Scientific Research Publishing

Advances in Pure Mathematics, 2026, 16(1), 19-28
https://www.scirp.org/journal/apm
ISSN Online: 2160-0384
ISSN Print: 2160-0368

Comparing Two Ranking Methods for Simultaneous Mean Differences

Lukian Peter Wernyj[1], **John Edward Kolassa**[2]

[1]School of Arts and Sciences, Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA
[2]Department of Statistics, Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA
Email: kolassa@stat.rutgers.edu

How to cite this paper: Wernyj, L.P. and Kolassa, J.E. (2026) Comparing Two Ranking Methods for Simultaneous Mean Differences. *Advances in Pure Mathematics*, **16**, 19-28.
https://doi.org/10.4236/apm.2026.161002

Received: November 12, 2025
Accepted: January 11, 2026
Published: January 14, 2026

Abstract

When comparing means of more than two groups, powers of two ranking methods for pairwise nonparametric simultaneous inference are compared, for various alternative hypotheses.

Keywords

Ranks, Honest Significant Difference, Simultaneous Inference, Wilcoxon Rank Sum Test

1. Introduction

Suppose that X_{ki} are samples from K potentially different populations, where k indexes groups, and i indexes the observation within the group. Let $M_k > 0$ represent the number of items in such a sample from the population k. That is, for fixed k, $X_{k1}, \cdots, X_{kM_k}$ are independent and identically distributed, each with a cumulative distribution function F_k. Here, $k \in \{1, \cdots, K\}$ indexes group, and M_k represents the number of observations in each group. In order to determine whether all populations are the same, test a null hypothesis

$$H_0 : F_1(x) = \cdots = F_K(x) \forall x, \tag{1}$$

vs. the alternative hypothesis H_A: there exists j, k, and x such that $F_j(x) \neq F_k(x)$. Tests considered in this paper, however, are most powerful against alternatives of the form $H_A : F_k(x) \leq F_j(x) \quad \forall x$ for some indices k, j, with strict inequality at some k, j, and x. Of particular interest, particularly for power calculations, are alternatives of the form

$$F_i(x - \theta_i) = F_j(x - \theta_j) \tag{2}$$

for some constants $\theta_1, \cdots, \theta_K$. While the shift alternative may seem restrictive, note that data ranks are used in this procedure, and that ranking is invariant to monotonic transformation of the data. Hence, for example, a multiplicative effect might be captured by such an alternative by taking logs of the original data, reducing this back to the additive model exhibited in this manuscript.

2. Background on Testing in Multiple Groups

This section reviews existing multiple comparison approaches for data with a distribution that is close to Gaussian, and for data in which the Gaussian assumption is inappropriate. This section is adapted from [1].

2.1. The Approach for Gaussian Data

Under the assumptions that the data X_{ki} are Gaussian and homoscedastic (that is, having equal variances) under both the null and alternative hypotheses, the null hypothesis H_0 in (1) is equivalent to the assumption that all observations have the same location parameter. This location could be measured by expectation or population median; we use the population median because methods discussed later in this paper apply to distributions with undefined expectation. One might test H_0 vs. H_A via analysis of variance. Let $\bar{X}_{k.} = \sum_{i=1}^{M_k} X_{ki} / M_k$, $\bar{X}_{..} = \sum_{k=1}^{K} \sum_{i=1}^{M_k} X_{ki} \Big/ \sum_{k=1}^{K} M_k$, and

$$W_A = \frac{\left[\sum_{k=1}^{K} M_k \left(\bar{X}_{k.} - \bar{X}_{..} \right)^2 \right] \Big/ (K-1)}{S^2}, \tag{3}$$

for

$$S^2 = \left[\sum_{k=1}^{K} \sum_{i=1}^{M_k} \left(X_{ki} - \bar{X}_{k.} \right)^2 \right] \Bigg/ \left(\sum_{k=1}^{K} M_k - K \right). \tag{4}$$

Reject the null hypothesis when W_A is large [2]. Because the quantities in square brackets in (3) and (4) have χ^2 distributions with degrees of freedom in their denominators, times the item variance, the critical value for this test is given by the F distribution.

