Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

An approximation to the exact distribution of the wilcoxon-mann-whitney rank sum test statistic

Ronald L. Iman [a]

[a] Sandia Laboratories , Albuquerque, N. M., 87115, Division 1223

Published online: 27 Jun 2007.

COMMUN. STATIST.-THEOR. METH., A5(7), 587-598 (1976)

AN APPROXIMATION TO THE EXACT DISTRIBUTION OF THE WILCOXON-MANN-WHITNEY RANK SUM TEST STATISTIC

Ronald L. Iman

Sandia Laboratories, Division 1223, Albuquerque, N. M. 87115

Key Words & Phrases: approximation to exact distribution; Siegel-Tukey; nonparametric.

ABSTRACT

An approximation to the exact distribution of the Wilcoxon rank sum test (Mann-Whitney U-test) and the Siegel-Tukey test based on a linear combination of the two-sample t-test applied to ranks and the normal approximation is compared with the usual normal approximation. The normal approximation results in a conservative test in the tails while the linear combination of the test statistics provides a test that has a very high percentage of agreement with tables of the exact distribution. Sample sizes $3 \leq m,n \leq 50$ were considered.

1. INTRODUCTION

Two-sample tests such as the Wilcoxon rank sum test (Mann-Whitney U-test) and the Siegel-Tukey test are among the most frequently used nonparametric tests. For tests of hypotheses, the users of these nonparametric techniques usually have to be satisfied with a limited number of exact percentage points for small sample sizes or with a normal approximation. The latter may have unsatisfactory accuracy for small significance levels unless sample sizes are large.

A test statistic that would be easy to compute and yet improve over the accuracy of the normal approximation for commonly used sample sizes would be desirable. It is the purpose of this paper to attempt to fill this need by showing that with only slightly more effort a much better approximation may be obtained. The approach presented herein uses a test statistic which is a monotone function of the normal approximation. This approximation is an extension to the two-sample case of the approximation given by Iman (1974) for the Wilcoxon signed ranks test. Wallace (1959) used an F statistic based on applying the one-way analysis of variance to ranks to approximate the exact distribution of the Kruskal-Wallis test statistic. Wallace's procedure resulted in a computationally difficult adjustment of the degrees of freedom to obtain his best approximation.

Each statistic used as an approximation is compared against the exact distribution of the nonparametric test statistic, and the results are summarized in table form in Section 2. Comments on the choice of weights used in the linear combination and a summary are given in Section 3.

2. THE TEST STATISTICS

Consider a set of n independent and identically distributed continuous random variables $X_1, X_2, \ldots, X_n$ from one population and a set of m independent and identically distributed continuous random variables $Y_1, Y_2, \ldots, Y_m$ from a second population. We shall assume that the two samples are mutually independent and assign ranks 1, 2,...,N, where $N = n + m$, to the combined sample of X's and Y's. Let $R(X_i)$ denote the rank assigned to X_i in the combined sample and let $R_x = \sum_{i=1}^{n} R(X_i)$. Likewise, $R(Y_j)$ denotes the rank assigned to Y_j in the combined sample. By using the formulas for $\sum_{i=1}^{s} i$ and $\sum_{i=1}^{s} i^2$ and some algebraic simplification, the usual two-sample t-test applied to the $R(X_i)$ and $R(Y_j)$ can be shown to yield the following form

$$T = \frac{R_x - n(N + 1)/2}{\left[\frac{nm(N + 1)(N - 1)}{12(N - 2)} - \frac{[R_x - n(N + 1)/2]^2}{N - 2}\right]^{1/2}} . \quad (2.1)$$

The Student's t-distribution with N - 2 degrees of freedom can be used to obtain approximate critical values for T.

