Canadian Journal of Pure and Applied Sciences
Vol. 15, No. 3, pp. 5307-5316, Oct 2021
Online ISSN: 1920-3853; Print ISSN: 1715-9997
Available online at www.cjpas.net

EXACT CRITICAL VALUES OF THE WILCOXON RANK SUM TEST

*Friday Ewere and Sunday Martins Ogbonmwan
Department of Statistics, Faculty of Physical Sciences, University of Benin, Benin City, Edo State, Nigeria

ABSTRACT

Generating the entire permutation sample space especially when sample sizes are not small have been a major problem in constructing an exact test of significance of a rank statistic. Recently, the use of softwares for computing statistical tests has become common. However, procedures on this software for calculating the significance levels for many nonparametric tests are based on asymptotic results. These asymptotic results are only reliable when sample sizes are large enough. Unfortunately, the definition of what constitute a large sample size for most statistics is quite vague. The aim of this paper is to formulate a method for obtaining the exact distribution of a rank statistic. The proposed method is based on combinatorics in the representation of the probability generating function of the test statistic. The proposed method bypasses the problem of actually carrying out a complete enumeration in a permutation test. Essentially, the exact critical values for the Wilcoxon Rank Sum (WRS) test statistic are produced. The asymptotic property of the WRS is carefully studied and the minimum sample size required for the application of the large sample approximation is provided.

Keywords: Wilcoxon Rank Sum Test, exact test, rank test, permutation test, combinatorics.

INTRODUCTION

The risk in decision making cannot be totally eliminated but it can be controlled if correct statistical procedures are employed. The unconditional permutation approach is a statistical procedure that ensures that the probability of a type I error is exactly α , thus ensuring that the resulting distribution of the test statistic is exact, see Agresti (1992), Good (2000), Pesarin (2001), Odiase and Ogbonmwan (2005) and Ogbonmwan *et al*. (2007). The unconditional exact permutation approach where row and column totals are allowed to vary with each permutation is very much unlike the conditional exact permutation approach of fixing the row and column totals (Headrick, 2003; Bagui and Bagui, 2004; Odiase and Ogbonmwan, 2005). Exact tests constructed by restricting attention to a conditional reference set of contingency tables with margins fixed at the values actually observed is not always true in nature. The unconditional permutation approach is computationally very demanding and more complex than the conditional approach (Agresti, 1992; Good, 2000; Opdyke, 2003). Earlier study Agresti (1992) observed that the results obtained from the conditional and unconditional permutation approaches can be quite discrepant.

Another study Scheffe (1943) showed that the permutation approach is the only possible technique of constructing exact tests of significance for a general class of problems. Hoeffding (1952) remarked that this permutation test is asymptotically as powerful as the best parametric test. There are several Monte Carlo methods that can be used in generating exact p-values. The most widely used is the bootstrap re-sampling technique developed by Efron (1979).

The Bayesian and the Likelihood approaches can be found in Bayarri and Berger (2004), Spiegelhalter (2004). All these alternative approaches to the unconditional permutation approach only give approximate results. Exact procedures are the best and should always be applied whenever it is practically possible, Lehmann (1986) and Good (2000). Permutation tests provide exact results especially when complete enumeration is possible, Pesarin (2001). A big challenge in using nonparametric test is the availability of computational formulas and tables of exact critical values. This continues to be a problem as revealed by a survey of 20 in-print general college statistics texts, Fahoome (2002). Many nonparametric tests have large sample approximations that can be used as an alternative to tabulated exact critical values. These approximations are useful substitutes if the sample size is sufficiently large and hence obviate the need for locating tables of exact critical values. However, there is no generally agreed upon definition of what constitutes a large sample size for most statistics (Bergmann *et al*., 2000; Fahoome, 2002). A complete enumeration of the permutation sample space for the purpose of constructing an exact test of significance is only possible when sample sizes are small

*Corresponding author e-mail: friday.ewere@uniben.edu

(Odiase and Ogbonmwan, 2005; Ewere and Ogbonmwan, 2020). The consideration given in this article produces the exact distribution of the Wilcoxon Rank Sum test statistic (WRS) by precisely tracking the number of permutations without carrying out a complete enumeration. Thus, providing the possibility of finding the exact distribution of the WRS for larger sample sizes. When sample sizes are large, the exact distribution of the WRS can be approximated by the normal approximation. We study the convergence of the normal approximation to the exact distribution of the WRS and provide the minimum sample size required for the application of the asymptotic distribution. Computations are done using the computer algebra Mathematica 6.0.

MATERIALS AND METHODS

Methodology

Exact distribution of the Wilcoxon rank sum test

The Wilcoxon Rank Sum Test is a nonparametric alternative to the two-sample t-test. Suppose we have two independent samples $X_{1,}X_2...,X_{n_1}$ and $Y_1,Y_2,...,Y_{n_2}$ of sizes n_1 and n_2 drawn from two continuous populations whose distributions are F and G respectively. We wish to test the null hypothesis $H_0 : F = G$. The alternatives could be $H_1 : F > G,\ \ H_1 : F < G,\ \ H_1 : F \neq G,$. The Wilcoxon Rank Sum Test statistic WRS[n_1, n_2] is the sum of the ranks from one of the samples. That is

$$\text{WRS}[n_1, n_2] = \sum_{i=1}^{n_1} r_{1i}, \qquad [1]$$

Where $r_{1i,}, i = 1(1)n_1$ represents the ranks of the first sample.

It rejects the null hypothesis H_0 if the sum of the ranks of the X_i`s in the combined ordered arrangement of the two samples is either too large or too small. Here, too large or two small implies WRS[n_1, n_2]$\geq W_{1-\alpha}$ or WRS[n_1, n_2]$\leq W_\alpha$ respectively. $W_{1-\alpha}$ and W_α are the upper-tail and lower-tail exact critical values respectively of the distribution of WRS[n_1, n_2] and α is the level of significance of the test. The values for $W_{1-\alpha}$ and W_α are provided in Table 1 The null distribution of WRS[n_1, n_2] is found by assuming that X_i and Y_j are identically distributed. This is true only when H_0 is true in the two-tailed test. If the X_i and the Y_j are independent and identically distributed, then every arrangement of the X's and Y's in the ordered combined sample is equally likely. This is the basic principle behind many rank tests.

