STATISTICA, anno LXXVII, n. 3, 2017

RANKED SET TWO-SAMPLE PERMUTATION TEST

Monjed H. Samuh [1]
Applied Mathematics and Physics Department, Palestine Polytechnic University, Palestine

1. INTRODUCTION

Ranked set sampling (RSS), a data collection scheme, was first implemented by McIntyre (1952) (see also McIntyre, 2005), as a good competitor to simple random sampling (SRS) scheme to estimate the mean of Australian pasture yields in agricultural experimentation. It is obviously applicable to other situations as well, see for example Philip and Lam (1997), Murray *et al.* (2000), and Murff and Sager (2006). For discussions of some other situations where RSS found applications, see Patil (1995), Barnett and Moore (1997) and Chen *et al.* (2004).

RSS can be useful when measurements are expensive or difficult to make but units from the population can be easily ranked by visual inspection. "In McIntyre's case, measuring the plots of pasture yields requires mowing and weighing crop yields, which is time consuming. However, a small number of plots can be even though sufficiently well ranked by eye without measurement. McIntyre's goal was to develop a sampling technique to reduce the number of necessary measurements to be made, maintaining the unbiasedness of the SRS mean and reducing the variance of the mean estimator by incorporating the outside information provided by visual inspection. Therefore, since the ranking of the plots could be done very cheap, he developed a technique to implement this advantage" (Rey, 2004).

Suppose the unidimensional study variable X has a probability density function (pdf) f and cumulative distribution function (cdf) F. Let $E(X)=\mu$ be the expectation of X and $Var(X)=\sigma^2<\infty$ its variance. RSS methodology consists of the following stages.

$$\text{Select} \rightarrow \text{Rank} \rightarrow \text{Identify} \rightarrow \text{Repeat}.$$

1. **Select:** Select m sets randomly, of size m elements each, from the study population.

[1] Corresponding Author. E-mail: monjedsamuh@ppu.edu

2. **Rank:** Rank the elements of each set in Step 1 with regard to X, by any negligible cost method or visually with no actual measurements.

3. **Identify:** Identify the i^{th} minimum from the i^{th} set, $i = 1, 2, \ldots, m$, for measurement. The acquired sample is called a ranked set sample of set size m.

4. **Repeat:** Independently repeat Steps 1-3 h cycles, if needed, to acquire an RSS of size $h \times m = n$.

It is worth to note that although $h \times m^2$ elements are sampled in the first stage, only $h \times m$ of them are considered for measurement. In case of perfect ranking (no error was made in the ranking mechanism) the measured elements are called the *order statistics* and they are not ordered; we denote the i^{th} order statistic acquired in the j^{th} cycle by $Y_{ji} = X_{j(i)}$, $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, h$. This version of RSS is a balanced RSS, in the sense that in each cycle the number of order statistics is fixed. In case of imperfect ranking the measured elements are called the *judgment order statistics*. In this paper, whenever we reference RSS we mean the balanced perfect RSS.

McIntyre (1952) claimed that the sample mean of RSS is an unbiased estimator of the population mean μ regardless of errors in ranking judgment. The mathematical theory of RSS was settled by Takahasi and Wakimoto (1968). They reported that the sample mean based on RSS is an unbiased estimator of μ, and has smaller variance than the sample mean of SRS based on m elements. Dell and Clutter (1972) developed the theory of RSS in the presence of ranking errors (imperfect ranking). Distribution function estimation in the context of RSS was studied by Stokes and Sager (1988). They showed that the RSS empirical distribution function is an unbiased estimator of the distribution function and more efficient than that from a SRS. Stokes (1995) studied the advantages of RSS in estimating the population parameters when the underlying distribution is known. MacEachern *et al.* (2002) developed an unbiased estimator of the population variance based on a RSS.

