Psicológica (2011), 32, 65-85.

Inheritance of Properties of Normal and Non-Normal Distributions After Transformation of Scores to Ranks

Donald W. Zimmerman*

Carleton University, Canada

This study investigated how population parameters representing heterogeneity of variance, skewness, kurtosis, bimodality, and outlier-proneness, drawn from normal and eleven non-normal distributions, also characterized the ranks corresponding to independent samples of scores. When the parameters of population distributions from which samples were drawn were different, the ranks corresponding to the same pairs of samples of scores inherited similar differences. This finding explains some known results concerning Type I error probabilities and the relative power of parametric and nonparametric tests for various non-normal densities.

When random samples are drawn from populations with certain known parameters, it is commonplace to expect that statistics calculated from the samples will not deviate too far from the population parameters. For example, if two populations have different means, one expects that two samples, one drawn from each population, will also have different means. Furthermore, if the scores in the samples are combined, or pooled together, and then transformed to a single series of ranks; and each score in the respective samples is replaced by its corresponding rank, one expects that the two sets of ranks will also have different means. This method, or a similar method, is employed in nonparametric rank tests such as the Wilcoxon-Mann-Whitney test, the Kruskal-Wallis test, and the Wilcoxon signed-ranks test.

When examining the properties of these nonparametric methods, researchers have focused on the fact that ranks preserve order, but do not preserve other characteristics of distributions, such as the differences, or intervals, between successive scores. A distribution of N ranks can be considered a rectangular distribution of integers on the interval (1, N). The fact that statistics calculated from ranks do not preserve the size of intervals in the original population of scores explains some of the favorable properties of nonparametric rank tests.

However, it is easy to overlook another feature of these nonparametric methods. If scores from two or more samples are combined into a single group and ranked, and scores in the respective samples are replaced by the ranks, the resulting groups of ranks do not necessarily have rectangular distributions. Although it is usually taken for granted that the *means* of separate groups of ranks reflect the means of the original samples, it is not always appreciated that other parameters, related to the shape or the degree of variability of the distributions of ranks, also can retain properties of the original samples.

* Correspondence: Donald W. Zimmerman, Ph.D. Professor Emeritus, Carleton University, Ottawa, Canada. Mailing address: 1978 134A Street. Surrey, BC, V4A 6B6 Canada. Phone: 604-531-9313. Fax: 604-541-4994. Email: dwzimm@telus.net

Consider, for example, differences in variances. If two samples of scores with decidedly heterogeneous variances, drawn from populations with heterogeneous variances, are transformed to ranks, those ranks can be expected to inherit, to some extent, the inequality of variances. A simple example makes this clear: Suppose the scores in one group are 1.5 and 20.5 and the scores in another group are 10 and 11. If all scores are combined and ranked, the resulting two sets of ranks are 1 and 4 compared to 2 and 3. So an initial difference in the variances of the scores is retained by the resulting ranks, although it is somewhat less. For larger *N*, a still smaller difference in variances of the original samples can also be preserved and noticeable in the ranks (see, for example, Zaremba, 1965; Zimmerman, 1996).

One might expect that other characteristics of distributions, including higher moments, also apply to ranks. Another simple example: Suppose scores in the first group are 10.1, 11, and 12.5, and those in the second group are 3, 20, and 22.5. Then the ranks in the first group are 2, 3, and 4, while the ranks in the second group are 1, 5, and 6. So in this case, not only variance but also *asymmetry* of scores, or skewness, is imparted to the ranks. The present study examined various sample statistics, calculated from both scores and ranks, using simulation methods to replicate the ranking process over large numbers of samples. In addition to variances, the study determined the extent to which skewness, kurtosis, bimodality, and outlier-proneness in distributions of scores also characterize corresponding distributions of ranks.

METHOD

The simulations in this study employed the *Mathematica* programming language, together with *Mathematica* statistical add-on packages[1]. First, the program generated two independent samples of N_1 and N_2 scores from one of the standard densities in the *Mathematica* package or from one of the mixtures described below. All distributions had mean 0 and variance 1. Next, the scores were combined into a single group and transformed to ranks ranging from 1 to $N_1 + N_2$. Finally, the scores in the original two groups were replaced by the ranks that had been assigned in the combined group. This ranking method is familiar in nonparametric significance tests, such as the Wilcoxon-Mann-Whitney test and the Kruskal-Wallis test.

The program calculated means, standard deviations, coefficients of skewness, and coefficients of kurtosis of the scores in the original groups of scores and the same statistics in the corresponding groups of ranks. Also, the program found ratios of standard deviations, s_1/s_2, for the two groups of scores, as well as ratios of standard deviations of ranks, r_1/r_2, for the respective groups of ranks. Similar ratios of both scores and ranks, were found for coefficients of skewness and kurtosis. The sampling procedure was replicated 50,000 times under each condition investigated. Also, the study found frequency distributions of both the original scores tabulated in class intervals, as well as the distributions of assigned ranks from 1 to $N_1 + N_2$, in the respective groups.

The study investigated six symmetric and six skewed distributions. Eight of these were standard continuous densities with known parameters. These included the normal, exponential, Laplace (double-exponential), lognormal, logistic, Gumbel (extreme-value), uniform (rectangular), and half-normal distributions. In addition, four mixtures of distributions, chosen to represent non-normal data often encountered in research, were included. A symmetric bimodal distribution consisted of samples from $N(1, 1)$ with probability .5 and from $N(-1, 1)$ with probability .5. A skewed bimodal distribution consisted of samples from $N(-.4, .2)$ with probability .25 and from

$N(.4, .1)$ with probability .75. To represent outliers in research data, a symmetric mixed-normal distribution consisted of samples from $N(0, 1)$ with probability .9 and from $N(0,10)$ with probability .1. Finally, a skewed mixed-normal distribution consisted of samples from $N(0,1)$ with probability .9 and from $N(5, .1)$ with probability .1. All mixtures were standardized to have mean 0 and variance 1.

The program also obtained frequency distributions of sample values of scores in two groups as well as the corresponding ranks over large numbers of replications of the procedure. These distributions were plotted graphically[2]. The program found similar distributions for several cases in which three or four distributions were combined before ranking, as done in the Kruskal-Wallis test, and the distributions were plotted graphically.

Sample values of skewness and kurtosis were found from the built-in functions of the *Mathematica* software. The population values of skewness and kurtosis shown in the tables were known for the standard probability densities included. The values for the bimodal and mixed-normal distributions created for purposes of the study were found by simulation, based on one million samples.

As a check on the accuracy of the method of finding random normal deviates, some simulations in the study were repeated using random numbers generated by the method of Marsaglia, Zaman, & Tsang (1990), described by Pashley (1993), together with normal deviates obtained by the method of Marsaglia & Bray (1964). These values were transformed using inverse distribution functions in order to obtain samples from the various non-normal densities. This procedure gave results close to those in Tables 1 and 2, so all subsequent simulations used random deviates obtained from continuous distributions in the *Mathematica* statistical add-on package.

RESULTS OF SIMULATIONS

Inheritance of heterogeneity of variance.

The results in Table 1 indicate that differences in variability in scores in two groups are preserved after transformation of the scores to ranks. The second column is the ratio of the standard deviations of two normally distributed populations. The third column is the mean ratio of standard deviations of two independent random samples of size 25 from those populations. The ratios in the samples of scores remained close to the ratios in the populations. The third column is the mean ratio of standard deviations in two groups of ranks, after the scores in both groups have been combined and the original scores in the two samples have been replaced by their ranks. The ratio was found for each individual pair of samples, and the table shows the mean of these ratios over all 50,000 pairs. The remaining two columns in the table give similar results for $N = 100$.

