NONPARAMETRIC TESTS FOR SCALE[1]

By Jerome Klotz

University of California, Berkeley

1. Summary. This paper is concerned with two sample rank tests for scale alternatives. The two samples are assumed to have continuous distribution functions with the difference in respective location parameters (medians) known.

Various rank tests are considered and compared from the point of view of limiting Pitman efficiency for normal and nonnormal alternatives. Among the tests considered is a test with efficiency one relative to the F-test for normal alternatives. Tables are given to facilitate its use.

Small sample power and efficiency for normal alternatives are computed for equal sample sizes of 5. The small sample efficiency values differ appreciably from the limiting value; this deficiency of power appears to derive from the use of ranks per se rather than from the use of a rank test that is not optimal among rank tests.

Lastly, a rank test is proposed for particular alternatives which is most powerful for rectangular densities. It is a simple test which is seen to have surprisingly good power for normal alternatives.

2. Introduction. Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be two samples with continuous cumulative distribution functions F and G. Set $F(x) = \Psi((x - \nu)/\sigma)$ and $G(x) = \Psi((x - \nu)/\tau)$ with Ψ a distribution function and σ and τ scale parameters. Assuming that the medians difference is known, we take ν to be the common unknown median for F and G without loss of generality. Under this assumption, the null hypothesis

$$H: \sigma = \tau$$

gives $F = G$ and statistics of the form

$$\sum_{i=1}^{N} W_{Ni} Z_{Ni} \tag{2.1}$$

have a distribution independent of Ψ under H. Here $N = m + n$, W_{Ni} are given numbers and $Z_{Ni} = 1$ if the ith smallest order statistic in the pooled X and Y sample is an X and is zero otherwise. Tests of the form (2.1) for scale alternatives have been given by A. M. Mood [9], J. E. Freund and A. R. Ansari [5], D. E. Barton and F. N. David [2], and more recently, by S. Siegel and J. W. Tukey [12]. The test of Mood uses weights $W_{Ni} = (i - (N + 1)/2)^2$ while the others use weights as follows for N even:

Received November 7, 1960; revised October 10, 1961.

[1] Doctoral dissertation prepared with the partial support of the Office of Naval Research (Nonr-222-43). This paper in whole or in part may be reproduced for any purpose of the United States Government.

TABLE OF W_{Ni}'s FOR FOUR STATISTICS

(2.2)

Statistic							
i:	1	2	$3 \cdots \frac{1}{2}N$	$\frac{1}{2}N + 1 \cdots N - 2$		$N - 1$	N
S	$\frac{1}{2}N$	$\frac{1}{2}N - 1$	$\frac{1}{2}N - 2 \cdots$	$\cdots \frac{1}{2}N - 2$		$\frac{1}{2}N - 1$	$\frac{1}{2}N$
W	1	2	$3 \cdots \frac{1}{2}N$	$\frac{1}{2}N \cdots 3$		2	1
T	1	4	$5 \cdots$	$\cdots 6$		3	2
T'	2	3	$6 \cdots$	$\cdots 5$		4	1

where S is the statistic of Barton and David, W the statistic of Freund and Ansari, and T and T' are the statistics of Siegel and Tukey.

Mood gives the value $15/2\pi^2 \cong .76$ for the limiting Pitman efficiency of his test relative to the F-test under normality. The value for the test of Siegel and Tukey was found by the author in [8] to be $6/\pi^2 \cong .61$, using the methods of Chernoff and Savage [4] to establish limiting normality for the statistic. The derivation of the efficiency of the test of Freund and Ansari given by Ansari and Bradley in [1] applies equally well to the tests of Siegel and Tukey and of Barton and David because of the following equivalence of the tests as seen from (2.2). For N even

$$(T + T' + 1)/4 = W = (N/2 + 1)N/2 - S.$$

The test of Mood, which has greater efficiency, gives more weight to the extreme ranks than does the Siegel-Tukey test. Thus the question arises whether or not still greater efficiency can be obtained if even more weight is given the extreme ranks. If one assigns weights $[\Phi^{-1}(i/(N + 1))]^2$ where Φ is the standard normal cumulative, limiting efficiency one can be obtained relative to the F-test for normal alternatives. Thus it is proposed to use the test statistic

$$\sum_{i=1}^{N} [\Phi^{-1}(i/(N + 1))]^2 Z_{Ni} \tag{2.3}$$

which shall be called the normal scores statistic. Large values of the statistic are significant for $\sigma > \tau$. The weights are simply the squares of the weights used in Van der Waerden's X-test [15].

The limiting efficiency can be calculated again using the methods of Chernoff and Savage. It should be noted that the efficiency depends only on the limiting form of the weights or on the function J defined in [4]. The weights in (2.3) were simply chosen for convenience and any other weights with the same J will have the same efficiency. In a recent paper by J. Capon [3] the weights $W_{Ni} = EU_{Ni}^2$ were proposed where U_{Ni} is the ith smallest order statistic from a sample of N standard normal variables. This test was shown to have limiting efficiency one relative to the F-test for normal alternatives. Thus the derivation for the normal scores test need not be given here because of the asymptotic equivalence of the two tests.

An interesting duality exists between the scale and location parameter problems. The correspondence of the Wilcoxon test to the Siegel-Tukey test, the Van der Waerden test to the normal scores test, and the c_1-test of Terry [14] to the test of Capon should be noted.

3. Efficiency comparisons for various distributions. B. V. Sukhatme [13] has shown that the test of Mood when compared with the F-test can have limiting Pitman efficiencies anywhere between 0 and ∞ for different distributions. He also shows that the same is true of his proposed test which assumes that the median is known. Since the efficacy expression of his test is the same as the test of Siegel and Tukey, the efficiency of the latter can also take on values anywhere between 0 and ∞. The author conjectures that the same is true for the normal scores test relative to the F-test but as yet has only been able to show that values between .47 and ∞ can be attained. (The value .47 was obtained using $\Psi(x) = \Phi(x^\alpha)$ and letting $\alpha \to \infty$.)