The probability distribution of WRS[n_1, n_2] may be obtained by considering the probability distribution of the sum of n_1 integers selected at random, without replacement, from among the integer from 1 to $n_1 + n_2$. The number of ways of selecting n_1 integers from a total number of $n_1 + n_2$ integers is

$$\Phi = \binom{n_1 + n_2}{n_1} = \frac{(n_1 + n_2)!}{n_1! n_2!} \qquad [2]$$

and each has probability Φ^{-1} of occurring. Hence the probability that WRS[n_1, n_2]=K_1 may be found by counting the number of different sets of n_1 integers from 1 to $n_1 + n_2$ that add up to the value K_1 and then dividing by Φ. However, as the sample sizes increase, it becomes difficult to obtain the distribution of WRS[n_1, n_2] because of the very large cardinality of the permutation sample spaces. As an example, if $n_1 = n_2 = 15$, there are 155,117,520 associated permutation sample spaces. In such instances, the large sample approximation is usually used. But, there is no clear definition of what constitute a large sample for WRS[n_1, n_2], (Fahoome, 2002; Ogbonmwan *et al*., 2007).

To calculate the probability that a statistic *X* based on ranks will take a value *x* which we denote as $\Pr ob(\{X = x\})$, it is therefore only necessary to obtain the number of cases satisfying the condition $X = x$. Following the idea of Baglivo *et al*. (1996), we formulate a combinatorial problem and develop generating functions to solve the problem formulated. This provides useful insight into the exact null distribution of the WRS statistic.

Combinatorial Problem

Suppose we have n observations which are ranked 1, 2, 3…,n. In how many different ways is it possible to divide these n observations among k samples such that the i^{th} sample T_i contains n_i observations and the sum of the ranks of these n_i observations in sample T_i is r_i with $n = \sum_{i=1}^{k} n_i$ and $r = \sum_{i=1}^{k} r_i = \frac{1}{2} n(n+1)$? Let the number be:

$$P[nlist, rlist] := P[\{n_{1,} n_{2,} ..., n_k\}, \{r_{1,} r_2 ..., r_k\}] \qquad [3]$$

We can calculate this number $P[nlist, rlist]$ by counting the relevant partitions. There are $\frac{(n_1 + n_2 + ... + n_k)!}{n_1! n_2! ..., n_k!}$ possible permutations of the n variates of the k samples of sizes $n_i, i = 1,2,...,k$ which are equally likely with probability $\left(\frac{n!}{n_1! n_2! ..., n_k!}\right)^{-1}$. The number $P[nlist, rlist]$ can easily be obtained for small n and k by counting the relevant partitions, for example, $P[\{3,2\},\{8,7\}] = 2$ which requires only 10 distinct arrangements (partitions). However, when n and k are not as small as in the above example, this method of obtaining $P[nlist, rlist]$ fails because of the large associated permutation sample spaces. For instance, when $n_1 = 10, n_2 = 7, n_3 = 3$, there are 22,170,720 distinct arrangements of the ranks. Admittedly, it is very difficult to carry out this enumeration manually in order to compute $P[nlist, rlist]$.

To overcome this problem of enumeration, we find the generating function for the number $P[nlist, rlist]$. To do this, Let $x[i]$ be a variable governing the number of observations in the i^{th} sample and $y[i]$ be a variable governing the sum of the ranks of the observations in the i^{th} sample. Then, the generating function for the number $P[nlist, rlist]$ is given as

$$p[n,k] = \prod_{j=1}^{n} \left(\sum_{i=1}^{k} x[i] y[i]^j \right) \quad [4]$$

See Ewere and Ogbonmwan (2010a)

Obviously, the numbers $P[nlist, rlist]$ are the coefficients of $\prod_{i=1}^{k} x[i]^{n_i} y[i]^{r_i}$ of the polynomial $p[n,k]$. Hence, $P[nlist, rlist]$ is obtained by selecting the coefficients of $\prod_{i=1}^{k} x[i]^{n_i} y[i]^{r_i}$. However, this method of enumeration is not as fast as one would expect due to the fact that the number of terms of the generating function in eqn [4] are of order k^n which is not too small even if n and k are not very large.

To improve on the computational efficiency of eqn [4], we let $nlist = \{n_1, n_2, ..., n_k\}$. In this case, the generating function $p[nlist]$ for the number $P[nlist, rlist]$ have number of terms whose order is only $Multinomial[n_1, n_2, ..., n_k]$ which is smaller than k^n. Clearly, this new generating function $p[nlist]$ are the coefficients of $\prod_{i=1}^{k} x[i]^{n_i}$ of the generating function $p[n,k]$. To speed up computations, the generating function $p[nlist]$ is defined recursively as:

$$p[nlist] = p[\{n_1, n_2, ..., n_k\}] = \sum_{i=1}^{k} y[i]^n \, p[\{n_1, n_2, ..., n_i - 1, ..., n_k\}] \quad [5],$$

See Ewere and Ogbonmwan (2010b).

In eqn [5], the number of ranked observations in the i^{th} sample is reduced by one during the exchangeability (rearrangement) process by allowing the n^{th} rank to be a member of the i^{th} sample. By being systematic and proceeding in this orderly fashion from one rearrangement to the next, we have substantially reduced the time required to examine a series of rearrangements. This idea has been suggested by Good (2000).