More detailed information about the shape of the entire distributions of these ratios over all replications of the sampling procedure also is informative. Relative frequency distributions of the ratios of standard deviations of scores and the ratios of standard deviations of the corresponding ranks are plotted in Figure 1, for normal distributions with $N = 100$. Three population ratios, σ_1/σ_2, of 1, 2, and 3 are shown in the figure. Here the degree of variability and shifts in the means one would expect from the means and standard deviations in Table 1 is obvious.

Table 1. Ratios of standard deviations of scores and ranks in samples from populations with equal means and varying ratios of standard deviations.

Population Distribution	σ_1/σ_2	$N_1 = N_2 = 25$ scores	$N_1 = N_2 = 25$ ranks	$N_1 = N_2 = 100$ scores	$N_1 = N_2 = 100$ ranks
	1	1.021	1.009	1.005	1.002
normal	2	2.045	1.444	2.010	1.437
	3	3.062	1.705	3.013	1.702
	1	1.075	1.008	1.016	1.001
exponential	2	2.153	1.590	2.040	1.586
	3	3.204	1.818	3.057	1.821
	1	1.056	1.012	1.014	1.003
Laplace	2	2.092	1.335	2.026	1.328
	3	3.150	1.547	3.036	1.542
	1	1.226	1.009	1.088	1.003
lognormal	2	2.446	1.623	2.174	1.620
	3	3.685	1.827	3.248	1.832
	1	1.033	1.009	1.008	1.002
logistic	2	2.072	1.407	2.016	1.398
	3	3.101	1.653	3.018	1.651
	1	1.042	1.008	1.010	1.002
Gumbel	2	2.083	1.464	2.020	1.458
	3	3.132	1.720	3.031	1.718
	1	1.008	1.008	1.002	1.002
bimodal	2	2.018	1.792	2.004	1.789
(symmetric)	3	3.027	2.125	3.006	2.135
	1	1.034	1.010	1.006	1.003
bimodal	2	2.057	1.808	2.013	1.813
(skewed)	3	3.106	2.086	3.017	2.104
	1	1.359	1.008	1.075	1.002
mixed-normal	2	2.711	1.341	2.154	1.334
(symmetric)	3	4.061	1.534	3.216	1.527
	1	1.056	1.009	1.010	1.003
mixed-normal	2	2.108	1.401	2.019	1.395
(skewed)	3	3.159	1.610	3.027	1.607
	1	1.011	1.010	1.002	1.002
uniform	2	2.019	1.603	2.004	1.601
	3	3.026	1.868	3.009	1.871
	1	1.032	1.010	1.007	1.002
half-normal	2	2.063	1.544	2.016	1.537
	3	3.099	1.796	3.022	1.795

In both Figure 1 and Table 1, the ratios of standard deviations of ranks are somewhat less than the same ratios of scores, although in all cases they change in a systematic way as the population ratios change. The pattern is much the same for all 12 distributions, and the values are quite close for both N = 25 and N = 100. These results are consistent with the well-established finding that "heterogeneity of variance," influences nonparametric rank tests in the same way that it influences t

and *F* tests, although to a lesser degree (see, for example, Pratt, 1964; Zaremba, 1965; Lehmann, 1975; Zimmerman, 1996) .

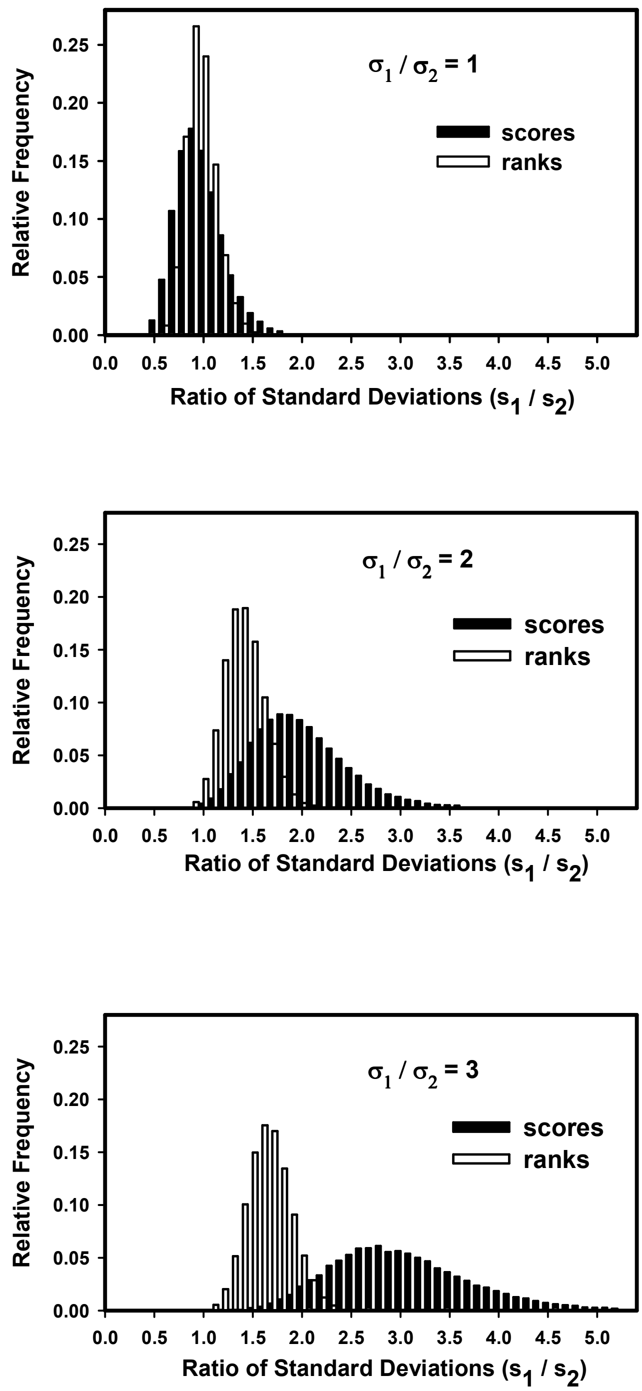

Fig. 1. Relative frequency distributions of the ratios of standard deviations of scores and ranks for three values of the corresponding population ratio.

Figure 2 plots relative frequency distributions of the scores and ranks themselves, over all 50,000 replications. Evidently, the shape of the distributions of ranks in each of the two groups is quite different from that of the distributions of scores. However, some characteristics of the scores, such as the difference in variability, are evident in the ranks. Also, the *symmetry* of each distribution of scores is retained by the corresponding distribution of ranks.

Table 2 presents results of a somewhat different approach, in which pairs of samples were selected so that only those in which the ratio of standard deviations of scores exceeded a cutoff value were retained and transformed to ranks. The cutoff ratios are shown as column headings, and the first column shows sample sizes. The first row for a given sample size is the mean ratio of standard deviations of ranks corresponding to the scores exceeding the cutoff. The second row for that sample size is the proportion of samples that exceeded the cutoff value. Clearly, the ratios of standard deviations of ranks reflected the ratios of standard deviations of scores for each cutoff ratio and varied systematically with the size of the cutoff ratio.

