > \mu_2 \ or \ \mu_1 - \mu_2 > 0 \tag{1}$$

$$H_0: \mu_1 = \mu_2 \ or \ \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 < \mu_2 \ or \ \mu_1 - \mu_2 < 0 \qquad (2)$$

$$H_0: \mu_1 = \mu_2 \ or \ \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 \neq \mu_2 \ or \ \mu_1 - \mu_2 \neq 0 \qquad (3)$$

The algorithm of testing hypotheses procedure for RSS is given as follows:

1. Two ranked set samples of sizes $n_1 = m_1 r_1$ and $n_2 = m_2 r_2$ are used to test the $H_0$ hypothesis in Eq. (1-3).
2. An appropriate estimator for $\mu_1 - \mu_2$ is calculated as $\bar{X}_{1(RSS)} - \bar{X}_{2(RSS)}$, where $\bar{X}_{k(RSS)} = \sum_{j=1}^{r_k}\sum_{i=1}^{m_k} X_{kij}^{(i)}/n_k$ is the ranked set sample mean for the *kth* population mean $\mu_k$ (k=1, 2).
3. The estimator of $V\left(\bar{X}_{1(RSS)} - \bar{X}_{2(RSS)}\right)$ is calculated as given below:

$$\hat{V}\left(\bar{X}_{1(RSS)} - \bar{X}_{2(RSS)}\right) = \frac{\sum_{i=1}^{m_1}\sigma_{Z_{1(i)}}^2}{m_1^2 r_1}\hat{\sigma}_{1(RSS)}^2 + \frac{\sum_{i=1}^{m_1}\sigma_{Z_{2(i)}}^2}{m_2^2 r_2}\hat{\sigma}_{2(RSS)}^2 \qquad (4)$$

where $\sigma_{Z_{k(i)}}^2$ is the variance of the *ith* order statistic of the standard normal distribution of size $m_k$ for the *kth* group $(k = 1, 2)$ (Arnold et al. 2008, Özdemir et al., 2017). Also, $\hat{\sigma}_{k(RSS)}^2$ is the estimator of $\hat{\sigma}_k^2$ $(k = 1, 2)$ which is the *kth* population variance.

4. The unbiased estimator of the population variance $\sigma_k^2$ $(k = 1, 2)$ under RSS can be written as follows (MacEachern et al., 2002):

$$\hat{\sigma}_{k(RSS)}^2 = \frac{1}{m_k r_k}\{(m-1)MST_k + (m_k r_k - m_k + 1)MSE_k\} \quad k = 1,2 \qquad (5)$$

where *MSE<sub>k</sub>* and *MST<sub>k</sub>* are the mean-square error and mean-square treatment, respectively, and they can be defined as follows:

$$MST_k = \frac{1}{m_k - 1}\sum_{i=1}^{m_k}\sum_{j=1}^{r_k}\left(X_{kij}^{(i)} - \bar{X}_{k(RSS)}\right)^2 - \frac{1}{m_k - 1}\sum_{i=1}^{m_k}\sum_{j=1}^{r_k}\left(X_{kij}^{(i)} - \bar{X}_{ki.}\right)^2 \qquad (6)$$

$$MSE_k = \frac{1}{m_k(r_k - 1)}\sum_{i=1}^{m_k}\sum_{j=1}^{r_k}\left(X_{kij}^{(i)} - \bar{X}_{ki.}\right)^2 \qquad (7)$$

where $\bar{X}_{ki.} = \sum_{j=1}^{r_k} X_{kij}^{(i)}/r_k \ \ (k = 1,2)$.

5. The test statistic for testing $H_0$ against $H_1$ in Eq. (1) under unknown variances can be defined as follows:

$$T_{RSS} = \frac{\bar{X}_{1(RSS)} - \bar{X}_{2(RSS)}}{\sqrt{\hat{V}\left(\bar{X}_{1(RSS)} - \bar{X}_{2(RSS)}\right)}} \tag{8}$$

If $T_{RSS} > T_\alpha$, ($T_\alpha$ is the critical value at significance level $\alpha$), $H_0$ is rejected for $H_1: \mu_1 > \mu_2$. For $H_1: \mu_1 < \mu_2$ if $T_{RSS} < -T_\alpha$, $H_0$ is rejected. In addition, for $H_1: \mu_1 \neq \mu_2$, if $T_{RSS} < -T_{\alpha/2}$ or $T_{RSS} > T_{\alpha/2}$, $H_0$ is rejected. We calculate the critical values $T_\alpha$ using R.

