Hacettepe Journal of Mathematics and Statistics
Volume 39 (4) (2010), 613 – 626

SELECTION OF ONE-STAGE SAMPLE SIZE IN CHEN-CHEN-CHANG'S $\tilde{R}$ TEST AND AN EVALUATION OF THE PERFORMANCE OF $\tilde{R}$

B. Altunkaynak*† and H. Gamgam*

Received 18 : 11 : 2009 : Accepted 28 : 04 : 2010

Abstract

In this study, in order to determine the one-stage sample size, power comparisons were made in the $\tilde{R}$ test suggested by Chen *et al.* for ordered alternatives, and the Jonckheere-Terpstra J test, which is one of the classic test statistics, was compared with $\tilde{R}$ in terms of effectiveness. The simulation concerning power comparisons showed that it is impossible to randomly select the one-stage sample size represented by n_0 in $\tilde{R}$, based on one-stage sampling, and that the one-stage sample size should be as large as possible. In addition, the recently-suggested $\tilde{R}$ test statistic for ordered alternatives was compared with J in terms of experimental type-I error and power. Whilst $\tilde{R}$ and J yielded almost the same results in terms of experimental type-I error, $\tilde{R}$ was found to give a worse performance than J with regard to power. The results obtained were validated for different sample sizes and different numbers of populations.

Keywords: Ordered alternatives, Simulation, Power, Exact critical values.

2000 AMS Classification: 62 G 10, 62 G 30.

1. Introduction

Let $X_{i1}, X_{i2}, \dots, X_{in_i}$, i=1,2,...,k, be independent samples from k populations with a mean μ_i and a variance σ_i^2, the size of which is n_i. In the test of the null hypothesis $H_0 : \mu_1 = \dots = \mu_k = \mu$ against H_1:*in which at least two of the* μ_i *are different*, under normality of population distributions and homogeneity of variances, an F statistic with degrees of freedom $k-1$ and $\sum_{i=1}^{k} n_i - k$ is used. Without these assumptions, the Kruskal-Wallis H or Median Test could be used instead of the F test. Without homogeneity of variances, Bishop [3], Bishop and Dudewicz [4, 5], Chen [8], Chen and Chen [9], Gamage

*Gazi University, Department of Statistics, Teknikokullar, 06500 Beşevler, Ankara, Turkey. E-mail: (B. Altunkaynak) bulenta@gazi.edu.tr (H. Gamgam) gamgam@gazi.edu.tr

†Corresponding Author.

et al. [12], Gerami and Zahedian [13], Lee and Ahn [17], Rice and Gaines [20], Weerahandi [22], Xu and Wong [24, 25] have suggested using different test statistics to test H_0 against H_1.

In some cases, the preliminary information the researcher has obtained could be suitable for forming H_1 as either

$$H_1 : \mu_1 \leq \mu_2 \leq \ldots \leq \mu_k$$

or

$$H_1 : \mu_1 \geq \mu_2 \geq \ldots \geq \mu_k,$$

with at least one inequality being strict.

Such a hypothesis is known as the *ordered alternative hypothesis*. To test H_0 against the ordered alternative hypothesis, Jonckheere [16] and Terpstra [21] independently suggested the same test. Subsequently, Archambault *et al.* [1], Odeh [18], Puri [19], Hettmansperger and Norton [15], Chacko [7], Bartholomew [2] and Chen *et al.* [10] studied the same subject.

The primary concern of this study is to determine the one-stage sample size in Chen-Chen-Chang's $\tilde{R}$ (referred to here and later as $\tilde{R}$). The reason why this study deals with $\tilde{R}$ is threefold. Firstly, $\tilde{R}$ is one of the latest tests suggested for ordered alternatives. Secondly, the article in which $\tilde{R}$ appears seems not to have reported sufficient power comparisons. Thirdly, and most importantly, Chen [8] maintains that it is possible to randomly select the one-stage sample size represented by n_0 in $\tilde{R}$; however, it is argued in this paper that this is not the case. Another concern is to compare the performances of $\tilde{R}$ with Jonckheere-Terpstra's J (referred to here and later as J), even when the one-stage sample size is selected as the most appropriate. To this end, in Section 2 $\tilde{R}$ and in Section 3 J, are introduced. Section 4 deals with the calculation of critical values for each of these. Then, in Section 5, power comparisons are made in $\tilde{R}$ to determine the one-stage sample size, and the tests are compared in terms of experimental type-I error and power using simulation for different values of k, n_i, μ_i and α. The last section presents conclusions and suggestions.

2. Chen-Chen-Chang's $\tilde{R}$ test

Suggested by Chen *et al.* [10], $\tilde{R}$ is based upon one-stage sampling. Further, there is also a test statistic based on two-stage sampling. However, Chen *et al.* [10] state that both one-stage and two-stage sampling have the same power. Therefore, this study uses only one-stage sampling.

Let $X_{i1}, X_{i2}, \ldots, X_{in_i}$, $i = 1, 2, \ldots, k$, be independent samples from k populations with a mean μ_i and a variance σ_i^2, the size of which is n_i ($n_i \geq 3$) without variance homogeneity, and define for each sample $2 \leq n_0 < n_i$, the known sample mean and the sample variance concerning the data including n_0 observation as

$$\bar{X}_i = \frac{1}{n_0} \sum_{j=1}^{n_0} X_{ij} \text{ and } S_i^2 = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (X_{ij} - \bar{X}_i)^2,$$

respectively. Assuming

$$z^* = \max \left\{ \frac{S_1^2}{n_1}, \frac{S_2^2}{n_2}, \ldots, \frac{S_k^2}{n_k} \right\},$$

then

$$t_i = \frac{\tilde{X}_i - \mu_i}{\sqrt{z^*}}, \ i = 1, 2, \ldots, k,$$

where $\tilde{X}_i$ represents the weighted sample mean [10]. In such a case, the test statistics is defined as

$$\tilde{R} = \max_{1\le r\le k} \frac{1}{r}\sum_{i=1}^{r}\left(t_i + \frac{\mu_i}{\sqrt{z^*}}\right) - \min_{1\le r\le k} \frac{1}{k-r+1}\sum_{i=r}^{k}\left(t_i + \frac{\mu_i}{\sqrt{z^*}}\right)$$