The exact null distribution of WRS[n_1, n_2] is obtained by finding the generating function of the test statistic. The generating function for WRS[n_1, n_2] can be written as:

$$G[WRS[n_1, n_2]] = \sum \mathrm{Prob}(\{WRS[n_1, n_2] = K_1\}) t^{K_1} \quad [6]$$

$$0 \le K_1 \le \frac{n(n+1)}{2} \text{ and } K_2 = \frac{n(n+1)}{2} - K_1 \quad [7]$$

Then

$$\mathrm{Prob}(\{WRS[n_1, n_2] = K_1\}) = \frac{P[\{n_1, n_2\},\{K_1, K_2\}]}{Binomial[n, n_1]} \quad [8]$$

where $n = n_1 + n_2$ and $Binomial\,[n, n_1] = \frac{(n_1 + n_2)!}{n_1! n_2!}$

Eqn [6] can be used to calculate the distributional characteristics of WRS [n_1, n_2].

The large sample approximation

When sample sizes are large, the time required to compute a permutation distribution can be prohibitive even if we are taking advantage of one of the optimal computing algorithm (Good, 2000). Fortunately, when sample sizes are large, we can make use of an asymptotic approximation in place of the exact distribution. However, asymptotic approximations are to be avoided except with very large samples (Good, 2000). This is because they can be grossly in error (Micceri, 1989; Mudholkar and Hutson, 1997).

Table 1. Exact Critical Values for the Wilcoxon Rank Sum test.

n_1	n_2	$W_{0.9000}$	$W_{0.9500}$	$W_{0.9750}$	$W_{0.9900}$	$W_{0.9950}$	$W_{0.9975}$	$W_{0.9990}$
2	2	-	-	-	-	-	-	-
	3	8 (3)	-	-	-	-	-	-
	4	10 (4)	-	-	-	-	-	-
	5	11 (5)	12 (4)	-	-	-	-	-
3	3	13 (7)	14 (6)	-	-	-	-	-
	4	16 (8)	17 (7)	-	-	-	-	-
	5	18 (9)	19 (8)	20 (7)	-	-	-	-
	6	20 (10)	21 (9)	22 (8)	-	-	-	-
4	4	22 (13)	24 (12)	25 (11)	-	-	-	-
	5	25 (15)	27 (13)	28 (12)	29 (11)	-	-	-
	6	28 (16)	30 (14)	31 (13)	32 (12)	33 (11)	-	-
	7	31 (17)	33 (15)	34 (14)	36 (12)	37 (11)	-	-
5	5	34 (21)	35 (20)	37 (18)	38 (17)	39 (16)	-	-
	6	37 (23)	39 (21)	41 (19)	42 (18)	43 (17)	44 (16)	-
	7	41 (24)	43 (22)	44 (21)	46 (19)	48 (17)	49 (16)	-
	8	44 (26)	46 (24)	48 (22)	50 (20)	52 (18)	53 (17)	54 (16)
6	6	47 (31)	49 (29)	51 (27)	53 (25)	54 (24)	55 (23)	-
	7	51 (33)	54 (30)	56 (28)	58 (26)	59 (25)	60 (24)	62 (22)
	8	55 (35)	58 (32)	60 (30)	62 (28)	64 (26)	65 (25)	67 (23)
	9	59 (37)	62 (34)	64 (32)	67 (29)	69 (27)	70 (26)	72 (24)
7	7	63 (42)	65 (40)	68 (37)	70 (35)	72 (33)	73 (32)	75 (30)
	8	67 (45)	70 (42)	73 (39)	76 (36)	77 (35)	79 (33)	81 (31)
	9	72 (47)	75 (44)	78 (41)	81 (38)	83 (36)	85 (34)	87 (32)
	10	76 (50)	80 (46)	83 (43)	86 (40)	88 (38)	90 (36)	92 (34)
8	8	80 (56)	84 (52)	86 (50)	90 (46)	92 (44)	93(43)	95(41)
	9	85 (59)	89 (55)	92 (52)	96 (48)	98 (46)	100(44)	102(42)
	10	91 (61)	95 (57)	98 (54)	102 (50)	104 (48)	106(46)	109(43)
	11	96 (64)	100 (60)	104 (56)	108 (52)	110 (50)	112(48)	115(45)
9	9	100(71)	104(67)	108(63)	111(60)	114(57)	116(55)	118(53)
	10	106(74)	110(70)	114(66)	118(62)	121(59)	123(57)	126(54)
	11	112(77)	116(73)	120(69)	125(64)	127(62)	130(59)	133(56)
	12	117(81)	122(76)	126(72)	131(67)	134(64)	137(61)	140(58)
10	10	122(88)	127(83)	131(79)	135(75)	138(72)	141(69)	144(66)
	11	128(92)	133(87)	138(82)	142(78)	146(74)	148(72)	152(68)
	12	135(95)	140(90)	145(85)	150(80)	153(77)	156(74)	160(70)
	13	141(99)	147(93)	151(89)	157(83)	160(80)	164(76)	167(73)
11	11	146(107)	152(101)	156(97)	161(92)	165(88)	168(85)	171(82)
	12	153(111)	159(105)	164(100)	169(95)	173(91)	176(88)	180(84)
	13	160(115)	166(109)	171(104)	177(98)	181(94)	184(91)	188(87)
	14	167(119)	173(113)	179(107)	185(101)	189(97)	193(93)	197(89)
12	12	172(128)	179(121)	184(116)	190(110)	194(106)	197(103)	201(99)
	13	180(132)	186(126)	192(120)	198(114)	202(110)	206(106)	210(102)
	14	187(137)	194(130)	200(124)	207(117)	211(113)	215(109)	220(104)
	15	194(142)	202(134)	208(128)	215(121)	220(116)	224(112)	229(107)
13	13	201(150)	208(143)	214(137)	220(131)	225(126)	229(122)	233(118)
	14	209(155)	216(148)	222(142)	229(135)	234(130)	238(126)	243(121)
	15	217(160)	224(153)	231(146)	238(139)	243(134)	248(129)	253(124)
14	14	231(175)	239(167)	245(161)	253(153)	258(148)	262(144)	268(138)
	15	240(180)	248(172)	255(165)	263(157)	268(152)	273(147)	278(142)
15	15	264(201)	272(193)	280(185)	288(177)	293(172)	298(167)	304(161)
20	20	458(362)	471(349)	482(338)	495(325)	504(316)	512(308)	521(299)
30	30	1002(898)	1026(868)	1047(846)	1071(822)	1088(807)	1102(793)	1120(776)
40	40	1754(1486)	1791(1449)	1823(1417)	1861(1379)	1886(1354)	1909(1331)	1937(1303)
50	50	2711(2339)	2764(2286)	2809(2241)	2861(2189)	2897(2153)	2929(2121)	2969(2081)
60	60	3875(3405)	3944(3345)	4003(3295)	4072(3241)	4118(3206)	4161(3175)	4214(3138)