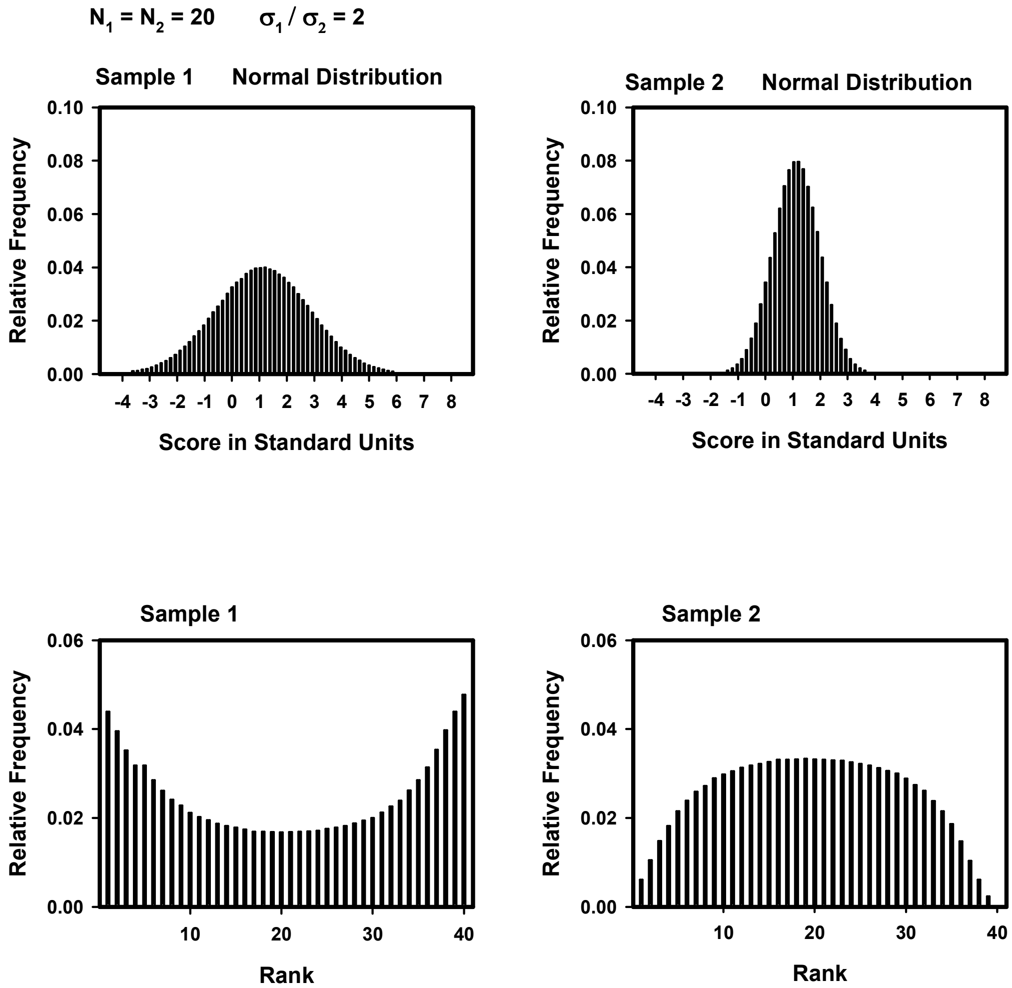

Fig. 2. Inheritance of heterogeneity of variance of distributions of scores by distributions of ranks.

Table 2. Mean ratio of standard deviation of ranks in group 1 and standard deviation of ranks in group 2 (s_1/s_2) conditional on the ratio of the standard deviation of scores in group 1 and the standard deviation of scores in group 2 exceeding the cutoff values shown in the first row. The rows labeled p are the proportions of all ratios of scores exceeding the cutoff values.

		Cutoff ratios of Standard Deviations of Scores					
$N_1 = N_2$		1.0	1.1	1.2	1.3	1.4	1.5
	p	.498	.404	.320	.252	.199	.153
8	s_1/s_2	1.218	1.251	1.305	1.349	1.389	1.414
	p	.498	.385	.296	.221	.163	.121
10	s_1/s_2	1.185	1.226	1.267	1.319	1.358	1.399
	p	.494	.360	.249	.171	.111	.078
15	s_1/s_2	1.134	1.176	1.228	1.265	1.305	1.347
	p	.505	.326	.187	.097	.056	.022
25	s_1/s_2	1.101	1.143	1.186	1.229	1.263	1.303
	p	.501	.258	.101	.034	.009	.003
50	s_1/s_2	1.067	1.109	1.150	1.197	1.251	1.286
	p	.496	.164	.035	.005	.001	0
100	s_1/s_2	1.047	1.087	1.129	1.168	1.233	---

Inheritance of differences in skewness.

Table 3 presents similar data comparing coefficients of skewness for scores and ranks.

Each of the distributions listed in the first column, from which sample 1 was drawn, was paired with a normal distribution, from which sample 2 was drawn. Therefore, the difference in skewness was zero when a symmetric distribution was paired with the normal distribution, but otherwise was non-zero.

The second column shows the known values of skewness for the various population distributions. The third and fourth columns are the mean coefficients of skewness for the two samples of scores, and the fifth and sixth columns are the mean coefficients for the ranks.

The coefficients obtained from samples were not far from the known values for the population distributions. First, for the symmetric population distributions, values for both scores and ranks were close to zero, consistent with the zero skewness of the populations. However, the skewed distributions, had substantial non-zero values for both scores and ranks, although the values for ranks were consistently less. Therefore, a pattern similar to the familiar one found for "heterogeneity of variance" occurred in the case of skewness. Figure 3 shows more detailed frequency distributions of scores and ranks obtained in samples from a Gumbel distribution and a normal distribution. Clearly, for this pair of distributions, the differences in skewness of the scores was inherited by the ranks.

Table 3. Mean coefficients of skewness of scores and ranks from two samples ($N_1 = N_2 = 100$). Samples from the distributions in the first column (Sample 1) were combined with samples from a normal distribution (Sample 2).

Population Distribution	Population Skewness (Sample 1)	Sample Skewness of Scores		Sample Skewness of Ranks	
		Sample 1	Sample 2	Sample 1	Sample 2
normal	0	.003	.004	.000	.000
exponential	2	1.787	.001	.344	−.284
Laplace	0	.002	−.001	−.001	.001
lognormal	6.185	3.185	−.001	.464	−.291
logistic	0	.000	.001	.002	−.002
Gumbel	1.140	1.033	.003	.163	−.152
bimodal (symmetric)	0	−.004	.002	−.003	.003
bimodal (skewed)	−1.206	−1.213	.001	−.535	.487
mixed-normal (symmetric)	0	−.022	.000	.001	−.002
mixed-normal (skewed)	1.370	1.365	−.002	.270	−.229
uniform	0	.001	−.005	.001	−.001
half-normal	.995	.947	.002	.197	−.193