### 2.2. Test for Two Independent Group Means in MRSS

Median Ranked Set Sampling (MRSS) was proposed by Muttlak (1997) to reduce ranking errors and increase the efficiency of estimators for unimodal symmetrical distributions. The sample selection procedure for one group in MRSS is as follows:

1. Steps 1 and 2 of the sample selection procedure described for RSS in section 2.1 is applied.
2. When the cluster size m is odd, the median unit from each cluster is selected for measurement. When the set size $m$ is even, the $(m/2)$th smallest unit from each of the first $(m/2)$ sets and the $(m/2 + 1)$th smallest unit from the each of the next $(m/2)$ sets are selected for measurement.
3. This selection steps can be repeated r times to obtain a median ranked set sample of size $n = mr$.

This selection procedure can be applying for each group separately. Using the selection procedure given above, the hypothesis for two normal population means with unknown variances for one tailed and two tailed hypotheses given in Eq.(1-3) are used respectively. The algorithm of testing hypotheses procedure for MRSS is given as follows:

1. Two median ranked set samples of sizes $n_1 = m_1 r_1$ and $n_2 = m_2 r_2$ are used to test the $H_0$ hypothesis in Eq. (1-3). According to whether $m_k$ is odd or even, median ranked set samples are selected from the groups as follows;

$$\begin{bmatrix} X_{k11}^{\left(\frac{m_k+1}{2}\right)} & X_{k12}^{\left(\frac{m_k+1}{2}\right)} & \cdots & X_{k1r_k}^{\left(\frac{m_k+1}{2}\right)} \\ X_{k21}^{\left(\frac{m_k+1}{2}\right)} & X_{k22}^{\left(\frac{m_k+1}{2}\right)} & & X_{k2r_k}^{\left(\frac{m_k+1}{2}\right)} \\ \vdots & & \ddots & \vdots \\ X_{km_k1}^{\left(\frac{m_k+1}{2}\right)} & X_{km_k2}^{\left(\frac{m_k+1}{2}\right)} & \cdots & X_{km_kr_k}^{\left(\frac{m_k+1}{2}\right)} \end{bmatrix}, m_k\ is\ odd$$

$$\begin{bmatrix} X_{k11}^{\left(\frac{m_k}{2}\right)} & X_{k12}^{\left(\frac{m_k}{2}\right)} & \cdots & X_{k1r_k}^{\left(\frac{m_k}{2}\right)} \\ \vdots & \vdots & \ddots & \vdots \\ X_{k(m_k/2)1}^{\left(\frac{m_k}{2}\right)} & X_{k(m_k/2)2}^{\left(\frac{m_k}{2}\right)} & \cdots & X_{k(m_k/2)r_k}^{\left(\frac{m_k}{2}\right)} \\ X_{k((m_k+2)/2)1}^{\left(\frac{m_k+2}{2}\right)} & X_{k((m_k+2)/2)1}^{\left(\frac{m_k+2}{2}\right)} & \cdots & X_{k((m_k+2)/2)r_k}^{\left(\frac{m_k+2}{2}\right)} \\ \vdots & \vdots & \ddots & \vdots \\ X_{km_k1}^{\left(\frac{m_k+2}{2}\right)} & X_{km_k2}^{\left(\frac{m_k+2}{2}\right)} & \cdots & X_{km_kr_k}^{\left(\frac{m_k+2}{2}\right)} \end{bmatrix}, m_k\ is\ even$$

where $X_{kij}^{(s)}$ is the $s$th order statistic in the $i$th set and $j$th cycle for the $k$th population $(s = 1,2,..,m_k, i = 1,2,\dots,m_k, j = 1,2,\dots,r_k, k = 1,2)$.