Several researchers have described the use of classic parametric statistics in the face of assumption violations as invalid (Keselman *et al.*, 1998; Leech and Onwuegbuzie, 2002; Wilcox, 2001; Grissom and Kim, 2005; Erceg-Hurn and Mirosevich, 2008). To better understand this, Royeen (1986) identified five published studies that used parametric statistics. For each study, the data were checked whether they met the assumptions for the parametric statistics used. In three of the five studies, the assumptions were not met. Next, the appropriate nonparametric statistic was computed on the data. For each of the three studies that did not meet the assumptions, there were large differences in the results yielded by the nonparametric statistic when compared with the published results from the parametric statistic. Thus, this examination demonstrates that if the assumptions are not met, the results can be very misleading.

The fundamental asymptotic result for the permutation distribution of the two-sample test statistic for a location parameter was first stated by Madow (1948) and formalized by Hoeffding (1951, 1952) who demonstrated convergence of the distribution of the studentized test statistic under the alternative as well as under the null hypothesis.

Let $T_n = T\left(X_{(1)},\ldots,X_{(n)}\right)$ be the test statistic and let μ_n and σ_n be its first and second moments respectively. Then the permutation distribution F_n of Z_n is $\frac{T_n - \mu_n}{\sigma_n}$ [9]

Eqn [9] is obtained by randomly rearranging the subscripts of the arguments of T_n and this converges to ϕ, the Gaussian (normal) distribution function.

To enable researchers who do not have access to the necessary tables of critical values to employ these tests, it is important to determine the minimum sample size in order to apply the large sample approximation for various statistics. Besides, the statistical assumption tests built into software such as SPSS often do a poor job of detecting violations from normality and homoscedasticity (Jaccard and Guilamo-Ramos, 2002).

In order to determine the minimum sample size required for the application of asymptotic results of the WRS test, we use the Bradley's (1978) conservative estimates of 0.045 < Type I error rate < 0.055 and 0.009 < Type I error rate < 0.011 as measures of robustness when nominal α was set at 0.05 and 0.01, respectively. Generally, the stringent criterion $0.9\alpha \le \alpha_0 \le 1.1\alpha$ where α_0 is the true probability of a type I error when one or more of a test's assumptions are violated and the null hypothesis is true seems more appropriate to illustrations of 'convergence' than the liberal criterion given by Cochran (1952), who considered actual significance levels less than 20% above the nominal level to be acceptable, Sullivan and D'Agostino (1992). The sample sizes were increased until the Type I error rates converged within these acceptable regions.

RESULTS AND DISCUSSION

The values in each cell of Table 1 represent the upper and lower critical values of the Wilcoxon Rank Sum test statistic with the lower critical values in brackets. We study the convergence of the asymptotic distribution (the normal distribution) to the exact distribution of the WRS test both numerically and graphically in Tables 1 through 7 and Figures 1 through 10. We determine the minimum sample size in order to apply the asymptotic distribution using the Bradley's conservative estimate.

Table 2. Exact and Asymptotic Type I error rates for WRS $\alpha = 0.01$

n_1	n_2	**Exact**	**Asymptotic**
5	5	0.00793651	0.0141401
6	6	0.00757576	0.0124873
7	7	0.00874126	0.0126737
8	8	0.00738151	0.0104313
9	9	0.00937886	0.0121705
10	10	0.00927169	0.0116711
11	11	0.00961538	0.0117428
12	12	0.0086356	0.0104607
13	13	0.00954935	0.0112432
14*	14*	0.00927701	0.0107985
15	15	0.00927686	0.0106667
16	16	0.00946554	0.0107518
17	17	0.00979409	0.0109963
18	18	0.00935297	0.0104547
19	19	0.00991213	0.0109589
20	20	0.00976838	0.0107452

Table 3. Exact and Asymptotic Type I error rates for WRS $\alpha = 0.025$

n_1	n_2	**Exact**	**Asymptotic**
4	4	0.014286	0.021654
5	5	0.015873	0.023601
6	6	0.0205628	0.027332
7	7	0.0189394	0.0238227
8	8	0.0249417	0.0293536
9*	9*	0.0199918	0.0234728
10	10	0.0216285	0.0246835
11	11	0.0236536	0.0263654

12	12	0.022451	0.0248236
13	13	0.0220583	0.0241705
14	14	0.0248678	0.0268157
15	15	0.0226669	0.0244071
16	16	0.0234075	0.0250082
17	17	0.0243428	0.0258237
18	18	0.0235424	0.0249051
19	19	0.0248279	0.0261017
20	20	0.0245452	0.0257312

Table 4. Exact and Asymptotic Type I error rates for WRS $\alpha = 0.05$

n_1	n_2	**Exact**	**Asymptotic**
4	4	0.0285714	0.0416323
5	5	0.047619	0.0585927
6	6	0.0465368	0.0546573
7	7	0.0486597	0.0551116
8*	8*	0.0414918	0.046446
9	9	0.046956	0.0511725
10	10	0.0446048	0.0481518
11	11	0.0439731	0.0470205
12	12	0.0443669	0.0470343
13	13	0.0454236	0.0477904
14	14	0.0469341	0.049052
15	15	0.0487629	0.0506708
16	16	0.0469062	0.0486272
17	17	0.0493437	0.0509127
18	18	0.0485296	0.0499635
19	19	0.0482046	0.0495223
20	20	0.0482498	0.049466