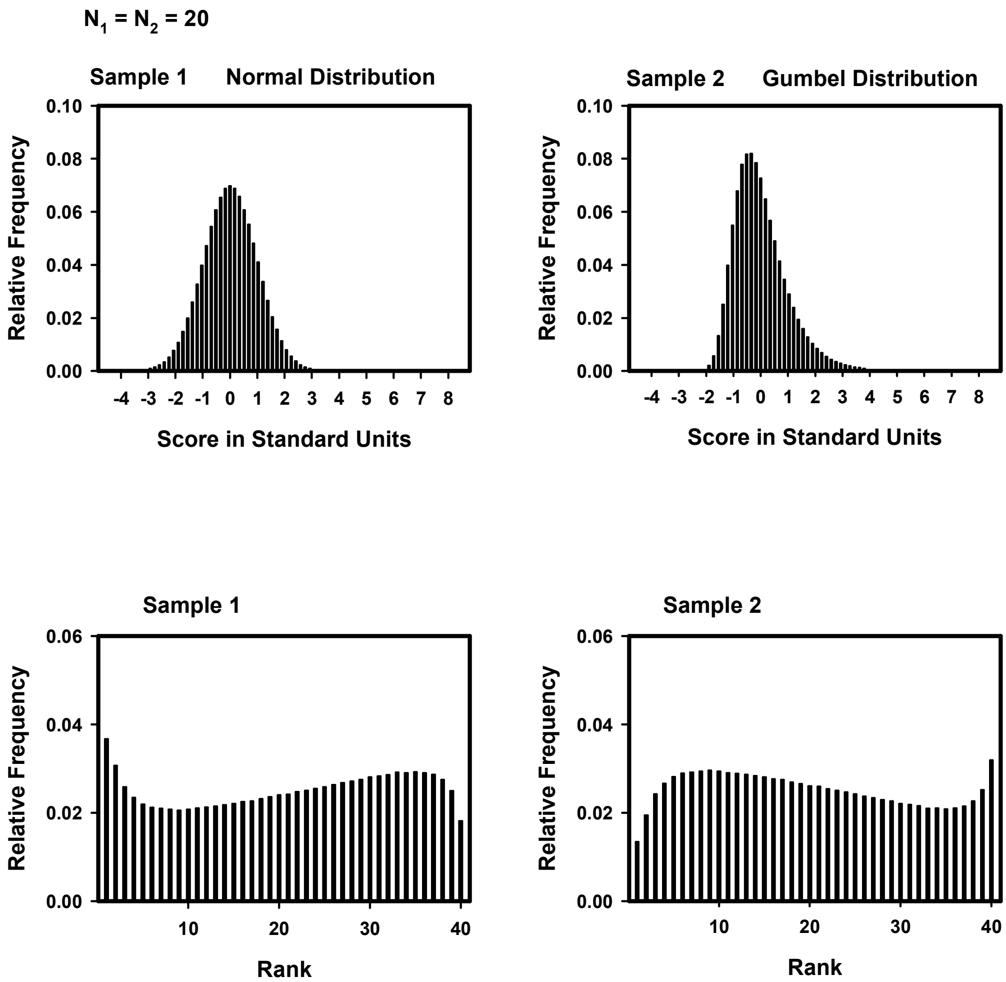

Fig. 3. Inheritance of skewness of distributions of scores by distributions of ranks.

Inheritance of differences in kurtosis.

Table 4 presents data for coefficients of kurtosis. Again the second column gives the kurtosis of the populations. The third column, labeled sample 1, is consistent with the values for the various populations, except for the anomalous case of the lognormal distribution, while the values in the fourth column, labeled sample 2, are near the 3.00 of the normal distribution. The fifth column gives mean ratios of coefficients obtained in sample 1 and sample 2. It should be emphasized that those ratios are not *ratios of the means* shown in the third column and those in the fourth column, but are the *means of the ratios* of the coefficients of kurtosis found in each individual sample.

Table 4. Mean coefficients of kurtosis and ratios of coefficients of kurtosis of scores and ranks from two samples (N1 = N2 = 100). Samples from the distributions in the first column (Sample 1) were combined with samples from a normal distribution (Sample 2).

Population Distribution	Population Kurtosis (Sample 1)	Sample Kurtosis of Scores			Sample Kurtosis of Ranks		
		Sample 1	Sample 2	Ratio	Sample 1	Sample 2	Ratio
normal	3	2.944	2.941	1.023	1.813	1.812	1.004
exponential	9	7.145	2.934	2.487	1.933	1.820	1.067
Laplace	6	5.321	2.944	.596	1.998	1.660	1.846
lognormal	113.936	17.577	2.942	6.088	2.187	1.651	1.332
logistic	4.200	3.908	2.945	1.355	1.872	1.757	1.069
Gumbel	5.400	4.719	2.933	1.643	1.865	1.803	1.030
bimodal (symmetric)	1.720	1.737	2.940	.604	1.578	2.117	.749
bimodal (skewed)	3.091	3.165	2.946	1.098	2.069	1.993	1.049
mixed-normal (symmetric)	25.348	20.333	2.928	7.089	2.441	1.417	1.728
mixed-normal (skewed)	4.553	4.682	2.948	1.622	2.108	1.678	1.261
uniform	1.800	1.827	2.938	.635	1.659	1.999	.833
half-normal	3.869	3.667	2.942	1.273	1.792	1.893	.951

Again, most of the values found for both scores and ranks reflect the values of the population distributions. Exceptions are the Laplace, uniform, and symmetric bimodal distributions, which the ratios of ranks are somewhat less than those of scores. In the three atypical cases, the initial values of kurtosis of the populations are less than those of the other distributions, and are less than the 3.00 of the normal distribution. In other words, once again, the degree of "heterogeneity of kurtosis," with respect to a normal density, was inherited by the ranks. Figure 4 shows in more detail relative frequency distributions of scores and ranks in samples from a Laplace distribution and a normal distribution, and the same thing is evident.

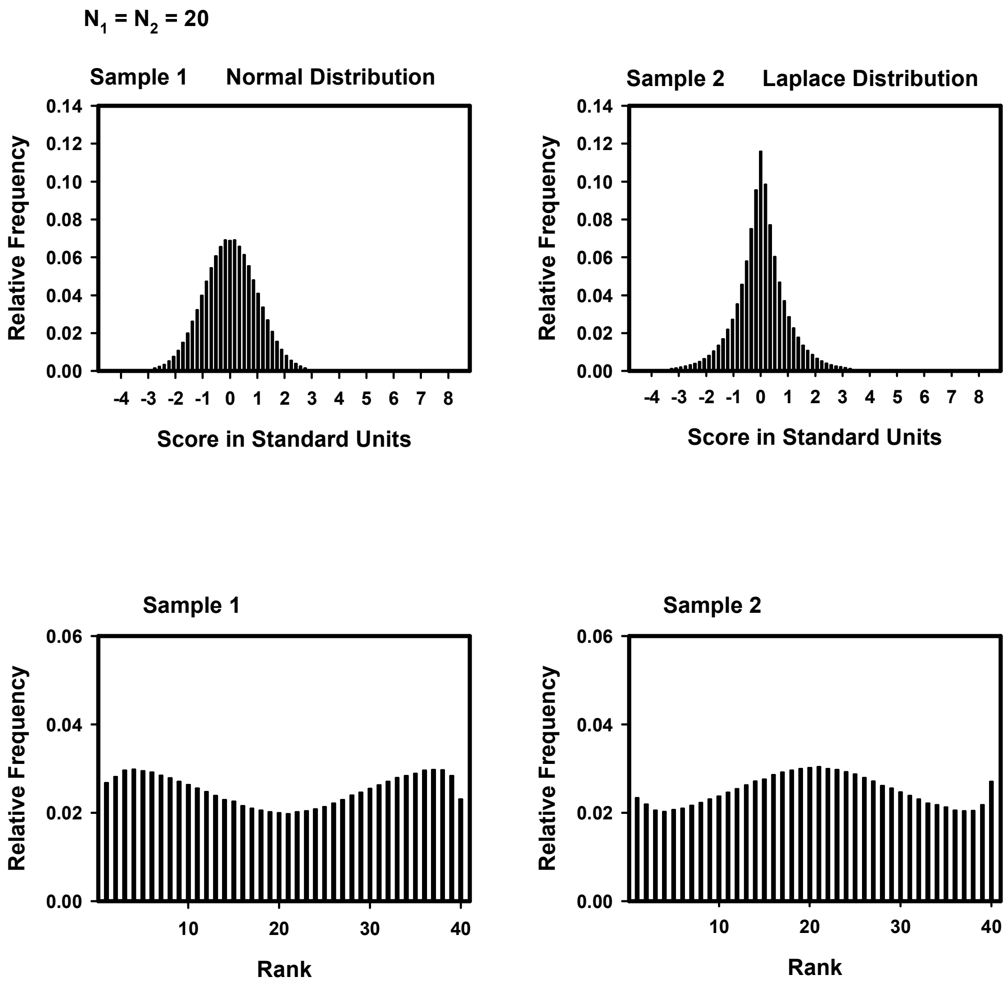

Fig. 4. Inheritance of kurtosis of distributions of scores by distributions of ranks.

