Thailand Statistician
April 2022; 20(2): 435-451
http://statassoc.or.th
Contributed paper

Adjusted Clustered Rank Tests for Clustered Data in Unbalanced Design

Prayad Sangngam*[a], Wipawan Laoarun [a], and Manachai Rodchuen [b]
[a] Department of Statistics, Faculty of Science, Silpakorn University, Nakorn Pathom, Thailand.
[b] Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand.
*Corresponding author; e-mail: sangngam_p@su.ac.th

Received: 1 October 2020
Revised: 3 May 2021
Accepted: 30 May 2021

Abstract

This paper proposes an adjusted rank test-T_1 to determine that two independent samples with clustered data in unbalanced design are drawn from the same population. For a large number of clusters, the paper finds that the test statistic-T_1 converges to normal distribution. In addition, the clustered rank sum test-T_2 is presented for three or more independent samples with clustered data. An adjusted rank test-T_3 is also proposed by adjusting the clustered rank sum test. For each adjusted rank test, the same critical value is used for data sets with equivalence between the numbers of clusters and cluster sizes, but the observations might differ. The critical values of two adjusted test statistics for some numbers of clusters and cluster sizes are given at the significance levels of 0.10 and 0.05. To compare the performances of the adjusted rank tests with the alternative tests, a simulation study is necessary. Results show that the two adjusted tests can maintain the size of the tests for all situations. For three samples, the Kruskal-Wallis test based on the observation mean of cluster gives the estimated size of about 25% at the true significance level of 5%. The adjusted test-T_1 has a higher power than the Wilcoxon test based on the observation mean of cluster. The power of both adjusted rank tests increases when the number of clusters, the number of observations per cluster, and the effect size all increase. However, the power of the adjusted tests decreases when the correlation coefficient between observations in a cluster increases.

Keywords: Clustered data, independent samples, clustered ranks sum test, central limit theorem.

1. Introduction

In many studies, researchers are interested in testing the null hypothesis that the two independent samples have been drawn from the same population or from populations with equal means. In parametric statistics, the two-sample independent t-test is widely used to test this hypothesis. The t-test requires that the random samples are drawn from normal distribution. If the assumptions of the t-test are violated, then the Wilcoxon rank sum test proposed by Wilcoxon (1945), which is the nonparametric procedure, can be used. The common assumption of the t-test and Wilcoxon rank sum test is that all observations are independent.

For many situations, the data are collected from clusters of correlated observations. The cluster may be a family, a litter, a laboratory, and a region of observation units. Examples of clustered data are the repeated measurements of blood pressure for a single object, the socio-economic characteristics of households in a block, and the body mass index of siblings. Wu et al. (1988) showed that using the F-test for clustered data leads to an inflated type I error rate. Thus, the probability of type I error is higher than a given significance level. In addition, the type I error rate increases as the intra-correlation increases.

In the parametric approach, many researchers have considered the different procedures for testing the hypothesis with correlated clustered data. Most of the theoretical research for clustered data assumes a parametric model. Wu et al. (1988) adjusted the F-test statistic by using coefficient of intra-correlation so that the adjusted statistic has the F distribution with the same degrees of freedom as those of the F-test statistic. Rao et al. (1993) proposed a two-stage generalized least squares test in regression analysis by transforming the observations to uncorrelated ones. The two mentioned tests depend on an unknown intra-correlation coefficient. However, Lahiri and Yan (2009) proposed an alternative test that does not require the estimation of the intra-correlation coefficient.

In the nonparametric approach, only a small body of literature exists for incorporating clustered data. Rosner and Grove (1999) considered the generalization of the Mann-Whitney U test, also called the Wilcoxon rank sum test, for clustered data. They used the estimates of correlation parameters to correct the estimated variance of the test statistic. The test has appropriate type I error rate in balanced design with as few as 20 clusters per group. Rosner et al. (2003) introduced a large sample randomization test (abbreviated by the RGL test) for clustered data by using the rank sum of observations. Rosner et al. (2006) extended the signed rank test to the clustered data setting. Rosner and Glynn (2009) presented the approach to estimate the power of the RGL test and the sample size. However, under the homogeneity of data sets with different observation values, the critical values of the RGL test may differ when testing hypotheses with the same significance level. The researcher who used the RGL test must find a critical value for one dataset. Sangngam and Laoarun (2021) presented adjusted rank tests for clustered data in balanced design. The adjusted rank test uses the same critical values for equivalent data sets, and it has more empirical power than the RGL test for a small number of clusters in samples.

Clustered data may occasionally occur in the unbalanced design. Therefore, this paper adjusts the RGL test statistic for clustered data using the rank of the mean of observation ranks, which is called the adjusted rank test-T_1 in unbalanced design. For a large number of clusters, the asymptotic distribution of the adjusted rank test-T_1 will be derived. To generalize the adjusted rank test, an adjusted rank test-T_3 is proposed for more than two samples with clustered data. Under small number of clusters, the critical values of both adjusted rank tests will be given. The type I error rate and the statistical power of the adjusted rank tests compared with the existing tests are considered in the simulation study.

2. Clustered Rank Sum Test for Two Samples

Let X_{ij} be the j^{th} observation in the cluster i for $1 \le i \le N$, $1 \le j \le c_i$, where c_i is the cluster size of the i^{th} cluster. The indicator δ_{ij} denotes the group of samples; $\delta_{ij} = 1$ if X_{ij} belongs to the first sample and $\delta_{ij} = 0$ if X_{ij} belongs to the second sample. The data can be presented in the form of $(\mathbf{X}, \boldsymbol{\delta}) = \left\{ \left(X_{ij}, \delta_{ij} \right) : 1 \le j \le c_i, 1 \le i \le N \right\}$. When there are some clusters with unequal cluster sizes from the other, the design is called an unbalanced design. We assume that clusters are independent while the observations within the cluster are not. The hypothesis to be tested is that there is no difference between the location parameters of the two populations.

Rosner et al. (2003) proposed the clustered rank sum test for clustered data abbreviated by RGL. Let R_{ij} be the rank of X_{ij} based on the combined two samples of all observations. The sum of ranks from the first sample is assigned to be the test statistic. Let $\delta_{ij} = \delta_i$ for all $1 \le j \le c_i$. The RGL test statistic is defined as

$$T = \sum_{i=1}^{N} \delta_i R_{i+}, \tag{1}$$

where $R_{i+} = \sum_{j=1}^{c_i} R_{ij}$ is the sum of observation ranks in the i^{th} cluster.