Table 5. Exact and Asymptotic Type I error rates for WRS $\alpha = 0.10$

n_1	n_2	**Exact**	**Asymptotic**
3	3	0.10000	0.137617
4	4	0.10000	0.124106
5	5	0.075397	0.0872629
6	6	0.089827	0.100092
7	7	0.082459	0.0898563
8*	8*	0.097436	0.103789
9	9	0.095125	0.100206
10	10	0.095158	0.099383
11	11	0.096593	0.10019
12	12	0.098904	0.102012
13	13	0.092844	0.0954888
14	14	0.096789	0.0991284
15	15	0.093634	0.095681
16	16	0.098187	0.100023
17	17	0.096608	0.098242
18	18	0.095819	0.0972845
19	19	0.095621	0.0969432
20	20	0.095876	0.0970745

Table 6. Exact and Asymptotic Type I error rates for WRS $\alpha = 0.005$

n_1	n_2	**Exact**	**Asymptotic**
5	5	0.00396825	0.00814685
6	6	0.004329	0.0081546
7	7	0.0034965	0.00635814
8	8	0.0034965	0.00585932
9	9	0.00388729	0.00592465
10	10	0.00446535	0.00630572
11	11	0.00416482	0.00573401
12	12	0.00414658	0.0055372
13	13	0.00430052	0.00556705
14	14	0.00457314	0.0057501
15	15	0.00493741	0.00604638
16	16	0.00477527	0.00578232
17	17	0.00474725	0.00567714
18	18	0.00481544	0.00568552
19	19	0.00495689	0.00577918
20	20	0.0047418	0.0054999
21	21	0.00462845	0.0053353
22	22	0.00493681	0.00562165
23	23	0.00493478	0.00558331
24	24	0.0049846	0.00560284
25*	25*	0.00478432	0.00536339
26	26	0.00492306	0.0054817
27	27	0.00482933	0.00535919
28	28	0.00478341	0.00528889
29	29	0.00477642	0.00526094
30	30	0.00480184	0.00526824
31	31	0.00485479	0.0053054
32	32	0.00493154	0.00536824
33	33	0.00483841	0.00525655
34	34	0.00495948	0.00536708
35	35	0.00491999	0.00531282

Table 7. Exact and Asymptotic Type I error rates for WRS $\alpha = 0.0025$

n_1	n_2	**Exact**	**Asymptotic**
6	6	0.0021645	0.00520282
7	7	0.00203963	0.00440432
8	8	0.002331	0.00432575
9	9	0.00199506	0.00353844
10	10	0.0019431	0.00325094
11	11	0.00205123	0.0032141
12	12	0.00225653	0.00332835
n_1	n_2	**Exact**	**Asymptotic**
13	13	0.00210757	0.00303857
14	14	0.00245278	0.00335493
15	15	0.00246981	0.00329566
16	16	0.00224209	0.00296805
17	17	0.00239608	0.00308823
18	18	0.00232243	0.00295661
19	19	0.00230915	0.00289976
20	20	0.0023406	0.00289799

21	21	0.00240712	0.00293864
22	22	0.00231233	0.00280479
23	23	0.00243746	0.00291553
24	24	0.00241165	0.00286333
25	25	0.00241661	0.00284695
26	26	0.00244654	0.00285945
27	27	0.00249737	0.00289584
28	28	0.00243136	0.0028082
29	29	0.00239359	0.0027524
30	30	0.0024972	0.00284885
31	31	0.00249536	0.0028336
32	32	0.00240266	0.00272278
33	33	0.00243653	0.00274742
34	34	0.00248261	0.00278551
35*	35*	0.00244438	0.00273515
36	36	0.00242278	0.00270304
37	37	0.0024153	0.00268644
38	38	0.00242002	0.0026832
39	39	0.00243542	0.00269159
40	40	0.0024603	0.00271038
41	41	0.0024937	0.00273834
42	42	0.00246273	0.00269937
43	43	0.0024433	0.00267283
44	44	0.00243394	0.00265723
45	45	0.00249759	0.00271778

Plots of the exact and asymptotic cumulative distribution functions of the WRS test statistic for different sample sizes.