Because of the lack of robustness of the F-test with departures from normality it seems of little interest to compare the various tests with the F-test for distributions other than normal. Of more interest is the comparison of the nonparametric tests for various distributions as has been done by J. L. Hodges, Jr. and E. L. Lehmann [7] for the corresponding location parameter problem. To do this we compute the ratios of the efficacies of the tests using the theorem of Pitman (see, for example, Noether [10]). Assuming that F has a density f and verifying the regularity conditions of [10], the efficacies are as follows:

(Normal scores)

$$\frac{2mn}{N}\left[\int_{-\infty}^{\infty} (\Phi^{-1}(F)/\varphi(\Phi^{-1}(F)))xf^2(x)\,dx\right]^2, \tag{3.1}$$

(Mood)

$$\frac{720mn}{N}\left[\int_{-\infty}^{\infty} (F(x) - \tfrac{1}{2})xf^2(x)\,dx\right]^2, \tag{3.2}$$

(Siegel and Tukey)

$$\frac{48mn}{N}\left[\int_{\nu}^{\infty} xf^2(x)\,dx - \int_{-\infty}^{\nu} xf^2(x)\,dx\right]^2, \tag{3.3}$$

where φ is the normal density function. The last two expressions can be found in [13]. If we compare the test of Siegel and Tukey to the normal scores test, the relative efficiency can take on any value between 0 and ∞. Consider the distribution (4.10) in [7] which is symmetric and defined for positive x by

$$\Psi(x) = \begin{cases}\Phi(x), & 0 \leqq x \leqq \epsilon \\ \Phi(y), & \epsilon < x < \infty,\ y = \epsilon + a(x - \epsilon),\ a > 0.\end{cases}$$

If we let $a \to \infty$ and then $\epsilon \to \infty$, the ratio of (3.3) to (3.1) converges to zero. Similarly, consider the symmetric distribution defined for positive x by $F(x) = \Phi(h^{-1}(x))$ where

$$h(u) = \begin{cases}au/\delta & \text{for } 0 \leqq u < \delta \\ u^2 + a - \delta^2 & \text{for } u \geqq \delta.\end{cases}$$

TABLE 1

Efficiency Comparisons for Different Densities of the Siegel-Tukey Test (S-T), Mood's Test (M) and the Normal Scores Test (n.s.)

Density	$S\text{-}T/n.s.$	$M/n.s.$	$S\text{-}T/M$
Exponential	0.	0.	0.600
Rectangular	0.	0.	0.806
Normal	0.608	0.760	0.800
Logistic	0.750	0.896	0.837
Double exp.	0.774	0.900	0.860
Cauchy	1.783	1.670	1.068

If we let $\delta \to 0$, the efficiency of the Siegel-Tukey test relative to the normal scores test converges to infinity.

The integrals have been evaluated (some numerically) for the normal, double exponential, Cauchy, logistic, exponential, and rectangular densities. For the last two densities the efficacy expression for the normal scores statistic gives infinite values. However, a limiting argument will justify using a value of zero for the efficiencies of the tests of Siegel and Tukey and of Mood relative to the normal scores test. The appropriate efficacy ratios give efficiencies as summarized in Table 1. The accuracy is believed to be at least 2 decimals.

Very roughly it appears to the author that the normal scores statistic should be used in preference to the Siegel-Tukey statistic when the extreme rankings give more dispersion information than do the central rankings. On the other hand, the Siegel-Tukey statistic should be used in preference for those distributions with heavy tails and more dispersion information in the central rankings. The similarity of this conclusion to that given for the corresponding location parameter problem in [7] can be seen.

4. Small sample power and efficiency. To calculate the power of a rank test, it is necessary to calculate the probabilities for alternatives of those orderings which lie in the rejection region of the test. The probability of an ordering for an alternative is calculated by integrating the joint density of the X and Y samples over that part of the N dimensional space defined by the ordering. For example, if we denote the densities by f and g respectively, then for a particular ordering

$$P[XYX \cdots YX] = m!n! \int_{x_1<y_1<x_2<\cdots<y_n<x_m} \cdots \int \prod_{i=1}^{m} f(x_i)\, dx_i \prod_{j=1}^{n} g(y_j)\, dy_j\,.$$

Fortunately, the problem of integration over N dimensions can be reduced to a problem of calculating N successive one dimensional indefinite integrals using a recursive scheme of Professor J. L. Hodges, Jr. If we denote an ordering of X's and Y's by π, the recursive scheme generates the probabilities of the orderings obtained by adjoining an X or a Y to the right, denoted by πX or πY. Let

$$A_\pi(u) = P \text{ [all } X\text{'s and } Y\text{'s} \leqq u \text{ and in the order } \pi].$$

Then

$$A_X(u) = F(u), \qquad A_Y(u) = G(u),$$

$$A_{\pi X}(u) = (m + 1) \int_{-\infty}^{u} A_\pi(v) f(v)\, dv,$$

and

$$A_{\pi Y}(u) = (n + 1) \int_{-\infty}^{u} A_\pi(v) g(v)\, dv,$$

where m and n are the numbers of X's and Y's, respectively, in the ordering π. The desired probability of π is $A_\pi(\infty)$. This scheme was programmed for a digital computer (I.B.M. 704) using normal densities for f and g. The probabilities of all orderings with $N \leqq 10$ were computed for selected σ/τ ratios (2, 3, and 4). The values are given in [8]. It should be noted that the program is applicable to other problems and has been used by Hodges and Lehmann for the location parameter problem. Table 2 gives some selected power values for the Siegel-Tukey test, the normal scores test and the most powerful rank test at the three chosen alternatives ($m = n = 4, 5$). For $\alpha = .06349$ and $m = n = 5$ efficiencies were calculated at various alternatives for the normal scores test relative to the F-test. The efficiency definition of Hodges and Lehmann [6] (pp. 329) is used with randomization between sample sizes $m = n$ for the F-test. The computed values indicate that this efficiency is a decreasing function of σ/τ.