2. An appropriate estimator for $\mu_1 - \mu_2$ can be given as $\bar{X}_{1(MRSS)} - \bar{X}_{2(MRSS)}$, where

$$\bar{X}_{k(MRSS)} = \begin{cases} \dfrac{1}{n_k}\left[\displaystyle\sum_{i=1}^{m_k}\sum_{j=1}^{r_k} X_{kij}^{\left(\frac{m_k+1}{2}\right)}\right], & if\ m_k\ is\ odd \\ \dfrac{1}{n_k}\left[\displaystyle\sum_{i=1}^{m_k/2}\sum_{j=1}^{r_k} X_{kij}^{\left(\frac{m_k}{2}\right)} + \sum_{i=\left(\frac{m_k+2}{2}\right)}^{m_k}\sum_{j=1}^{r_k} X_{kij}^{\left(\frac{m_k+2}{2}\right)}\right], & if\ m_k\ is\ even \end{cases}$$

is the $k$th median ranked set sample mean. The estimation of $V\left(\bar{X}_{1(MRSS)} - \bar{X}_{2(MRSS)}\right)$ can be determined as follows;

$$\hat{V}\left(\bar{X}_{1(MRSS)} - \bar{X}_{2(MRSS)}\right) = \frac{\hat{\sigma}_1^2}{m_1 r_1} + \frac{\hat{\sigma}_2^2}{m_2 r_2}$$

Here;

$$\hat{\sigma}_k^2 = \begin{cases} \hat{\sigma}^2_{k\left(\frac{m_k+1}{2}\right)} & if\ m_k\ is\ odd \\ \frac{\hat{\sigma}^2_{k\left(\frac{m_k}{2}\right)}+\hat{\sigma}^2_{k\left(\frac{m_k+2}{2}\right)}}{2} & if\ m_k\ is\ even \end{cases} ,k=1,2 \tag{9}$$

where $\hat{\sigma}^2_{k\left(\frac{m_k}{2}\right)}$, $\hat{\sigma}^2_{k\left(\frac{m_k+1}{2}\right)}$, $\hat{\sigma}^2_{k\left(\frac{m_k+2}{2}\right)}$ are the estimation of variances for the $(m_k/2)th, (m_{k+1}/2)th, (m_{k+2}/2)th$ order statistics of the $kth$ population $(k=1,2)$ with the sample size $m_k$. $\hat{\sigma}^2_{k\left(\frac{m_k}{2}\right)}$, $\hat{\sigma}^2_{k\left(\frac{m_k+1}{2}\right)}$, $\hat{\sigma}^2_{k\left(\frac{m_k+2}{2}\right)}$ are defined as follows:

$$\hat{\sigma}^2_{k\left(\frac{m_k+1}{2}\right)} = \sum_{i=1}^{r_k}\sum_{j=1}^{m_k}\frac{\left(X_{kij}^{\left(\frac{m_k+1}{2}\right)} - \bar{X}_{k\left(\frac{m_k+1}{2}\right)}\right)^2}{m_k r_k - 1} \quad where \quad \bar{X}_{k\left(\frac{m_k+1}{2}\right)} = \sum_{i=1}^{r_k}\sum_{j=1}^{m_k}\frac{X_{kij}^{\left(\frac{m_k+1}{2}\right)}}{m_k r_k},$$

$$\hat{\sigma}^2_{k\left(\frac{m_k}{2}\right)} = \sum_{i=1}^{r_k}\sum_{j=1}^{m_k}\frac{\left(X_{kij}^{\left(\frac{m_k}{2}\right)} - \bar{X}_{k\left(\frac{m_k}{2}\right)}\right)^2}{(m_k r_k/2) - 1} \quad where\ \bar{X}_{k\left(\frac{m_k}{2}\right)} = \sum_{i=1}^{r_k}\sum_{j=1}^{m_{k/2}}\frac{X_{kij}^{\left(\frac{m_k}{2}\right)}}{m_k r_k/2}$$

and

$$\hat{\sigma}^2_{k\left(\frac{m_k+2}{2}\right)} = \sum_{i=1}^{r_k}\sum_{j=1}^{m_k}\frac{\left(X_{kij}^{\left(\frac{m_k+2}{2}\right)} - \bar{X}_{k\left(\frac{m_k+2}{2}\right)}\right)^2}{(m_k r_k/2) - 1}\ where\ \bar{X}_{k\left(\frac{m_k+2}{2}\right)} = \sum_{i=1}^{r_k}\sum_{j=m_{k/2}+1}^{m_k}\frac{X_{kij}^{\left(\frac{m_k+2}{2}\right)}}{m_k r_k/2}$$

