Journal of Modern Applied Statistical Methods

Volume 1 | Issue 1 Article 3

5-1-2002

Combining Two Nonparametric Tests Of Location

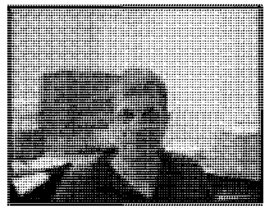
R. Clifford Blair
University of South Florida

Journal Of Modern Applied Statistical Methods
Winter 2002, Vol. 1, No. 1, 13-18

1538 - 9472/02/$30.00

Combining Two Nonparametric Tests Of Location

R. Clifford Blair
Department of Epidemiology and Biostatistics
College of Public Health, &
Jaeb Center For Health Research
University Of South Florida



A distribution-free test is proposed whose power is similar to that of the Wilcoxon Rank-Sum or Terry-Hoeffding Normal Scores tests depending on which of these two tests is more powerful in a given data analysis situation, regardless of the population. This new statistic is distribution-free, and adds no new assumptions to those associated with the constituent tests. A table of critical values for the new statistic is given and some of its Type I error and power properties are examined.

Key words: Nonparametric tests, Shift in location, Wilcoxon rank-sum, Terry-Hoeffding, Normal scores

Introduction

Researchers are sometimes presented with situations in which two (or more) statistical tests appear to be equally appropriate for a given data analysis problem. In choosing between these tests the researcher may consider such factors as ease of computation, acceptability by peers, and availability of tables of critical values. Among the more important factors to influence such a choice would be the relative power of the statistics under consideration. Ceteris Paribus, one would desire to use the most powerful test available.

Unfortunately, it does not usually occur in such cases that one test is more powerful than its competitor among all plausible population models that may be appropriate for the data in the sample. Instead, one test or another may be more powerful than its primary competitor under a given set of circumstances. Thus, for example, one test might be preferred when the population has a light-tailed distribution, but may give way to its rival statistic when the distribution is heavy-tailed.

Factors that influence a test's power may be difficult to assess from available data. Moreover, these factors may interact in such complex patterns as to preclude any clear indication as to which test might be more powerful in a given situation. For certain inferential tests, the dilemma of test choice can be avoided through use of a "maximum" statistic (Cox, 1977). In essence, a maximum statistic is obtained by computing two or more statistics on a given data set, and choosing as the test statistic the one with the smallest associated p-value.

R. Clifford Blair is Professor and Interim Chair, Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, Tampa, Florida. His areas of expertise are in computer-intensive statistical methods, multiple end point analysis, and control of family-wise error.

Consider that two independent samples layout. Two robust and powerful competitors are the Wilcoxon Rank-Sum test (W) and its normal scores counterpart, the Terry-Hoeffding (NS) (Terry, 1952) counterpart. Both procedures are used to test the null hypothesis that samples are from a common population. Asymptotic results suggest that these two tests may manifest substantial power differences, with the magnitudes and advantages of such differences depending on the shape of the population. The Asymptotic Relative Efficiencies (AREs) indicate that, in general, when alternatives are expressed as simple shifts in location, the normal scores test is more efficient than the rank test when sampling is from a light-tailed distribution. However, the normal scores test is at a disadvantage when the populations are heavy-tailed. (For details, see Chernoff and Savage, 1958; Hodges & Lehmann, 1961; Lehmann, 1959; Mikulski, 1963; and Terry, 1952.)

Thus, the purpose of this paper is to present a simple maximum statistic that can be used in lieu of a choice between the Wilcoxon Rank-Sum and Terry-Hoeffding tests.

Methodology

The proposed statistic is obtained by computing both the rank-sum and the normal scores statistics, and choosing as the test statistic the one with the smaller p-value. In order to facilitate development of the sampling distribution of this maximum statistic, it is helpful to express W and NS in a common metric. In this case, both W and NS may be easily expressed in the form of a t statistic. (There are other possibilities, but existing software makes this choice computationally simpler.)