In the context of testing hypothesis, Koti and Jogesh Babu (1996) derived the exact distribution of the sign test statistic based on RSS. They reported that the test is more powerful than the counterpart sign test statistic of SRS. Liangyong and Xiaofang (2010) used the sign test based on RSS for testing hypotheses about the quantiles of a population distribution. Bohn and Wolfe (1992, 1994) suggested the RSS analogue of the classical two-sample Wilcoxon test and studied its relative properties under perfect and imperfect judgement. Öztürk (1999) studied the effect of the RSS on two-sample sign test statistic. Öztürk and Wolfe (2000) presented an optimal RSS allocation scheme for a two-sample RSS median test. They derived the exact distribution of the two-sample median test statistic in the context of RSS and tabled it for some sample sizes. Samuh (2012) and Amro and Samuh (2017) investigated the two-sample permutation test with the context of RSS and multistage RSS. In this paper, a new testing procedure for the two-sample design within RSS is introduced and compared with the test statistics proposed by Samuh (2012).

The rest of the paper is structured as follows. The independent two-sample model is introduced in Section 2. The classical independent t-test is reviewed in Section 2.1. The two-sample permutation test is presented in Section 2.2. Permutation test for two-sample ranked set samples with three proposed test statistics is discussed in Section 3. Simulation study that shows the benefits of permutation test of the ranked set two-sample design is provided in Section 4. Finally, concluding remarks are reported in Section 5.

2. THE INDEPENDENT TWO-SAMPLE DESIGN

Let us consider the testing problems for one-sided alternative hypotheses as produced by treatments with non-negative effect size δ. Particularly, the fixed additive effects model, which is written as

$$X_i^t = \mu + \delta + \sigma E_i^t,\ i = 1, \ldots, n_t; \quad X_i^c = \mu + \sigma E_i^c,\ i = 1, \ldots, n_c, \tag{1}$$

where μ is a common location parameter, E_i^k $(k = t, c)$ are random error deviates with location parameter zero and scale parameter one, σ is a scale parameter independent on experimental units and treatment levels, and δ is the effect size (treatment effect) which is typically unknown. In practice, for comparing $\mathbf{X}^t = (X_1^t, \ldots, X_{n_t}^t)$ to $\mathbf{X}^c = (X_1^c, \ldots, X_{n_c}^c)$, and without loss of generality, $\mu = 0$ is assumed. Therefore, the dataset can be also written as $\mathbf{X}(\boldsymbol{\delta}) = (\mathbf{E}^t + \boldsymbol{\delta}, \mathbf{E}^c)$ where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{n_t})$ (For simplicity, set $\delta_i = \delta > 0$, for each $i = 1, \ldots, n_t$), $\mathbf{E}^t = (E_1^t, \ldots, E_{n_t}^t)$, and $\mathbf{E}^c = (E_1^c, \ldots, E_{n_c}^c)$. The hypothesis of interest is

$$H_0 : \boldsymbol{\delta} = 0 \text{ against } H_1 : \boldsymbol{\delta} > 0. \tag{2}$$

2.1. *Classic independent two-sample t-test*

Suppose in Equation (1) $E_1^t, \ldots, E_{n_t}^t, E_1^c, \ldots, E_{n_c}^c$ be independent and identically distributed normal random deviates. Then the likelihood ratio test statistic for testing the null hypothesis in Equation (2) is given by

$$T = \frac{\bar{X}^t - \bar{X}^c}{S_p \sqrt{\frac{1}{n_t} + \frac{1}{n_c}}},$$

where $\bar{X}^k = \sum_{i=1}^{n_k} X_i^k / n_k$ $(k = t, c)$, and $S_p^2 = \sum_{k \in \{t,c\}} \sum_{i=1}^{n_k} (X_i^k - \bar{X}^k)^2 / (n_t + n_c - 2)$ is the pooled variance.

Under H_0, T is distributed as Student's distribution with $\nu = n_t + n_c - 2$ degrees of freedom. And hence, H_0 is rejected when $T > t_\nu^\alpha$, where t_ν^α is the upper α critical value, and α is the level of significance. The statistical power level is given by

$$W(\delta; n_t, n_c, \alpha) = 1 - F_t(t_\nu^\alpha, \nu, ncp),$$

where F_t is the cumulative distribution of the Student t, $\nu = n_t + n_c - 2$ is the degrees of freedom, and $ncp = \delta\left(S_p^2(\frac{1}{n_t} + \frac{1}{n_c})\right)^{-1/2}$ is the non-centrality parameter. Note that W is a function of δ for a given sample sizes (n_t, n_c), and preassigned level of significance α. The power level measures how likely to get a significance result given that the alternative hypothesis is true.