In the event that the null hypothesis of equal distributions is rejected, one naturally asks which distributions differ. Testing for group differences pairwise (perhaps using the two-sample *t*-test) allows for $K(K-1)/2$ chances to find a significant result. If each test is done at nominal size, this will inflate the family-wise error rate, or the proportion of experiments that provide any incorrect result. This family-wise error rate will be bounded by the nominal size used for each separate test multiplied by the number of possible comparisons performed ($K(K-1)/2$), but such a bound, called the Bonferroni bound, will usually result in a very conservative bound.

Define the Studentized mean differences

$$V_{jk} = \left(Z_j - \bar{Z}_k \right) \Big/ \left[S \sqrt{1/M_j + 1/M_k} \right], \tag{5}$$

where $Z_i = \bar{X}_{i.}$ for $i \in \{1, \cdots, K\}$. The Studentized mean differences V_{jk} and

V_{lm}, for $j \neq k$ and $l \neq m$ group indices, are dependent, because they share the same pooled standard deviation estimate S. The Studentized mean differences are more strongly dependent when there is overlap between $\{j,k\}$ and $\{l,m\}$, because they share one of the means. Hence, the family-wise error probability is not the product of separate probabilities.

Tukey's Honest Significant Difference method [3] [4] allows one to test for separate differences while controlling for simultaneous errors. If M_j are all equal, the distribution of $T = \sqrt{2}\max_{1\le j,k\le K}|V_{jk}|$ is called the Studentized range distribution with K and $m = \sum_{j=1}^{K}(M_j - 1)$ degrees of freedom. If M_j are not all equal, the distribution of T has the Studentized range distribution with these degrees of freedom, approximately [5]. Let $Q_{K,m}$ be the cumulative distribution function of this distribution, and let $q_{K,m,\alpha} = Q_{K,m}(1-\alpha)$ be its $1-\alpha$ quantile.

When X_{ji} are independent observations approximately distributed as Gaussian with expectation μ_j and variance σ^2, for $j \in \{1,\cdots,K\}$ and $i \in \{1,\cdots,M_j\}$, one then applies this distribution with $Y_j = (\bar{X}_j - \mu_j)/\sigma$ and U/m the standard sample variance S^2, divided by σ^2 (Note that Y_j are not yet standardized by the square root of group size). If one then sets

$$P_{jk} = \bar{Q}_{K,N-K}\left(\sqrt{2}|V_{jk}|\right), \tag{6}$$

for $N = \sum_{k=1}^{K} M_k$, then for any $\alpha \in (0,1)$,

$$\mathrm{P}\left[P_{jk} \le \alpha \text{ for any } j \neq k \text{ such that } \mu_j = \mu_k\right] \le \alpha, \tag{7}$$

and the collection of tests that rejects the hypothesis $\mu_i = \mu_j$ if $P_{jk} \le \alpha$ provides a simultaneous test size less than or equal to α. Then the minimum of these,

$$\min_{1\le j<k\le K} P_{jk}, \tag{8}$$

is a valid p-value for the null hypothesis of equality of distribution across these groups, by (7).

Furthermore, if

$$\mathcal{C}_{jk} = \bar{X}_k - \bar{X}_j \pm q_{K,N-K,\alpha} S \sqrt{1/M_j + 1/M_k}\Big/\sqrt{2}, \tag{9}$$

then

$$\mathrm{P}\left[\mu_k - \mu_j \notin \mathcal{C}_{jk} \text{ for some } j,k\right] \le \alpha. \tag{10}$$

2.2. The Rank Approach for Non-Gaussian Data

When the underlying data cannot reasonably be modeled as Gaussian, one might rank the data and substitute rank averages for observation averages in (5) to obtain pairwise p-values corrected for multiple comparisons. A central limit theorem applies to these rank averages [6]. Furthermore, the variance of these rank average differences may be calculated exactly [7]. These rankings contributing to a standardized difference V_{jk} may be done in one of two ways [7]:

1. Procedure I: The rank averages may be taken by summing ranks of items, combining all K groups and averaging ranks in groups j and k; that is, Z_j is the sum of ranks in group j, where the ranks are taken from among all observations.