In the case of W, this accomplished by replacing original observations with their respective ranks (with ranking being carried out without regard to group) and computing the usual independent samples t statistic on those ranks. The resulting rank transformation statistic (t_W) is a monotone function of W (Conover & Iman, 1981). Its sampling distribution is well approximated by a t distribution with $n_1 + n_2 - 2$ degrees of freedom (Iman, 1974).

Similarly, an expression for NS may be obtained by replacing observations with their respective normal scores, which are defined as the expected values of the order statistics under normality (Owen, 1962). The t statistic is then computed on these normal scores. The resulting statistic (t_{NS}) is a monotone function of NS, and it too may be referred to the t distribution (Bradley, 1968).

Thus, the new test statistics (t_{max}) is defined as

$$t_{max} = \begin{cases} tW; |tW| > |tNS| \\ tNS; |tNS| > |tW| \end{cases}$$

In the event $|t_W| = |t_{NS}|$, then either statistic may be used.

Sampling Distribution

The exact sampling distribution of tmax may be obtained by forming all possible permutations of the integers 1 to $n_1 + n_2$ (where n_1 and n_2 represent the number of observations in each of two samples), computing t_{max} on each set of integers, and forming the cumulative distribution of the values obtained. In this study, the cumulative distribution of t_{max} was estimated by randomly permuting the integers n to 2n 500,000 times with t_{max} being computed on each permutation.

Table 1 provides values for n1 = n2 = n = 5(1)40, 40(5)60, 60(10)120. It should be noted that the sampling distribution of t_{max} is discrete, and therefore, it was not always possible to find critical values (c) such that $p(t_{max} \leq c) = \alpha$. As a result, values of c were chosen so that c was as large as possible, while maintaining the inequality.

It can also be seen that in some instances, the magnitude of c increases when n is increased, contrary to what is usually expected. This occurs because t_W (t_{NS}) may be the test statistic for one particular value of n, and t_{NS} (t_W) for the situation where n is increased. This does not lead to a violation of the above stated inequality, however, so that the test level is maintained.

Results

Type I Errors

The results of a Monte Carlo study are compiled in Table 2. The entries reflect the Type I error rates for t_W, t_{NS}, and t_{max} when samples are of various sizes. Data were generated by randomly permuting the integers from 1 to $n_1 + n_2$, with the three statistics being computed on each permutation. Entries in the table for t_W, t_{NS}, and $t_{max(a)}$ were obtained by referring the three statistics to the appropriate critical values in a t table, using $n_1 + n_2 - 2$ degrees of freedom. Entries for $t_{max(b)}$ were obtained, in the case of $n_1 = n_2$, by referencing t_{max} to the critical values in Table 1. In the case of $n_1 \neq n_2$, entries for $t_{max(b)}$ were obtained by referring t_{max} to the critical value in Table 1 using $n = .5(n_1 + n_2)$. (Recall that the critical values in Table 1 were obtained under the condition of $n_1 = n_2 = n$.) Twenty thousand repetitions of the experiment were carried out for each condition studied.

Several points should be made regarding the results of these simulations. (1) The t distribution provides a good approximation for the distribution of t_W and t_{WS} is reaffirmed. (2) Critical values from Table 1 produce Type I error rates for t_{max} near nominal levels both in the case of $n_1 = n_2$ and in the case of $n_1 \neq n_2$. (3) Referencing t_{max} to a t distribution with $n_1 + n_2 - 2$ degrees of freedom results in only modest Type I error inflations. This result implies that the researcher who is willing to tolerate minor Type I inflations need not rely on the special table of critical values provided when conducting a test based on t_{max}.

Power

Let $P_{tW}(\alpha)$ and $P_{tNS}(\alpha)$ denote the power of the t_W and t_{NS} tests, respectively, when carried out individually at the α level of significance. Let $\alpha^* > \alpha$ denote the effective level of significance of the maximum test when the critical value is chosen in this way. Because the maximum test rejects the null hypothesis when either t_W or t_{NS} is significant, it follows that the power of the maximum test a level of significance α^* has a lower bound $\max(P_{tW}(\alpha), P_{tNS}(\alpha))$. As indicated by the simulation above, α^* can be expected to be only slightly larger than α. Therefore, the power of the maximum test when conducted at level of significance α should never be much less than the power of the better of the two individual tests when each is conducted at level of significance α.