The standard normal approximation to the exact distribution of the Wilcoxon rank sum test statistic uses the statistic

$$Z = \frac{R_x - n(N + 1)/2}{[nm(N + 1)/12]^{1/2}} . \quad (2.2)$$

Substitution of (2.2) into (2.1) gives the following simpler computational form for T, which is a monotone increasing function of Z,

$$T = \frac{Z}{[(N - 1 - Z^2)/(N - 2)]^{1/2}} . \quad (2.3)$$

We now introduce a third test statistic based on a linear combination of (2.2) and (2.3) as

$$J = (Z + T)/2, \quad (2.4)$$

which is also a monotone increasing function of Z and has the following computational form

$$J = \frac{Z}{2}\left[1 + \left(\frac{N - 2}{N - 1 - Z^2}\right)^{1/2}\right] . \quad (2.5)$$

A comment on the weights for this linear combination of Z and T is given in Section 3.

Critical values for the J statistic are given approximately by

$$J_\alpha \cong (Z_\alpha + t_\alpha(N - 2))/2, \quad (2.6)$$

where Z_α is the upper alpha level critical table value of the standard normal distribution and $t_\alpha(N - 2)$ is the upper alpha level critical table value of the Student's t-distribution with N - 2 degrees of freedom. The critical values from (2.6) are given in

Table I for N = 7(1)100. Table I is presented for reasons of convenience and accuracy. The entries in the table were computed using an inverse t-distribution algorithm given by Hill (1970)

TABLE I

Approximate Critical Values of J_α from (2.6)

N	α: .05	.025	.01	.005	N	α: .05	.025	.01	.005
7	1.8299	2.2653	2.8456	3.3040	54	1.6598	1.9833	2.3633	2.6248
8	1.7940	2.2034	2.7345	3.1416	55	1.6595	1.9829	2.3626	2.6238
9	1.7697	2.1623	2.6621	3.0377	56	1.6592	1.9824	2.3619	2.6229
10	1.7522	2.1330	2.6114	2.9656	57	1.6589	1.9820	2.3612	2.6220
11	1.7390	2.1111	2.5739	2.9128	58	1.6587	1.9816	2.3606	2.6212
12	1.7287	2.0940	2.5451	2.8726	59	1.6584	1.9812	2.3600	2.6204
13	1.7204	2.0805	2.5222	2.8408	60	1.6582	1.9808	2.3594	2.6196
14	1.7136	2.0694	2.5037	2.8152	61	1.6580	1.9805	2.3588	2.6188
15	1.7079	2.0602	2.4883	2.7941	62	1.6577	1.9801	2.3582	2.6181
16	1.7031	2.0524	2.4754	2.7763	63	1.6575	1.9798	2.3577	2.6173
17	1.6989	2.0457	2.4644	2.7613	64	1.6573	1.9795	2.3572	2.6167
18	1.6954	2.0399	2.4549	2.7483	65	1.6571	1.9791	2.3567	2.6160
19	1.6922	2.0349	2.4466	2.7370	66	1.6569	1.9788	2.3562	2.6153
20	1.6895	2.0304	2.4394	2.7271	67	1.6567	1.9786	2.3557	2.6147
21	1.6870	2.0265	2.4329	2.7184	68	1.6566	1.9783	2.3553	2.6141
22	1.6848	2.0230	2.4272	2.7106	69	1.6564	1.9780	2.3548	2.6135
23	1.6828	2.0198	2.4220	2.7036	70	1.6562	1.9777	2.3544	2.6130
24	1.6810	2.0169	2.4173	2.6973	71	1.6560	1.9774	2.3540	2.6124
25	1.6794	2.0143	2.4131	2.6916	72	1.6559	1.9772	2.3536	2.6119
26	1.6779	2.0119	2.4093	2.6864	73	1.6557	1.9770	2.3532	2.6113
27	1.6765	2.0098	2.4057	2.6816	74	1.6556	1.9767	2.3528	2.6108
28	1.6752	2.0077	2.4025	2.6773	75	1.6554	1.9765	2.3524	2.6104
29	1.6741	2.0059	2.3995	2.6733	76	1.6553	1.9763	2.3521	2.6099
30	1.6730	2.0042	2.3967	2.6695	77	1.6551	1.9760	2.3517	2.6094
31	1.6720	2.0026	2.3942	2.6661	78	1.6550	1.9758	2.3514	2.6090
32	1.6711	2.0011	2.3918	2.6629	79	1.6549	1.9756	2.3511	2.6085
33	1.6702	1.9997	2.3896	2.6599	80	1.6547	1.9754	2.3507	2.6081
34	1.6694	1.9984	2.3875	2.6572	81	1.6546	1.9752	2.3504	2.6077
35	1.6686	1.9972	2.3856	2.6545	82	1.6545	1.9750	2.3501	2.6073
36	1.6679	1.9961	2.3837	2.6521	83	1.6544	1.9748	2.3498	2.6069
37	1.6672	1.9950	2.3820	2.6498	84	1.6542	1.9746	2.3495	2.6065
38	1.6666	1.9940	2.3804	2.6477	85	1.6541	1.9745	2.3492	2.6061
39	1.6660	1.9931	2.3789	2.6456	86	1.6540	1.9743	2.3490	2.6057
40	1.6654	1.9922	2.3775	2.6437	87	1.6539	1.9741	2.3487	2.6054
41	1.6649	1.9913	2.3761	2.6419	88	1.6538	1.9739	2.3484	2.6050
42	1.6644	1.9905	2.3748	2.6401	89	1.6537	1.9738	2.3482	2.6047
43	1.6639	1.9898	2.3736	2.6385	90	1.6536	1.9736	2.3479	2.6043
44	1.6634	1.9890	2.3724	2.6369	91	1.6535	1.9735	2.3477	2.6040
45	1.6630	1.9883	2.3713	2.6355	92	1.6534	1.9733	2.3474	2.6037
46	1.6625	1.9877	2.3702	2.6341	93	1.6533	1.9732	2.3472	2.6034
47	1.6621	1.9870	2.3692	2.6327	94	1.6532	1.9730	2.3470	2.6031
48	1.6618	1.9864	2.3683	2.6314	95	1.6531	1.9729	2.3467	2.6028
49	1.6614	1.9859	2.3673	2.6302	96	1.6530	1.9727	2.3465	2.6025
50	1.6610	1.9853	2.3665	2.6290	97	1.6530	1.9726	2.3463	2.6022
51	1.6607	1.9848	2.3656	2.6279	98	1.6529	1.9725	2.3461	2.6019
52	1.6604	1.9843	2.3648	2.6268	99	1.6528	1.9723	2.3459	2.6016
53	1.6601	1.9838	2.3640	2.6258	100	1.6527	1.9722	2.3457	2.6014