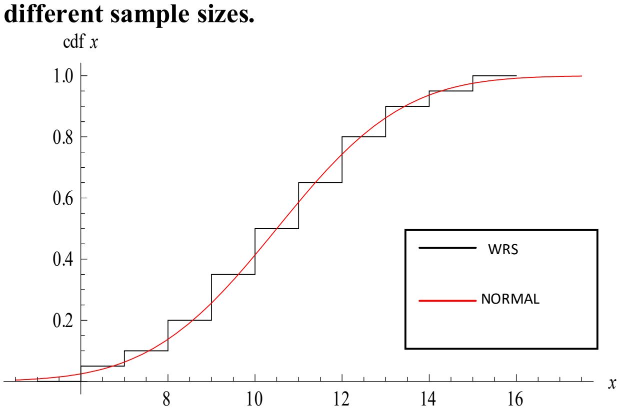

Fig. 1. $n_1 = n_2 = 3$

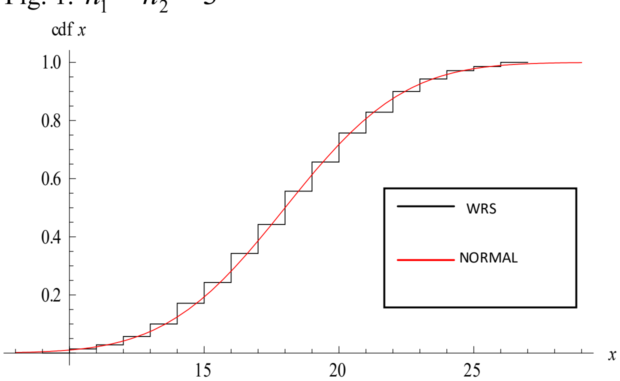

Fig. 2. $n_1 = n_2 = 4$

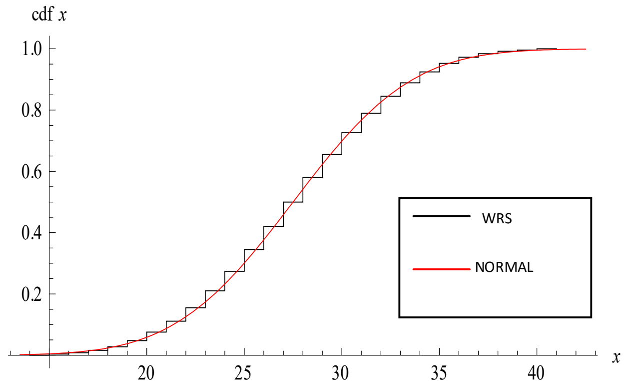

Fig. 3. $n_1 = n_2 = 5$

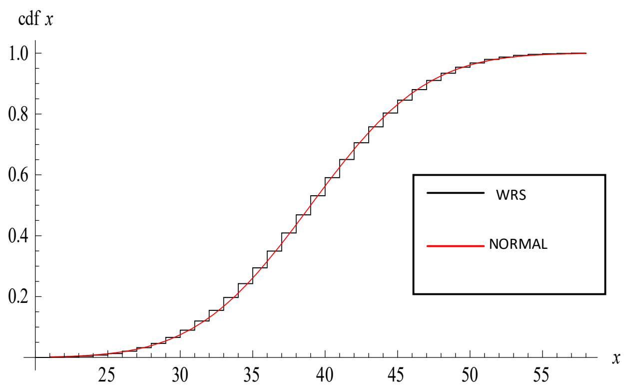

Fig. 4. $n_1 = n_2 = 6$

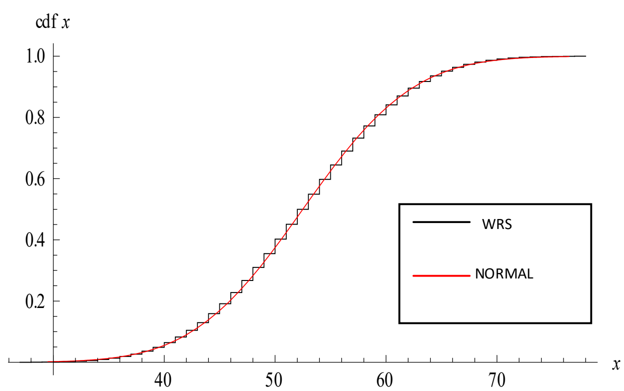

Fig. 5. $n_1 = n_2 = 7$

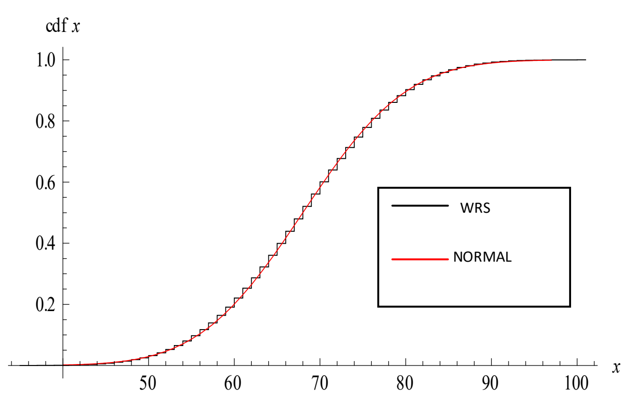

Fig. 6. $n_1 = n_2 = 8$

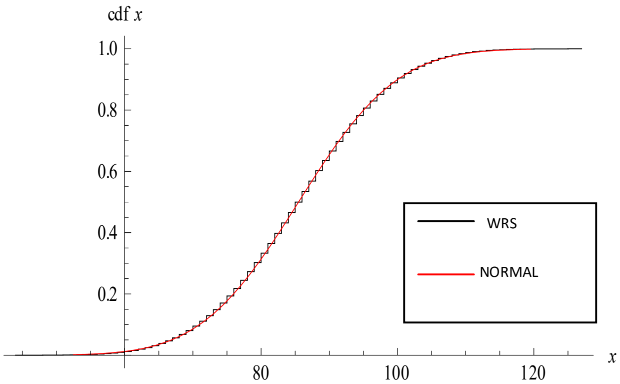

Fig. 7. $n_1 = n_2 = 9$

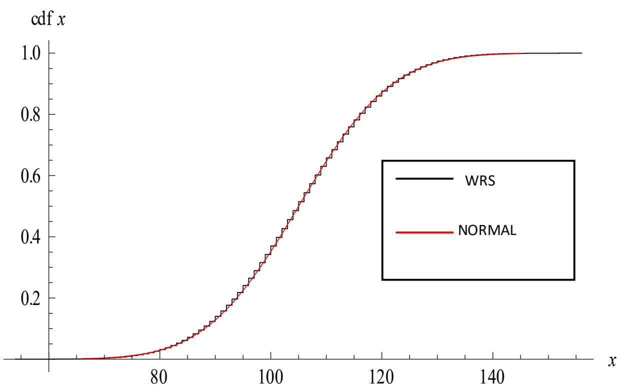

Fig. 8. $n_1 = n_2 = 10$

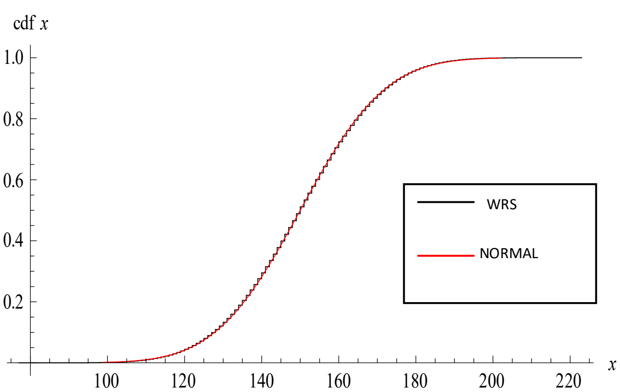

Fig. 9. $n_1 = n_2 = 12$

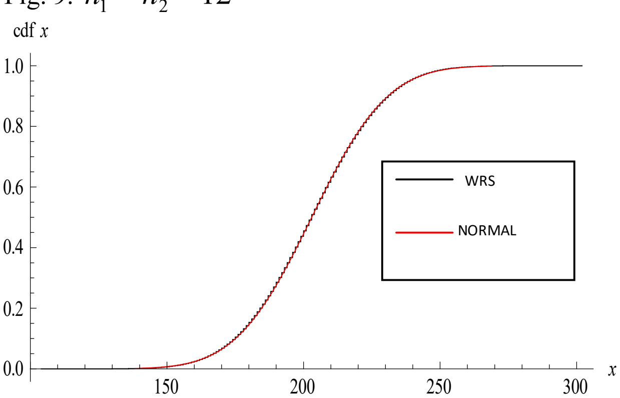

Fig. 10. $n_1 = n_2 = 14$

Table 1 provides exact critical values for the WRS test statistic. The exact critical values for the test statistic has been provided for some combination of sample sizes as it is not practical to present the entire distribution here, but these are easily obtained from the computer program available on request from the authors.

When sample sizes are large, it is reasonable to use the large sample approximation. But the question has always been "How large is large? Providing exact critical values for the WRS test statistic for small and larger samples do not only negate the need for approximations in many additional settings, but they also allow us to study the approximation and make more reasonable inference about the usefulness of the approximation. What we are more concerned about is what the performance of the approximation in the smaller-sample conditions might imply about conditions for which there are still no exact tables.

Tables 2 through 7 shows the exact and asymptotic Type I error rates for the WRS test statistic for nominal level of significance $\alpha = 0.01$, 0.025, 0.05, 0.10, 0.005 and 0.0025 respectively. We use these results to determine the minimum sample sizes necessary to use the large sample approximation of the critical value of the WRS statistic. We superimpose the curve of the normal distribution to that of the WRS in Figures 1 through Figures 10. The exact distribution of the WRS test becomes more stable and closer to the normal distribution as the group sample size increases as depicted in Figures 1 through 10. The normal approximation of the exact critical values of the WRS test is adequate when $n_1 = n_2 = 14$ (14 per sample) for $\alpha = 0.01$, $n_1 = n_2 = 9$ for $\alpha = 0.025$, $n_1 = n_2 = 8$ for $\alpha = 0.05$ and 0.10. A minimum of 25 and 35 per sample is adequate to apply the normal approximation for $\alpha = 0.005$ and $\alpha = 0.0025$ respectively. The minimum sizes per sample required to apply the asymptotic distribution have been reported and are asterisked in Tables 2 through 7. These recommendations in this paper are based on the results that converged using the Bradley's (1978) conservative estimates.