Under H_0, the null distribution of $\tilde{R}$ is

$$\tilde{Q} = \max_{1\le r\le k} \frac{1}{r}\sum_{i=1}^{r} t_i - \min_{1\le r\le k} \frac{1}{k-r+1}\sum_{i=r}^{k} t_i.$$

From the distribution of $\tilde{Q}$, in the case of $P(\tilde{Q} > q_{\alpha,k,v}) = \alpha$, if $\tilde{R}_c > q_{\alpha,k,v}$, H_0 is rejected. Here $\tilde{R}_c$ represents the value calculated from the sample and $q_{\alpha,k,v}$ is the critical value. For the various combinations of $v = 3, 5, 9, 14, 19, 24, 29, 59$ and $k = 3, 4, 6, 10$, the critical values $q_{\alpha,k,v}$ were calculated by Chen *et al.* [10] using Monte Carlo simulation .

3. Jonckheere-Terpstra's J test

Let $X_{i1}, X_{i2}, \ldots, X_{in_i}$ be independent samples of size n_i from populations with continuous cumulative distribution function $F_i(x)$, $i = 1, 2, \ldots, k$.

For the i and j th samples, the Mann-Whitney U statistic is defined as

$$U_{ij} = \sum_{s=1}^{n_i}\sum_{t=1}^{n_j} D(X_{jt} - X_{is}),$$

where

$$D(u) = \begin{cases} 1 & u > 0, \\ 0 & u \le 0, \end{cases}$$

see [14]. The statistic suggested by Jonckheere [16] and Terpstra [21] to test H_0 against the ordered alternative, is the sum of the Mann-Whitney statistics, that is

$$J = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} U_{ij}.$$

When $k = 3$, $n_1, n_2, n_3 = 2, 3, \ldots, 8$ there are already existing tables containing critical values derived from the exact distributions of J [11]. However, these tables are very limited in nature with respect to the size of population and the sample size. For instance, the existing tables do not contain exact critical values for $k > 3$; asymptotic distributions are used, therefore. Bucchianico [6] and Wiel [23] have established that using asymptotic distributions, especially when the sample size is not large enough, is misleading as it brings about flawed results. Hence, they have extended the existing tables using algorithms based upon computer algebra and by calculating exact critical values for various non-parametric tests.

In the simulation study in Section 5, the approach suggested by Wiel [23] is used to calculate the critical values for J.

If $J_c > J'_\alpha$, where J_c is the value calculated from the sample and J'_α is the critical value for J, then H_0 against the ordered alternative is rejected.

4. Calculation of critical values

The critical values $q_{\alpha,k,v}$ were obtained through Monte Carlo simulation for different sizes of population and sample sizes. In the simulation study, for the different values of k and n, the Q value was calculated by using the values of $t_i = Z/\sqrt{Y/v}$, $i = 1, 2, \ldots, k$, where Z is the random variable of the standard normal distribution, and Y is the chi-square with $v = n_0 - 1$ degrees of freedom. The probability distribution of Q was formed after 10000 simulation runs. This process was replicated 20 times. When $\alpha = 0.01, 0.05, 0.10$ and $P(\tilde{Q} > q_{\alpha,k,v}) = \alpha$, the average values of 20 critical points were calculated. This average was used as the critical value of $\tilde{R}$ in the simulation study.

There are existing critical values for J for a limited number of cases of k and n, such as $k = 3$ and $n_1, n_2, n_3 = 2, 3, \ldots, 8$ [11]. To calculate the critical values of J'_α, Wiel [23] has suggested a method based on the probability generating function of this statistic, and this method could be used for a wide range of values of k and n. Wiel [23] has stated that the critical values generated previously are very limited and that using a normal distribution is not proper if the sample size is not very large. Thus, the critical values generated using the method suggested by Wiel were used in the simulation of this study.

5. Simulation study

In this section we first describe a simulation study to determine the value of n_0 which was carried out before the comparisons. Chen *et al.* [10] have pointed out that n_0 could have any value ($2 \leq n_0 < n_i$). This simulation was done to find out the best value n_0 could take. To obtain power values, for $k = 3$ as $n_i = 3(1)10(5)20(10)30$, for $k = 4$ as $n_i = 3(1)10(5)20$, for $k = 5$ as $n_i = 3(1)10(5)15$, for $k = 10$ as $n_i = 3(1)7$, the values of X_{ij} were randomly generated from a normal distribution with a mean of $95 + 5i$ and a variance of $(2+2i)^2$ $(i = 1, 2, \ldots, k)$. With this, the rejection ratio of the null hypothesis in 10000 iterations, that is the power of test, was calculated. In this procedure, the value of n_0 was taken from the extremes ($n_0 = 2$ and $n_0 = n - 1$). The results are illustrated in Table 1.