2.2. *Two-sample permutation test*

Suppose now in Equation (1) $E_1^t, \ldots, E_{n_1}^t, E_1^c, \ldots, E_{n_2}^c$ are exchangeable random deviates. For the considered testing problem in Equation (2) a suitable test statistic should be chosen such that, without loss of generality, large values of it are considered to be against H_0. For more details about the choice of the test statistic in the permutation framework see page 84 of Pesarin and Salmaso (2010). One may choose $T = \bar{X}^t - \bar{X}^c$ as a test statistic. For determining the p-value, an appropriate reference distribution is needed which is called the permutation distribution. Indeed, the following steps are used to carry out the permutation test for two-sample design.

1. Randomly assign experimental units to one of the two samples with n_t units to the first sample (treatment group) and n_c units to the second sample (control or placebo group). Then, the observed data sets, $\mathbf{X}^t$ and $\mathbf{X}^c$, are obtained and the test statistic is evaluated, $T_0 = T(\mathbf{X})$.
2. Permute the $n = n_t + n_c$ observations between the two groups. Write down the set of all possible permutations, $\mathcal{X}$. The cardinality of this support is $n!$.
3. For each permutation $\mathbf{X}^* \in \mathcal{X}$, compute the test statistic, $T^* = T(\mathbf{X}^*)$. The cardinality of related support is $n!/n_t!n_c!$.
4. The true p-value is calculated as
$$\lambda_T(\mathbf{X}) = \frac{\text{number of } T^*\text{'s} \geq T_0}{\binom{n}{n_t}}.$$
5. If a preassigned level of significance, α, has been set, declare the test to be significant if the p-value is not larger than this level.

Since it is tedious or even practically impossible to write down and enumerate the whole members of permutation sample space $\mathcal{X}$, conditional Monte Carlo simulation (Algorithm 1) can be used to approximate the p-value at any desired accuracy.

Note that $\hat{\lambda}_T(\mathbf{X})$ is an unbiased estimate of the true $\lambda_T(\mathbf{X})$ and, due to the Glivenko-Cantelli theorem (Shorack and Wellner, 1986), as B diverges it is strongly consistent. Moreover, the standard error for $\hat{\lambda}_T(\mathbf{X})$ is
$$\sqrt{\lambda_T(\mathbf{X})(1 - \lambda_T(\mathbf{X}))/B}.$$

Algorithm 1 Conditional Monte Carlo (CMC)

1. For the given dataset $\mathbf{X}$, compute the observed test statistic, $T_0 = T(\mathbf{X})$.
2. From $\mathscr{X}$ take a random permutation $\mathbf{X}^*$ of $\mathbf{X}$, and compute the corresponding permutation test statistic $T^* = T(\mathbf{X}^*)$.
3. Independently repeat Step 2 a large number of times, say B, giving B values for T^*, say $\{T_b^*,\ b = 1,\ldots,B\}$.
4. The estimated permutation p-value is

$$\hat{\lambda}_T(\mathbf{X}) = \frac{\sum_{b=1}^{B} \mathbb{I}(T_b^* \geq T_0)}{B},$$

where $\mathbb{I}(\cdot)$ is the indicator function.

Algorithm 2 Power Function of Permutation Test

1. Choose an arbitrary value of the effect δ.
2. Draw one set of n error deviates $\mathbf{E}$ from the underlying distribution, and then add δ to the first n_t error deviates to define the data set $\mathbf{X}(\delta) = (\mathbf{E}^t + \delta, \mathbf{E}^c)$.
3. Use the CMC algorithm to estimate the permutation p-value $\hat{\lambda}_T(\mathbf{X}(\delta))$.
4. Independently repeat Steps 2 and 3 a large number of times, say R, giving R estimated p-values, say $\{\hat{\lambda}_T(\mathbf{X}_r(\delta)),\ r = 1,\ldots,R\}$.
5. Finally, the estimated power level is given by

$$\hat{W}(\delta; n_t, n_c, \alpha, T, P) = \frac{\sum_{r=1}^{R} \mathbb{I}[\hat{\lambda}_T(\mathbf{X}_r(\delta)) \leq \alpha]}{R},$$

where T is the chosen test statistic and P is the underlying distribution.