Normal scores test ($m = n = 5$, $\alpha = .06349$)

σ/τ	1.5	2	3	4
β	.16950	.27951	.45007	.55977
e	.803	.776	.683	.640

The small sample values are quite disappointing in view of the large sample value of 1. The question naturally arises whether the loss is due to a failure in the normal scores test at this small sample size and if so whether a substantial improvement can be made with the use of a different rank test. Light can be thrown on this question if we investigate the most powerful rank test for the various alternatives. Using the likelihood ratio principle, power values are calculated along with corresponding efficiencies.

L. R. test ($m = n = 5$, $\alpha = .06349$)

σ/τ	1.5	2	3	4
β	.17082	.28848	.48116	.60770
e	.810	.786	.719	.688

TABLE 2

Power of the Siegel-Tukey Test, the Normal Scores Test, and the Most Powerful Rank Test (Likelihood Ratio Test) for Normal Distributions at Sample Sizes $m = n = 4$, $m = n = 5$, and σ/τ Ratios of 2, 3, and 4

Siegel-Tukey Test

Sample size			Power = P[T ≤ t]		
m = n = 4	t	α	σ/τ = 2	σ/τ = 3	σ/τ = 4
	10	0.01429	0.06619	0.11788	0.15823
	11	0.02857	0.10627	0.16671	0.20740
	12	0.05714	0.20661	0.33305	0.42497
	13	0.10000	0.30159	0.44745	0.53929
m = n = 5	15	0.00397	0.02886	0.06262	0.09365
	16	0.00794	0.05771	0.12524	0.18729
	17	0.01587	0.09492	0.18218	0.25198
	18	0.02778	0.14395	0.25151	0.32742
	19	0.04762	0.21193	0.34967	0.44221
	20	0.07540	0.29554	0.46232	0.56864
	21	0.11111	0.37637	0.55258	0.65469

Normal Scores Test

Sample size			Power = P[S ≥ s]		
m = n = 4	s	α	σ/τ = 2	σ/τ = 3	σ/τ = 4
	4.151	0.01429	0.06619	0.11789	0.15823
	3.752	0.07143	0.24669	0.38188	0.47413
	3.586	0.12857	0.39213	0.55640	0.64952
m = n = 5	5.582	0.00794	0.05771	0.12524	0.18729
	5.338	0.01587	0.09861	0.19000	0.26219
	5.229	0.02381	0.13506	0.24418	0.32270
	5.122	0.03175	0.16857	0.29314	0.37717
	4.878	0.06349	0.27951	0.45007	0.55977
	4.769	0.09524	0.37710	0.57145	0.68336
	4.634	0.12698	0.45052	0.64313	0.74268

Most Powerful Rank Test (L.R. Test)

Sample size				
m = n = 4	α	σ/τ = 2	σ/τ = 3	σ/τ = 4
	0.01429	0.06619	0.11789	0.15823
	0.04286	0.16653	0.28422	0.37580
	0.07143	0.24673	0.38189	0.47413
	0.10000	0.32128	0.47345	0.56742
m = n = 5	0.00794	0.05771	0.12524	0.18729
	0.01587	0.09861	0.19441	0.28616
	0.02381	0.13506	0.25917	0.36105
	0.03175	0.17112	0.31334	0.42156
	0.03968	0.20463	0.36230	0.47604
	0.04762	0.23340	0.40495	0.52467
	0.05556	0.26151	0.44374	0.56678
	0.06349	0.28848	0.48116	0.60770
	0.07143	0.31213	0.50625	0.62922
	0.07937	0.33524	0.53014	0.64932
	0.08730	0.35623	0.55106	0.66738
	0.09524	0.37707	0.57185	0.68455
	0.10317	0.39698	0.59237	0.70141

The power and efficiency are only slightly greater for the L. R. test as compared to the normal scores test. It appears that the loss in efficiency is inherent in the use of ranks for small samples with a rather high price being paid for the insurance obtained with rank statistics. This is in marked contrast to the corresponding location parameter problem where the small sample efficiency is quite high.

5. Exact null distribution for the normal scores statistic. Under the null hypothesis $F = G$ each of the ${}_NC_m$ different orderings of the m X's and n Y's are equally likely. The distribution can be obtained by enumerating the orderings and ranking them according to the associated value of the statistic. If we let S denote the statistic, then $P[S \leqq x] = r/{}_NC_m$, where r is the number of orderings with a value for the statistic of x or smaller. The enumeration, although straightforward can become quite lengthy as N increases. It is thus desirable to reduce the enumeration as much as possible. It is sufficient to consider only distinct values of the statistic and count the number of orderings corresponding to each value. To do this the symmetry of the weights is used. We have from the form of the weights $W_{N1} = W_{NN}$, $W_{N2} = W_{NN-1}$, $\cdots$. Assume N is even $(= 2k)$. Then the number of different values of the statistic is equal to the number of distinguishable ways in which m identical balls can be placed into k different cells with each cell having room for at most two. Each cell corresponds to a pair of identical weights (W_{N1}, W_{NN}), (W_{N2}, W_{NN-1}), $\cdots$, (W_{Nk}, W_{Nk+1}). Cell number i is said to have 0, 1, or 2 balls in it if the ordering of the X's and Y's is such that $Z_{Ni} + Z_{NN-i+1} = 0$, 1, or 2, respectively. For each arrangement of the m balls in the k cells there corresponds 2^i orderings, where i is the number of cells with exactly one ball. If we count the total number of distinct values of the statistic by fixing the numbers of 0's, 1's, and 2's, counting the number of distinguishable permutations, and then sum over all partitions of 0's, 1's and 2's which add up to m, we have a total of

$$(5.1) \qquad \binom{k}{m} + \binom{k}{m-2, 1} + \binom{k}{m-4, 2} + \cdots + \binom{k}{m-2l, l}.$$