3. When the variances are unknown, the test statistic can be defined as follows:

$$T_{MRSS} = \frac{\bar{X}_{1(MRSS)} - \bar{X}_{2(MRSS)}}{\sqrt{\hat{V}\left(\bar{X}_{1(MRSS)} - \bar{X}_{2(MRSS)}\right)}}, \tag{10}$$

4. Median ranked set samples are generated under null hypothesis $M$ times. T<sub></sub>MRSS values are calculated for each generated samples using Eq.(10) (e.g. $T^1_{MRSS}, T^2_{MRSS}, \ldots, T^M_{MRSS}$) . $p$-value for $H_1 : \mu_1 > \mu_2$ is calculated as follows:

$$p = \frac{1}{M}\sum_{i=1}^{M} I\left(T^i_{MRSS} > T_{MRSS}\right)$$

where $I(.)$ is the indicator function. For $H_1 : \mu_1 < \mu_2$, $r, p$ value can be obtained as

$$p = \frac{1}{M}\sum_{i=1}^{M} I\left(T_{MRSS}^{i} < T_{MRSS}\right)$$

and for $H_1 : \mu_1 \neq \mu_2$ can be obtained as

$$p = 2 * min\left(\frac{1}{M}\sum_{i=1}^{M} I\left(T_{MRSS}^{i} > T_{MRSS}\right), \frac{1}{M}\sum_{i=1}^{M} I\left(T_{MRSS}^{i} < T_{MRSS}\right)\right)$$

If $p < \alpha$ $H_0$ is rejected.

## 3. Using RSStest package

The RSStest package is available on CRAN and can be loaded via the following commands:

```
# Call the RSStest package
library(RSStest)
# Call the huxtable package
library(huxtable)
```

The package consists of two test procedures for two independent sample cases which are described the functions for RSS and MRSS. The functions for test procedures provide hypothesis testing procedures based on RSS and MRSS. The functions that are included in the packages are given in Table 1.

**Table 1.** The functions for the test procedure for two groups in **RSStest** package

| Function | Description |
|---|---|
| datagen_MRSS | This function generates random samples from normal population using MRSS with mean μ and standard deviation $\sigma$ using cycle size $r$ and set size $m$. |
| datagen_RSS | This function generates random samples from normal population using RSS with mean $\mu$ and standard deviation $\sigma$ using cycle size $r$ and set size $m$. |
| teststat_RSS | This function tests for the difference of two population means using RSS given in Özdemir et al. (2017). |
| teststat_MRSS | This function tests for the difference of two population means using median ranked set sampling given in Özdemir et al. (2019). |
| imperfectRSS | This function chooses ranked set samples from specific finite population using auxiliary variable with cycle sizes $r_1$ and $r_2$ and set sizes $m_1$ and $m_2$. |
| imperfectMRSS | This function chooses median ranked set samples from specific finite population using auxiliary variable with cycle sizes $r_1$ and $r_2$ and set sizes $m_1$ and $m_2$. |

datagen_RSS function gives a normal ranked set sample with set size $m$ and cycle size $r$ with mean $\mu$ (mu) and standard deviation $\sigma$ (s).

```
datagen_RSS(mu=0,s=1,m=2,r=3)
#>                [,1]      [,2]
#> Cycle 1 -2.3476834 0.6352628
#> Cycle 2 -0.9810088 2.0333810
#> Cycle 3 -0.7962410 0.9587357
```

datagen_MRSS function gives a normal median ranked set sample with set size $m$ and cycle size $r$ with mean $\mu$ (mu) and standard deviation $\sigma$ (s).

```
datagen_MRSS(mu=0,s=1,m=2,r=3)
#>                [,1]       [,2]
#> Cycle 1 -1.8112827 -0.2220520
#> Cycle 2  0.1339985  1.9532652
#> Cycle 3  0.1156633  0.2734261
```