The clustered rank sum method assumes that the observations in a given cluster are exchangeable. The exact distribution of the RGL test statistic is considered based on random permutation conditioning on the sum of observation ranks in the i^{th} cluster, R_{i+}. To derive the distribution of the cluster test under null hypothesis, the clusters are partitioned by using the cluster size. Let $G_{\max}$ be the maximum of cluster sizes. The test statistic-T can be written as

$$T = \sum_{g=1}^{G_{\max}} \sum_{i \in I_g} \delta_i R_{i+} = \sum_{g=1}^{G_{\max}} W_g,$$

where I_g is the set of indices of cluster size g from the first sample, and $W_g = \sum_{i \in I_g} \delta_i R_{i+}$. Let N_g be the number of clusters of size g. Let m_g and n_g be the number of cluster of size g from the first and second samples, respectively. If N is small, the distribution of statistic-T conditioning on R_{i+} can be generated by combining all possible permutations of R_{i+} in W_g for a given cluster of size g .

If N is large, the computation of the permutation is intensive. The RGL asymptotic statistic is given by

$$Z = \frac{T - \sum_{g=1}^{G_{\max}} m_g \left(R_{++g} / N_g \right)}{\sqrt{Var(T)}} \xrightarrow{d} N(0,1), \tag{2}$$

where $R_{++g} = \sum_{i \in I_g} R_{i+}$ and $Var(T) = \sum_{g=1}^{G_{\max}} \frac{m_g n_g}{N_g (N_g - 1)} \sum_{i \in I_g} \left(R_{i+} - \frac{R_{++,g}}{N_g} \right)^2$.

Under some conditions, the test statistic Z has an asymptotic standard normal distribution. For an unbalanced number of clustered sizes between two samples, the RGL test may result in inefficiency (Rosner et al. 2003). If some clusters from either sample have a different size from other clusters, then the permutation of these clusters will be ignored as no permutation of these clusters can be made.

If there is homogeneity of data sets of $(\mathbf{X}, \boldsymbol{\delta})$'s with different observation values, the set of R_{i+} in each data set may be different. Thus, if N is small, the different critical values of test statistic-T for these data sets might be used at the same nominal level.

3. Adjusted Rank Sum Test for Two Samples

Under the presented data of $(\mathbf{X},\boldsymbol{\delta})$, let L be the number of distinct cluster sizes. To compute the proposed test statistic, the data are stratified into L strata by cluster size. Assume that the first stratum contains the observations with the smallest cluster size, the second stratum contains the observations with the next higher cluster size, and so on. The L^{th} stratum contains the observations with the largest cluster size. Let N_h be the number of clusters from the stratum h. Let m_h and n_h be the number of cluster in the stratum h from the first and second samples, respectively. We denote that X_{hij} is the j^{th} observation in the cluster i from the h^{th} stratum for $1 \le j \le c_h$, where c_h is the cluster size of the h^{th} stratum. Let R_{hij} be the observation rank of X_{hij} based on the combined samples of all observations. The indicator δ_{hi} denotes the group of the samples; $\delta_{hi} = 1$ if R_{hij} belongs to the first sample and $\delta_{hi} = 0$ if R_{hij} belongs to the second sample.

The cluster mean of ranks from the i^{th} cluster in the h^{th} stratum is calculated as $\bar{R}_{hi} = \frac{1}{c_h}\sum_{j=1}^{c_h} R_{hij}$ for all $h = 1,2,\ldots,L$ and $1 \le i \le N_h$. We define the cluster mean as the mean of observation ranks in a cluster. The new ranks are assigned to the cluster means for each stratum, namely, R_{hi}^{*}. In the first stratum, we assign rank 1 for the cluster with the smallest cluster mean, rank 2 for the cluster with the next higher cluster mean, and so on. The rank N_1 is assigned to the cluster with the highest cluster mean. In the second stratum, we assign rank $N_1 + 1$ for the cluster with the smallest cluster mean, rank $N_1 + 2$ for the next higher cluster mean, and so on. The rank $N_1 + N_2$ is assigned to the cluster with the highest cluster mean. The ranks are continuously assigned to the next stratum. In the L^{th} stratum, the rank $\sum_{h=1}^{L-1} N_h + 1$ is assigned to the cluster with the smallest cluster mean, the rank $\sum_{h=1}^{L-1} N_h + 2$ is assigned to the cluster with the next higher cluster mean, and so on. The rank $\sum_{h=1}^{L} N_h = N$ is assigned to the cluster with the largest cluster mean. In case of ties at each stratum, the average of the ranks is assigned to those clusters. The adjusted rank test statistic-T_1 is proposed by

$$T_1 = \sum_{h=1}^{L}\sum_{i=1}^{N_h} \delta_{hi} R_{hi}^{*} = \sum_{h=1}^{L} W_h^{*}, \tag{3}$$

where $W_h^{*} = \sum_{i=1}^{N_h} \delta_{hi} R_{hi}^{*}$ is the sum of cluster ranks in the first sample from stratum h. If the total number of cluster N is small, the distribution of T_1 conditioning on R_{hi}^{*} can be generated by combining all possible permutations of R_{hi}^{*} in W_h^{*} for a given stratum h. The total number of permutation is $\prod_{h=1}^{L}\binom{N_h}{m_h}$. Given stratum h and under the null hypothesis, a subset of the clustered sum ranks is randomly assigned to the first sample (Rosner et al. 2003), so the subset of the new clustered ranks is also equally likely to be observed from the first sample. When the hypothesis about the difference between two location parameters is tested, the null hypothesis will be rejected for either a sufficiently small or a sufficiently large value of T_1. Therefore, we reject null hypothesis at a significance level α if the computed statistical value of T_1 is less than or equal to the critical value of $t^*_{\alpha/2}$ or greater than or equal to the critical value of $t^*_{1-\alpha/2}$. These critical values $t^*_{\alpha/2}$ and $t^*_{1-\alpha/2}$ are the $(\alpha/2)100\%^{\text{th}}$ and $(1-\alpha/2)100\%^{\text{th}}$ quantiles of T_1, respectively. From the adjusted rank test-T_1, when there are many data sets with the equal numbers of L and of (m_h, n_h) for $h = 1,2,...,L$, the same critical value will be used to test the hypothesis at the same significance level α.

If the total number of clusters is large, the asymptotic normal distribution of the adjusted rank test statistic-T_1 will be established. Note that the indicator function δ_{hi} is correlated, the Central Limit Theorem cannot be directly applied. The distribution of W^*_h in stratum h is equivalent to the distribution of the sample total of R^*_{hi} when the subset of R^*_{hi} is drawn by simple random sampling without replacement with sample size m_h from the finite population size N_h. In addition, the samples between strata are independent. In stratum h, the probability that the i^{th} cluster belongs to the first sample is equal to $\frac{m_h}{N_h}$. Thus, $P(\delta_{hi}=1)=\frac{m_h}{N_h}$, $P(\delta_{hi}=0)=\frac{n_h}{N_h}$, and $Cov(\delta_{hi},\delta_{hj}) = -\frac{m_h n_h}{N_h^2(N_h-1)}$.
The expected value of the adjusted rank test statistic-T_1 and its variance are given by

$$E(T_1) = \sum_{h=1}^{L} \frac{m_h}{2}\left(N_h + 2N^*_{h-1} + 1\right), \tag{4}$$

and

$$Var(T_1) = \sum_{h=1}^{L} \frac{m_h n_h S_h^2}{N_h}, \tag{5}$$

respectively, where $N_0 = 0$, $N^*_h = \sum_{l=1}^{h} N_l$, $S_h^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(R^*_{hi} - \bar{R}_h^2\right)$, and $\bar{R}_h + N^*_{h-1} + \frac{N_h+1}{2}$.