6. To obtain the power function as a function of δ, Steps 1-5 are repeated for different values of δ.

Therefore, a $100(1-\alpha)\%$ approximate confidence interval for $\lambda_T(\mathbf{X})$ is

$$\hat{\lambda}_T(\mathbf{X}) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}_T(\mathbf{X})\left(1-\hat{\lambda}_T(\mathbf{X})\right)}{B}}.$$

The p-value of the permutation test is conditional upon the observations for each sample, but the power is the proportion of p-values that are less than or equal α over repeated sampling from the underlying population. Therefore, the power level can be estimated by the use of standard Monte Carlo simulation. Algorithm 2 is used for estimating the power.

3. Ranked set two-sample permutation test

Apparently, RSS methodology produces a data set as follows

$$\begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ Y_{21} & Y_{22} & \cdots & Y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{h1} & Y_{h2} & \cdots & Y_{hm} \end{pmatrix},$$

where $Y_{ji} = X_{j(i)}$, $j = 1,\ldots,h$, $i = 1,\ldots,m$. Let $\mathbf{Y}^t = \{Y^t_{ji}\}$ and $\mathbf{Y}^c = \{Y^c_{ji}\}$ denote the treatment and control groups, respectively. Note that the two groups are independent. For each group, $(\mathbf{Y}^t, \mathbf{Y}^c)$, the data are all mutually independent and, in addition, the data in the same column are identically distributed. Therefore, the exchangeability assumption holds within columns and hence the permutation test can be applied. So, to maintain the distribution diversity, the data in the i^{th} column of $\mathbf{Y}^t$ has to be permuted by the data in the i^{th} column of $\mathbf{Y}^c$ for each i in $\{1,2,\ldots,m\}$. To carry out the permutation test for this ranked set two-sample design Algorithm 3 is used.

The test statistics considered in this paper are as follows.

1. First, the difference between overall means of the two groups; that is
$$T_1 = \bar{Y}^t - \bar{Y}^c, \tag{3}$$
where $\bar{Y}^k = \sum_{i=1}^{n_k} \sum_{j=1}^{h} Y^k_{ji}/hm$ $(k = t, c)$ (Assuming balanced design, $n_t = n_c = mh$).

2. Second, the sum of the studentized statistics for all columns of the two matrices; that is
$$T_2 = \sum_{i=1}^{m} \left(\frac{\bar{Y}^t_i - \bar{Y}^c_i}{\hat{\sigma}_i} \right), \tag{4}$$
where $\hat{\sigma}^2_i = \sum_{k\in\{t,c\}} \sum_{j=1}^{h} (Y^k_{ji} - \bar{Y}^k_i)^2/(2mh-2)$, and $\bar{Y}^k_i = \sum_{j=1}^{h} Y^k_{ji}/h$.

Algorithm 3 Ranked set two-sample permutation test

1. For the ranked set two-sample data sets $(\mathbf{Y}^t, \mathbf{Y}^c)$, compute the test statistic, T_0.
2. Form a new matrix $\mathbf{Y} = \mathbf{Y}^t \uplus \mathbf{Y}^c$ by concatenating $\mathbf{Y}^t$ and $\mathbf{Y}^c$ vertically (Note that the two matrices have m columns).
3. Randomly permute $\mathbf{Y}$ column by column to get $\mathbf{Y}^*$.
4. Split $\mathbf{Y}^*$ into $\mathbf{Y}^{t*}$ and $\mathbf{Y}^{c*}$ such that $\mathbf{Y}^{t*}$ and $\mathbf{Y}^{c*}$ contain the same number of rows as in $\mathbf{Y}^t$ and $\mathbf{Y}^c$, respectively.
5. Compute the test statistic $T^* = T(\mathbf{Y}^*)$ based on $\mathbf{Y}^* = \mathbf{Y}^{t*} \uplus \mathbf{Y}^{c*}$.
6. Independently repeat Steps 3-5 a large number of times, say B, giving B test statistics, say $\{T_b^*,\ b = 1, \ldots, B\}$.
7. The estimated p-value is

$$\hat{\lambda}(\mathbf{Y}) = \frac{\sum_{b=1}^{B} \mathbb{I}(T_b^* \geq T_0)}{B}.$$