Table 5 includes comparisons in which both samples 1 and 2 are taken from non-normal densities that possibly differ in skewness and kurtosis. The first two columns list the respective densities. The outcome is similar to that found in Table 4: A difference in skewness of samples of scores is reflected by a similar, although somewhat less, difference in skewness of ranks. Furthermore, a difference in kurtosis of samples of scores is reflected in similar and lesser difference in kurtosis of ranks, and the ratios show a similar trend.

Table 5. Mean coefficients of skewness and kurtosis of scores and ranks in samples from non-normal distributions (N_1 = N_2 = 100). Samples from distributions in the first column were combined with Samples from distributions in the second column.

Population Distribution		Sample Skewness of Scores		Sample Skewness of Ranks	
1	2	Sample 1	Sample 2	Sample 1	Sample 2
Gumbel	uniform	1.025	.001	.175	−.133
exponential	Laplace	1.793	−.003	.340	−.340
Laplace	uniform	−.017	.001	.000	.000
lognormal	exponential	3.201	1.796	.204	−.203
lognormal	Laplace	3.198	.006	.465	−.386
logistic	exponential	.000	1.797	−.306	.338

Population Distribution		Sample Kurtosis of Scores			Sample Kurtosis of Ranks		
1	2	Sample 1	Sample 2	Ratio	Sample 1	Sample 2	Ratio
Gumbel	uniform	4.681	1.825	2.576	2.038	1.653	1.238
exponential	Laplace	7.196	5.325	1.484	1.801	2.058	.880
Laplace	uniform	5.314	1.824	2.925	2.198	1.545	1.427
lognormal	exponential	17.662	7.231	2.929	1.942	1.580	1.233
lognormal	Laplace	17.628	5.311	3.648	2.050	1.888	1.093
logistic	exponential	3.909	7.225	.648	1.902	1.884	1.014

Inheritance of outlier-proneness.

Outliers have played a large role in alteration of Type I and Type II errors of many familiar parametric tests, and one desirable feature of nonparametric tests is their reduction of the influence of outliers through ranking. The fact that various non-normal densities are outlier-prone and others are outlier-resistant was investigated by Neyman & Scott (1971) and Green (1976). However, contrary to expectation, it is not true that the influence of outliers is entirely eliminated by ranking. Of course, maximum and minimum scores are transformed into maximum and minimum ranks.

When ranks are separated in two groups are frequency distributions are plotted as done in the figures in the present study, the common influence of outliers on distributions of both scores and ranks becomes evident.

Figure 5 plots relative frequency distributions of samples of scores and ranks from a mixed-normal distribution, compared to samples from a normal distribution. The mixed-normal distribution was obtained by sampling from *N*(0,1) with probability .9 and from *N*(5, .1) with probability .1, where these values are given in units of a standard deviation. That is, in the mixed-normal case an outlying score, approximately 5σ above the mean, occurred with probability .1. The result is a clustering of scores far above the mean, as expected, but also a clustering of anomalous ranks corresponding to the outlying scores.

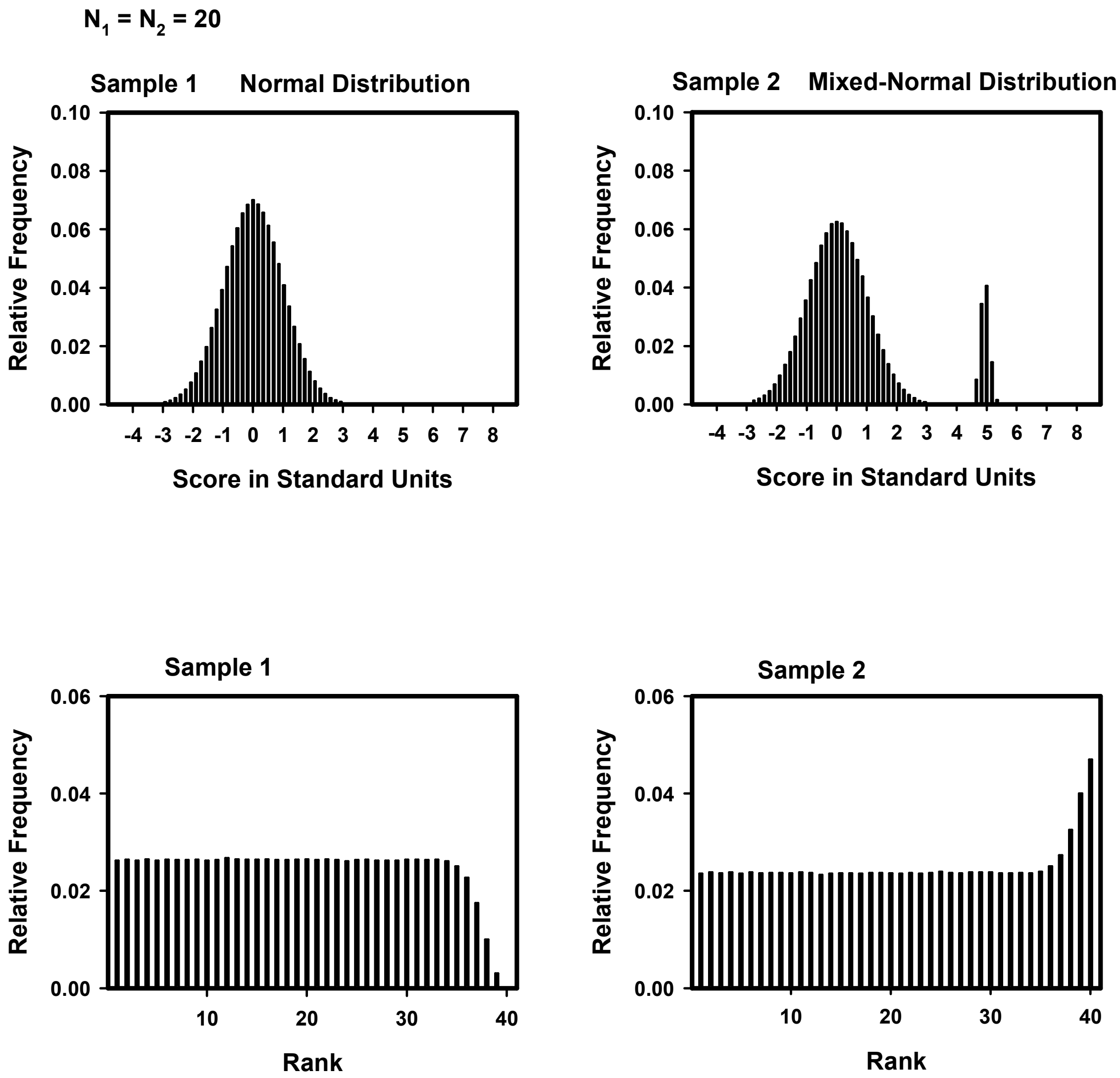

Fig. 5. Inheritance of susceptibility to outliers of distributions of scores by distributions of ranks.