Theorem 1 *Suppose that we have two independent samples* $\mathbf{X}$ *and* $\mathbf{Y}$ *stratified into* L *strata by cluster size. The numbers of clusters of* $\mathbf{X}$ *and* $\mathbf{Y}$ *consist of* $m = \sum_{h=1}^{L} m_h$ *and* $n = \sum_{h=1}^{L} n_h$, *respectively where* m_h *and* n_h *are numbers of clusters in stratum* h *of* $\mathbf{X}$ *and* $\mathbf{Y}$, *respectively. The cluster sizes in stratum* h *are equal to* c_h *for* $h = 1,2,\ldots,L$. *Let* T_1, $E(T_1)$, *and* $Var(T_1)$ *be defined in (3), (4), and (5), respectively. Under null hypothesis that the two samples are drawn from the same population, the sufficient conditions for the standardized statistic*

$$Z_1 = \frac{T_1 - \sum_{h=1}^{L}\left[m_h\left(\sum_{l=0}^{h-1} N_l + \frac{N_h+1}{2}\right)\right]}{\sqrt{\sum_{h=1}^{L}\frac{m_h n_h S_h^2}{N_h}}} \tag{6}$$

to be asymptotically normal distributed according to $N(0,1)$ *are that (a) the number of strata* (L) *is finite, (b) for all strata* $\frac{m_h}{N_h} \to \xi_h, 0 < \xi_h < 1, h = 1,2,\ldots,L$ *as* $N \to \infty$ *provided that* m_h *and* n_h *trends to infinity.*

Proof. Let $\{\Pi_N\}_{N=L}^{\infty}$ be a sequence of finite populations portioned into L strata. Π_N consists of N_h elements of R_{hi}^*'s where $N_h \geq 1$ and $\sum_{h=1}^{L} N_h = N$. Suppose that from each stratum of the population Π_N, a random sample without replacement of size m_h is drawn from a population with stratum size N_h. The selection of samples between strata is independent. The total sample size is equal to $m = \sum_{h=1}^{L} m_h$. The adjusted rank test is defined as the sum of the sampled ranks:

$$T_N = \sum_{h=1}^{L} \sum_{i=1}^{m_h} R_{hi}^*.$$

Given that the terms of T_N are dependent, the central limit theorem cannot be applied. We will construct another sequence, which is asymptotically equivalent to T_N in terms of corollary 2 from Lehmann (1975, p.349). The statistic S_N will be constructed such that $\left[S_N - E(S_N)\right] / \sqrt{Var(S_N)} \to N(0,1)$ and $E(S_N - T_N)^2 / Var(S_N) \to 0$ as $N \to \infty$. Theorem 5 of Lehmann (1975, p.345) will be used to create the statistic S_N.

Let U_{hi} be independent random variables with uniform distribution on the interval (0,1) for $i = 1, 2, ..., N_h$ and $h = 1, 2, ..., L$. In stratum h, the process of drawing a random sample of size m_h from the population Π_N can be defined by including the rank R_{hi}^* if and only if U_{hi} is one of the m_h smallest U_{hi} that is if $O_{hi} < m_h$ where O_{hi} is the rank of U_{hi}. Given that each set of m_h of U_{hi} is equally likely to create the set of m_h smallest of U_{hi}, each of the $\binom{N_h}{m_h}$ possible sample is equally likely. We define the independent variables $K_{hi} = \begin{cases} 1 & \text{if } U_{hi} < \dfrac{m_h}{N_h} \\ 0 & \text{otherwise,} \end{cases}$ and define the statistic $S_N = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \left[\left(R_{hi}^* - \bar{R}_h \right) K_{hi} + m_h \bar{R}_h \right]$. Using algebra, we can prove the asymptotic normality of S_N provided that conditions (a) and (b) hold.

To complete the proof of Theorem 1, we must prove that $E\left[S_N - T_N\right]^2 / Var(S_N) \to 0$ as $N \to \infty$. Let $a_{Nh}(u_h) = \begin{cases} 1 & \text{if } \quad u_h \leq \dfrac{m_h}{N_h} \\ 0 & \text{otherwise,} \end{cases}$ be an indicator function for stratum h. Let $J_{hi} = a_{Nh}\left(\dfrac{R_{hi}}{N_h}\right)$ and $K_{hi} = a_{Nh}(U_{hi}) = a_{Nh}\left(U_{(\mathrm{R}_{hi})}\right)$ where $U_{(h1)} < U_{(h2)} < ... < U_{(hN_h)}$ denote the ordered U_{hi} and R_{hi} is the rank of U_{hi}. After extensive algebra and using sampling theory together with the condition (a), we have

$$\frac{E\left(S_N-T_N\right)^2}{Var\left(S_N\right)} \le \frac{\sum_{h-1}^{L}\left[\sqrt{\frac{m_h n_h}{N_h}}\frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(R_{hi}^{*}-\overline{R}_h\right)^2\right]}{\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(R_{hi}^{*}-\overline{R}_h\right)^2\frac{m_h n_h}{N_h^2}}$$

$$=\frac{\sum_{h-1}^{L}S_h^2\sqrt{\frac{m_h n_h}{N_h}}}{\sum_{h=1}^{L}S_h^2\left(\frac{N_h-1}{N_h}\right)\frac{m_h n_h}{N_h}} \quad \text{where } S_h^2=\frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(R_{hi}^{*}-\overline{R}_h^2\right)$$

$$\le L\cdot\max\left\{\frac{S_h^2\sqrt{\frac{m_h n_h}{N_h}}}{S_h^2\left(\frac{N_h-1}{N_h}\right)\frac{m_h n_h}{N_h}}\right\}$$

$$=L\cdot\frac{S_{h*}^2\sqrt{\frac{m_{h*}n_{h*}}{N_{h*}}}}{S_{h*}^2\left(\frac{N_h-1}{N_h}\right)\frac{m_{h*}n_{h*}}{N_{h*}}}$$

$$=L\cdot\left(\frac{N_{h*}}{N_{h*}-1}\right)\sqrt{\frac{N_{h*}}{m_{h*}n_{h*}}},$$

where h^* is the stratum that contains the maximum value. Without loss of generality, assume that $m_{h*}>n_{h*}$, so that $m_{h*}>\frac{1}{2}N_{h*}$. Then $L\cdot\left(\frac{N_{h*}-1}{N_{h*}}\right)\sqrt{\frac{m_{h*}n_{h*}}{N_{h*}}}\le L\cdot\left(\frac{N_{h*}-1}{N_{h*}}\right)\sqrt{\frac{2}{m_{h*}}}\to 0$ as m_{h*} and N_{h*} tend to infinity .This completes the proof of Theorem 1.