2. Procedure II: The rank averages may be taken by summing ranks among only the two groups j and k, then averaging; that is, Z_j is the sum of ranks in group j, where the ranks are taken from among only groups j and k. In this case, the rank sum also depends on the group that it is compared with, in a way that is not reflected in the notation.

In either case, (5) is then employed with $S = \sqrt{N(N+1)/12}$, where for procedure I $N = \sum_{i=1}^{K} M_i$, and for procedure II, $N = M_j + M_k$ [1], and mean ranks Z_j are calculated as above. This manuscript demonstrates that the choice between these two ranking procedures impacts the power of the resulting test.

Furthermore, (10) continues to apply, and so (9) continues to have simultaneous coverage probabilities for rank mean expectations by group, although inference on rank mean expectations is seldom of interest.

For example, consider a data set consisting of three groups of three observations each, in which all entries in group 1 are smaller than all in group 2, and in turn, all entries in group 2 are smaller than all those of group 3. We compare the values of V_{13} for each of these two procedures. For procedure I, ranks in groups 1 and 3 are 1, 2, 3, and 7, 8, 9, respectively. Then in this case, $Z_1 = 2$, $Z_3 = 8$, and $S = \sqrt{9 \times 10/12} = 2.7386$. Hence $V_{13} = (8-2)\Big/\left(2.7386 \times \sqrt{1/3 + 1/3}\right) = 2.6833$. For procedure II, ranks in groups 1 and 3 are 1, 2, 3, and 4, 5, 6, respectively. Then in this case, $Z_1 = 2$, $Z_3 = 5$, and $S = \sqrt{6 \times 7/12} = 1.8708$. Hence, $V_{13} = (5-2)\Big/\left(1.8708 \times \sqrt{1/3 + 1/3}\right) = 1.9640$. In our example, at each stage, each of the two procedures compares the ranks of 6 items, but Procedure II uses only these six observations, and Procedure I uses all nine observations.

3. Methods

Various examples of multiple observations divided into groups are investigated. For each set of group sizes, various group location parameters are posited; these sets of locations are associated with the null hypothesis, and with various alternatives having powers of a magnitude to be interesting. For each sample size and group location values, 100,000 data sets consisting of the indicated numbers of variables in each group, and with the indicated location values, are randomly generated. These observations are drawn from a normal distribution and from a Laplace distribution with location defined by expectation. For each of these data sets, the two ranking procedures were applied, the Z_j and S are as in §2.2, V_{jk} are calculated as in (5), and *p*-values as in (6) are combined as in (8). Powers (including test sizes, if the group means are all identical) are calculated as the proportion of the data sets with a *p*-value less than the test size. After a simulation to check the size of the test with a nominal target size, succeeding simulations are done by resetting the test size using the appropriate quantile of the null distribution *p*-

values. Alternatively, using mixed cumulants up to order four for group rank sums [8], a Cornish-Fisher expansion [9] [10] may be used to improve the approximate test size without Monte Carlo sampling. That is, one could calculate the series in the normal deviate ([10], Section 3.12) and set the skewness to zero, and reverse the series to obtain

$$\frac{z}{1-\rho_4/8}-\frac{\rho_4 z^3/8}{3\left(1-\rho_4/8\right)^4} \tag{11}$$

for ρ_4 the (excess) kurtosis of each z statistic.

Powers are the same if the groups are rearranged, and if all of the separate shifts θ_k are offset by a constant. Hence, we need only consider alternatives with $\theta_1 = 0$, and with θ_k non-decreasing in k. All powers are calculated with 100,000 Monte Carlo samples.