Here $l = [m/2]$ is the largest integer less than or equal to $m/2$ and

$$\binom{k}{i, j} = \frac{k!}{i!j!(k-i-j)!}$$

is the trinomial coefficient which gives the number of distinguishable permutations of i 1's, j 2's, and $k - i - j$ 0's. From the consideration that the total number of orderings must be $\binom{N}{m}$ we have the following equality for N even

$$(5.2) \qquad \sum_{j=0}^{l} \binom{k}{m-2j, j} 2^{m-2j} = \binom{N}{m}.$$

For N odd a slight difficulty arises due to the fact that there is but one middle weight w_{Nk+1} where $k = [N/2]$. Thus one of the cells is a half cell (can take only

0 or 1 ball). However, this difficulty can be overcome if we drop this cell and proceed as in the even case using m balls and then $m - 1$ balls since the middle weight is zero. If it were not zero the value would have to be added to all the values of the statistic which were computed using $m - 1$ balls. We note that this method of enumeration can constitute a considerable saving. In particular, when $N = 20$ and $m = 10$, the number of distinct values of the statistic given by (5.1) is 8,953 which is less than five per cent of the total number of orderings $\binom{20}{10} = 184{,}756$. It should be noted that this scheme is applicable to other statistics using symmetric weights. In particular, it can be applied to the statistics of Mood and of Capon and with a slight modification to the statistic of Barton and David (here the half cell for N odd is the end cell and is not zero). The scheme was programmed for a binary digital computer (I.B.M. 704) with $N = 4(1)20$ and $m \leqq n$. Table 3 gives the numerical approximations to the weights used in the computation. Table 4 gives selected critical values and corresponding exact probabilities for $N = 8(1)20$ (see Table 4 at end of article).

TABLE 3

Normal Scores Weights w_{Ni} for $i \leqq N/2$ used in the Tabulation of the Distribution
$$w_{Ni} = [\Phi^{-1}(i/N + 1)]^2$$

N \ i	1	2	3	4	5	6	7	8	9	10
4	0.7083	0.0642								
5	0.9361	0.1855	0.							
6	1.1406	0.3204	0.0324							
7	1.3225	0.4550	0.1015	0.						
8	1.4908	0.5848	0.1855	0.0195						
9	1.6435	0.7083	0.2750	0.0642	0.					
10	1.7827	0.8253	0.3655	0.1216	0.0130					
11	1.9127	0.9361	0.4550	0.1855	0.0443	0.				
12	2.0335	1.0404	0.5421	0.2524	0.0861	0.0093				
13	2.1462	1.1406	0.6266	0.3204	0.1340	0.0324	0.			
14	2.2530	1.2343	0.7084	0.3776	0.1855	0.0642	0.0070			
15	2.3532	1.3225	0.7871	0.3880	0.2389	0.1015	0.0236	0.		
16	2.4508	1.4090	0.8629	0.5206	0.2931	0.1424	0.0497	0.0054		
17	2.5376	1.4908	0.9360	0.5848	0.3474	0.1855	0.0796	0.0195	0.	
18	2.6244	1.5675	1.0060	0.6474	0.4015	0.2299	0.1340	0.0397	0.0044	
19	2.7060	1.6435	1.0741	0.7083	0.4550	0.2750	0.1485	0.0642	0.0158	0.
20	2.8241	1.7135	1.1332	0.7676	0.5076	0.3203	0.1855	0.0918	0.0324	0.0036

6. The rectangular case. Using the calculated probabilities of the orderings for the normal case with $\sigma/\tau = 3$ and $m = n = 5$, a listing was formed. The orderings were sorted according to the associated probability value and from this arrangement a pattern was suggested. The most significant orderings seemed to be grouped according to fixed values of the statistics U and V where

$$U = \text{number of } X\text{'s less than the minimum } Y,$$

$$V = \text{number of } X\text{'s greater than the maximum } Y.$$

In addition, the probabilities in these groupings, although not constant, were of the same order of magnitude. This suggests that for some distributions a good rank test may be constructed using only the statistics U and V. In this connection it should be noted that the test proposed by S. Rosenbaum in [11] uses the statistic $U + V$.

If we consider the problem for rectangular densities it will be shown that (U, V) forms a sufficient rank statistic for $\sigma \geqq \tau$. Thus, if we use the rectangular density to construct a test it is hoped that the test will be useful for other distributions. In particular, it is hoped that the test will have good power when the samples are small for the normal case.

To fix ideas let $X_1, \cdots, X_m$: $R(-\sigma, \sigma)$ and $Y_1, \cdots, Y_n$: $R(-\tau, \tau)$. To see that (U, V) is sufficient for the rank orderings we show that the conditional probabilities of the orderings are constant for given values $U = u$, $V = v$. Evaluating the probabilities by first conditioning with respect to the values of $\min Y$ and $\max Y$ we see that the remaining $N - u - v - 2$ X and Y variables between $\min Y$ and $\max Y$ are independently and uniformly distributed between these values. Since the remaining X and Y variables in this range are identically distributed as well as being independent, their particular ordering does not affect the value of the probabilities which are calculated by integrating out the conditioning variables $\min Y$ and $\max Y$. Thus the probabilities are constant for fixed values of (U, V).