which guarantees six decimal place accuracy. Hence, all entries in the table may be considered accurate to the four decimal places shown.

The tables given by Wilcoxon, Katti, and Wilcox (1970) were used to check the accuracy of the approximate critical values for Z, T and J. These tables cover 1176 pairs of values of (m,n) from (3,3) to (50,50). However, the entire exact distribution of the Wilcoxon rank sum test statistic is not tabled, but rather a pair of values, L and L - 1, of R_x is given which have corresponding exact alpha levels that form bounds around the desired alpha level. Four values of alpha, .05, .025, .01 and .005, are tabled for each pair of (m,n). Hence, 4 x 1176 = 4704 cases can be considered. However, 16 of these cases associated with small pairs of sample sizes had to be eliminated from consideration.

α = .005 : (3,3), (3,4), (3,5), (3,6), (3,7), (3,8), (4,4), (4,5)
α = .01 : (3,3), (3,4), (3,5), (3,6), (4,4)
α = .025 : (3,3), (3,4)
α = .05 : (3,3)

This was because these pairs do not have enough points in the corresponding sample space of the test statistic to provide a test for some or all of the values of alpha considered. Hence, we consider the 4704 - 16 = 4688 remaining cases. As an example of how the Wilcoxon, et al (1970) tables are constructed, consider the case where α = .05, n = 17 and m = 22. The Wilcoxon, et al tables give a pair of values of R_x, L - 1 = 398 and L = 399, with corresponding exact alpha levels of .0520 and .0490. Hence, if we were to reject the hypothesis of interest when R_x = 399, the probability of making a Type I error is .0490. Values of R_x of 400, 401, 402,... have corresponding probabilities of Type I error less than .0490 but which are not given by Wilcoxon, et al. Also, values of R_x of ..., 395, 396, 397 have corresponding probabilities of Type I error greater than .0520.