Inheritance of bimodality.

Figure 6 plots frequency distributions of scores and ranks for a normal distribution compared to a symmetric bimodal distribution having two distinct clusters of scores. Figure 7 is the same for a skewed bimodal distribution, in which the concentration of scores around the upper mode exceeds the concentration around the lower mode. The symmetric bimodal distribution consisted of samples from $N(1, 1)$ with probability .5 and from $N(-1, 1)$ with probability .5 and the skewed bimodal distribution consisted of samples from $N(-.4, .2)$ with probability .25 and from $N(.4, .1)$ with probability .75. It is apparent in the figures that the distribution of ranks again displays features of the original distributions. Note that the asymmetry of the distribution of scores in Figure 6 also characterizes the ranks, and that the symmetry of the distribution of scores in Figure 7 is preserved by the ranks.

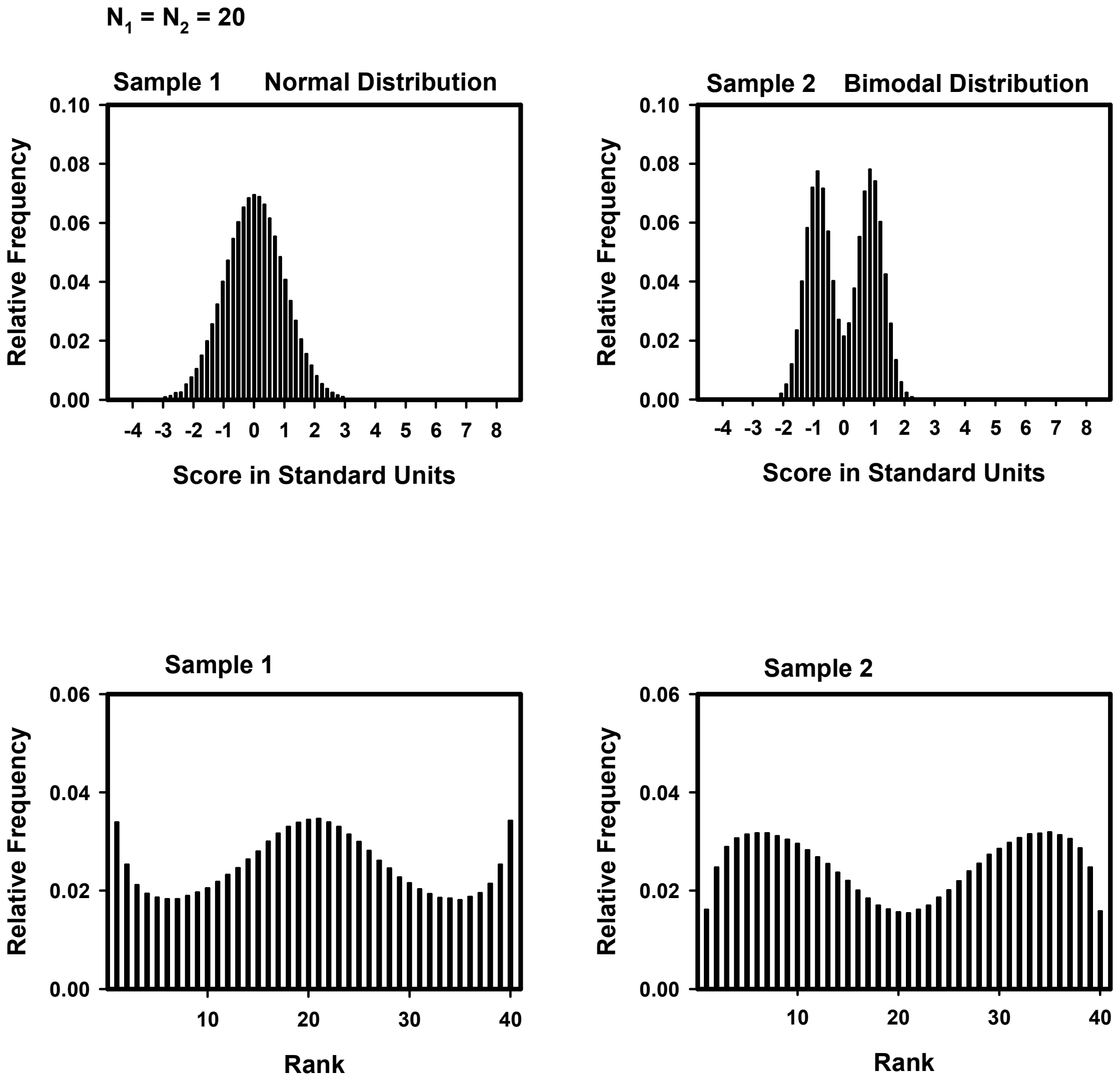

Fig. 6. Inheritance of bimodality (symmetric).