We proceed to calculate the most powerful rank test using the likelihood ratio principle. Under the hypothesis we have

$$P[U = u, V = v] = \binom{N - u - v - 2}{n - 2} \Big/ \binom{N}{n}. \tag{6.1}$$

To calculate the joint probability under the alternative we write $U = S + Z$ and $V = T + W$, where

$$\begin{aligned} S &= \text{number of } X\text{'s in the interval } [-\sigma, -\tau), \\ Z &= \text{number of } X\text{'s in the interval } [-\tau, \min Y), \\ W &= \text{number of } X\text{'s in the interval } (\max Y, \tau], \\ T &= \text{number of } X\text{'s in the interval } (\tau, \sigma). \end{aligned}$$

Then

$$P[U = u, V = v] = \sum_{z=0}^{u} \sum_{w=0}^{v} P[Z = z, W = w \mid S = u - z, T = v - w]$$

$$\cdot P[S = u - z, T = v - w]$$

$$(6.2) \qquad = \sum_{z=0}^{u} \sum_{n=0}^{v} \frac{\binom{N - u - v - 2}{n - 2}}{\binom{N - u - v + z + w}{n}} \cdot \binom{m}{u - z, v - w} p^{u-z} q^{v-w} r^{m-u-v+z+w},$$

where $p = q = (\sigma - \tau)/(2\sigma) = (1 - \tau/\sigma)/2$ and $r = 1 - p - q$. The likelihood ratio (the ratio of (6.2) to (6.1) is equivalent to

$$(6.3) \qquad \sum_{a=0}^{u} b(a, N, p) \cdot B(v, N - a, q/(1 - p)),$$

using the notation for the binomial probability and its cumulative. The expression (6.3) was used to rank the values of (U, V) for $\sigma/\tau = 3$ and $m = n = 5$. The problem is symmetric in U and V and so it suffices to list only the pairs (u, v) for $u \leqq v$ provided the probabilities for the symmetric pair are included. Using the ordering to form a test and applying it to the normal case, good results were obtained. In fact, for the normal alternative $\sigma/\tau = 3$ at the first four smallest natural levels the power is the same as that of the most powerful rank test! The results are given in Table 5.

For large N it is desirable to have an asymptotic approximation. Let $\sigma_N/\tau_N \to 1$ with the speed of $1/N$. If we write $\sigma_N/\tau_N = 1 + 2\lambda/N$ we have $Np = \lambda/(1 + 2\lambda/N) \to \lambda$. Thus, the binomial converges to the Poisson and the asymptotic approximation to the likelihood ratio test is equivalent to rejecting

TABLE 5

Power for normal distributions ($\sigma/\tau = 3$) of the most powerful rank test in the rectangular case for $m = n = 5$

U	V	α	Power	Power of L.R. test at this level
2	3	0.00794	0.12524	0.12524
1	4	0.01587	0.19441	0.19441
2	2	0.03175	0.31334	0.31334
1	3	0.06349	0.48116	0.48116
0	5	0.07143	0.48285	0.50625
1	2	0.15079	0.68127	0.69535

for those values (u, v) for which

$$\mathcal{P}_\lambda(u)\mathcal{P}_\lambda(v) > \text{Const.},$$

where $\mathcal{P}_\lambda(u)$ is the cumulative Poisson distribution with parameter λ. This test was again applied for $m = n = 5$ and $\sigma_N/\tau_N = 3$ $(\lambda = 10)$. The (u, v) rankings were the same as in the previous case with the same results applicable.

7. Acknowledgment. I wish to acknowledge my debt to Professors Siegel and Tukey for permitting me to view their manuscript in advance of publication, and to the University of California Computer Center for machine time. I wish to thank Professors E. L. Lehmann, W. F. Taylor, L. LeCam, and the referee for very helpful suggestions and criticisms. Finally, I wish to express my very deep gratitude to Professor Hodges for recommending the problem and for his constant help, criticism, and inspiration.

REFERENCES

[1] Ansari, A. R. and Bradley, R. A. (1960). Rank-sum tests for dispersions. *Ann. Math. Statist.* **31** 1174–1189.

[2] Barton, D. E., and David, F. N. (1958). A test for birth-order effects. *Ann. Eugenics.* **22** 250–257.

[3] Capon, J. (1961). Asymptotic efficiency of certain locally most powerful rank tests. *Ann. Math. Statist.* **32** 88–100.

[4] Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.* **29** 972–994.

[5] Freund, J. E. and Ansari, A. R. (1957). Two-way rank sum test for variances. Virginia Polytechnic Institute Report supported by the Office of Ordnance Research and National Science Foundation.

[6] Hodges, Jr., J. L. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the *t*-test. *Ann. Math. Statist.* **27** 324–335.

[7] Hodges, Jr., J. L. and Lehmann, E. L. (1961). Comparison of the normal scores and Wilcoxon test. *Proc. Fourth Berkeley Symp. Math. Stat. Prob.* **1** 307–317. Univ. of California Press.

[8] Klotz, J. H. (1961). Nonparametric tests for scale. Unpublished doctoral dissertation, University of California, Berkeley.

[9] Mood, A. M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *Ann. Math. Statist.* **25** 514–522.

[10] Noether, G. E. (1955). On a theorem of Pitman. *Ann. Math. Statist.* **26** 64–68.

[11] Rosenbaum, S. (1953). Tables for a nonparametric test of dispersion. *Ann. Math. Statist.* **24** 663–668.

[12] Siegel, S. and Tukey, J. W. (1960). A nonparametric sum of ranks procedure for relative spread in unpaired samples. *J. Amer. Stat. Assn.* **55** 429–445.

[13] Sukhatme, B. V. (1957). On certain two-sample nonparametric tests for variances. *Ann. Math. Statist.* **28** 188–194.