Figures 1-3 depict the results of a Monte Carlo study designed to compare the power of t_W, t_{NS}, and t_{max}. These figures indicate, respectively, the results for the normal, uniform, and Cauchy distributions. In this study, t_W and t_{NS} were referred to the appropriate t distribution, while t_{max} was referred to the values found in Table 1. Tests were conducted at the $\alpha = .05$ (two-tailed) level of significance. In the cases of the normal and uniform distributions, the alternative condition was constructed by adding a constant equal to $.5\,\sigma$ to the scores in one group. In the case of the Cauchy distribution, which has infinite standard deviation, an arbitrary constant of 1.00 was used. Ten thousand repetitions of the experiment were carried out for each condition studied.

Figure 1 shows that there was generally little difference in the power of the three tests when sampling was from a normal population. Differences that did occur favored t_{NS} and t_{max}. In the case of the uniform distribution, Figure 2 shows that t_{NS} was the most powerful test, with t_{max} showing power similar to, but slightly less than, that of

Table 1. Two-tailed Critical Values For t_{max}.

	Level of Significance					
n	.400	.200	.100	.050	.020	.010
5	.9727	1.4572	1.8312	2.4958	3.7812	5.0000
6	.9571	1.3978	1.8503	2.2804	3.3211	3.8468
7	.9525	1.3919	1.8688	2.2926	2.7464	3.6623
8	.9431	1.4124	1.8066	2.2630	2.7940	3.4503
9	.9238	1.4076	1.8122	2.2005	2.6706	3.0822
10	.9027	1.3942	1.8023	2.1603	2.6616	2.9771
11	.9232	1.3744	1.7861	2.1732	2.6429	2.9582
12	.9207	1.3830	1.7872	2.1517	2.5778	2.9292
13	.9180	1.3836	1.7925	2.1339	2.5852	2.8869
14	.9162	1.3744	1.7655	2.1349	2.5764	2.8852
15	.9210	1.3682	1.7688	2.1310	2.5580	2.8463
16	.9161	1.3762	1.7613	2.1052	2.5327	2.8400
17	.9104	1.3790	1.7582	2.1188	2.5340	2.8243
18	.9154	1.3778	1.7589	2.1025	2.5128	2.8039
19	.9177	1.3737	1.7533	2.0954	2.5135	2.8152
20	.9178	1.3675	1.7486	2.0909	2.5080	2.8043
21	.9164	1.3628	1.7426	2.0826	2.4981	2.7824
22	.9137	1.3755	1.7544	2.0942	2.4956	2.7761
23	.9168	1.3640	1.7497	2.0835	2.4950	2.7876
24	.9205	1.3729	1.7485	2.0911	2.5003	2.7775
25	.9202	1.3654	1.7506	2.0763	2.4780	2.7598
26	.9142	1.3657	1.7402	2.0776	2.4769	2.7582
27	.9219	1.3692	1.7479	2.0776	2.4821	2.7618
28	.9163	1.3709	1.7476	2.0790	2.4823	2.7575
29	.9220	1.3712	1.7492	2.0817	2.4730	2.7401
30	.9200	1.3702	1.7440	2.0709	2.4652	2.7421
31	.9210	1.3683	1.7387	2.0743	2.4694	2.7491
32	.9237	1.3654	1.7441	2.0742	2.4606	2.7357
33	.9189	1.3621	1.7385	2.0703	2.4573	2.7361
34	.9189	1.3633	1.7399	2.0710	2.4545	2.7302
35	.9211	1.3665	1.7419	2.0707	2.4645	2.7340
36	.9225	1.3628	1.7364	2.0697	2.4594	2.7300
37	.9183	1.3648	1.7370	2.0693	2.4532	2.7288
38	.9237	1.3688	1.3730	2.0691	2.4561	2.7286
39	.9203	1.3623	1.7361	2.0654	2.4495	2.7129
40	.9229	1.3642	1.7345	2.0619	2.4419	2.7104
45	.9221	1.3656	1.7341	2.0664	2.4419	2.7073
50	.9221	1.3640	1.7339	2.0588	2.4444	2.7018
55	.9170	1.3595	1.7309	2.0532	2.4338	2.7012
60	.9216	1.3643	1.7320	2.0587	2.4378	2.7061
70	.9205	1.3653	1.7293	2.0487	2.4245	2.6837
80	.9210	1.3619	1.7289	2.0473	2.4205	2.6747
90	.9237	1.3635	1.7280	2.0494	2.4269	2.6840
100	.9212	1.3627	1.7288	2.0492	2.4212	2.6847
110	.9195	1.3605	1.7257	2.0419	2.4136	2.6653
120	.9205	1.3602	1.7247	2.0421	2.4150	2.6691