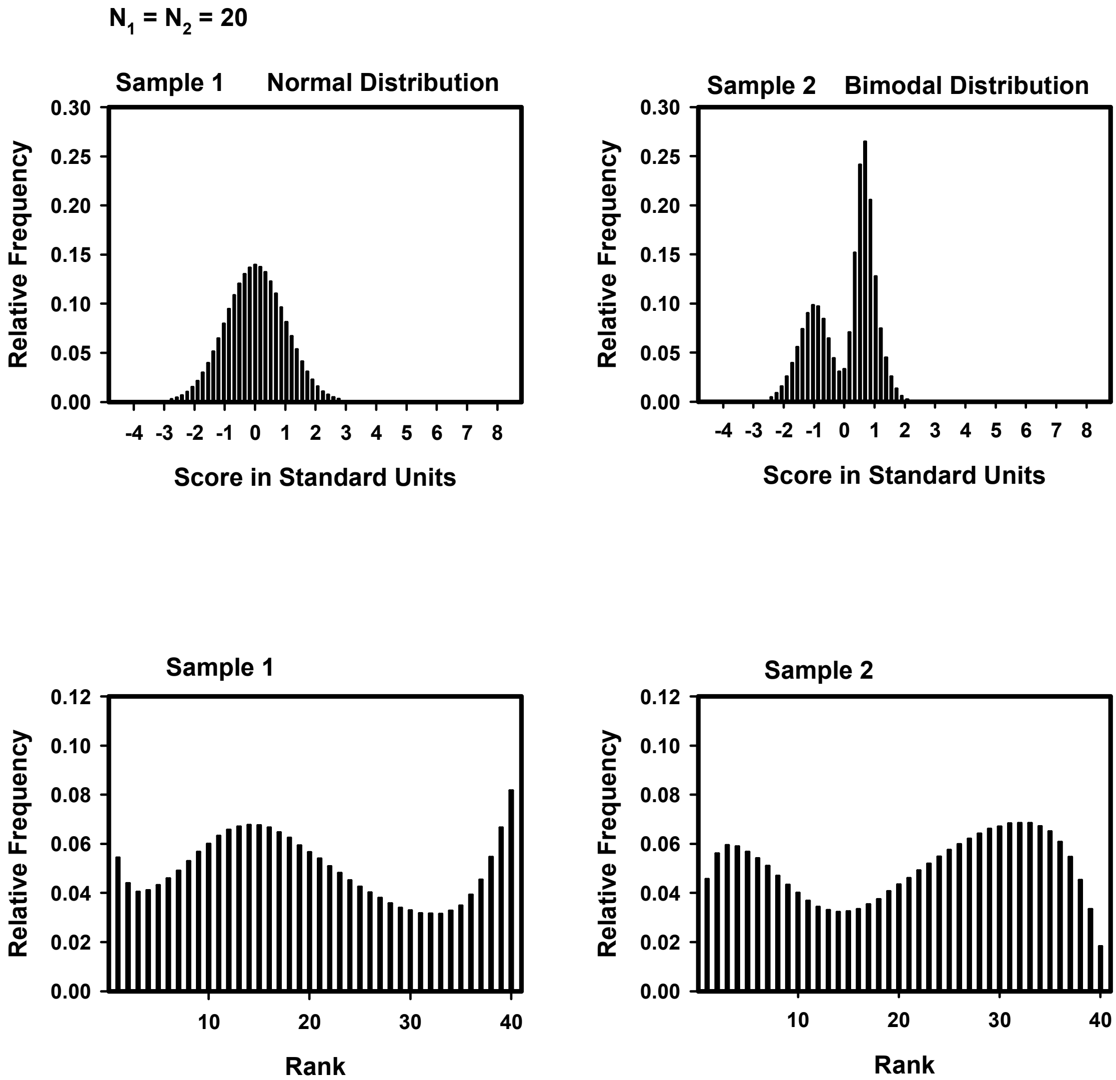

Fig. 7. Inheritance of bimodality (skewed).

Properties of three or more sets of ranks from different score distributions.

Figures 8, 9, 10, and 11 plot distributions of ranks in three or more groups, resulting from the ranking procedure employed in the Kruskal-Wallis test, an extension of the one used in the Wilcoxon-Mann-Whitney test. That is, the scores in several groups were all pooled together and ranked as a single group, after which the scores in the original groups were replaced by their ranks in the combined group.

Figure 8 is the result for three symmetric distributions, and Figure 9 is the result for two symmetric and one skewed distribution. Figure 10 is the result for four symmetric distributions, and Figure 11 is for two symmetric and two skewed distributions. In each case, when all three or all four distributions were symmetric, the resulting distributions of ranks were symmetric. And when one or more distributions was skewed, the distributions of ranks were skewed.

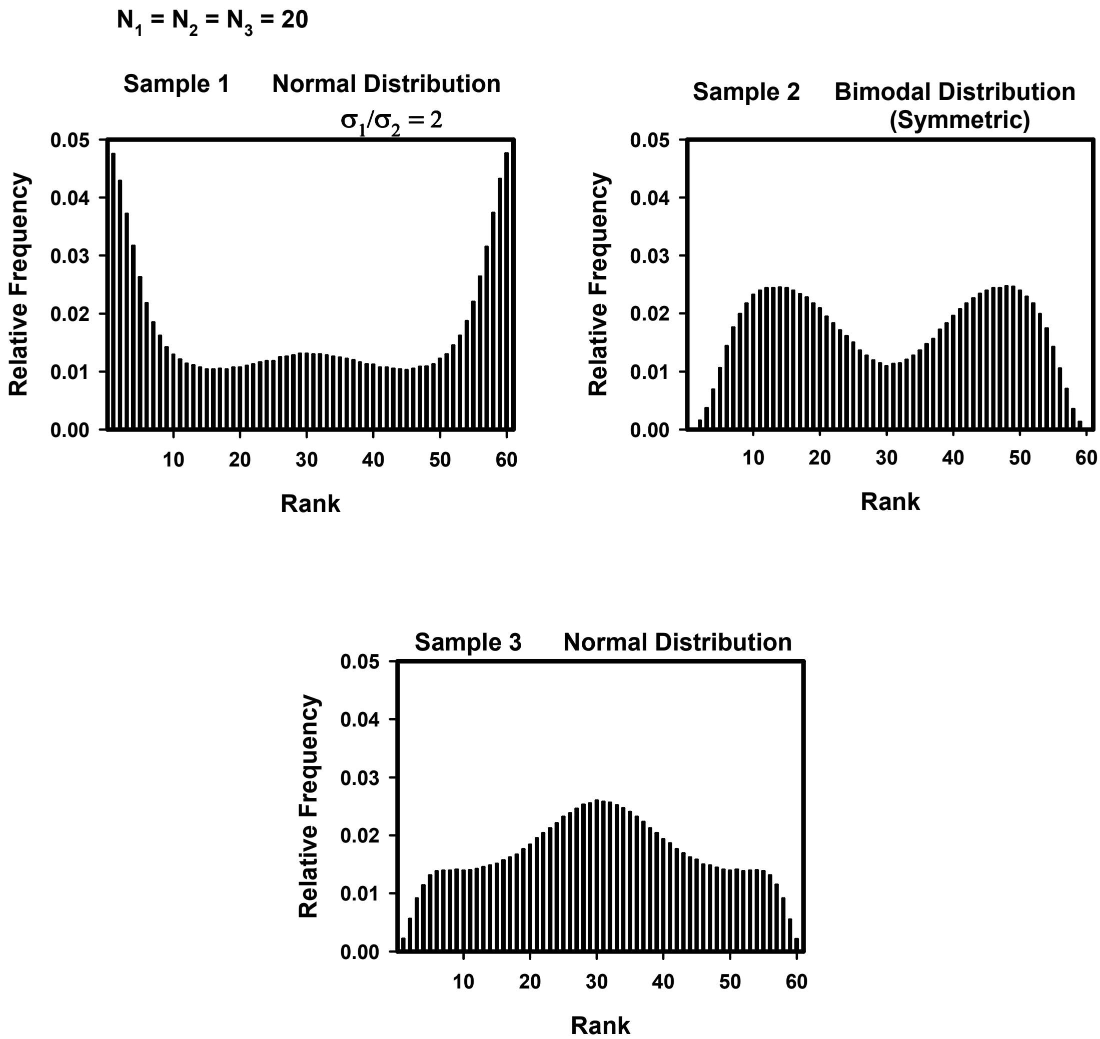

Fig. 8. Relative frequency distributions of ranks for three distributions of scores (all symmetric)

FURTHER DISCUSSION

Nonparametric tests, such as the Wilcoxon-Mann-Whitney test and the Kruskal-Wallis test maintain the nominal significance level in the case of many non-normal distributions for which *t* and *F* tests are not robust. Conversion of scores to ranks is an essential feature of these methods. Moreover, it has been found that in many cases the power of these nonparametric tests exceeds that of their parametric counterparts, often substantially. Although these familiar rank tests are nonparametric distribution-free, other properties of ranks have to be taken into consideration to account for power advantages.