Table 1. Power values for $\tilde{R}$ calculated using the data generated from $X_{ij} \sim N((95+5i), (2+2i)^2)$, $(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n)$, when $n_0 = 2$ and $n_0 = n - 1$

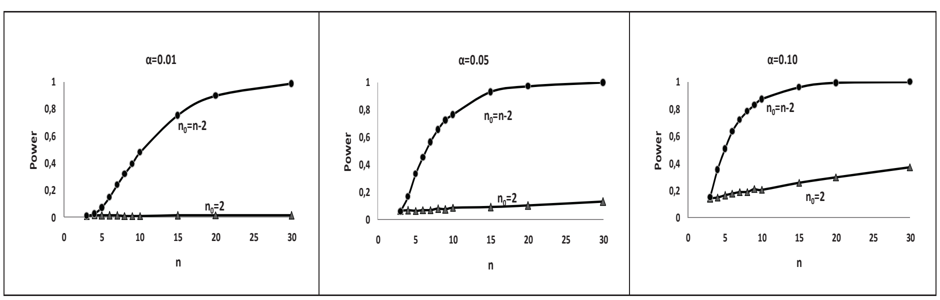
		α					
		0.01		0.05		0.10	
k	n	$n_0 = 2$	$n_0 = n-1$	$n_0 = 2$	$n_0 = n-1$	$n_0 = 2$	$n_0 = n-1$
3	3	0.0097	0.0102	0.0628	0.0569	0.1400	0.1463
	4	0.0130	0.0230	0.0643	0.1661	0.1464	0.3512
	5	0.0108	0.0690	0.0638	0.3289	0.1621	0.5047
	6	0.0112	0.1458	0.0680	0.4516	0.1761	0.6320
	7	0.0106	0.2362	0.0688	0.5630	0.1862	0.7188
	8	0.0091	0.3189	0.0768	0.6529	0.1882	0.7817
	9	0.0100	0.3931	0.0708	0.7219	0.2103	0.8297
	10	0.0084	0.4773	0.0833	0.7619	0.2036	0.8702
	15	0.0111	0.7510	0.0878	0.9276	0.2559	0.9600
	20	0.0129	0.8958	0.0994	0.9712	0.2971	0.9903
	30	0.0120	0.9843	0.1282	0.9968	0.3722	0.9987

Table 1. Continued

		α					
		0.01		0.05		0.10	
k	n	$n_0 = 2$	$n_0 = n - 1$	$n_0 = 2$	$n_0 = n - 1$	$n_0 = 2$	$n_0 = n - 1$
4	3	0.0112	0.0087	0.0559	0.0613	0.1344	0.1294
	4	0.0105	0.0196	0.0634	0.1806	0.1416	0.3962
	5	0.0105	0.0816	0.0636	0.3864	0.1447	0.6031
	6	0.0125	0.1945	0.0625	0.5657	0.1569	0.7308
	7	0.0092	0.3268	0.0623	0.6796	0.1673	0.8117
	8	0.0131	0.4587	0.0596	0.7669	0.1695	0.8742
	9	0.0098	0.5610	0.0641	0.8398	0.1717	0.9160
	10	0.0083	0.6440	0.0712	0.8817	0.1916	0.9415
	15	0.0132	0.8842	0.0794	0.9795	0.2288	0.9898
	20	0.0104	0.9735	0.0824	0.9948	0.2607	0.9985
5	3	0.0108	0.0090	0.0524	0.0529	0.1204	0.1188
	4	0.0104	0.0189	0.0555	0.1933	0.1409	0.4171
	5	0.0077	0.0802	0.0575	0.4467	0.1399	0.6602
	6	0.0084	0.2386	0.0583	0.6415	0.1460	0.8000
	7	0.0123	0.4268	0.0569	0.7661	0.1618	0.8736
	8	0.0119	0.5364	0.0646	0.8351	0.1577	0.9221
	9	0.0099	0.6598	0.0563	0.8978	0.1554	0.9560
	10	0.0078	0.7461	0.0623	0.9291	0.1720	0.9699
	15	0.0092	0.9535	0.0708	0.9908	0.1917	0.9965
10	3	0.0108	0.0121	0.0495	0.0532	0.1151	0.1132
	4	0.0091	0.0184	0.0560	0.2000	0.1156	0.4475
	5	0.0077	0.1184	0.0554	0.5487	0.1156	0.7753
	6	0.0107	0.3352	0.0523	0.7958	0.1194	0.9167
	7	0.0094	0.5743	0.0542	0.9042	0.1247	0.9648

It was found that whilst the test power values were extremely bad when $n_0 = 2$, they were better when $n_0 = n - 1$. This result is valid for different values of k and n. For ease of illustration, the results for $\alpha = 0.01, 0.05, 0.10$ when $k = 3$ are demonstrated in Figure 1.

Figure 1. The power graphics for $\tilde{R}$ formed with the data generated from $X_{ij} \sim N((95 + 5i), (2 + 2i)^2)$, $(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n)$, when $k = 3$



As seen in the graphics, the value of n_0 must be selected as near to n as the condition $n_0 < n_i$ permits, contrary to what Chen *et al.* [10] have suggested. Considering the $v = n_0 - 1$ degrees of freedom, such a result is not surprising.

Following the simulation study concerning the identification of n_0, J and $\tilde{R}$ were compared in terms of experimental type-I error and power when $\alpha = 0.01, 0.05, 0.10$. Initially, the critical values for J and $\tilde{R}$ were calculated by using the method explained in Section 4, for $k = 3$ as $n_i = 3(1)10(5)20(10)30$, for $k = 4$ as $n_i = 3(1)10(5)20$, for $k = 5$ as $n_i = 3(1)10(5)15$, and for $k = 10$ as $n_i = 3(1)7$. The values obtained concerning experimental type-I error and power are presented in the tables. For ease of illustration, we take $n_i = n$ for all groups.

The comparisons regarding experimental type-I error were made for two different cases, and the results tabled separately. In the first case, the mean and variance for all the populations were the same in the different combinations of values of k and n. Here, the values of X_{ij} were randomly generated from normal distributions with a mean of 100 and a variance of 16, and the rejection ratio of the null hypothesis in 10000 iterations, that is the values of the experimental type-I error, were calculated for both tests (Table 2).

In the second case, the values of the experimental type-I error were calculated under unequal variance. Here, the values of X_{ij} were randomly generated from a normal distributions with a mean of 100 and a variance of $(2+2i)^2$, $i = 1, 2, \ldots, k$ (Table 3).