4. Adjusted Rank Test for Three or More Samples

In this section, we propose the clustered rank sum test and adjusted rank test for three or more independent samples with clustered data in unbalanced design. Assume that the observations consist of $k\ge 3$ samples. The null hypothesis is that the k samples have been drawn from the same population. Let X_{ijr} be the r^{th} observation in the cluster j of the i^{th} sample for $i=1,2,\ldots,k$, $j=1,2,...,n_i$, and $k=1,2,...,c_{ij}$, where n_i is the number of clusters in the i^{th} sample and c_{ij} is the cluster size of the j^{th} cluster in the i^{th} sample**.** We also consider the case of unbalanced design, i.e., the unequal numbers of clustered sizes for some $i=1,2,...,k$ and $j=1,2,...,n_i$.

To compute the clustered rank sum test statistics, we stratify the observations into L strata by cluster size. Assume that the first stratum contains the observations with the smallest cluster size, the second stratum contains the observations with the next higher cluster size, and so on. The L^{th} stratum contains the observations with the largest cluster size. Let n_{ih} be the number of cluster in the stratum h from the i^{th} sample and c_h be the cluster size in stratum $h.$ The total number of clusters in the h -stratum is denoted by $N_h=\sum_{i=1}^{k}n_{ih}$. We denote that X_{ihjr} is the r^{th} observation in the cluster j from the h^{th} stratum of sample i. Let R_{ihjr} be the observation rank of X_{ihjr} based on all observations from k samples. If two or more observations are equal, assign each a rank of the mean of the rank positions. The sum of ranks from the i-sample is defined as $T_i = \sum_{h=1}^{L}\sum_{j=1}^{n_{ih}}\sum_{r=1}^{c_h} R_{ihjr} = \sum_{h=1}^{L}\sum_{j=1}^{n_{ih}} R_{ihj+}$ where $R_{ihj+} = \sum_{r=1}^{c_h} R_{ihjr}$.

Assume that the observations in a given cluster are exchangeable. Under the null hypothesis, the distribution of the statistic-T_i conditioning on R_{ih+} can be generated by combining all possible permutations of R_{ihj+} in T_i for a given cluster of size c_h. The total number of permutation is equal to $\frac{N_h!}{n_{1h}!n_{2h}!\cdots n_{kh}!}$. Moreover, we can derive that the expected value of the statistic T_i is equal to $E(T_i) = \sum_{h=1}^{L}\sum_{j=1}^{N_h}\frac{n_{ih}}{N_h}R_{ihj+}$. If the null hypothesis is true, we expect that the sum of ranks is equal to its expected value. The clustered rank sum test statistic-T_2 is defined as the weighted sum of squares of deviations of sums of ranks from its expected value:

$$T_2 = \sum_{i=1}^{k}\frac{\left(T_i - \sum_{h=1}^{L}\sum_{j=1}^{N_h}\frac{n_{ih}}{N_h}R_{ihj+}\right)^2}{\sum_{h=1}^{L}\sum_{j=1}^{N_h}\frac{n_{ih}}{N_h}R_{ihj+}}. \tag{7}$$

If N is small, the exact distribution of statistic-T_2 conditioning on R_{ihj+} can be generated by combining all possible permutations of R_{ihj+} for a given cluster of size c_h. The total number of permutation for all strata is equal to $\prod_{h=1}^{L}\frac{N_h!}{n_{1h}!n_{2h}!\cdots n_{kh}!}$. To conduct an α-level test of the null hypothesis that the k samples are drawn from the same population, the test statistic of T_2 can be compared with the $(1-\alpha)100\%^{\text{th}}$ percentile of T_2, so that the null hypothesis is rejected if the statistic of T_2 is greater than or equal to this percentile. If there are many equivalent datasets but different observations, the $(1-\alpha)100\%^{\text{th}}$ percentile of statistic-T_2 may vary between datasets. To obtain the same percentile of a test statistic, we propose an adjusted rank test-T_3 as follows.

From the sum of observation ranks, let $\bar{R}_{ihj} = \frac{1}{c_h}\sum_{r=1}^{c_h} R_{ihjr}$ be the mean of ranks from the j^{th} cluster in stratum h of the i^{th} sample. Similar to Section 3, for the first stratum, we assign rank 1 for the cluster with the smallest cluster mean, rank 2 for the cluster with the next higher cluster mean, and so on. The highest cluster mean is assigned with the rank N_1. For the second stratum, we assign rank $N_1 + 1$ through $N_1 + N_2$ for the cluster with the smallest cluster mean to the cluster with the highest cluster mean. The ranks are continuously assigned to the next stratum. For the L^{th} stratum, the rank $\sum_{h=1}^{L-1} N_h + 1$ is assigned to the cluster with the smallest cluster mean, and so on. The rank $\sum_{h=1}^{L} N_h = N$ is assigned to the cluster with the largest cluster mean. The average of the ranks is assigned to clusters with the same cluster mean. Let Z_{ihj} be the rank of $\overline{R}_{ihj}$. The sum of new ranks from the i-sample is defined as $T_i' = \sum_{h=1}^{L}\sum_{j=1}^{n_{ih}} Z_{ihj}$ for $i = 1, 2, ..., k.$ Under null hypothesis, the expected value of the statistic-T_i' is equal to $E(T_i') = \sum_{h=1}^{L} \frac{n_{ih}}{2}\left(N_h + 2N_{h-1}^* + 1\right)$ where $N_0 = 0$ and $N_h^* = \sum_{l=1}^{h} N_l.$ The adjusted test statistic-T_3 is defined as

$$T_3 = \sum_{i=1}^{k} \frac{\left[T_i' - \sum_{h=1}^{L} \frac{n_{ih}}{2}\left(N_h + 2N_{h-1}^* + 1\right)\right]^2}{\sum_{h=1}^{L} \frac{n_{ih}}{2}\left(N_h + 2N_{h-1}^* + 1\right)}. \tag{8}$$

In stratum h, the probability of observing each the rank values from the i-sample $\left(z_{ih1}, z_{ih2}, ..., z_{ihn_{ih}}\right)$ is also equally likely from the $\frac{N_h!}{n_{1h}!n_{2h}!...n_{kh}!}$ of all possible permutations. To test the hypothesis at α-level test, the test statistic of T_3 be compared to the critical value t_α' which is the $(1-\alpha)100\%^{\text{th}}$ percentile of T_3 so that the null hypothesis is rejected if the statistic of T_3 greater than or equal to t_α'. Moreover, if there are many equivalence datasets, the test statistic-T_3 use the same critical value at the same size of the test.