Comparison of the three statistics of this section does not allow an exact alpha level to be given each time due to the limitations of the Wilcoxon, et al tables. Therefore, our approach, say for the Z statistic, was to find the value of R_x which when substituted in (2.2) will yield a calculated Z value just large enough to exceed the critical table value, Z_α. If this value of R_x is found at L + 1, L + 2, L + 3,..., the test will be conservative, while values at ..., L - 4, L - 3, L - 2 indicate a liberal test. Consider the following example to illustrate how each of the three statistics were compared. Let n = 11 and m = 16 with α = .01. From the Wilcoxon, et al (1970) tables we find L = 201 with an exact alpha of .0099 and L - 1 = 200 with an exact alpha of .0114. We can then perform the following calculations from (2.2), (2.3) and (2.5).

R_x	Exact alpha	Z	T	J
L+1 = 202	< .0099	2.3686	2.6228	2.4957
L = 201	.0099	2.3193	2.5537	2.4365
L-1 = 200	.0114	2.2699	2.4857	2.3778
L-2 = 199	> .0114	2.2206	2.4189	2.3197

From the standard normal table we have $Z_{.01}$ = 2.3263. Examination of the Z column shows the critical table value to first be exceeded at L + 1 = 202 with an exact alpha less than .0099, hence, a conservative test for Z. It is worth noting at this point that a normal approximation with the usual continuity correction will yield smaller calculated values than those in the Z column. Since the Z statistic is conservative, a corrected Z statistic would be more conservative; hence, a corrected normal approximation is not considered in this paper. The Student's t-table has $t_{.01}$(11 + 16 - 2) = 2.4851 and the T column shows this table value is first exceeded at L - 1 = 200 with an exact alpha of .0114. From Table I we have $J_{.01}$(11 + 16) = 2.4057. The J column indicates this value is first exceeded at L = 201. Hence, in this case the J statistic has selected the boundary value from the tables of Wilcoxon, et al (1970) that has a corresponding exact alpha level closest to the required alpha of .01.

Each of the three statistics, Z, T and J, were compared in a similar manner for all 4688 cases. Results of the frequency of the selection of the boundaries ..., L + 2, L + 1, L, L - 1, L - 2,... by each of the three statistics is given in Table II.

Table II is set up according to $k = \min(m,n) = 3, 5, 10, 15, 20$ and 25 and for each of the four alpha levels covered by Wilcoxon, et al (1970).

Examination of Table II indicates that all three statistics perform well with respect to selecting one of the pair of tabled values, L and L - 1, given by Wilcoxon, et al (1970) when $\alpha = .025$ and .05. However, at $\alpha = .005$ and .01 we note Z and T going in opposite directions with respect to selection of boundary values. The J statistic takes advantage of this situation and therefore has a high percentage of selection of the tabled boundaries, L and L - 1, given by Wilcoxon, et al (1970). Table III gives the percentage of selection of the table boundaries, L and L - 1, for each of the three statistics over all four α levels. It should be noted that selection of one of the table boundaries, L or L - 1, by any of the three statistics does not necessarily imply that the boundary has been selected that has an exact alpha closest to the required alpha. In the previous example we required $\alpha = .01$, and we had choices of .0099 and .0114 for alpha. Hence, if one's philosophy is to select the alpha closest to the required alpha, regardless of whether it makes the test conservative or liberal, then the statistics can be compared on this basis. Table III also gives the results of such comparisons.