Table 2. Experimental type-I errors of J and $\tilde{R}$ when $n_0 = n - 1$, calculated using the data generated from $X_{ij} \sim N(100, 16)$, ($i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, n$)

		α					
		0.01		0.05		0.10	
k	n	J	$\tilde{R}$	J	$\tilde{R}$	J	$\tilde{R}$
3	3	0.0046	0.0102	0.0347	0.0517	0.0990	0.0993
	4	0.0104	0.0074	0.0457	0.0462	0.0830	0.0982
	5	0.0081	0.0096	0.0471	0.0489	0.0848	0.1020
	6	0.0092	0.0100	0.0501	0.0529	0.0901	0.1005
	7	0.0111	0.0084	0.0469	0.0512	0.0936	0.0951
	8	0.0091	0.0110	0.0519	0.0519	0.1010	0.0988
	9	0.0090	0.0097	0.0441	0.0547	0.0963	0.1025
	10	0.0088	0.0099	0.0492	0.0488	0.0916	0.0962
	15	0.0083	0.0095	0.0464	0.0489	0.1046	0.0980
	20	0.0083	0.0096	0.0544	0.0505	0.0981	0.0985
	30	0.0095	0.0107	0.0481	0.0482	0.0998	0.0995
4	3	0.0080	0.0095	0.0382	0.0470	0.0908	0.0951
	4	0.0097	0.0116	0.0432	0.0513	0.0910	0.0991
	5	0.0097	0.0120	0.0509	0.0470	0.0963	0.0980
	6	0.0103	0.0104	0.0491	0.0511	0.0994	0.1037
	7	0.0102	0.0102	0.0507	0.0513	0.0964	0.1024
	8	0.0085	0.0113	0.0452	0.0523	0.1026	0.1024
	9	0.0098	0.0077	0.0500	0.0477	0.0984	0.0943
	10	0.0088	0.0085	0.0508	0.0483	0.1004	0.0975
	15	0.0081	0.0100	0.0479	0.0529	0.0981	0.1007
	20	0.0084	0.0088	0.0499	0.0539	0.1018	0.0933

Table 2. Continued

		α					
		0.01		0.05		0.10	
k	n	J	$\tilde{R}$	J	$\tilde{R}$	J	$\tilde{R}$
5	3	0.0075	0.0126	0.0493	0.0504	0.0869	0.0913
	4	0.0101	0.0106	0.0430	0.0485	0.0932	0.1050
	5	0.0098	0.0117	0.0492	0.0504	0.0899	0.1047
	6	0.0092	0.0119	0.0484	0.0498	0.1016	0.0989
	7	0.0103	0.0101	0.0440	0.0519	0.0955	0.0940
	8	0.0102	0.0098	0.0495	0.0539	0.0958	0.1008
	9	0.0105	0.0087	0.0469	0.0497	0.0995	0.0962
	10	0.0097	0.0099	0.0481	0.0507	0.0947	0.1001
	15	0.0104	0.0094	0.0460	0.0487	0.0985	0.1030
10	3	0.0099	0.0092	0.0509	0.0481	0.1005	0.1062
	4	0.0108	0.0097	0.0488	0.0476	0.0961	0.1071
	5	0.0093	0.0080	0.0522	0.0500	0.1013	0.1018
	6	0.0106	0.0099	0.0468	0.0465	0.0962	0.0995
	7	0.0076	0.0107	0.0488	0.0507	0.1024	0.1033

It can be concluded from Table 2 that for the cases where the variance is the same, the values of the experimental type-I error of both tests are very close to nominal type-I error, and this is also valid in different values of k and n.

Table 3. Experimental type-I errors of J and $\tilde{R}$ when $n_0 = n - 1$, calculated using the data generated from $X_{ij} \sim N(100, (2+2i)^2)$, $(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n)$

		α					
		0.01		0.05		0.10	
k	n	J	$\tilde{R}$	J	$\tilde{R}$	J	$\tilde{R}$
3	3	0.0048	0.0117	0.0388	0.0514	0.0961	0.1014
	4	0.0123	0.0097	0.0474	0.0499	0.0845	0.0997
	5	0.0087	0.0086	0.0448	0.0469	0.0903	0.0973
	6	0.0119	0.0088	0.0526	0.0481	0.0942	0.1034
	7	0.0093	0.0100	0.0495	0.0530	0.0894	0.0925
	8	0.0120	0.0104	0.0494	0.0498	0.1021	0.1048
	9	0.0103	0.0091	0.0485	0.0491	0.0935	0.1011
	10	0.0097	0.0117	0.0529	0.0513	0.1002	0.0967
	15	0.0095	0.0099	0.0482	0.0529	0.1012	0.1050
	20	0.0119	0.0107	0.0522	0.0496	0.1021	0.1035
	30	0.0106	0.0105	0.0531	0.0503	0.1019	0.0989
4	3	0.0079	0.0090	0.0403	0.0489	0.0933	0.1023
	4	0.0106	0.0092	0.0460	0.0514	0.0958	0.0989
	5	0.0097	0.0111	0.0523	0.0472	0.0980	0.1007
	6	0.0109	0.0098	0.0495	0.0473	0.0976	0.1028
	7	0.0099	0.0096	0.0518	0.0494	0.1035	0.0986
	8	0.0099	0.0109	0.0479	0.0522	0.1047	0.0973
	9	0.0106	0.0107	0.0524	0.0474	0.0992	0.0996
	10	0.0122	0.0125	0.0535	0.0527	0.1028	0.1051
	15	0.0108	0.0089	0.0540	0.0456	0.1019	0.1070
	20	0.0103	0.0102	0.0485	0.0528	0.0992	0.1003