3. The third proposal is based on partial tests. To this end, the null hypothesis in Equation (2) is partitioned into m independent sub-hypotheses as follows.

$$H_{0i} : \boldsymbol{\delta}_i = 0 \text{ against } H_{1i} : \boldsymbol{\delta}_i > 0, \quad i = 1, \ldots, m, \tag{5}$$

where $\boldsymbol{\delta}_i = \mu_i^t - \mu_i^c$, μ_i^t and μ_i^c are the population means of the i^{th} order statistic in the treatment and control group respectively. Thus, the differences between the column means of the treatment and control groups are considered as test statistics; that is,

$$T_{3i} = \bar{Y}_i^t - \bar{Y}_i^c,\ i = 1, \ldots, m. \tag{6}$$

Now, Algorithm 1 is used and this leads m independent p-values $(\lambda_1, \ldots, \lambda_m)$. Finally, these p-values have to be combined for testing the overall hypothesis in Equation (2). The following approaches are considered for combining p-values.

   1. The Fisher approach (Fisher, 1934).
   It is based on the statistic $X^2 = -2\sum_{i=1}^{m} \log \lambda_i$. Under the null hypothesis, $X^2 \sim \chi^2_{(2m)}$. So, the combined p-value is given by

$$\lambda_F = P(\chi^2_{(2m)} > X^2). \tag{7}$$

2. The Liptak approach (Liptak, 1958).
   It is based on the statistic $L = \sum_{i=1}^{m} \Phi^{-1}(1-\lambda_i)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Under the null hypothesis, $L/\sqrt{m} \sim N(0,1)$. So, the combined p-value is given by

$$\lambda_L = P\left(Z > \frac{L}{\sqrt{m}}\right), \tag{8}$$

   where Z is the standard normal random variable.
3. The logistic approach (Mudholkar and George, 1979). It is based on the logit statistic

$$t = C^{-1} \sum_{i=1}^{m} \log[(1-\lambda_i)/\lambda_i],$$

   where $C = \sqrt{m\pi^2(5m+2)/3(5m+4)}$. Under the null hypothesis, t follows approximately a Student's distribution with degrees of freedom $5m+4$. So, the combined p-value is given by

$$\lambda_M = P\left(T_{(5m+4)} > t\right). \tag{9}$$

4. SIMULATION STUDY

A simulation study is carried out to assess the significance level and the power of the proposed test statistics for the two-sample RSS design and to compare it with the usual two-sample permutation design and the classical two-sample t-test. Different configurations are considered in the simulation study. For each combination of $m = \{3, 4\}$ and $h = \{5, 10\}$, four different distributions are considered; uniform distribution $U(-1,1)$, normal distribution $N(0,1)$, skew normal distribution $SN(0,1,-5)$, and exponential distribution $Exp(1)$. We also performed several other combinations, not reported here, and the results follow the same behavior. The simulation study is performed based on $R = 5000$ datasets. The permutation is based on $B = 1000$ replications. To examine the significance level of the tests we set $\delta = 0$, while to investigate the power behavior we select values of δ in the set $\{0.2, 0.4, 0.6, 0.8\}$. The nominal significance level was set to $\alpha = 0.05$. The results of the study are reported in Tables 1 and 2.

It is worth to point out that the power levels of the proposed test statistics are obtained for the same generated two-sample RSS. Moreover, balanced designs are considered in computing the power levels; that is, each sample of the two-sample RSS was with set size m and number of cycles h. Also, the size of each sample in the two-sample SRS is $h \times m$. So that power comparisons between the considered test statistics under RSS and SRS are done by maintaining the same number of experimental units in both schemes (to insure that the two schemes have the same cost).

TABLE 1
Empirical significance level and power from the simulation study – Uniform and normal distributions.