Examination of Tables II and III makes it clear that much of the poor performance reflected in the low percentage of times that Z gives the test closest to the required alpha level is due to its almost total failure to select either L or L - 1 at the smallest values of alpha. Therefore, Table IV was constructed to compare only Z and J at the commonly used alpha levels of .025 and .05.

3. CONCLUSIONS

The three statistics, Z, T and J, given respectively by (2.2), (2.3) and (2.5) are easily seen to be asymptotically equivalent and as such each could be used to approximate the exact distribution of the Wilcoxon rank sum test statistic (Mann-Whitney U statistic). However, Tables II and III indicate that the J statistic

TABLE II

Frequency of Selection of Boundaries for the Statistics Z, T and J from (2.2), (2.3) and (2.5), Respectively, for all Values of (m,n) such that $k \leq m,n \leq 50$.

		$\alpha = .005$			$\alpha = .01$		
k	R_x	Z	T	J	Z	T	J
3	L+5	2	0	0	0	0	0
	L+4	13	1	5	0	0	0
	L+3	40	9	15	1	0	0
	L+2	340	22	44	31	5	11
	L+1	724	58	147	657	36	72
	L	49	161	753	482	225	803
	L-1	0	597	204	0	825	285
	L-2	0	320	0	0	80	0
5	L+3	18	0	1	0	0	0
	L+2	315	3	22	8	0	0
	L+1	703	34	124	613	8	35
	L	45	139	734	460	183	767
	L-1	0	585	200	0	810	279
	L-2	0	320	0	0	80	0
10	L+2	235	0	0	0	0	0
	L+1	604	0	50	480	0	3
	L	22	50	637	381	59	612
	L-1	0	495	174	0	722	246
	L-2	0	316	0	0	80	0
15	L+2	180	0	0	0	0	0
	L+1	482	0	8	373	0	0
	L	4	4	515	293	15	454
	L-1	0	361	143	0	571	212
	L-2	0	301	0	0	80	0
20	L+2	141	0	0	0	0	0
	L+1	354	0	2	285	0	0
	L	1	0	374	211	2	330
	L-1	0	220	120	0	415	166
	L-2	0	276	0	0	79	0
25	L+2	119	0	0	0	0	0
	L+1	232	0	0	212	0	0
	L	0	0	262	139	0	231
	L-1	0	133	89	0	281	120
	L-2	0	218	0	0	70	0

Table II (continued)

k	R_x	α = .025 Z	T	J	α = .05 Z	T	J
3	L+5	0	0	0	0	0	0
	L+4	0	0	0	0	0	0
	L+3	0	0	0	0	0	0
	L+2	0	0	0	0	0	0
	L+1	1	0	0	0	0	0
	L	964	460	693	492	605	535
	L-1	209	714	481	683	570	640
	L-2	0	0	0	0	0	0
5	L+3	0	0	0	0	0	0
	L+2	0	0	0	0	0	0
	L+1	0	0	0	0	0	0
	L	884	404	630	467	575	508
	L-1	197	677	451	614	506	573
	L-2	0	0	0	0	0	0
10	L+2	0	0	0	0	0	0
	L+1	0	0	0	0	0	0
	L	697	281	478	369	470	410
	L-1	164	580	383	492	391	451
	L-2	0	0	0	0	0	0
15	L+2	0	0	0	0	0	0
	L+1	0	0	0	0	0	0
	L	535	209	361	290	374	325
	L-1	131	457	305	376	292	341
	L-2	0	0	0	0	0	0
20	L+2	0	0	0	0	0	0
	L+1	0	0	0	0	0	0
	L	408	149	269	218	285	247
	L-1	88	347	227	278	211	249
	L-2	0	0	0	0	0	0
25	L+2	0	0	0	0	0	0
	L+1	0	0	0	0	0	0
	L	294	101	194	155	208	179
	L-1	57	250	157	196	143	172
	L-2	0	0	0	0	0	0