Table 3. Continued

		α					
		0.01		0.05		0.10	
k	n	J	$\tilde{R}$	J	$\tilde{R}$	J	$\tilde{R}$
5	3	0.0075	0.0088	0.0493	0.0484	0.0891	0.0966
	4	0.0092	0.0117	0.0482	0.0497	0.1018	0.1020
	5	0.0092	0.0108	0.0538	0.0494	0.1014	0.1074
	6	0.0118	0.0092	0.0546	0.0462	0.1004	0.0984
	7	0.0122	0.0098	0.0510	0.0484	0.1023	0.0983
	8	0.0108	0.0093	0.0493	0.0506	0.0986	0.1039
	9	0.0122	0.0117	0.0524	0.0517	0.0997	0.1117
	10	0.0124	0.0119	0.0527	0.0518	0.1068	0.0967
	15	0.0098	0.0103	0.0562	0.0539	0.1065	0.0956
10	3	0.0104	0.0096	0.0506	0.0494	0.1030	0.0950
	4	0.0123	0.0104	0.0552	0.0527	0.0969	0.0967
	5	0.0105	0.0086	0.0494	0.0496	0.1071	0.1058
	6	0.0118	0.0122	0.0549	0.0494	0.1060	0.0995
	7	0.0113	0.0083	0.0527	0.0506	0.1046	0.1046

Similarly, it can be concluded from Table 3 that for the cases where the variance is different, the values of the experimental type-I error of both tests are very close to the nominal type-I error again, and this is also valid for different values of k and n.

In the power of test comparisons, two cases were considered. The first was when variances were equal, but means were not. Here, firstly the values of X_{ij} were randomly generated from normal distributions with a mean of $(99 + i)$, $i = 1, 2, \ldots, k$, and a variance of 16, and the rejection ratio of the null hypothesis in 10000 iterations, that is the power of the test was calculated. Then, the values of X_{ij} were randomly generated from normal distributions with a mean of $(95 + 5i)$, $i = 1, 2, \ldots, k$, and a variance of 16 and the rejection ratio of the null hypothesis in 10000 iterations, that is the power of the test was calculated (Tables 4-5). The reason why the means were taken in this way was to check the sensitivity of both tests to variations in the mean.

Table 4 shows the power values when the variance was the same while there were minor variations in the means. It is obvious from the table that the power values of J are bigger than those of $\tilde{R}$. This does not change even when the number of population or the sample size becomes larger. Figure 2 (a) shows the related result when $k = 3$ and $\alpha = 0.05$.

Table 5 demonstrates the power values for the case where there was a bigger variation in the means than those in Table 4, with the variance being the same. It is seen that the power values of J tended to be bigger than those of $\tilde{R}$. When $k = 3$ and $\alpha = 0.05$ (Figure 2 (b)), the difference between the values of J and those of $\tilde{R}$ is more visible, especially in small sample sizes. When the sample size becomes larger, the power values of both tests approximate to 1, but this approximation is faster in J.

Table 4. Power of J and $\tilde{R}$ when $n_0 = n - 1$, calculated using the data generated from $X_{ij} \sim N((99 + i), 16)$, $(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n)$

		α					
		0.01		0.05		0.10	
k	n	J	$\tilde{R}$	J	$\tilde{R}$	J	$\tilde{R}$
3	3	0.0154	0.0115	0.1069	0.0506	0.2256	0.1073
	4	0.0422	0.0111	0.1472	0.0738	0.2340	0.1507
	5	0.0419	0.0165	0.1736	0.0921	0.2640	0.1803
	6	0.0558	0.0243	0.1970	0.1149	0.3062	0.2185
	7	0.0659	0.0269	0.2147	0.1379	0.3188	0.2489
	8	0.0807	0.0358	0.2300	0.1600	0.3588	0.2652
	9	0.0877	0.0487	0.2497	0.1711	0.3728	0.2882
	10	0.0898	0.0495	0.2756	0.1945	0.4062	0.3229
	15	0.1471	0.0976	0.3574	0.2819	0.5039	0.4224
	20	0.1966	0.1406	0.4501	0.3589	0.5872	0.5167
	30	0.3218	0.2361	0.5864	0.4996	0.7202	0.6455
4	3	0.0557	0.0105	0.1753	0.0502	0.3267	0.1072
	4	0.0897	0.0120	0.2459	0.0772	0.3711	0.1651
	5	0.1104	0.0181	0.3158	0.1203	0.4395	0.2302
	6	0.1280	0.0287	0.3492	0.1618	0.4968	0.2819
	7	0.1643	0.0469	0.3933	0.2040	0.5570	0.3410
	8	0.1855	0.0634	0.4322	0.2380	0.5897	0.3909
	9	0.2189	0.0873	0.4764	0.2797	0.6100	0.4187
	10	0.2411	0.0966	0.5122	0.3073	0.6581	0.4570
	15	0.3914	0.2072	0.6612	0.4740	0.7979	0.6257
	20	0.5317	0.3115	0.7756	0.6166	0.8730	0.7531
5	3	0.1003	0.0111	0.3345	0.0527	0.4731	0.1109
	4	0.1601	0.0109	0.4103	0.0868	0.5823	0.1834
	5	0.2250	0.0245	0.5064	0.1439	0.6398	0.2829
	6	0.2867	0.0419	0.5602	0.2206	0.7114	0.3789
	7	0.3591	0.0674	0.6243	0.2753	0.7567	0.4490
	8	0.4018	0.1064	0.6811	0.3540	0.8009	0.5143
	9	0.4487	0.1444	0.7373	0.4007	0.8408	0.5818
	10	0.5154	0.1702	0.7654	0.4834	0.8687	0.6298
	15	0.7266	0.3960	0.9014	0.6819	0.9509	0.8183
10	3	0.8902	0.0109	0.9799	0.0486	0.9914	0.1070
	4	0.9701	0.0144	0.9951	0.1378	0.9986	0.3431
	5	0.9934	0.0611	0.9993	0.4062	0.9999	0.6598
	6	0.9982	0.2012	1.0000	0.6759	1.0000	0.8438
	7	0.9999	0.4028	1.0000	0.8240	1.0000	0.9211

Table 5. Power of J and $\tilde{R}$ when $n_0 = n - 1$, calculated using the data generated from $X_{ij} \sim N((95+5i), 16)$, $(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n)$