[14] Terry, M. E. (1952). Some rank order tests which are most powerful against specific parametric alternatives. *Ann. Math. Statist.* **23** 346–366.

[15] Van der Waerden, B. L. (1956). The computation of the *X*-distribution. *Proc. Third Berkeley Symp. Math. Stat. Prob.* **1** 207–208. Univ. of Calif. Press.

Table 4

Normal Scores Distribution

$P[S \leq s]$ N	m	.005	.01	.025	.05	.10	.90	.95	.975	.99	.995
						Nominal levels					
8	2			.0390 .03571				2.0756 .96429			
8	3			.2245 .03571	.3905 .07143	.6238 .10714	2.6604 .89286	3.1671 .96429			
8	4		.4100 .01429		.8093 .07143	.9753 .12857	3.3526 .87143	3.5859 .92857		3.7519 .98571	
9	1					0. .11111					
9	2				.0642 .05556	.1284 .08333			2.3518 .97222		
9	3		.1284 .01191		.3392 .05952	.5500 .09524	2.6268 .89286	3.3512 .95238	3.5620 .97619		
9	4		.4034 .01587	.6142 .03175	.8367 .05556	1.0475 .11905	3.6262 .90476	4.0595 .96032		4.2703 .99206	
10	2			.0260 .02222		.1346 .11111	2.1482 .88889		2.6080 .97778		
10	3		.1476 .01667	.2562 .03333	.3915 .05000	.5001 .11667	2.9735 .91667	3.5784 .95000	3.6870 .96667	3.9309 .98333	
10	4	.2692 .00476		.5131 .02381	.7570 .04762	1.0815 .10952	3.8086 .89524	4.4037 .95714	4.5123 .97619		4.7562 .99524
10	5	.6347 .00794	.8786 .01587	.9872 .02381	1.3384 .06349	1.5823 .10318	4.5253 .89683	4.8778 .96825	5.1217 .97619	5.2290 .98413	5.3376 .99206

$P[S \leq s]$ N	m	.005	.01	.025	.05	.10	.90	.95	.975	.99	.995
						Nominal levels					
11	1					0. .09091					
11	2			.0443 .03636	.0886 .05455	.1855 .09091	2.3677 .90909		2.8488 .98182		
11	3	.0886 .00606		.2298 .03030	.3710 .04849	.5436 .09697	3.0343 .88485	3.8254 .95152	4.0109 .97576	4.2804 .98788	
11	4	.2741 .00606	.4153 .01212	.5436 .02121	.6864 .04546	1.0955 .09697	4.0109 .89394	4.4659 .95152	4.8058 .97273	4.9470 .98485	5.2165 .99697
11	5	.4596 .00217	.7291 .01082	.9986 .02597	1.2810 .04978	1.6209 .11039	4.8058 .89827	5.2165 .94589	5.4020 .98052	5.7419 .99134	5.8831 .99567
12	2		.0186 .01515		.0954 .07576	.1722 .09091	2.5756 .92424			3.0739 .98485	
12	3		.1047 .00909	.2710 .02727	.3478 .06364	.5909 .10000	3.3263 .90909	3.6160 .94546	4.1531 .97273	4.6091 .99091	
12	4	.1908 .00202	.3571 .01010	.6002 .02828	.7236 .04647	1.0562 .10303	4.2004 .90303	4.6564 .94748	5.1512 .97374	5.3598 .98990	5.6495 .99798
12	5	.6095 .00505	.7329 .01010	.8992 .02020	1.2312 .05051	1.6406 .10354	4.9088 .89647	5.4459 .95455	5.7356 .97727	6.1571 .98990	6.2339 .99495
12	6	.9853 .00541	1.1516 .00974	1.4836 .02381	1.7733 .04762	2.1827 .10390	5.6983 .89610	5.9880 .95238	6.4095 .97619	6.6992 .99026	6.7760 .99459
13	1		0. .07692								
13	2			.0324 .02564	.0648 .03846	.1664 .11539	2.4666 .88462	2.7728 .93590		3.2868 .98718	
13	3	.0648 .00350	.1664 .01748	.1988 .02448	.3528 .04895	.5884 .10490	3.4208 .89860	3.9134 .95455	4.4264 .97203	4.6128 .98601	4.9190 .99301

s P[S ≦ s]		Nominal levels									
N	m	.005	.01	.025	.05	.10	.90	.95	.975	.99	.995
13	4	.3004 .00559	.3852 .00979	.5192 .02657	.7748 .04476	1.0810 .10490	4.4264 .89930	4.9332 .94965	5.4330 .97483	5.5670 .98741	5.7534 .99301
13	5	.5192 .00389	.7056 .99032	.9594 .02409	1.2150 .04895	1.6274 .10490	5.1870 .90054	5.6806 .94950	6.0738 .97281	6.3800 .99145	6.6862 .99534
13	6	.9736 .00525	1.1458 .00991	1.4520 .02389	1.7724 .05012	2.2540 .10839	5.8984 .89918	6.3800 .94988	6.7076 .97494	7.0280 .99068	7.2326 .99476
14	1					.0070 .14286					
14	2		.0140 .01099		.0712 .05495	.1925 .10989	2.6306 .90110	2.9614 .94506		3.4873 .98901	
14	3	.0782 .00550	.1354 .01099	.1995 .01648	.3780 .04945	.5701 .10989	3.6698 .89560	3.8649 .93956	4.5702 .97253	4.8836 .98901	5.2144 .99451
14	4	.2637 .00500	.3850 .00999	.5771 .02697	.7556 .05095	1.0223 .10090	4.5200 .89511	4.9478 .94905	5.4300 .97403	5.9258 .99101	6.1179 .99501
14	5	.5200 .00400	.6985 .01199	.9721 .02597	1.2214 .05095	1.5920 .09990	5.2847 .89910	5.8687 .95005	6.3004 .97303	6.6342 .99001	6.9816 .99501
14	6	.8976 .00466	1.1576 .01032	1.4069 .02531	1.7407 .04962	2.2296 .09990	6.0542 .90043	6.5771 .94872	6.9886 .97336	7.3592 .99001	7.6900 .99534
14	7	1.4139 .00525	1.5924 .00991	1.9630 .02389	2.3919 .04895	2.8333 .09962	6.7626 .90035	7.2313 .95105	7.6019 .97611	8.0540 .99009	8.1248 .99476
15	1				0. .06667		1.3225 .86667				
15	2		.0236 .01905	.0472 .02857	.1015 .04762	.2030 .09524	2.7412 .91429	3.1403 .95238		3.6757 .99048	
15	3	.0472 .00220	.1251 .01099	.2266 .02198	.3404 .04396	.5014 .09451	3.6993 .89231	4.0637 .94945	4.7300 .97363	5.0944 .99121	5.4935 .99560