imperfectRSS function gives a ranked set sample from a finite population for the imperfect ranking case. The df is the dataframe of interest, the cat is the grouping variable, catname is the categories of the grouping variables, aux is the auxiliary variable, the var is variable of interest, r1 and r2 are the cycle sizes and m1 and m2 are the set sizes.

```
imperfectRSS(df=otolith,cat="sex",catname=c("F","M"),aux="fish.length",
             var="age",r1=3,r2=3,m1=4,m2=3)
#> [[1]]
#>      [,1] [,2] [,3] [,4]
#> [1,]    5    2    5    6
#> [2,]    5    2    4    5
#> [3,]    3    5    5    4
#>
#> [[2]]
#>      [,1] [,2] [,3]
#> [1,]    3    4    5
#> [2,]    2    4    4
#> [3,]    5    3    3
```

imperfectMRSS function gives a median ranked set sample from a finite population for imperfect ranking case. The df is the dataframe of interest, cat is the grouping variable, catname is the categories of the grouping variables, aux is the auxilary variable, the var is variable of interest, r1 and r2 are the cycle sizes and m1 and m2 are the set sizes.

```
imperfectMRSS(df=otolith,cat="sex",catname=c("F","M"),aux="fish.length",
              var="age",r1=3,r2=3,m1=4,m2=3)
#> [[1]]
#>      [,1] [,2] [,3] [,4]
#> [1,]    3    3    6    5
#> [2,]    2    2    4    4
#> [3,]    4    4    5    5
#>
#> [[2]]
#>      [,1] [,2] [,3]
#> [1,]    3    5    4
#> [2,]    4    2    5
#> [3,]    5    4    3
```

$X_1$ and $X_2$ are ranked set samples obtained from group 1 and group 2, respectively. The parameter, “alternative” is determined for the types of the different alternative hypothesis. Such as “left” for $\mu_1 < \mu_2$, “right” for $\mu_1 > \mu_2$ and “two-tailed” for $\mu_1 \neq \mu_2$. The default value of alternative is “two-tailed”. The “table” is Boolean parameter when TRUE option gives a complete table of the analysis, and FALSE option gives only test statistic and $r$ values ($r = 1$ indicates the null hypothesis is rejected, $r = 0$ indicates the null hypothesis is accepted) that indicates the test is rejected or not. The default value of the table is TRUE. Alpha is the level of significance of the test. The default value of alpha is 5%.

```
x1=matrix(c(1,2.3, 3.4,4.5,5.6,4 ),nrow=3)
x2=matrix(c(2,3.2, 4.2,6.5,4.6,6 ),nrow=3)
teststat_RSS(x1,x2,alternative="left",table=T)
#> ------------------------------------------------------------
#>  Group_No Set_Size C.Size S.Size S.Mean S.Var T.Stat C.Value
#> --------- -------- ------ ------ ------ ----- ------ --------
#>         1        2      3      6   3.47   2.4  -1.27   -1.78
#>
#>         2        2      3      6   4.42  2.56
#> ------------------------------------------------------------
teststat_RSS(x1,x2,alternative="left",table=F)
#> [1]  1.000000 -1.265721
```

In the output, group number, set size, cycle size, sample size, sample mean, sample variance, test statistic and critical value are given respectively. Since the test statistic (-1.27) is not smaller than the critical value (-1.78) population mean of group 1 is not significantly smaller than the mean of group 2 at significance level 5%.