5. Critical Values

In this section, we generate the critical values for the adjusted rank test statistic-T_1 and the approximate critical values for the adjusted rank test-T_3 at alpha values of 0.10 and 0.05. The exact significance levels of the statistic-T_1 are also presented in Table 1. We consider that the number of strata is equal to 2. In each stratum, we set the numbers of clusters in the samples to be equal. The numbers of clusters in the first stratum are equal to 3, 4, and 5; the numbers of clusters in the second stratum are set to be 3, 4, 5, and 6.

The critical values $t_{\alpha/2}^*$ and $t_{1-\alpha/2}^*$ for the adjusted rank test statistic-T_1 are determined by cutting the most extreme $(\alpha/2)100\%$ and $(1-\alpha/2)100\%$ of the exact distribution of the test statistic-T_1, where α is the level of significance. These critical values are obtained from the enumeration of all possible distinct permutations of the rank (R_{hi}^*) in T_1 for a given stratum h. An exact significance level is also obtained by enumeration of the statistic values of T_1, which extend to the critical value.

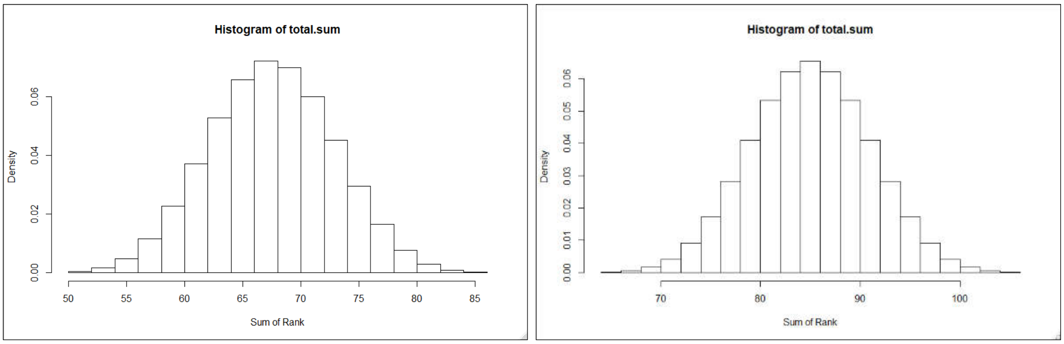
The approximate critical values t_α' of the adjusted rank test statistic-T_3 are also obtained in Table 2 for $k = 3$. These values are constructed by generating the random ranks in each stratum h from N_{h-1}^* to N_h^* into k samples of sizes n_{1h}, n_{2h}, n_{3h}. We then calculate the rank sum of each sample and compute the test statistic-T_3. The procedure is replicated a number of 1,000,000 times. The percentile values of 0.90 and 0.95 are selected to be the critical values at alpha values of 0.10 and 0.05, respectively.

Table 1 Critical values of the adjusted rank statistic- T_1 at significance levels of 0.10 and 0.05 with $L=2$

Number of Clusters	$\alpha=0.10$		$\alpha=0.05$	
	$t^*_{0.05}$	$t^*_{0.95}$	$t^*_{0.025}$	$w^*_{0.975}$
$m_1=3, n_1=3$ $m_2=3, n_2=3$	33 (0.04500)	45 (0.95500)	32 (0.02000)	46 (0.98000)
$m_1=3, n_1=3$ $m_2=4, n_2=4$	45 (0.04571)	60 (0.95429)	43 (0.01286)	62 (0.98714)
$m_1=3, n_1=3$ $m_2=5, n_2=5$	58 (0.03611)	78 (0.96389)	57 (0.02242)	79 (0.97758)
$m_1=3, n_1=3$ $m_2=6, n_2=6$	74 (0.04946)	97 (0.95054)	72 (0.02408)	99 (0.97592)
$m_1=4, n_1=4$ $m_2=3, n_2=3$	45 (0.04571)	60 (0.95429)	43 (0.01286)	62 (0.98714)
$m_1=4, n_1=4$ $m_2=4, n_2=4$	59 (0.04102)	77 (0.95898)	58.475 (0.02510)	77.525 (0.97490)
$m_1=4, n_1=4$ $m_2=5, n_2=5$	75 (0.04535)	96 (0.95465)	73 (0.01984)	98 (0.98016)
$m_1=4, n_1=4$ $m_2=6, n_2=6$	92 (0.03980)	118 (0.96020)	90 (0.01991)	120 (0.98009)
$m_1=5, n_1=5$ $m_2=3, n_2=3$	58 (0.03611)	78 (0.96389)	57 (0.02242)	79 (0.97758)
$m_1=5, n_1=5$ $m_2=4, n_2=4$	75 (0.04535)	96 (0.95465)	73 (0.01984)	98 (0.98016)
$m_1=5, n_1=5$ $m_2=5, n_2=5$	93 (0.04472)	117 (0.95528)	91 (0.02206)	119 (0.97794)
$m_1=5, n_1=5$ $m_2=6, n_2=6$	113 (0.04951)	140 (0.95049)	110 (0.02001)	143 (0.97999)

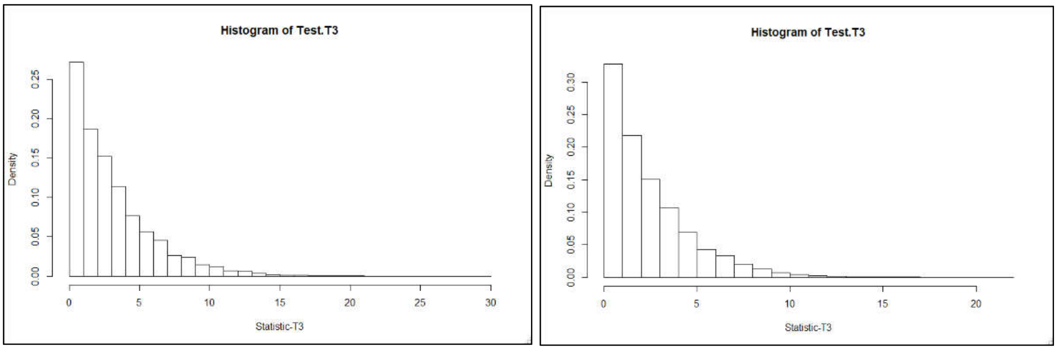
The values in bracket are $P(T_1 \le t^*_{\alpha/2})$ and $P(T_1 \le t^*_{1-\alpha/2})$.