4. Results

Power for the rank-based statistics for the two methods for normal data are given in **Tables 1-4**, and for Laplace data in **Tables 5-8**. These tables represent the powers for pairwise comparisons using the Tukey Studentized range distribution under the two different ranking regimes. **Table 1** presents results for small samples (group sizes 10, 10, and 10, and for 10, 5, and 5), and **Table 2** presents results for large samples (group sizes 100, 100, and 100, and for 100, 50, and 50), and for various patterns of mean values for normal data. Cases in which the means were all the same represent the null hypothesis. In null cases with nominal size 0.05 (lines 1 and 8 in both tables), the power represents the test size. These sizes in all cases were smaller than the target, by differing amounts, and the nominal size was recalibrated using (11) to yield tests with comparable true sizes (lines 2 and 9 in both tables). Tests were also recalibrated using the Monte Carlo null *p*-value distribution to yield tests with comparable true sizes (lines 3 and 10 in both tables). Lines 4 through 7 and 11 through 14 were calculated for various alternative hypotheses, using the empirical *p*-value threshold of lines 3 and 10, so that the powers compared correspond to tests of the same level (0.05).

Table 1. Powers for simultaneous testing with two ranking schemes, small samples, three groups, and normal data.

Group Means	Group Sizes	Nominal Size		Rejection Proportion	
		Proc. I	Proc. II	Proc. I	Proc. II
0 0 0	10 10 10	0.05	0.05	0.04659	0.03969
0 0 0	10 10 10	0.05*	0.05*	0.05060	0.04991
0 0 0	10 10 10	0.05651	0.06276	0.04989	0.04969
0 0.8 1.6	10 10 10	0.05651	0.06276	0.84262	0.82559
0 0 1.5	10 10 10	0.05651	0.06276	0.88530	0.87771
0 1.5 1.5	10 10 10	0.05651	0.06276	0.88593	0.87869

Continued

0 0.4 1.6	10 10 10	0.05651	0.06276	0.86419	0.85325
0 0 0	10 5 5	0.05	0.05	0.03999	0.03698
0 0 0	10 5 5	0.05*	0.05*	0.04452	0.04829
0 0 0	10 5 5	0.06042	0.07284	0.04997	0.04884
0 0.8 1.6	10 5 5	0.06042	0.07284	0.62835	0.60967
0 0 1.5	10 5 5	0.06042	0.07284	0.61739	0.58191
0 1.5 1.5	10 5 5	0.06042	0.07284	0.71357	0.70668
0 0.4 1.6	10 5 5	0.06042	0.07284	0.63105	0.60571

*Standardized group means corrected according to (11).

Table 2. Powers for simultaneous testing with two ranking schemes, large samples, three groups, and normal data.

Group Means	Group Sizes	Nominal Size		Rejection Proportion	
		Proc. I	Proc. II	Proc. I	Proc. II
0 0 0	100 100 100	0.05	0.05	0.04717	0.04716
0 0 0	100 100 100	0.05*	0.05*	0.05000	0.04971
0 0 0	100 100 100	0.05298	0.05291	0.04998	0.04991
0 0.3 0.6	100 100 100	0.05298	0.05291	0.96666	0.96566
0 0 0.5	100 100 100	0.05298	0.05291	0.94789	0.94696
0 0.4 0.4	100 100 100	0.05298	0.05291	0.81032	0.80856
0 0.1 0.5	100 100 100	0.05298	0.05291	0.90891	0.90752
0 0 0	100 50 50	0.05	0.05	0.04644	0.04617
0 0 0	100 50 50	0.05*	0.05*	0.04837	0.04844
0 0 0	100 50 50	0.05372	0.05399	0.05000	0.04976
0 0.3 0.6	100 50 50	0.05372	0.05399	0.86445	0.86322
0 0 0.5	100 50 50	0.05372	0.05399	0.76230	0.76049
0 0.4 0.4	100 50 50	0.05372	0.05399	0.65429	0.65309
0 0.1 0.5	100 50 50	0.05372	0.05399	0.71685	0.71563

*Standardized group means corrected according to (11).

Table 3 and Table 4 represent the same values for data sets with four groups.

Table 3. Powers for simultaneous testing with two ranking schemes, small samples, four groups, and normal data.