		α					
		0.01		0.05		0.10	
k	n	J	$\tilde{R}$	J	$\tilde{R}$	J	$\tilde{R}$
3	3	0.3518	0.0097	0.7697	0.0684	0.9214	0.1792
	4	0.7117	0.0297	0.9209	0.3362	0.9638	0.5688
	5	0.8415	0.1782	0.9670	0.6320	0.9879	0.8119
	6	0.9277	0.4189	0.9899	0.8051	0.9970	0.9066
	7	0.9647	0.6481	0.9957	0.9081	0.9988	0.9498
	8	0.9856	0.7738	0.9980	0.9500	0.9998	0.9773
	9	0.9932	0.8730	0.9998	0.9711	0.9995	0.9891
	10	0.9973	0.9329	0.9996	0.9859	1.0000	0.9929
	15	0.9999	0.9967	1.0000	0.9994	1.0000	0.9998
	20	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000
	30	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	3	0.9215	0.0104	0.9914	0.0707	0.9982	0.1795
	4	0.9907	0.0530	0.9992	0.4937	0.9998	0.7681
	5	0.9987	0.4119	0.9999	0.8696	1.0000	0.9467
	6	1.0000	0.8097	1.0000	0.9645	1.0000	0.9832
	7	1.0000	0.9296	1.0000	0.9894	1.0000	0.9969
	8	1.0000	0.9792	1.0000	0.9977	1.0000	0.9990
	9	1.0000	0.9919	1.0000	0.9991	1.0000	0.9995
	10	1.0000	0.9977	1.0000	0.9996	1.0000	0.9998
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	3	0.9987	0.0116	1.0000	0.0711	1.0000	0.1747
	4	1.0000	0.0747	1.0000	0.6931	1.0000	0.8944
	5	1.0000	0.7146	1.0000	0.9664	1.0000	0.9882
	6	1.0000	0.9559	1.0000	0.9948	1.0000	0.9985
	7	1.0000	0.9930	1.0000	0.9996	1.0000	0.9994
	8	1.0000	0.9986	1.0000	0.9998	1.0000	0.9999
	9	1.0000	0.9998	1.0000	0.9999	1.0000	1.0000
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	3	1.0000	0.0098	1.0000	0.0677	1.0000	0.2181
	4	1.0000	0.5409	1.0000	0.9936	1.0000	0.9980
	5	1.0000	0.9991	1.0000	1.0000	1.0000	1.0000
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

In the second case, the power of the test was calculated under unequal variance. The values of X_{ij} were randomly generated from normal distributions with a mean of $(99+i)$, $i = 1, 2, \ldots, k$, and a variance of $(2+2i)^2$, $i = 1, 2, \ldots, k$, and the power of the test was calculated. Then, the values of X_{ij} were randomly generated from normal distributions with a mean of $(95+5i)$, $i = 1, 2, \ldots, k$ and a variance of $(2+2i)^2$, $i = 1, 2, \ldots, k$, and power values were calculated (Tables 6-7). The MATLAB code for both tests were developed by the present authors.

Table 6. Power of J and $\tilde{R}$ when $n_0 = n - 1$, calculated using the data generated from $X_{ij} \sim N((99+i),(2+2i)^2)$, ($i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, n$)

		α					
		0.01		0.05		0.10	
k	n	J	$\tilde{R}$	J	$\tilde{R}$	J	$\tilde{R}$
3	3	0.0122	0.0118	0.0768	0.0554	0.1736	0.1096
	4	0.0256	0.0095	0.1000	0.0635	0.1733	0.1272
	5	0.0269	0.0140	0.1199	0.0683	0.1963	0.1438
	6	0.0346	0.0180	0.1362	0.0797	0.2164	0.1740
	7	0.0369	0.0213	0.1391	0.0951	0.2214	0.1793
	8	0.0453	0.0228	0.1514	0.1040	0.2566	0.1834
	9	0.0480	0.0265	0.1496	0.1071	0.2578	0.1969
	10	0.0452	0.0287	0.1701	0.1154	0.2710	0.2157
	15	0.0720	0.0427	0.2176	0.174	0.4145	0.2524
	20	0.0894	0.0546	0.2655	0.1820	0.5133	0.2881
	30	0.1408	0.0828	0.3422	0.2313	0.6449	0.3540
4	3	0.0091	0.0100	0.0547	0.0465	0.1073	0.1071
	4	0.0118	0.0122	0.0760	0.0582	0.1643	0.1296
	5	0.0180	0.0167	0.1126	0.0789	0.2366	0.1604
	6	0.0340	0.0182	0.1551	0.0880	0.2898	0.1823
	7	0.0464	0.0217	0.1928	0.1050	0.3357	0.1933
	8	0.0584	0.0256	0.2292	0.1173	0.3944	0.2190
	9	0.0814	0.0293	0.2630	0.1268	0.4315	0.2316
	10	0.0959	0.0328	0.3110	0.1432	0.4657	0.2381
	15	0.2061	0.0541	0.4680	0.1840	0.6350	0.3060
	20	0.3135	0.0739	0.6220	0.2277	0.7447	0.3505
5	3	0.0103	0.0095	0.0497	0.0529	0.1107	0.1048
	4	0.0121	0.0121	0.0911	0.0611	0.1778	0.1297
	5	0.0243	0.0143	0.1498	0.0788	0.2797	0.1587
	6	0.0395	0.0216	0.2169	0.0988	0.3704	0.1824
	7	0.0600	0.0219	0.2819	0.1038	0.4493	0.2109
	8	0.0925	0.0270	0.3280	0.1286	0.5171	0.2271
	9	0.1410	0.0324	0.4135	0.1391	0.5772	0.2391
	10	0.1716	0.0378	0.4459	0.1530	0.6341	0.2622
	15	0.3799	0.0678	0.6891	0.2076	0.8187	0.3427
10	3	0.0083	0.0100	0.0544	0.0520	0.1092	0.0950
	4	0.0191	0.0132	0.1365	0.0605	0.3405	0.1326
	5	0.0684	0.0133	0.4236	0.0818	0.6637	0.1691
	6	0.2137	0.0223	0.6823	0.1069	0.8386	0.2123
	7	0.4010	0.0244	0.8296	0.1374	0.9236	0.2364