It is instructive here to point out the contributions of other authors as regards the minimum sample size for the application of the large sample approximation for the WRS statistic. Mann and Whitney (1947) considered the case of unpaired data with samples of equal sizes and reported tables only up to sample sizes of 8, that is, $n_1 = n_2 = 8$ and concluded that "*at this point the distribution is almost normal*". They probably failed to recognize the influence upon robustness of α, Bradley (1978). And very early, it was thus acknowledged that the normal approximation was acceptable for rather small sample sizes. Conover (1971) recommended that one or both sample must exceed 20. Gibbons (1971) placed the lower limit at 12 per sample. Another study Sprent (1989) also suggested that one or both samples must exceed 20. While, Deshpande *et al*. (1995) stated that the combined sample size should be at least 20 to use a large sample approximation of the critical value of the WRS statistic. Fahoome (2002) recommended a minimum of 15 and 29 per sample for $\alpha = 0.05$ and $\alpha = 0.01$ respectively. Ogbonmwan *et al*. (2007) suggested a minimum of 7 per sample to apply the asymptotic results for $\alpha = 0.05$. The definition of what constitute a large sample for the WRS test is quite vague. We have attempted to address this vagueness in this article. The results in Table 2 clearly demonstrate that the large sample approximation of the critical value prevents the statistic from converging with nominal $\alpha = 0.01$ if Bergmann et. al (2000) are correct with their perception of common practices using as few as 11 per sample.

CONCLUSION

In this study, a straight forward but logical approach has been adopted in developing a procedure for constructing an exact test of significance. With this approach, the exact critical values of the Wilcoxon Rank Sum test have been accurately generated, thereby ensuring that the probability of making a type I error is exactly α. Numerical evaluations and graphical illustrations of the asymptotic property of the WRS test have been presented. We feel, however, that such representations (numerical and graphical) could be valuable tools when introducing these statistics. This claim is well exemplified by the following quotation from Bellera *et al*. (2010):

We examined a convenience sample of 12 introductory statistics textbooks and three nonparametric statistics textbooks available to students at the McGill University Science Library. None of these books included graphical display of the distributions of the Wilcoxon Statistics. Only one book (Lehmann, 1998) included recursive formulas.

REFERENCES

Agresti, A. 1992. A survey of exact inference for contingency tables. Statistical Science. 7: 131-177.

Baglivo, J., Pagano, M. and Spino, C. 1996. Permutation distributions via generating function with applications to sensitivity analysis of discrete data. Journal of the American Statistical Association. 91:1037-1046.

Bagui, S. and Bagui, S. 2004. An algorithm and code for computing exact critical values for the Kruskal-Wallis nonparametric one-way ANOVA. Journal of Modern Applied Statistical Methods. 3:498-503.