			Uniform					Normal				
m	h	Test stat $\delta \rightarrow$	0.0	0.2	0.4	0.6	0.8	0.0	0.2	0.4	0.6	0.8
3	5	SRS	0.056	0.234	0.584	0.874	0.984	0.056	0.128	0.281	0.495	0.678
		T_1	0.050	0.361	0.832	0.991	1.000	0.044	0.187	0.436	0.711	0.904
		T_2	0.053	0.346	0.812	0.987	1.000	0.044	0.181	0.421	0.691	0.891
		T_3(Fisher)	0.049	0.309	0.769	0.979	0.999	0.040	0.159	0.367	0.637	0.850
		T_3(Liptak)	0.048	0.342	0.810	0.986	1.000	0.043	0.173	0.410	0.684	0.890
		T_3(Logit)	0.048	0.332	0.803	0.985	0.999	0.040	0.173	0.402	0.678	0.882
	10	SRS	0.051	0.371	0.842	0.990	1.000	0.053	0.181	0.449	0.749	0.917
		T_1	0.050	0.604	0.982	1.000	1.000	0.052	0.284	0.683	0.938	0.997
		T_2	0.049	0.602	0.982	1.000	1.000	0.049	0.281	0.682	0.934	0.996
		T_3(Fisher)	0.050	0.564	0.975	1.000	1.000	0.051	0.260	0.637	0.913	0.993
		T_3(Liptak)	0.050	0.602	0.982	1.000	1.000	0.049	0.279	0.681	0.936	0.996
		T_3(Logit)	0.052	0.597	0.982	1.000	1.000	0.050	0.275	0.674	0.930	0.996
4	5	SRS	0.058	0.283	0.683	0.943	0.996	0.055	0.157	0.342	0.586	0.796
		T_1	0.048	0.514	0.959	1.000	1.000	0.051	0.249	0.599	0.881	0.984
		T_2	0.047	0.503	0.953	1.000	1.000	0.051	0.237	0.579	0.869	0.980
		T_3(Fisher)	0.043	0.452	0.934	0.999	1.000	0.048	0.206	0.519	0.830	0.966
		T_3(Liptak)	0.046	0.504	0.951	1.000	1.000	0.048	0.231	0.578	0.865	0.980
		T_3(Logit)	0.046	0.497	0.948	1.000	1.000	0.048	0.224	0.560	0.861	0.977
	10	SRS	0.045	0.449	0.921	0.999	1.000	0.055	0.228	0.546	0.843	0.972
		T_1	0.050	0.782	0.999	1.000	1.000	0.053	0.393	0.851	0.992	1.000
		T_2	0.047	0.791	0.999	1.000	1.000	0.049	0.391	0.848	0.992	1.000
		T_3(Fisher)	0.048	0.748	0.998	1.000	1.000	0.051	0.362	0.820	0.988	1.000
		T_3(Liptak)	0.049	0.790	0.999	1.000	1.000	0.050	0.395	0.848	0.993	1.000
		T_3(Logit)	0.049	0.785	0.999	1.000	1.000	0.051	0.390	0.849	0.993	1.000

TABLE 2

Empirical significance level and power from the simulation study – Skew normal and exponential distributions.