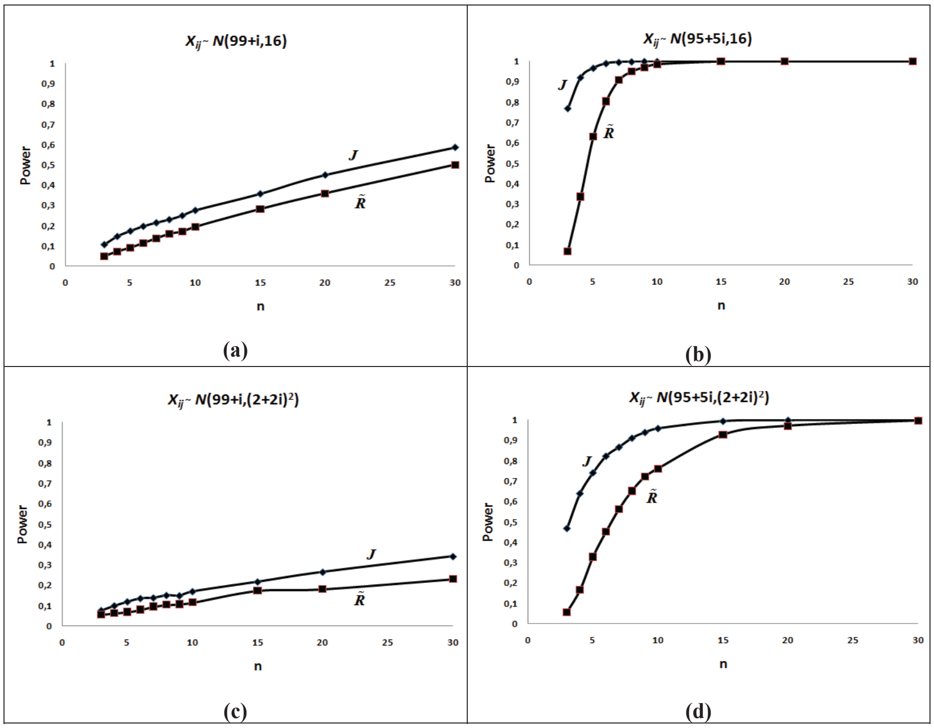
Table 7. Power of J and $\tilde{R}$ when $n_0 = n - 1$, calculated using the data generated from $X_{ij} \sim N((95+5i),(2+2i)^2)$, $(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n)$

		α					
		0.01		0.05		0.10	
k	n	J	$\tilde{R}$	J	$\tilde{R}$	J	$\tilde{R}$
3	3	0.1479	0.0102	0.4694	0.0569	0.6851	0.1463
	4	0.3518	0.0230	0.6402	0.1661	0.7545	0.3512
	5	0.4428	0.0690	0.7418	0.3289	0.8434	0.5047
	6	0.5703	0.1458	0.8233	0.4516	0.9088	0.6320
	7	0.6501	0.2362	0.8673	0.5630	0.9335	0.7188
	8	0.7347	0.3189	0.9117	0.6529	0.9609	0.7817
	9	0.7926	0.3931	0.9398	0.7219	0.9700	0.8297
	10	0.8357	0.4773	0.9590	0.7619	0.9840	0.8702
	15	0.9678	0.7510	0.9945	0.9276	0.9986	0.9600
	20	0.9935	0.8958	0.9991	0.9712	0.9996	0.9903
	30	1.0000	0.9843	1.0000	0.9968	1.0000	0.9987
4	3	0.4375	0.0087	0.7212	0.0613	0.8674	0.1294
	4	0.6501	0.0196	0.8610	0.1806	0.9345	0.3962
	5	0.7640	0.0816	0.9379	0.3864	0.9733	0.6031
	6	0.8573	0.1945	0.9671	0.5657	0.9870	0.7308
	7	0.9197	0.3268	0.9849	0.6796	0.9959	0.8117
	8	0.9579	0.4587	0.9921	0.7669	0.9980	0.8742
	9	0.9767	0.5610	0.9966	0.8398	0.9990	0.9160
	10	0.9865	0.6440	0.9985	0.8817	0.9997	0.9415
	15	0.9994	0.8842	1.0000	0.9795	1.0000	0.9898
	20	1.0000	0.9735	1.0000	0.9948	1.0000	0.9985
5	3	0.6665	0.0090	0.9080	0.0529	0.9518	0.1188
	4	0.8448	0.0189	0.9648	0.1933	0.9883	0.4171
	5	0.9399	0.0802	0.9893	0.4467	0.9966	0.6602
	6	0.9716	0.2386	0.9967	0.6415	0.9997	0.8000
	7	0.9895	0.4268	0.9986	0.7661	0.9995	0.8736
	8	0.9960	0.5364	0.9998	0.8351	1.0000	0.9221
	9	0.9981	0.6598	1.0000	0.8978	1.0000	0.9560
	10	0.9999	0.7461	1.0000	0.9291	1.0000	0.9699
	15	1.0000	0.9535	1.0000	0.9908	1.0000	0.9965
10	3	0.9984	0.0121	0.9999	0.0532	1.0000	0.1132
	4	0.9999	0.0184	1.0000	0.2000	1.0000	0.4475
	5	1.0000	0.1184	1.0000	0.5487	1.0000	0.7753
	6	1.0000	0.3352	1.0000	0.7958	1.0000	0.9167
	7	1.0000	0.5743	1.0000	0.9042	1.0000	0.9648

It is obvious from Table 6 and Table 7 that $\tilde{R}$ yielded worse results than the classic J under unequal variances in terms of power. It is remarkable that especially for Table 6, where there are minor variations in the means, as the sample size and the size of the population became larger, the power values of the two tests increased, but the difference between J and $\tilde{R}$ got larger and larger, with $\tilde{R}$ changing more slowly. However, when the variation in the means was larger, there was a small improvement in $\tilde{R}$. Still, J in all cases yielded better results.