s P[S ≦ s]		Nominal levels									
N	m	.005	.01	.025	.05	.10	.90	.95	.975	.99	.995
15	4	.2861 .00513	.3640 .01099	.5131 .02711	.6808 .04908	1.0149 .10037	4.5643 .90037	5.1180 .95018	5.5950 .97363	6.1304 .98974	6.4169 .99634
15	5	.5367 .00533	.6741 .01066	.8894 .02531	1.1511 .04962	1.5391 .10390	5.3862 .89977	6.0242 .95005	6.4405 .97503	6.9175 .99034	7.2040 .99567
15	6	.8535 .00500	1.0373 .00999	1.3900 .02458	1.7001 .04995	2.1457 .09950	6.1969 .90030	6.7584 .94965	7.2040 .97423	7.6267 .98981	7.9783 .99521
15	7	1.3010 .00482	1.5627 .01088	1.8795 .02378	2.3134 .04988	2.8516 .09946	6.9311 .89899	7.4429 .94779	7.8645 .97483	8.3415 .98990	8.5501 .99549
16	1					.0054 .12500	1.4090 .87500				
16	2		.0108 .00833	.0551 .04167	.0994 .05000	.1478 .08333	2.8180 .89167	3.3137 .95833		3.8598 .99167	
16	3	.0605 .00357	.1532 .1071	.1975 .02500	.3482 .05357	.5757 .10357	3.8652 .89643	4.1766 .94286	4.9070 .97500	5.2688 .98929	5.7645 .99643
16	4	.2472 .00495	.3536 .00989	.5349 .02582	.7181 .04890	1.0520 .09835	4.7281 .89725	5.2742 .94945	5.7894 .97418	6.3603 .99011	6.6274 .99506
16	5	.5403 .00550	.6910 .00962	.9669 .02518	1.2165 .05037	1.6119 .10073	5.5673 .89927	6.2359 .95009	6.6534 .97482	7.2232 .99084	7.4666 .99496
16	6	.9239 .00487	1.1093 .00999	1.4032 .02523	1.7752 .04958	2.2080 .10003	6.4165 .89960	7.0000 .94955	7.4942 .97515	7.9872 .99051	8.2899 .99513
16	7	1.3865 .00507	1.5815 .01014	1.9619 .02465	2.3566 .04965	2.8740 .10000	7.1794 .90087	7.7514 .94983	8.1842 .97465	8.6819 .98986	8.9253 .99493
16	8	1.9230 .00466	2.1567 .00979	2.5680 .02510	3.0142 .04957	3.5603 .09977	7.8916 .90023	8.4377 .95043	8.8864 .97490	9.2952 .99021	9.5386 .99534
17	1				0. .05882		1.4908 .88235				

TABLE 4 (*Continued*)

P[S ≤ s]		Nominal levels									
N	m	.005	.01	.025	.05	.10	.90	.95	.975	.99	.995
17	2		.0195 .01471	.0390 .02206	.0796 .03677	.1855 .08824	2.8850 .89706	3.1224 .93382	3.4736 .96324	4.0284 .99265	
17	3	.0991 .00735	.1186 .01029	.2245 .02353	.3669 .05000	.5524 .09853	4.0479 .90147	4.3758 .94853	4.9644 .97500	5.5192 .99118	5.6600 .99412
17	4	.2846 .00630	.3642 .01050	.5524 .02647	.7379 .04958	1.0351 .09958	4.9606 .90294	5.5022 .94958	6.0074 .97521	6.5960 .99076	6.9134 .99580
17	5	.5497 .00501	.6921 .01018	.9490 .02537	1.2168 .05220	1.6161 .09971	5.7047 .89948	6.3781 .95055	6.9134 .97479	7.4982 .98982	7.7356 .99499
17	6	.8930 .00485	1.1048 .00978	1.4023 .02545	1.7535 .05050	2.1942 .09955	6.5543 .90037	7.1785 .95063	7.6875 .97487	8.2423 .98998	8.5176 .99491
17	7	1.3197 .00504	1.5796 .00998	1.9804 .02499	2.3380 .05008	2.8695 .09991	7.3400 .89994	7.9485 .94987	8.4537 .97547	9.0123 .99003	9.3060 .99501
17	8	1.8293 .00485	2.1204 .00991	2.5498 .02493	2.9765 .05006	3.5610 .09971	8.1096 .89967	8.6888 .95027	9.1745 .97511	9.6534 .98996	9.9445 .99519
18	1					.0044 .11111	1.5675 .88889				
18	2	.0088 .00654		.0441 .03268	.0794 .03922	.1737 .09150	3.0259 .90850	3.2718 .94118	3.6304 .96732		4.1919 .99346
18	3	.0838 .00490	.1428 .00735	.2724 .02451	.3683 .04902	.5752 .10049	4.1410 .89706	4.4218 .94608	5.1979 .97794	5.6503 .99020	5.8962 .99510
18	4	.2784 .00458	.3727 .00882	.5443 .02386	.7698 .05163	1.0710 .09967	4.9252 .89771	5.5184 .94739	6.1261 .97549	6.8207 .99020	7.0462 .99543
18	5	.5773 .00490	.7136 .01027	.9458 .02474	1.2270 .05135	1.6209 .09967	5.8497 .90103	6.4891 .95028	7.0506 .97526	7.6623 .98996	7.9563 .99510
18	6	.9098 .00479	1.0814 .00959	1.4365 .02489	1.7801 .05010	2.2330 .10165	6.6748 .90002	7.3562 .95007	7.9049 .97517	8.4741 .98998	8.8680 .99510