Table 2. Type I Error Rates Of t_w, t_{NS}, And t_{max} For Various Sample Sizes.

		Level Of Significance		
n1,n2	Statistics	.100	.050	.010
6,6	t_W	.095	.067	.015
	t_{NS}	.107	.058	.015
	$t_{max(a)}$	.111	.067	.015
	$t_{max(b)}$	.097	.049	.008
3,9 t_W	t_W	.099	.064	.010
	t_{NS}	.099	.054	.010
	$t_{max(a)}$	.109	.064	.010
	$t_{max(b)}$	.109	.064	.010
10,10 t_W	t_W	.105	.052	.012
	t_{NS}	.099	.050	.012
	$t_{max(a)}$	.114	.057	.013
	$t_{max(b)}$	.099	.049	.011
5,15 t_W	t_W	.097	.051	.011
	t_{NS}	.100	.050	.011
	$t_{max(a)}$	.112	.058	.013
	$t_{max(b)}$	.106	.050	.010
20,20	t_W	.104	.050	.010
	t_{NS}	.102	.052	.011
	$t_{max(a)}$	.117	.060	.013
	$t_{max(b)}$	.102	.052	.010
10,30	t_W	.101	.047	.010
	t_{NS}	.098	.047	.009
	$t_{max(a)}$	.116	.055	.011
	$t_{max(b)}$	.101	.049	.010
40,40	t_W	.100	.049	.010
	t_{NS}	.101	.049	.010
	$t_{max(a)}$	.117	.058	.012
	$t_{max(b)}$	.100	.050	.010
20,60	t_W	.101	.052	.009
	t_{NS}	.101	.051	.010
	$t_{max(a)}$	.116	.061	.012
	$t_{max(b)}$	.101	.052	.009
60,60	t_W	.101	.049	.010
	t_{NS}	.101	.048	.010
	$t_{max(a)}$	.117	.058	.012
	$t_{max(b)}$	.102	.048	.010
30,90	t_W	.100	.049	.011
	t_{NS}	.100	.050	.011
	$t_{max(a)}$	.116	.058	.011
	$t_{max(b)}$	.100	.049	.011