			Skew normal					Exponential				
m	h	Test Stat $\delta \rightarrow$	0.0	0.2	0.4	0.6	0.8	0.0	0.2	0.4	0.6	0.8
3	5	SRS	0.052	0.214	0.537	0.833	0.958	0.054	0.143	0.316	0.525	0.712
		T_1	0.054	0.335	0.755	0.964	0.999	0.051	0.197	0.433	0.699	0.857
		T_2	0.054	0.359	0.796	0.978	1.000	0.048	0.298	0.682	0.916	0.982
		T_3(Fisher)	0.047	0.321	0.756	0.970	0.999	0.044	0.282	0.663	0.902	0.975
		T_3(Liptak)	0.048	0.351	0.791	0.975	1.000	0.044	0.283	0.649	0.884	0.967
		T_3(Logit)	0.047	0.348	0.787	0.974	1.000	0.045	0.287	0.662	0.895	0.974
	10	SRS	0.048	0.344	0.802	0.980	1.000	0.051	0.206	0.474	0.756	0.916
		T_1	0.052	0.514	0.958	1.000	1.000	0.050	0.281	0.662	0.896	0.980
		T_2	0.056	0.564	0.974	1.000	1.000	0.048	0.451	0.901	0.993	1.000
		T_3(Fisher)	0.057	0.534	0.967	1.000	1.000	0.046	0.465	0.904	0.992	1.000
		T_3(Liptak)	0.059	0.563	0.974	1.000	1.000	0.048	0.460	0.899	0.993	1.000
		T_3(Logit)	0.058	0.561	0.975	1.000	1.000	0.048	0.467	0.907	0.994	1.000
4	5	SRS	0.052	0.256	0.642	0.913	0.988	0.045	0.180	0.385	0.600	0.800
		T_1	0.058	0.450	0.907	0.999	1.000	0.054	0.243	0.578	0.827	0.947
		T_2	0.054	0.492	0.942	0.999	1.000	0.052	0.438	0.877	0.986	0.999
		T_3(Fisher)	0.049	0.434	0.919	0.999	1.000	0.042	0.423	0.860	0.985	0.998
		T_3(Liptak)	0.050	0.482	0.938	0.999	1.000	0.046	0.425	0.851	0.976	0.998
		T_3(Logit)	0.049	0.477	0.937	0.999	1.000	0.046	0.434	0.863	0.981	0.999
	10	SRS	0.052	0.415	0.881	0.997	1.000	0.053	0.234	0.578	0.852	0.966
		T_1	0.047	0.697	0.995	1.000	1.000	0.052	0.360	0.791	0.969	0.996
		T_2	0.049	0.769	0.999	1.000	1.000	0.051	0.655	0.986	1.000	1.000
		T_3(Fisher)	0.054	0.737	0.997	1.000	1.000	0.049	0.667	0.988	1.000	1.000
		T_3(Liptak)	0.050	0.772	0.999	1.000	1.000	0.051	0.658	0.986	1.000	1.000
		T_3(Logit)	0.051	0.767	0.998	1.000	1.000	0.049	0.672	0.988	1.000	1.000

We noticed from Tables 1 and 2 that the permutation test statistics within RSS control the type I error probability at the nominal significance level ($\alpha = .05$). Also, for fixed m and h, the power levels of the permutation test statistics within RSS are strictly higher than the power level of the permutation test statistic within SRS for all given δ. Moreover, the power levels based on RSS increase as δ, m and h increase. Among the proposed test statistics, T_1 is the best for symmetric distributions, while for asymmetric distribution T_2 is the best. Among the considered combining functions, the Liptak combining function is the best for the uniform, normal, and skew normal distributions, while for the exponential distribution the logit combining function is the best. In addition, all considered combining functions behave the same in terms of power levels for large sample size. Finally, under normality assumption, Table 3 reports the exact power levels for the parametric one-sided two-sample t-test when the sample sizes $n_c = n_t = 15$. Apparently, the power levels of the permutation test under SRS and the parametric t-test are equivalent, while the power levels of the permutation test statistics based on RSS are higher than in the parametric t-test.

TABLE 3
The power levels of the parametric one-sided two-sample t-test, the size of each sample is 15, and the nominal significance level is $\alpha = 0.05$.

δ	0.0	0.2	0.4	0.6	0.8
Power	0.05	0.133	0.282	0.483	0.689

5. Conclusion

In this paper, the permutation test is discussed within the context of RSS. Three test statistics are suggested and compared with the usual permutation test statistic and the classical t-test. In summary, our simulation results assert that the power levels of the permutation test statistics using RSS are higher than the power levels of the classical t-test statistics and the permutation test statistic using SRS. Subsequently, it is recommended to apply permutation test within the framework of RSS in lieu of SRS.

Acknowledgements

The author would like to thank the Editor and the referee whose comments and suggestions were very useful in improving a first draft of the paper.

References

L. Amro, M. H. Samuh (2017). *More powerful permutation test based on multistage ranked set sampling*. Communications in Statistics-Simulation and Computation, 46, no. 7, pp. 5271–5284.

V. BARNETT, K. MOORE (1997). *Best linear unbiased estimates in ranked-set sampling with particular reference to imperfect ordering*. Journal of Applied Statistics, 24, no. 6, pp. 697–710.

L. L. BOHN, D. A. WOLFE (1992). *Nonparametric two-sample procedures for ranked-set samples data*. Journal of the American Statistical Association, 87, no. 418, pp. 552–561.

L. L. BOHN, D. A. WOLFE (1994). *The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the Mann-Whitney-Wilcoxon statistic*. Journal of the American Statistical Association, 89, no. 425, pp. 168–176.