Figure 1 shows the distribution of the adjusted rank test statistic- T_1 for different numbers of clusters. We find that the distribution of the adjusted rank test statistic- T_1 is symmetric. When the number of clusters of m_h and n_h increase, the mean and variance of this test statistic also increase. Figure 2 shows the distribution of the adjusted rank test statistic- T_3 for $k=3$ samples at different numbers of clusters. The adjusted rank test statistic- T_3 is presents a right-skewed distribution.



(a) (b)

Figure 1 Distribution of the adjusted rank statistic-T_1 (a) for $m_1 = 3, n_1 = 3,\ m_2 = 5, n_2 = 5$ and (b) for $m_1 = 4, n_1 = 4,\ m_2 = 5, n_2 = 5$



(a) (b)

Figure 2 Distribution of the adjusted rank statistic-T_3 with $k = 3$ samples (a) for $(n_{11}, n_{12}, n_{13}) = (3,3,3)$, $(n_{21}, n_{22}, n_{23}) = (5,5,5)$, and (b) for $(n_{11}, n_{12}, n_{13}) = (3,3,3)$, $(n_{21}, n_{22}, n_{23}) = (6,6,6)$

6. Simulation Study

In this section, we study the properties of the adjusted rank tests consisting of the probability of type I error and the power of the tests via simulation study. The properties of the adjusted rank test-T_1 are compared with those of the RGL test-T and the Wilcoxon rank sum test based on the cluster means of the observations denoted by T_W. The properties of the adjusted rank test-T_3 are also compared with the Kruskal-Wallis test based on the cluster means of the observations denoted by T_{KW}. The control of probability of type I error is evaluated based on the criterion of Bradley (1978). If the type I error rate belongs to (0.0250, 0.0750) for a significance level of 0.05, then the test can protect the probability of type I error.

Table 2 Critical values of the adjusted rank statistic-T_3 at significance levels of 0.10 and 0.05 with $L = 2$ for $k = 3$ samples

Number of Clusters	$t'_{0.90}$	$t'_{0.95}$
$(n_{11},n_{21},n_{31})=(3,3,3)$ $(n_{12},n_{22},n_{32})=(3,3,3)$	3.61403	4.52632
$(n_{11},n_{21},n_{31})=(3,3,3)$ $(n_{12},n_{22},n_{32})=(4,4,4)$	4.44156	5.63636
$(n_{11},n_{21},n_{31})=(3,3,3)$ $(n_{12},n_{22},n_{32})=(5,5,5)$	5.54000	7.02000
$(n_{11},n_{21},n_{31})=(3,3,3)$ $(n_{12},n_{22},n_{32})=(6,6,6)$	7.00000	8.87302
$(n_{11},n_{21},n_{31})=(4,4,4)$ $(n_{12},n_{22},n_{32})=(3,3,3)$	4.44156	5.63636
$(n_{11},n_{21},n_{31})=(4,4,4)$ $(n_{12},n_{22},n_{32})=(4,4,4)$	4.82000	6.02000
$(n_{11},n_{21},n_{31})=(4,4,4)$ $(n_{12},n_{22},n_{32})=(5,5,5)$	5.44444	7.00000
$(n_{11},n_{21},n_{31})=(4,4,4)$ $(n_{12},n_{22},n_{32})=(6,6,6)$	6.59355	8.40000
$(n_{11},n_{21},n_{31})=(5,5,5)$ $(n_{12},n_{22},n_{32})=(3,3,3)$	5.54000	7.02000
$(n_{11},n_{21},n_{31})=(5,5,5)$ $(n_{12},n_{22},n_{32})=(4,4,4)$	5.44444	7.00000
$(n_{11},n_{21},n_{31})=(5,5,5)$ $(n_{12},n_{22},n_{32})=(5,5,5)$	5.89677	7.58710
$(n_{11},n_{21},n_{31})=(5,5,5)$ $(n_{12},n_{22},n_{32})=(6,6,6)$	6.67380	8.48128

The study is constructed under two and three samples with $L=2$ strata. The numbers of observations in stratum (c_1,c_2) are equal to (2, 3), (3, 4) and (4, 5). In case of two samples, we set the equal number of clusters in the samples $m_1=n_1$, $m_2=n_2$ consisting of (3, 3) and (4, 4). In case of three samples, we also set equal numbers of clusters, $n_{11}=n_{21}=n_{31}$, $n_{12}=n_{22}=n_{32}$ with (3, 3, 3) and (4, 4, 4).

We generate data $X_{ihjr} = \exp\left(Y_{ihjr}\right) + \left(i-1\right)d,$ where $\mathbf{Y}_{ihj} = \left(Y_{ihj1}, Y_{ihj2}, \ldots, Y_{ihjc_h}\right)$ is an independent multivariate normal with mean vector $\mathbf{0}$ and exchangeable covariance matrix $\Sigma = (1-\rho)\mathbf{I} + \rho\mathbf{1},$ where $\mathbf{I}$ is the identity matrix of size $c_h \times c_h$ and $\mathbf{1}$ is the $c_h \times c_h$ matrix of all elements equal to 1.

For each case, the coefficient of correlation between observations in a cluster (ρ) is set to be 0.1, 0.3, 0.5, 0.7, and 0.9. The effect size (d) is equal to 0.0, 0.3, and 0.5. For each situation, the rejection rate is computed from 10,000 replicates. The results are summarized in Tables 3-6.

Table 3 Estimated probability of type I error $(d = 0.0)$ of the test statistics $T_1,$ T and T_W tests at the significance level of 0.05

(c_1, c_2)	ρ	$(m_1,n_1)=(3,3),(m_2,n_2)=(3,3)$			$(m_1,n_1)=(4,4),(m_2,n_2)=(4,4)$		
		T_1	T	T_W	T_1	T	T_W
	0.1	0.0346	0.0416	0.0436	0.0418	0.0425	0.0490
	0.3	0.0360	0.0427	0.0420	0.0436	0.0423	0.0474
(2, 3)	0.5	0.0357	0.0421	0.0430	0.0420	0.0418	0.0464
	0.7	0.0362	0.0424	0.0416	0.0452	0.0431	0.0482
	0.9	0.0378	0.0446	0.0418	0.0451	0.0441	0.0482
	0.1	0.0342	0.0415	0.0393	0.0485	0.0484	0.0532
	0.3	0.0342	0.0405	0.0409	0.0491	0.0476	0.0523
(3, 4)	0.5	0.0363	0.0411	0.0396	0.0494	0.0489	0.0508
	0.7	0.0359	0.0437	0.0397	0.0504	0.0488	0.0524
	0.9	0.0384	0.0434	0.0396	0.0515	0.0500	0.0534
	0.1	0.0413	0.0473	0.0419	0.0467	0.0440	0.0521
	0.3	0.0405	0.0473	0.0426	0.0492	0.0439	0.0496
(4, 5)	0.5	0.0418	0.0469	0.0412	0.0495	0.0447	0.0493
	0.7	0.0422	0.0487	0.0416	0.0494	0.0457	0.0495
	0.9	0.0415	0.0495	0.0412	0.0495	0.0475	0.0487

Table 3 shows that the empirical probabilities of type I error of the adjusted rank test statistic-T_1 are within the Bradley's criterion of the interval (0.025, 0.075). Under $(m_h, n_h) = (3,3)$ and (4,4) in Table 1, the exact significance levels are equal to 0.040 and 0.052, respectively. The empirical probability of type I error is close to the exact significance levels. The RGL test-T and the Wilcoxon test based on observation mean test-T_W can also control the probability of type I error.