$X_1$ and $X_2$ are median ranked set samples obtained from group 1 and group 2, respectively. The parameter, “alternative” is determined for the types of the different alternative hypothesis. Such as “left” for $\mu_1 < \mu_2$, “right” for $\mu_1 > \mu_2$ and “two-tailed” for $\mu_1 \neq \mu_2$. The default value of alternative is “two-tailed”. The “table” is boolean parameter when TRUE option gives a complete table of the analysis, and FALSE option gives only test statistic, $p$-value and $r$ values ($r = 1$ indicates the null hypothesis is rejected, $r = 0$ indicates the null hypothesis is accepted) that indicates the test is rejected or not. The default value of the table is TRUE. Alpha is the level of significance of the test. The default value of alpha is 5%. Also, since the method is based on bootstrap, the parameter tn is the number of Bootstrap samples. The default value of tn is 1000.

```
x1=matrix(c(1,2.3, 3.4,4.5,5.6,4 ),nrow=3)
x2=matrix(c(2,3.2, 4.2,6.5,4.6,6 ),nrow=3)
teststat_MRSS(x1,x2,table=T)
#> ---------------------------------------------------------------
#>  Group_No Set_Size C.Size S.Size S.Mean S.Var T.Stat p_value
#> --------- -------- ------ ------ ------ ----- ------ --------
#>         1        2      3      6   3.47  1.06  -1.59   0.153
#>
#>         2        2      3      6   4.42  1.09
#> ---------------------------------------------------------------
teststat_MRSS(x1,x2,table=F)
#> [1]  0.000000  0.161000 -1.587627
```

In the output, group number, set size, cycle size, sample size, sample mean, sample variance, test statistic and $p$ values are given respectively. Since the $p$ value is not smaller than the significance level 5%, population mean of group 1 is not significantly smaller than the mean of group 2.

## 4. Real Data Example

In this section, the problem of estimating fish age using RSS and MRSS methods is discussed. While fish age is accurately determined using otolith bones, this method is both time consuming and costly as it results in fish mortality. Therefore, alternative variables that are highly correlated with fish age, such as length or weight, can be used for ranking. In this study, the data of Doğan and Dursun (2017), who examined the relationship between total weight and fish age of male and female Capoeta trutta caught from Keban Dam Lake (Elazığ) between November 2011 and December 2012, were used. In addition to the fish age, fish length and fish weight, this data also contains the fish sex and length, width and weight of right otolith bone. In this section, we present examples illustrating the RSStest package.

```
data("otolith")
```

The purpose of this example is to compare the age of the male and female Capoeta trutta population of Keban Dam Lake. To determine the fish age, fish length is used as the auxiliary variable. In this case, the ranked set sample and median ranked set sample are selected using the length of fish. After that, age values of the final median ranked set sample are exactly measured according to the otolith bone. In the following outputs, test results of MRSS and RSS are given for two-tailed hypotheses for set sizes 3 and cycle size 4 for group 1 and cycle size 3 for group 2.

```
XRSS=imperfectRSS(otolith,"sex",c("F","M"),"fish.length","age",4,3,3,3)
x1RSS=XRSS[[1]]
x2RSS=XRSS[[2]]
teststat_RSS(x1RSS, x2RSS, alpha = 0.05, alternative = "two-tailed", table =
TRUE)
#> -------------------------------------------------------------
#>  Group_No Set_Size C.Size S.Size S.Mean S.Var T.Stat C.Value
#> --------- -------- ------ ------ ------ ----- ------ --------
#>         1        3      4      12   4.67   1.7   4.47   -2.08
#>
#>         2        3      3      9   3.11 0.815           2.08
#> -------------------------------------------------------------
XMRSS=imperfectMRSS(otolith,"sex",c("F","M"),"fish.length","age",3,3,4,2)
x1MRSS=XMRSS[[1]]
x2MRSS=XMRSS[[2]]
teststat_MRSS(x1MRSS, x2MRSS, alpha = 0.05, alternative = "two-tailed", table =
TRUE)
#> -------------------------------------------------------------
#>  Group_No Set_Size C.Size S.Size S.Mean S.Var T.Stat  p_value
#> --------- -------- ------ ------ ------ ----- ------ --------
#>         1        4      3      12   4.25   1.03   1.52   0.153
#>
#>         2        2      3      6    3.33   1.67
#> -------------------------------------------------------------
```

As seen from the outputs, using RSS and MRSS the age of male and female fish are not significantly different at the significance level 5%.