Z. CHEN, Z. BAI, B. K. SINHA (2004). *Ranked set sampling: theory and applications*, vol. 176. Springer Verlag, New York.

T. R. DELL, J. L. CLUTTER (1972). *Ranked set sampling theory with order statistics background*. Biometrics, 28, no. 2, pp. 545–555.

R. A. FISHER (1934). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

K. M. KOTI, G. JOGESH BABU (1996). *Sign test for ranked-set sampling*. Communications in Statistics-Theory and Methods, 25, no. 7, pp. 1617–1630.

Z. LIANGYONG, D. XIAOFANG (2010). *Optimal ranked set sampling design for the sign test*. Chinese Journal of Applied Probability and Statistics, 26, no. 3, pp. 225–233.

T. LIPTAK (1958). *On the combination of independent tests*. Magyar Tud Akad Mat Kutato Int Kozl, 3, pp. 171–197.

S. N. MACEACHERN, Ö. ÖZTÜRK, D. A. WOLFE, G. V. STARK (2002). *A new ranked set sample estimator of variance*. Journal of the Royal Statistical Society: Series B, 64, no. 2, pp. 177–188.

G. MCINTYRE (1952). *A method for unbiased selective sampling, using ranked sets*. Australian Journal of Agricultural Research, 3, no. 4, pp. 385–390.

G. MCINTYRE (2005). *A method for unbiased selective sampling, using ranked sets*. The American Statistician, 59, no. 3, pp. 230–232.

G. S. MUDHOLKAR, E. O. GEORGE (1979). *The logit statistic for combining probabilities - an overview*. In J. S. RUSTAGI (ed.), *Optimizing Methods in Statistics*, Academic Press, New York, pp. 345–365.

E. J. T. MURFF, T. W. SAGER (2006). *The relative efficiency of ranked set sampling in ordinary least squares regression*. Environmental and Ecological Statistics, 13, no. 1, pp. 41–51.

R. A. MURRAY, M. S. RIDOUT, J. V. CROSS (2000). *The use of ranked set sampling in spray deposit assessment*. Aspects of Applied Biology, 57, pp. 141–146.

Ö. ÖZTÜRK (1999). *Two-sample inference based on one-sample ranked set sample sign statistics*. Journal of Nonparametric Statistics, 10, no. 2, pp. 197–212.

Ö. ÖZTÜRK, D. A. WOLFE (2000). *Optimal allocation procedure in ranked set two-sample median test*. Journal of Nonparametric Statistics, 13, no. 1, pp. 57–76.

G. P. PATIL (1995). *Editorial: Ranked set sampling*. Environmental and Ecological Statistics, 2, no. 4, pp. 271–285.

F. PESARIN, L. SALMASO (2010). *Permutation Tests for Complex Data: Theory, Application and Software*. John Wiley & Sons, Ltd., Chichester.

L. H. PHILIP, K. LAM (1997). *Regression estimator in ranked set sampling*. Biometrics, 53, no. 3, pp. 1070–1080.

D. S. REY (2004). *The Informational Order in Ranked Set Sampling Experiments*. Ph.D. thesis, Mathematisches Institut der Universität Göttingen.

M. H. SAMUH (2012). *Some Advances in Permutation Testing*. Ph.D. thesis, Department of Statistical Science, Padua University, Italy.

G. R. SHORACK, J. A. WELLNER (1986). *Empirical Processes with Applications to Statistics*. Wiley Series in Probability & Mathematical Statistics, New York.

L. STOKES (1995). *Parametric ranked set sampling*. Annals of the Institute of Statistical Mathematics, 47, no. 3, pp. 465–482.

S. L. STOKES, T. W. SAGER (1988). *Characterization of a ranked-set sample with application to estimating distribution functions*. Journal of the American Statistical Association, 83, no. 402, pp. 374–381.

K. TAKAHASI, 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, no. 1, pp. 1–31.

SUMMARY

In this paper, ranked set two-sample permutation test of comparing two-independent groups in terms of some measure of location is presented. Three test statistics are proposed. The statistical power of these new test statistics are evaluated numerically. The results are compared with the statistical power of the usual two-sample permutation test under simple random sampling and with the classical independent two-sample t-test.

Keywords: Permutation test; Ranked set sampling; Statistical power; Type I error.