Bayarri, MJ. and Berger, JO. 2004. The interplay of Bayesian and frequentist analysis. Statistical Science. 19:58-80.

Bellera, CA., Julien, M. and Hanley, JA. 2010. Normal approximations to the Distributions of the Wilcoxon Statistics: Accurate to What N? Graphical Insights. Journal of Statistics Education. 18(2):1-17.

Bergmann, R., Ludrook, J. and Spooren, WPJM. 2000. Different outcomes of the Wilcoxon-Mann-Whitney test from different statistics packages. American Statistician. 54:72-77.

Bradley, JV. 1978. Robustness? British Journal of Mathematical and Statistical Psychology. 31:144-152.

Cochran, WG. 1952. The χ^2 test of goodness of fit. Ann Math Stat. 23:315-345.

Conover, WJ. 1971. Practical Nonparametric statistics. Wiley and Sons Inc., New York, USA.

Deshpande, JV, Gore, AP. and Shanubhogue, A. 1995. Statistical analysis of non-normal data. John Wiley and Sons, Inc., New York, USA.

Efron, B. 1979. Bootstrap methods: another look at the Jackknife. The Annals of Statistics. 7:1-26.

Erceg-Hurn, DM. and Mirosevich, VM. 2008. Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. American Psychologist. 63(7):591- 601.

Ewere, F. and Ogbonmwan, SM. 2010[a]. A method for generating the permutation distribution of ranks in a k-sample experiment. Journal of the Nigerian Association of Mathematical Physics. 16:579-584.

Ewere, F. and Ogbonmwan, SM. 2010[b]. A fast algorithm for generating the permutation distribution of ranks in a k-sample experiment. Journal of the Nigerian Association of Mathematical Physics. 16:585-590.

Ewere, F. and Ogbonmwan, SM. 2020. Exact Critical Values of the Kruskal Wallis Test. Canadian Journal of Pure and Applied Sciences. 14(3):5105-5115.

Fahoome, G. 2002. Twenty Nonparametric Statistics and their large sample Approximations. Journal of Modern Applied Statistical Methods. 1:248-268.

Gibbons, JD. 1971. Nonparametric statistical inference. McGraw-Hill book Company. New York, USA.

Good, P. 2000. Permutation Tests: A practical Guide to Re-sampling methods for Testing Hypothesis (2nd edi.). Springer-Verlag, New York, USA.

Grissom, RJ. and Kim, JJ. 2005. Effect sizes for research: A broad practical approach. Mahwah, NJ: Erlbaum.

Headrick, TC. 2003. An algorithm for generating exact critical values for the Kruskal-Wallis One-Way ANOVA. Journal of Modern Applied Statistical Methods. 2:268-271.

Hoeffding, W. 1951. Combinatorial central limit theorem. Ann. Math. Statist. 22:556-558.

Hoeffding, W. 1952. Large sample power of tests based on permutations of observations. The Annals of Mathematical Statistics. 23:169-192.

Jaccard, J. and Guilamo-Ramos, V. 2002. Analysis of variance frameworks in clinical child and adolescent psychology: Advanced issues and recommendations. Journal of Clinical Child Psychology. 31:278-294.

Keselman, HJ., Huberty, CJ., Lix, LM., Olejnik, S., Cribbie, RA. and Donahue, B. 1998. Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. Review of Educational Research. 68:350-386.

Leech, NL. and Onwuegbuzie, AJ. 2002. A call for greater use of nonparametric statistics. Paper presented at the Annual Meeting of the Mid South Educational Research Association. Retrievedfromhttp://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED471346.

Lehmann, E. 1998. Nonparametrics-Statistical Methods Based on Ranks (Revised 1st edi.). San Francisco: Holden-Day Inc.

Lehmann, EL. 1986. Testing Statistical Hypothesis. (2 *nd* edi.). Wiley, New York, USA.

Madow, WG. 1948. On the limiting distribution of estimates based on samples from finite universes. Ann. Math. Stat. 19:534-545.

Mann, HB. and Whitney, DR. 1947. On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics. 18:50-60.

Miccceri, T. 1989. The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin. 105:156-166.

Mudholkar, GS. and Hutson, AD. 1997. Continuity corrected approximations for and “exact” inference with Pearson’s chi-square. J. Statist. Plan. Infer. 23:61-78.

Odiase, JI. and Ogbonmwan, SM. 2005. An algorithm for generating unconditional exact permutation distribution for a two-sample experiment. Journal of Modern Applied Statistical Methods. 4:319-332.

Ogbonmwan, SM, Odiase, JI. and Aitusi, DN. 2007: Exact permutation critical values for the Wilcoxon Rank Sum Test. International Journal of Natural and Applied Sciences. 3(1):90- 95.

Opdyke, JD. 2003. Fast permutation tests that maximize power under conventional Monte Carlo sampling for pairwise and multiple comparisons. Journal of Modern Applied Statistical Methods. 2:27-49.

Pesarin, F. 2001. Multivariate Permutation Tests. John Wiley and Sons. New York, USA.

Royeen, CB. 1986. A comparison of parametric versus nonparametric statistics. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA., USA.

Scheffe, H. 1943. Statistical Inference in the nonparametric case. The Annals of Mathematical Statistics. 14:305-332.

Spiegelhalter, DJ. 2004. Incorporating Bayesian ideas into health-care evaluation. Statistical Science. 19:156-174.

Sprent, P. 1989. Applied nonparametric statistical methods. Chapman and Hall. London.

Sullivan, LM. and D’Agostino, RB. 1992. Robustness of the t Test applied to data distorted from normality by floor effects. Journal of Dental Research. 71(12):1938-1943.

Wilcox, RR. 2001. Fundamentals of modern statistical methods. Springer. New York, USA.

Received: August 29, 2021; Revised: Sept 1, 2021;
Accepted: Sept 24, 2021