Group Means	Group Sizes	Nominal Size		Rejection Proportion	
		Proc. I	Proc. II	Proc. I	Proc. II
0 0 0 0	10 10 10 10	0.05	0.05	0.04251	0.03553
0 0 0 0	10 10 10 10	0.05*	0.05*	0.04767	0.04487

Continued

0 0 0 0	10 10 10 10	0.05637	0.06381	0.04921	0.04535
0 0.5 1 1.5	10 10 10 10	0.05637	0.06381	0.76404	0.72186
0 0 0 1.5	10 10 10 10	0.05637	0.06381	0.88791	0.86183
0 0 1.3 1.3	10 10 10 10	0.05637	0.06381	0.84704	0.82712
0 0.4 1.6 1.6	10 10 10 10	0.05637	0.06381	0.91723	0.89995
0 0 0 0	10 10 5 5	0.05	0.05	0.03939	0.02972
0 0 0 0	10 10 5 5	0.05*	0.05*	0.04584	0.04197
0 0 0 0	10 10 5 5	0.06149	0.07130	0.04998	0.04380
0 0.5 1 1.5	10 10 5 5	0.06149	0.07130	0.54143	0.46631
0 0 0 1.5	10 10 5 5	0.06149	0.07130	0.61259	0.53204
0 0 1.3 1.3	10 10 5 5	0.06149	0.07130	0.62185	0.56592
0 0.4 1.6 1.6	10 10 5 5	0.06149	0.07130	0.73711	0.67012

*Standardized group means corrected according to (11).

Table 4. Powers for simultaneous testing with two ranking schemes, large samples, four groups, and normal data.

Group Means	Group Sizes	Nominal Size		Rejection Proportion	
		Proc. I	Proc. II	Proc. I	Proc. II
0 0 0 0	100 100 100 100	0.05	0.05	0.04695	0.04626
0 0 0 0	100 100 100 100	0.05*	0.05*	0.04734	0.04719
0 0 0 0	100 100 100 100	0.05316	0.05368	0.04999	0.04974
0 0.2 0.4 0.6	100 100 100 100	0.05316	0.05368	0.96221	0.96127
0 0 0 0.4	100 100 100 100	0.05316	0.05368	0.81041	0.80833
0 0 0.4 0.4	100 100 100 100	0.05316	0.05368	0.88194	0.88166
0 0.1 0.4 0.4	100 100 100 100	0.05316	0.05368	0.80170	0.80085
0 0 0 0	100 100 50 50	0.05	0.05	0.04546	0.04476
0 0 0 0	100 100 50 50	0.05*	0.05*	0.04594	0.04625
0 0 0 0	100 100 50 50	0.05451	0.05530	0.04999	0.04988
0 0.2 0.4 0.6	100 100 50 50	0.05451	0.05530	0.84864	0.84584
0 0 0 0.4	100 100 50 50	0.05451	0.05530	0.55391	0.55065
0 0 0.4 0.4	100 100 50 50	0.05451	0.05530	0.70854	0.70591
0 0.1 0.4 0.4	100 100 50 50	0.05451	0.05530	0.61591	0.61289

*Standardized group means corrected according to (11).

Table 5 and Table 6 are analogous to Table 1 and Table 2 with Laplace data; that is, they represent powers for tests for three-group data sets of various sizes.

Table 5. Powers for simultaneous testing with two ranking schemes, small samples, three groups, and Laplace data.