TABLE III

Summary of the Performances of the Statistics Z, T and J

	Percent of cases that L or L-1 is selected by:			Percent of cases that each is closest to required alpha:		
k	Z	T	J	Z	T	J
3	61.41	88.67	93.73	42.86	69.01	73.14
5	61.68	89.71	95.79	43.04	69.70	74.86
10	61.70	88.50	98.46	42.80	69.08	77.24
15	61.15	85.70	99.70	42.83	67.19	79.09
20	60.69	82.11	99.90	42.44	64.87	79.94
25	59.90	79.49	100.00	42.02	61.89	80.48

TABLE IV

Percent of Cases that Z and J are Closest to the Required Alpha Level

	$\alpha = .025$		$\alpha = .05$	
k	Z	J	Z	J
3	71.89	93.70	91.57	95.40
5	72.25	94.82	92.32	96.30
10	71.89	96.40	93.15	97.44
15	72.67	97.75	93.84	98.50
20	71.57	98.59	94.15	99.19
25	71.79	98.86	93.45	99.15

has more desirable properties for the small values of (m,n) and alpha considered. For example, when alpha is small, the Z-test is conservative and the J-test would be more powerful with the probability of Type I error closer to what the experimenter expects when he selects his alpha level. This improved accuracy coupled with ease of computation and the fact that critical values can readily be found for J without bulky or nonexistent (for many users) tables make the J approximation attractive.

The nonparametric Siegel-Tukey test statistic for equality of variances has the same exact distribution as the Wilcoxon rank sum test, see Conover (1971). Hence, the form of the J statistic as given by (2.5) can be used with equal effectiveness in approximating the exact distribution of the Siegel-Tukey test statistic.

On selection of the weights a and b for the linear combination $J = aZ + bT$, certainly for the convenience of the user, $a = b = 0.5$ is appealing. Examination of the frequencies of boundary selection given in Table II indicates for alphas of .025 and .05 the boundary values of L and L - 1 are always selected by the J statistic. Hence, no adjustment of the weights need be considered in these cases. Further examination of Table II indicates there may be some justification for increasing the weights in favor of T when $k = \min(m,n)$ is less than or equal to 10 and alpha is less than or equal to .01. Some calculations with changed weights showed a slight improvement in this area. For example, a J statistic given by $.45\ Z + .55\ T$ increased the boundary selection of L and L - 1 from the 93.73% of Table III to 94.26% when $k = 3$. Further changing of the weights can increase these percentages slightly. However, too much of an increase of the weights in favor of T can result in a lower percentage of the selection of the alpha level closest to the required alpha level as given by Table III. It is the author's feeling that the slight percentage gain in the selection of the boundaries L and L - 1 is not enough to offset the convenience of using the average of Z and T for computing J.

ACKNOWLEDGMENT

Research of this article was supported by a Faculty Research Fellowship and Grant while the author was at Western Michigan University.

BIBLIOGRAPHY

[1] Conover, W. J. (1971). Practical Nonparametric Statistics. New York: John Wiley and Sons, Inc.

[2] Hill, G. W. (1970). Algorithm 396, Student's t-quantiles. Commun. Assoc. Comput. Mach. 13, 619-20.

[3] Iman, Ronald L. (1974). Use of a t-statistic as an approximation to the exact distribution of the Wilcoxon signed ranks test statistic. Commun. Statist. 3 (8), 795-806.

[4] Wallace, David L. (1959). Simplified beta-approximations to the Kruskal-Wallis H-test. J. Amer. Statist. Assoc. 54, 225-30.

[5] Wilcoxon, F., Katti, S. K., and Wilcox, R. A. (1970). Critical values and probability levels for the Wilcoxon Rank Sum Test. Selected Tables in Mathematical Statistics, Vol. 1 (Ed., H. L. Harter and D. B. Owen). Providence: American Mathematical Society, 171-235.

Received January 1976; corrected February 1976.

Recommended by Truman Lewis, Texas Tech University.

Refereed by James M. Davenport, Texas Tech University.