P[S ≤ s]		Nominal levels									
N	m	.005	.01	.025	.05	.10	.90	.95	.975	.99	.995
18	7	1.3622 .00490	1.5877 .01006	1.9527 .02489	2.3670 .05028	2.8772 .09980	7.4960 .90001	8.1552 .95010	8.6915 .97499	9.2451 .98988	9.6097 .99510
18	8	1.8561 .00501	2.1055 .01017	2.5507 .02507	2.9908 .05000	3.5680 .09989	8.2723 .90000	8.9153 .95048	9.4221 .97484	9.9482 .98985	10.2715 .99500
18	9	2.4143 .00502	2.6972 .01033	3.1854 .02501	3.6695 .04982	4.2603 .10029	9.0461 .89971	9.6366 .95019	10.1210 .97499	10.6092 .98968	10.9837 .99498
19	1				0. .05263	.0158 .15789	1.6435 .89474				
19	2		.0158 .01170	.0642 .02924	.0800 .05263	.1643 .09357	2.9810 .89474	3.4143 .94737	3.7801 .97076		4.3495 .99415
19	3	.0800 .00516	.1442 .01032	.2285 .02890	.3550 .05263	.5658 .09907	4.1226 .89783	4.6245 .95356	5.4120 .97317	5.6870 .98968	6.1203 .99587
19	4	.2443 .00490	.3550 .00980	.5035 .02399	.7520 .04979	1.0585 .09907	5.0694 .89809	5.6870 .94840	6.1420 .97523	7.0555 .99020	7.2040 .99536
19	5	.5193 .00507	.6784 .00955	.9416 .02503	1.2118 .05100	1.5967 .09890	5.9428 .90050	6.5869 .94986	7.2156 .97523	7.7855 .99011	8.1938 .99475
19	6	.8900 .00505	1.0700 .00991	1.3918 .02503	1.7261 .04990	2.2112 .09992	6.7880 .90004	7.4914 .95002	8.0608 .97494	8.7235 .99005	9.0396 .99499
19	7	1.2918 .00504	1.5352 .01000	1.9219 .02507	2.3168 .05003	2.8519 .09979	7.6389 .90004	8.3327 .94997	8.9021 .97499	9.4787 .98998	9.9015 .99500
19	8	1.7886 .00500	2.0484 .01014	2.4986 .02506	2.9367 .05011	3.5400 .10000	8.4422 .90000	9.1052 .94991	9.6321 .97515	10.2439 .99002	10.5923 .99500
19	9	2.3234 .00502	2.6351 .00997	3.1212 .02506	3.6138 .05008	4.2576 .09986	9.2045 .90010	9.8432 .94997	10.3815 .97494	10.9207 .99006	11.2331 .99497
20	1					.0036 .10000	1.7135 .90000				

TABLE 4 (*Concluded*)

$P[S \leq s]$ N	m	.005	.01	.025	.05	.10	.90	.95	.975	.99	.995
		Nominal levels									
20	2	.0072 .00526		.0360 .02632	.0954 .05263	.1891 .10000	3.1444 .90526	3.5917 .95263	3.9573 .97368		4.5376 .99474
20	3	.0990 .00526	.1278 .01228	.2215 .02632	.3563 .05088	.5724 .10000	4.1946 .90000	4.7249 .95263	5.3052 .97544	5.8337 .98947	6.2511 .99474
20	4	.2539 .00475	.3421 .00970	.5418 .02663	.7291 .05036	1.0458 .09887	5.1370 .89969	5.8373 .95005	6.4366 .97441	7.1017 .99030	7.4535 .99463
20	5	.5399 .00516	.6390 .00980	.9269 .02516	1.2052 .05005	1.5973 .00985	6.0829 .89990	6.7685 .95021	7.4310 .97485	8.0566 .99007	8.5175 .99497
20	6	.8602 .00501	1.0530 .01011	1.3985 .02482	1.7233 .05044	2.2114 .10008	6.9540 .89990	7.7011 .94992	8.2939 .97495	8.9572 .98996	9.3525 .99489
20	7	1.2700 .00501	1.4984 .00999	1.9061 .02464	2.3036 .05003	2.8538 .09997	7.8174 .90005	8.5606 .94992	9.1436 .97500	9.7737 .99002	10.1955 .99502
20	8	1.7539 .00503	2.0060 .00996	2.4731 .02504	2.9324 .04999	3.5259 .09994	8.6506 .89996	9.3624 .95005	9.9378 .97494	10.5647 .98995	10.9408 .99500
20	9	2.2852 .00501	2.5859 .00998	3.0891 .02501	3.5917 .04998	4.2359 .09996	9.4383 .89992	10.1264 .95030	10.6880 .97497	11.2839 .98995	11.6513 .99500
20	10	2.8655 .00502	3.2110 .00996	3.7519 .02498	4.2944 .04999	4.9631 .09999	10.1930 .90001	10.8643 .95001	11.4072 .97502	11.9482 .98997	12.2872 .99498