Group Means	Group Sizes	Nominal Size		Rejection Proportion	
		Proc. I	Proc. II	Proc. I	Proc. II
0 0 0	10 10 10	0.05	0.05	0.04564	0.04039
0 0 0	10 10 10	0.05*	0.05*	0.04874	0.05027
0 0 0	10 10 10	0.05651	0.05226	0.04900	0.04039
0 0.8 1.6	10 10 10	0.05651	0.05226	0.66474	0.59927
0 0 1.5	10 10 10	0.05651	0.05226	0.71054	0.66668
0 1.5 1.5	10 10 10	0.05651	0.05226	0.70925	0.66502
0 0.4 1.6	10 10 10	0.05651	0.05226	0.68515	0.63167
0 0 0	10 5 5	0.05	0.05	0.04061	0.03670
0 0 0	10 5 5	0.05*	0.05*	0.04463	0.04900
0 0 0	10 5 5	0.05831	0.07284	0.04963	0.04807
0 0.8 1.6	10 5 5	0.05831	0.07284	0.45513	0.44639
0 0 1.5	10 5 5	0.05831	0.07284	0.44273	0.43269
0 1.5 1.5	10 5 5	0.05831	0.07284	0.52228	0.52774
0 0.4 1.6	10 5 5	0.05831	0.07284	0.45151	0.44357

*Standardized group means corrected according to (11).

Table 6. Powers for simultaneous testing with two ranking schemes, large samples, three groups, and Laplace data.

Group Means	Group Sizes	Nominal Size		Rejection Proportion	
		Proc. I	Proc. II	Proc. I	Proc. II
0 0 0	100 100 100	0.05	0.05	0.04581	0.04563
0 0 0	100 100 100	0.05*	0.05*	0.04827	0.04838
0 0 0	100 100 100	0.05426	0.05452	0.04998	0.04990
0 0.3 0.6	100 100 100	0.05426	0.05452	0.90582	0.90268
0 0 0.5	100 100 100	0.05426	0.05452	0.87476	0.87443
0 0.4 0.4	100 100 100	0.05426	0.05452	0.69634	0.69591
0 0.1 0.5	100 100 100	0.05426	0.05452	0.81530	0.81338
0 0 0	100 50 50	0.05	0.05	0.04502	0.04430
0 0 0	100 50 50	0.05*	0.05*	0.04787	0.04768
0 0 0	100 50 50	0.05490	0.05530	0.05	0.04985
0 0.3 0.6	100 50 50	0.05490	0.05530	0.75054	0.74666
0 0 0.5	100 50 50	0.05490	0.05530	0.63985	0.63937
0 0.4 0.4	100 50 50	0.05490	0.05530	0.53717	0.53724
0 0.1 0.5	100 50 50	0.05490	0.05530	0.59376	0.59167

*Standardized group means corrected according to (11).

Table 7 and Table 8 represent the same values for data sets with four groups.

Table 7. Powers for simultaneous testing with two ranking schemes, small samples, four groups, and Laplace data.

Group Means	Group Sizes	Nominal Size		Rejection Proportion	
		Proc. I	Proc. II	Proc. I	Proc. II
0 0 0 0	10 10 10 10	0.05	0.05	0.04185	0.03508
0 0 0 0	10 10 10 10	0.05*	0.05*	0.04732	0.04530
0 0 0 0	10 10 10 10	0.05916	0.06381	0.04920	0.04523
0 0.5 1 1.5	10 10 10 10	0.05916	0.06381	0.59109	0.54373
0 0 0 1.5	10 10 10 10	0.05916	0.06381	0.70854	0.67148
0 0 1.3 1.3	10 10 10 10	0.05916	0.06381	0.67588	0.65696
0 0.4 1.6 1.6	10 10 10 10	0.05916	0.06381	0.76339	0.73845
0 0 0 0	10 10 5 5	0.05	0.05	0.03835	0.02980
0 0 0 0	10 10 5 5	0.05*	0.05*	0.04574	0.04229
0 0 0 0	10 10 5 5	0.06149	0.07130	0.04899	0.04421
0 0.5 1 1.5	10 10 5 5	0.06149	0.07130	0.40029	0.35396
0 0 0 1.5	10 10 5 5	0.06149	0.07130	0.44137	0.39886
0 0 1.3 1.3	10 10 5 5	0.06149	0.07130	0.46143	0.43477
0 0.4 1.6 1.6	10 10 5 5	0.06149	0.07130	0.55077	0.50646

*Standardized group means corrected according to (11).

Table 8. Powers for simultaneous testing with two ranking schemes, large samples, four groups, and Laplace data.