Table 4 Estimated power of the tests for $T_1,$ T and T_M at the significance level of 0.05

d	(c_1, c_2)	ρ	$(m_1,n_1)=(3,3),(m_2,n_2)=(3,3)$	$(m_1,n_1)=(4,4),(m_2,n_2)=(4,4)$

			T_1	T	T_M	T_1	T	T_M
		0.1	0.1085	0.1157	0.0757	0.1424	0.1523	0.1016
		0.3	0.0949	0.1028	0.0731	0.1229	0.1289	0.0977
	(2, 3)	0.5	0.0870	0.0919	0.0724	0.1109	0.1131	0.0964
		0.7	0.0803	0.0858	0.0739	0.1007	0.1044	0.0966
		0.9	0.0745	0.0804	0.0743	0.0933	0.0970	**0.0972**
		0.1	0.1265	0.1380	0.0736	0.1860	0.1967	0.1132
		0.3	0.0990	0.1109	0.0714	0.1470	0.1564	0.1065
0.3	(3, 4)	0.5	0.0877	0.0962	0.0689	0.1268	0.1300	0.1039
		0.7	0.0779	0.0878	0.0697	0.1115	0.1131	0.1018
		0.9	0.0729	0.0804	0.0708	0.1025	0.1041	0.1021
		0.1	0.1470	0.1569	0.0798	0.2100	0.2249	0.1136
		0.3	0.1091	0.1185	0.0733	0.1547	0.1616	0.1054
	(4, 5)	0.5	0.0890	0.1013	0.0715	0.1268	0.1336	0.1008
		0.7	0.0808	0.0908	0.0739	0.1103	0.1150	0.0997
		0.9	0.0769	0.0837	0.0760	0.1015	0.1044	**0.1027**
		0.1	0.2091	0.2318	0.1300	0.2974	0.3263	0.1919
		0.3	0.1746	0.1937	0.1255	0.2498	0.2709	0.1842
	(2, 3)	0.5	0.1533	0.1729	0.1238	0.2121	0.2341	0.1817
		0.7	0.1368	0.1545	0.1226	0.1923	0.2052	0.1841
		0.9	0.1273	0.1404	0.1230	0.1763	0.1831	**0.1854**
		0.1	0.2611	0.2903	0.1382	0.3798	0.4093	0.2146
		0.3	0.1998	0.2254	0.1287	0.2955	0.3165	0.1966
0.5	(3, 4)	0.5	0.1605	0.1854	0.1247	0.2441	0.2583	0.1918
		0.7	0.1389	0.1595	0.1236	0.2132	0.2218	0.1872
		0.9	0.1202	0.1425	0.1226	0.1902	0.1964	0.1898
		0.1	0.2920	0.3337	0.1448	0.4430	0.4760	0.2195
		0.3	0.2094	0.2406	0.1290	0.3221	0.3420	0.2000
	(4, 5)	0.5	0.1654	0.1901	0.1253	0.2589	0.2757	0.1885
		0.7	0.1427	0.1619	0.1230	0.2192	0.2307	0.1871
		0.9	0.1222	0.1435	**0.1233**	0.1941	0.1962	0.1882

Table 4 shows that the RGL test-T can give the highest estimated power than the other tests for almost all situations. The RGL test-T is slightly more powerful than the adjusted rank test-T_1. The adjusted rank test-T_1 gives the higher estimated empirical power than the test-T_M in almost all cases. When the effect size is fixed, the empirical power of the RGL and adjusted rank tests are slightly different when the number of clusters increases. Moreover, the empirical powers of both tests increase as the effect size, the number of clusters, or the cluster size increases when the other primer two factors are fixed at the same level of correlation. However, when all three factors are given (i.e., the effect size, the number of clusters, and the cluster size), the empirical power of both tests decreases as the observations in the given cluster are highly correlated.

Table 5 Estimated probability of type I error ($d = 0.0$) of the test statistics-T_3 and T_{KW} tests at the significance level of 0.05

(c_1, c_2)	ρ	$(n_{11}, n_{21}, n_{31}) = (3,3,3)$ $(n_{12}, n_{22}, n_{32}) = (3,3,3)$		$(n_{11}, n_{21}, n_{31}) = (4,4,4)$ $(n_{12}, n_{22}, n_{32}) = (4,4,4)$	
		T_3	T_{KW}	T_3	T_{KW}
(2, 3)	0.1	0.0461	0.2486	0.0472	0.2668
	0.3	0.0455	0.2594	0.0540	0.2626
	0.5	0.0453	0.2523	0.0515	0.2598
	0.7	0.0472	0.2465	0.0523	0.2702
	0.9	0.0503	0.2570	0.0565	0.2666
(3, 4)	0.1	0.0502	0.2558	0.0538	0.2561
	0.3	0.0507	0.2562	0.0549	0.2574
	0.5	0.0477	0.2520	0.0541	0.2638
	0.7	0.0486	0.2514	0.0525	0.2624
	0.9	0.0472	0.2598	0.0554	0.2627
(4, 5)	0.1	0.0457	0.2570	0.0528	0.2612
	0.3	0.0470	0.2511	0.0523	0.2562
	0.5	0.0454	0.2515	0.0540	0.2532
	0.7	0.0490	0.2586	0.0567	0.2605
	0.9	0.0456	0.2531	0.0505	0.2571

In Table 5, the adjusted rank test-T_3 protects better against type I error because its empirical type I errors belong to the interval (0.025, 0.075). The empirical type I errors of the adjusted rank test-T_3 are close to the significance level of 0.05, whereas the type I error rates of the Kruskal-Wallis test based on the cluster means-T_{KW} exceed the nominal level by about five times. The average of the empirical type I error of the Kruskal-Wallis test based on the cluster means-T_{KW} is equal to 0.2576.