## 5. Conclusion

RSS and MRSS are effective sampling methods compared to SRS, especially in situations where the measurement of a unit is expensive. Especially ranking sampling units are cheap, these methods are much better than SRS. Comparing two independent group means is one of the most common methods in statistical theory. This procedure has various application areas such as medical sciences, agriculture, biology, engineering and social sciences, etc. In this paper, we present the RSStest package for researchers to investigate the difference between means of two independent groups. RSStest is the first package in R that gives hypothesis tests for the equality of means using RSS and MRSS. In this package, we give two test functions for this purpose using RSS and MRSS. These functions include sample statistics such as sample mean and variance, test statistics, critical value and $p$ value of the tests. We also created a set of functions to obtain a ranked set sample and median ranked set sample for the perfect and imperfect ranking case using an auxiliary variable. Furthermore, we create a pseudo-data generating functions for RSS and MRSS and we provide an original data set for the usage of these methods. In the future study, this package can be improved for equality of more than two means and some other useful modifications of RSS. For example, in the presence of outliers in the data, extreme RSS can be used to increase the power of the test.

## Authors' Contributions

All authors contributed equally to the study.

## Statement of Conflicts of Interest

There is no conflict of interest between the authors.

## Statement of Research and Publication Ethics

The author declares that this study complies with Research and Publication Ethics.

## References

Arnold, B. C., N. Balakrishnan, and H. N. Nagaraja. (2008). A first course in order statistics. SIAM.

Abu-Dayyeh W. and H. A. Muttlak. (1996). Using ranked set sampling for hypothesis tests on the scale parameter of the exponential and uniform distributions. NCBA&E.

Dell, T. and J. Clutter. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, 545–555.

Dogan Y. and ¸S. Dursun. (2017). Keban baraj gölü'nde ya¸sayan capoeta trutta (heckel, 1843)'da otolitbiyometrisi-balık boyu ilişkisi. *Fırat Üniversitesi Fen Bilimleri Dergisi,* 29(2):33–38.

Gokpinar, F., M. Ebegil, Y. Ozdemir, and E. Gokpinar. (2023). Package 'RSStest'. R Foundation for Statistical Computing, Vienna, Austria, URL https://cran.r project.org/web/packages/RSStest/index.html.

Hatefi, A., M. J. Jozani, and O. Ozturk. (2015). Mixture model analysis of partially rank-ordered set samples: age groups of fish from length-frequency data. *Scandinavian Journal of Statistics*, 42(3):848–871.

MacEachern, S. N., Ö. Öztürk, D. A. Wolfe, and G. V. Stark. (2002). A new ranked set sample estimator of variance. *Journal of the Royal Statistical Society: Series B* (Statistical Methodology), 64(2):177–188.

McIntyre G. (1952). A method for unbiased selective sampling, using ranked sets. *Australian journal of agricultural research*, 3(4):385–390.

Muttlak H. (1997). Median ranked set sampling. *Journal of Applied Statistics* , 6:245–255.

Muttlak, H. A. and W. Abu-Dayyeh. (1998). Testing some hypotheses about the normal distribution using ranked set sample: a more powerful test. *Journal of Information and Optimization Sciences*, 19(1):1–11.

Özdemir, Y. A., M. Ebegil, and F. Gökpinar. (2017). A test statistic based on ranked set sampling for two normal means. *Communications in Statistics: Simulation and Computation*, 46(10):8077–8085.

Özdemir, Y. A., M. Ebegil, and F. Gökpinar. (2019). A test statistic for two normal means with median ranked set sampling. *Iranian Journal of Science and Technology, Transactions A: Science* , 43(3):1109–1126.

Ozturk, O. and N. Balakrishnan. (2009). Exact two-sample nonparametric test for quantile difference between two populations based on ranked set samples. *Annals of the Institute of Statistical Mathematics*, 61(1): 235–249.

Sevinc, B., Cetintav, B., Esemen, M., and Gürler, S. (2019). RSSampling: A pioneering package for ranked set sampling. *R Journal*, *11*(1).

Shen, W.H. (1994). Use oe ranked set sampling for test of a normal mean. *Calcutta Statistical Association Bulletin* , 44(3-4):183–194.

Takahasi, K. and K. Wakimoto. (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.

Tseng, Y.L. and S.W. Wu. (2007). Ranked-set-sample-based tests for normal and exponential means. *Communications in Statistics: Simulation and Computation*, 36(4):761–782.

Wang, Y.G., Y. Ye, and D. A. Milton. (2009). Efficient designs for sampling and subsampling in fisheries research based on ranked sets. *ICES Journal of Marine Science*, 66(5):928–934.