Group Means	Group Sizes	Nominal Size		Rejection Proportion	
		Proc. I	Proc. II	Proc. I	Proc. II
0 0 0 0	100 100 100 100	0.05	0.05	0.0465	0.04551
0 0 0 0	100 100 100 100	0.05*	0.05*	0.04737	0.04705
0 0 0 0	100 100 100 100	0.05375	0.05402	0.04999	0.04961
0 0.2 0.4 0.6	100 100 100 100	0.05375	0.05402	0.89152	0.88673
0 0 0 0.4	100 100 100 100	0.05375	0.05402	0.69003	0.68782
0 0 0.4 0.4	100 100 100 100	0.05375	0.05402	0.77689	0.77758
0 0.1 0.4 0.4	100 100 100 100	0.05375	0.05402	0.68210	0.68072
0 0 0 0	100 100 50 50	0.05	0.05	0.04492	0.04407
0 0 0 0	100 100 50 50	0.05*	0.05*	0.04710	0.04704
0 0 0 0	100 100 50 50	0.05543	0.05643	0.04995	0.04999
0 0.2 0.4 0.6	100 100 50 50	0.05543	0.05643	0.73151	0.72566
0 0 0 0.4	100 100 50 50	0.05543	0.05643	0.43957	0.43947
0 0 0.4 0.4	100 100 50 50	0.05543	0.05643	0.58381	0.58489
0 0.1 0.4 0.4	100 100 50 50	0.05543	0.05643	0.50080	0.49963

*Standardized group means corrected according to (11).

5. Conclusions

In all eight of these examples, with small and large samples, balanced and unbalanced groups, three and four groups, and normal and Laplace data, Procedure I has a true size closer to the desired nominal size than does Procedure II, that the inversion of the Cornish Fisher expansion (11) provides an adequate control for the non-normality of rank mean differences, and that after adjusting nominal size to have approximately the same actual size, Procedure I provides powers greater than does Procedure II. Some of these differences were small, but all were in favor of Procedure I. Based on the simulation presented here, we recommend the use of Procedure I and not Procedure II; this is based on a limited set of simulations, and comparisons of asymptotic relative efficiency, outside the scope of this manuscript, might completely resolve this issue and demonstrate the origin of this difference.

Funding

This work was supported by the NSF award DMS-2449488.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Kolassa, J.E. (2020) An Introduction to Nonparametric Statistics. Chapman and Hall/CRC.

[2] Fisher, R.A. (1925) Statistical Methods for Research Workers. Oliver and Boyd.

[3] Tukey, J.W. (1993) The Problem of Multiple Comparisons: Introduction and Parts A, B, and C. Princeton University.

[4] Tukey, J.W. (1993) Reminder Sheets for "Allowances for Various Types of Error Rates". In: Braun, H.I., Ed., *The Collected Works of John W. Tukey Volume VIII*: *Multiple Comparisons*, 1948-1983, Chapman and Hall, 335-339.

[5] Kramer, C.Y. (1956) Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications. *Biometrics*, **12**, 307-310. https://doi.org/10.2307/3001469

[6] Erdös, P. and Réyni, A. (1959) On the Central Limit Theorem for Samples from a Finite Population. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, **4**, 49-61.

[7] Dunn, O.J. (1964) Multiple Comparisons Using Rank Sums. *Technometrics*, **6**, 241-252. https://doi.org/10.1080/00401706.1964.10490181

[8] Gebhart, L. and Kolassa, J. (2024) Adjustments for Kurtosis and Continuity on the Prentice Test. *Advances in Pure Mathematics*, **14**, 101-117. https://doi.org/10.4236/apm.2024.142005

[9] Cornish, E.A. and Fisher, R.A. (1938) Moments and Cumulants in the Specification of Distributions. *Revue de l'Institut International de Statistique*/*Review of the International Statistical Institute*, **5**, 307-320. https://doi.org/10.2307/1400905

[10] Kolassa, J.E. (2006) Series Approximation Methods in Statistics. 3rd Edition, Springer-Verlag.