Figure 1. Power of tW, tMS and tMAX When Sampling Is From a Normal Distriubtuon

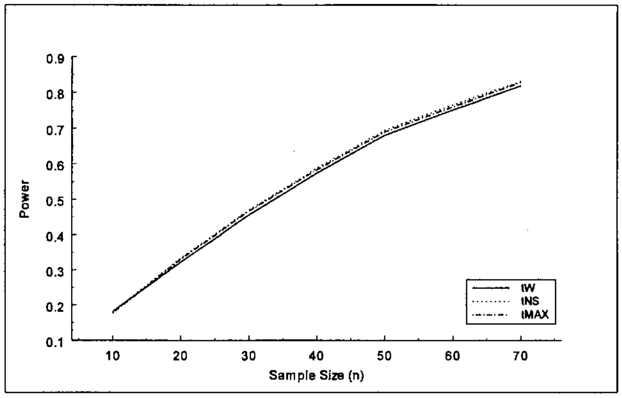


Figure 2. Power of tW, tMS and tMAX When Sampling Is From a Uniform Distriubtuon

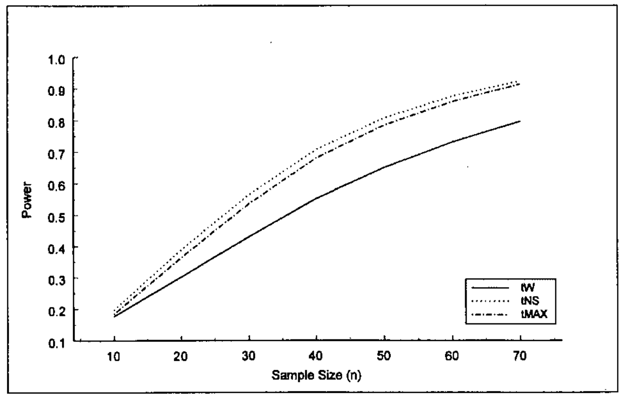
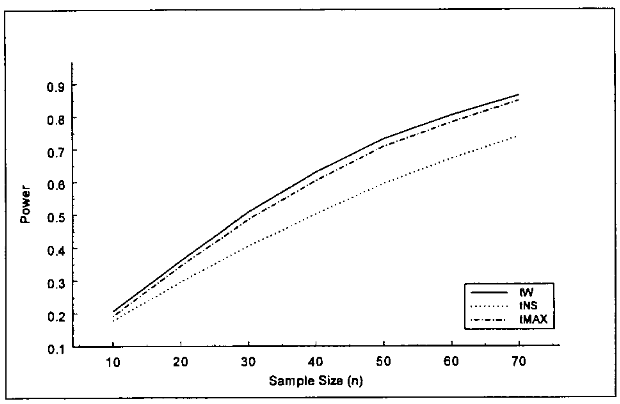


Figure 3. Power of tW, tMS and tMAX When Sampling Is From a Cauchy Distriubtuon



t_{NS}. Under the heavy-tailed Cauchy distribution, t_W was the most powerful statistic, with t_{max} once again demonstrating power similar to, but slightly less than, that of the most powerful test.

Conclusion

It is usually difficult for researchers to obtain sufficient information from a given data set so as to make reasonable choices between suitable statistical tests. It is important, therefore, that tests have power to detect broad classes of alternatives with high probability. The technique demonstrated here is a simple method for constructing such tests.

A major advantage of the test presented here lies in the fact that this test is automatically adaptive to the weight in the tail of the population from which the data were sampled. This is contrasted with various adaptive estimation procedures which require a preliminary estimate of tail weight.

It should also be noted that the maximum method may be extended to a wide variety of testing situations. For example, more than two statistics my be formed into a maximum test, with component tests being both parametric and non-parametric. A large number of other possibilities exist.

References

Bradley, J. V. (1968). *Distribution-free statistical tests*. Englewood Cliffs, NJ: Prentice-Hall.

Chernoff, H., & Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Annals of Mathematical Statistics*, *28*, 972-994.

Conover, W. J., & Iman, R. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, *35*, 124-129.

Cox, D. R. (1977). The role of significance tests. *Scandinavian Journal of Statistics*, *4*, 49-70.

Hodges, J. L., & Lehmann, E. L. (1961). Comparison of the normal scores and Wilcoxon tests. In (J. Neyman, ed.) *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1*. Berkeley and Los Angeles, CA: University of California Press.

Iman, R. L. (1974). An approximation to the exact distribution of the Wilcoxon-Mann-Whitney rank sum test statistic. *Communications in Statistics*, *5*, 587-598.

Lehmann, E. L. (1959). *Testing statistical hypotheses*. NY: Wiley.

Mikulski, P. W. (1963). On the efficiency of optimal nonparametric procedures in the two sample case. *Annals of Mathematical Statistics*, *34*, 22-32.

Owen, D. B. (1962). *Handbook of statistical tables*. Reading, MA: Addison-Wesley.

Terry, M. E. (1952). Some rank order tests which are most powerful against specific parametric alternatives. *Annals of Mathematical Statistics*, *23*, 346-366.