In Table 6, the empirical power of the Kruskal-Wallis test based on the cluster means-T_{KW} is not included in the table because it fails to maintain the probability of type I error. Given the effect size (d), the cluster size (c_1, c_2), and the numbers of clusters, the empirical power of the test-T_3 decreases as the coefficient of correlation between observations increases. When the effect size (d), the cluster size (c_1, c_2), and the correlation coefficient are fixed, the empirical power of the test-T_3 increases as the number of clusters increases. Fixing the cluster size, the correlation coefficient, and the number of clusters, the empirical power of the adjusted rank test-T_3 increases when the effect size increases. Finally, when the effect size, the coefficient of correlation, and the number of clusters are fixed, the empirical power of the test-T_3 increases as the cluster size increases. To increase the empirical power of the test-T_3, we can increase the number of observations. In case of (n_{1h}, n_{2h}, n_{3h}) = (3,3,3), (c_1, c_2) = (3,4), and (n_{1h}, n_{2h}, n_{3h}) = (4,4,4), (c_1, c_2) = (2, 3), the total number of observation is about 60. The empirical power from the situation of (n_{1h}, n_{2h}, n_{3h}) = (4, 4, 4), is higher than that from the case of (n_{1h}, n_{2h}, n_{3h}) = (3,3,3), (c_1, c_2) = (3,4). Thus, under the same number of observations, increasing the number of clusters will result in higher empirical power than increasing the cluster size.

Table 6 Estimated power of the tests for T_3 and T_{KW} at a significance level of 0.05

d	(c_1, c_2)	ρ	$(n_{11}, n_{21}, n_{31}) = (3,3,3)$ $(n_{12}, n_{22}, n_{32}) = (3,3,3)$		$(n_{11}, n_{21}, n_{31}) = (4,4,4)$ $(n_{12}, n_{22}, n_{32}) = (4,4,4)$	
			T_3	T_{KW}	T_3	T_{KW}
0.3	(2, 3)	0.1	0.2302	-	0.3345	-
		0.3	0.2000	-	0.2809	-
		0.5	0.1756	-	0.2432	-
		0.7	0.1520	-	0.2185	-
		0.9	0.1426	-	0.1992	-
	(3, 4)	0.1	0.2970	-	0.4289	-
		0.3	0.2254	-	0.3279	-
		0.5	0.1746	-	0.2537	-
		0.7	0.1543	-	0.2262	-
		0.9	0.1451	-	0.2006	-
	(4, 5)	0.1	0.3504	-	0.5035	-
		0.3	0.2494	-	0.3453	-
		0.5	0.1936	-	0.2767	-
		0.7	0.1592	-	0.2340	-
		0.9	0.1445	-	0.1997	-
0.5	(2, 3)	0.1	0.4850	-	0.6674	-
		0.3	0.4166	-	0.5705	-
		0.5	0.3625	-	0.4930	-
		0.7	0.3081	-	0.4470	-
		0.9	0.2802	-	0.4023	-
	(3, 4)	0.1	0.6072	-	0.7855	-
		0.3	0.4763	-	0.6573	-
		0.5	0.3790	-	0.5337	-
		0.7	0.3260	-	0.4626	-
		0.9	0.2860	-	0.4011	-
	(4, 5)	0.1	0.6932	-	0.8601	-
		0.3	0.5187	-	0.6899	-
		0.5	0.4046	-	0.5646	-
		0.7	0.3230	-	0.4721	-
		0.9	0.2800	-	0.4148	-

7. Conclusions

An attractive feature of rank transformation is its ability to deal with any problem of skewness because all ranks are equally far apart from each other. By ranking the data, the influence of outliers is mitigated: regardless of how extreme an outlier is; it receives the same rank as if it were only slightly larger than the second-largest observation. To test the equality of location parameters from the two or more independent samples with clustered data in unbalanced design, this paper proposed the adjusted rank tests by using the rank transformation of the cluster means of observation ranks. An assumption of the adjusted rank tests is that all observations are exchangeable in a cluster. The adjusted rank tests use the same critical value for data sets with equivalence such as the numbers of clusters and cluster sizes at the same significance level. The critical values of both adjusted rank tests for some numbers of clusters with two strata are given at the nominal levels of 0.10 and 0.05. The efficiency of the adjusted rank tests were compared with those of the alternative tests via simulation study. For all situations, the adjusted rank tests can maintain the size of the test because the adjusted rank tests use the permutation of ranks with exact significance level close to the nominal level 0.05. The Kruskal-Wallis test based on the cluster means of observations -T_{KW} cannot protect type I error. The adjusted rank test-T_1 has more empirical power than the Wilcoxon test based on the cluster means of observations. The empirical power of the adjusted rank test-T_1 is slightly different from the empirical power of the RGL test-T. The empirical power of both adjusted rank tests increases as the cluster size increases, and the number of clusters increases as the effect size increases. However, the empirical power of both adjusted rank tests decreases as the correlation coefficient of observations in clusters increases. To increase the power of the adjusted rank tests by increasing the number of observations, we suggest that increasing the number of observations by increasing the number of clusters will result in the power of the test to be higher than that by increasing the cluster size.

Acknowledgements

This work was supported by Silpakorn University Research, Innovation and Creativity Administration Office, Thailand. Special thanks are extended to the referees for their suggestions to improve the manuscript.

References

Bradley JV. Robustness?. Br J Math Stat Psychol. 1978; 31: 144-152.

Lahiri P, Yan L. A new alternative to the standard F test for clustered data. J Stat Plan Inference. 2009; 139: 3430-3441.

Lehman EL. Nonparametric: Statistical methods based on ranks. San Francisco: Holden-Day; 1975.

Rao JNK, Sutradhar BC, Yue K. General least squares F-test in regression analysis with two-stage cluster samples. J Am Stat Assoc. 1993; 88: 1388-1391.

Rosner B, Glynn RJ, Lee ML. Incorporation of clustering effect for the Wilcoxon rank sum test: a large sampled approach. Biometrics. 2003; 59(4): 1089-1098.

Rosner B, Grove D. Use of the Mann-Whitney U-test for clustered data. Stat Med. 1999; 18: 1387-1400.

Rosner B, Glynn RJ. Power and sample size estimation for the Wilcoxon rank sum test with application to comparisons of C statistics from alternative prediction models. Biometrics. 2009; 65(1): 188-197.

Rosner B, Glynn RJ, Lee ML. The Wilcoxon signed rank test for paired comparisons of clustered data. Biometrics. 2006; 62(1): 185-192.

Sangngam P, Laoarun W. Adjusted rank tests for clustered data in balanced design. SEHS. 2021; 15: 21020002.

Wilcoxon F. Individual comparisons by ranking methods. Biometrics. 1945; 1(6): 80-83.

Wu CFJ, Holt D, Holmes DJ. The effect of two-stage sampling of the F-statistic. J Am Stat Assoc 1988; 83(401): 150-159.