For the ease of illustration of these cases, the results for $k = 3$ and $\alpha = 0.05$ are presented in Figure 2 (c) and 2 (d).

Figure 2. The power values of J and $\tilde{R}$ when $k = 3$ and $\alpha = 0.05$



6. Conclusion

The results obtained in this study can be summarized for two cases. The first is about the selection of n_0 in $\tilde{R}$. Chen *et al.* [10] have stated that n_0 could be selected randomly $(2 \leq n_0 < n_i)$. However, this study has established that the selection of n_0 is important. When $n_0 = 2$, the power of the test is extremely bad, and this changes only slightly when the sample size becomes larger. No variation in test power is apparent for the different values of k and n, especially for $\alpha = 0.01$.

It has been observed that as n_i increases, there is little increase in the power of test for $\alpha = 0.05$ and $\alpha = 0.10$. On the other hand, it has been discovered that if n_0 is chosen as large as possible, such as $n_0 = n_i - 1$, the results obtained are better than those when $n_0 = 2$. Since $v = n_0 - 1$, this is a statistically expected outcome.

The second case is about the results obtained from the comparison of $\tilde{R}$ and J in terms of experimental type-I error and power. Both tests yielded almost the same results in terms of experimental type-I error. For different values of k, n_i and nominal type-I error, the values of the experimental type-I error for both tests obtained were near the nominal type-I error. For all values of k, n_i and nominal type-I error, the power values of $\tilde{R}$ were worse than those of J. Therefore, $\tilde{R}$ exhibits a worse performance than the classic J, even though it is a more recent test recommended for ordered alternatives.

References

[1] Archambault, W. A. T., Mack, G. A. and Wolfe, D. A. *K-sample rank tests using pair-specific scoring functions*, Canad. J. Statist. **5**, 195–207, 1977.

[2] Bartholomew, D. J. *Ordered tests in the analysis of variance*, Biometrika **48**, 325–332, 1961.

[3] Bishop, T. A. *Heteroscedastic ANOVA, MANOVA and multiple comparisons* (Unpublished Ph.D. Thesis, The Ohio State University, Ohio, 1976).

[4] Bishop, T. A. and Dudewicz, E. J. *Exact analysis of variance with unequal variances: Test procedures and tables*, Technometrics **20**, 419–430, 1978.

[5] Bishop, T. A. and Dudewicz, E. J. *Heteroscedastic ANOVA*, Sankhya **43(B)**, 40–57, 1981.

[6] Bucchianico, A. D. *Computer algebra, combinatorics, and the Wilcoxon-Mann-Whitney statistic*, J. Stat. Plan. Inf. **79**, 349–364, 1999.

[7] Chacko, V. J. *Testing homogeneity against ordered alternatives*, Ann. Math. Statist. **34**, 945–956, 1963.

[8] Chen, S. *One-stage and two-stage statistical inference under heteroscedasticity*, Communications in Statistics: Simulation and Computation **30** (4), 991–1009, 2001.

[9] Chen, S. and Chen, J. H. *Single-stage analysis of variance under heteroscedasticity*, Communications Comm. Statist. Simulation **27** (3), 641–666, 1998.

[10] Chen, S., Chen, J. H., and Chang, H. F. *A one-stage procedure for testing homogeneity of means against an ordered alternative under unequal variances*, Communications in Statistics: Simulation and Computation **33** (1), 49–67, 2004.

[11] Daniel, W. W. *Applied nonparametric statistics* (Houghton Mifflin, Boston, 1978).

[12] Gamage, J., Mathew, T., and Weerahandi, S. *Generalized p-values and generalized confidence regions for the multivariate Behrens-Fisher problem and MANOVA*, J. Multivariate Anal. **88**, 177–189, 2004.

[13] Gerami, A. and Zahedian, A. *Comparing the means of normal populations with unequal variances* (Proceedings of the 53rd Session of International Statistical Institute, Seoul, Korea 2001).

[14] Gibbons, J. D. *Nonparametric Statistical Inference* (McGraw-Hill Book Company, New York, 1971).

[15] Hettmansperger. T. P. and Norton, R. M. *Tests for patterned alternatives in k-sample problems*, Journal of the American Statistical Association **82**, 292–299, 1987.

[16] Jonckheere, A. R. *A distribution-free k-sample test against ordered alternatives*, Biometrika **41**, 133–145, 1954.

[17] Lee, S. and Ahn, C. H. *Modified ANOVA for unequal variances*, Communications in Statistics: Simulation and Computation **32**, 987–1004, 2003.

[18] Odeh, R. E. *On Jonckheere's k-sample test against ordered alternatives*, Technometrics **13**, 912–918, 1971.

[19] Puri, M. L. *Some distribution-free k-sample rank tests of homogeneity against ordered alternatives*, Comm. Pure Appl. Math. **18**, 51–63. 1965.

[20] Rice, W. R. and Gaines, S. D. *One-way analysis of variance with unequal variances*, Proc. Nat. Acad. Sci. **86**, 8183–8184, 1989.

[21] Terpstra, T. J. *The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking*, Indigationes Mathematicae **14**, 327–333, 1952.

[22] Weerahandi, S. *ANOVA under unequal error variances*, Biometrics **51**, 589–599, 1995.

[23] Wiel, M. A. *Exact distributions of distribution-free test statistics* (Ph.D. Thesis, Eindhoven University of Technology, The Netherlands, 2000).

[24] Xu, L. and Wang, S. A. *New generalized p-value and its upper bound for ANOVA under unequal error variances*, Commun Stat Theory Methods **37**, 1002–1010, 2008.

[25] Xu, L. and Wang, S. A. *A new generalized p-value for ANOVA under heteroscedasticity*, Statistics & Probability Letters **78** (8), 963–969, 2008.