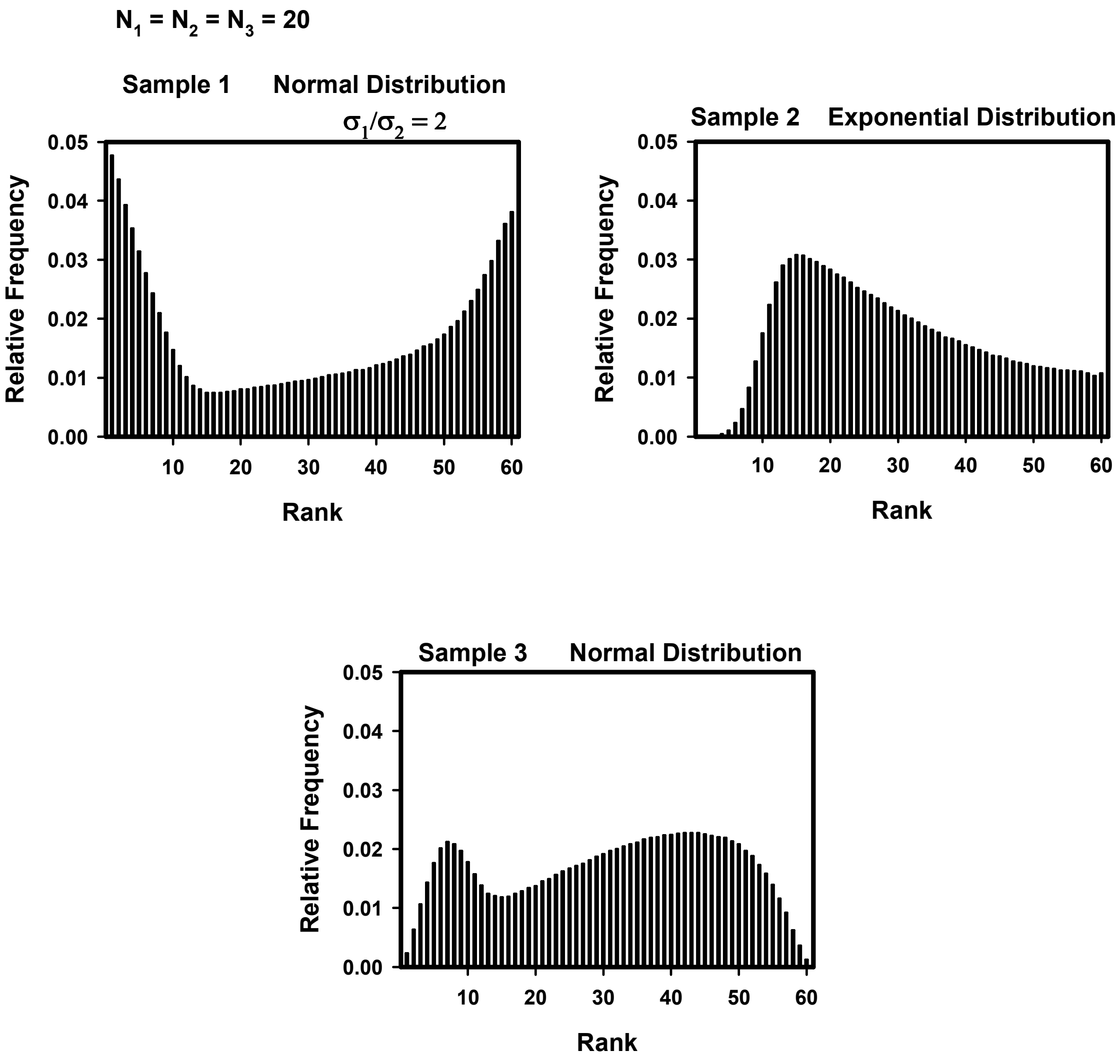

Fig. 9. Relative frequency distributions of ranks for three distributions of scores (2 symmetric, 1 skewed).

The fact that ranks themselves can have *distributions* that are not necessarily rectangular is frequently overlooked. If a single group of N scores are transformed to ranks, the distribution of the ranks is necessarily rectangular on the integers 1, …, N. However, if scores in two or more groups representing different experimental treatments are combined and ranked together, then the separate groups of combined ranks are not necessarily rectangular. The present study revealed marked similarities between parameters characterizing the distributions of the original scores and statistics calculated from the ranks.

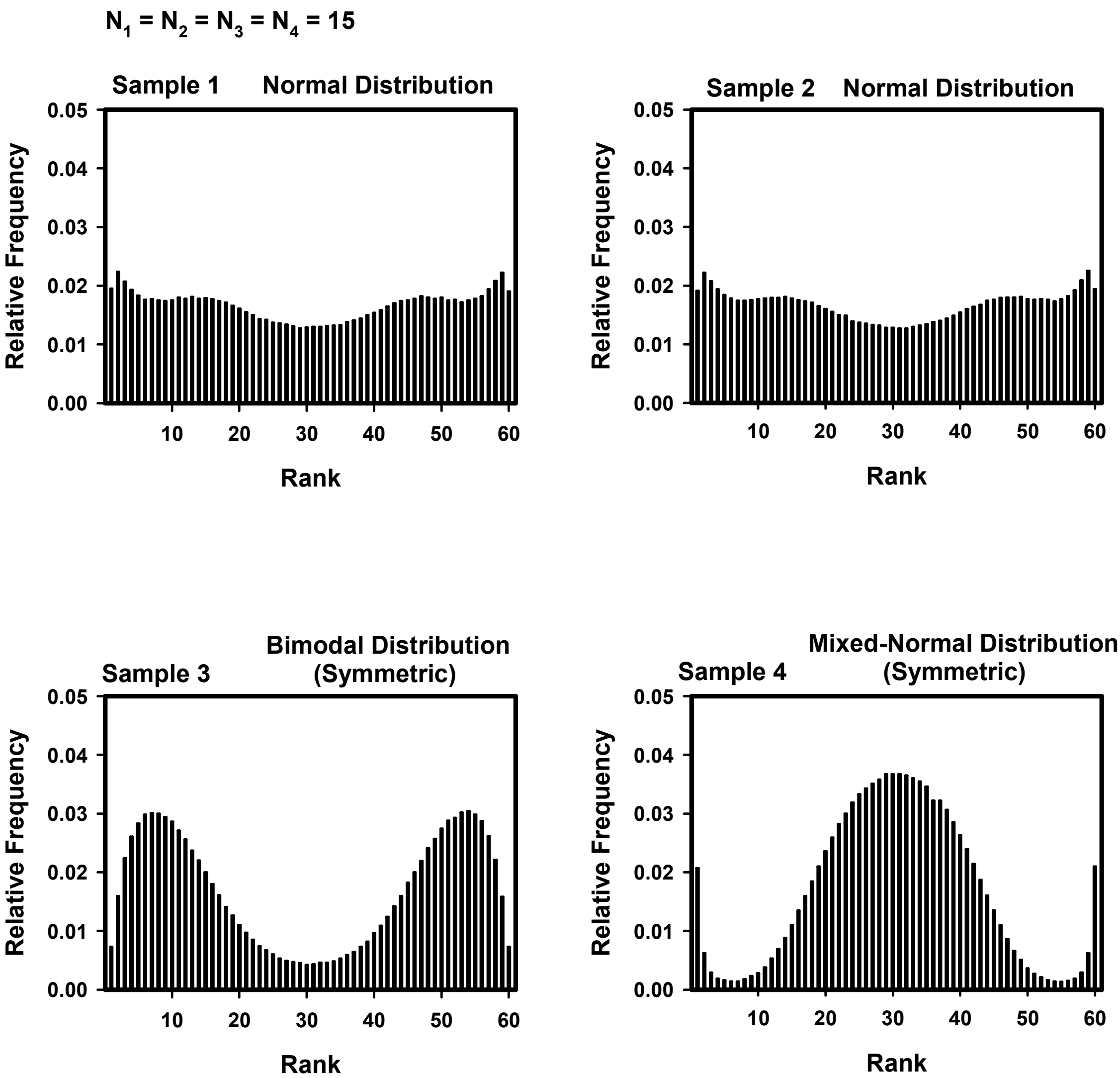

Fig. 10 Relative frequency distributions of ranks for four distributions of scores (all symmetric).

It is known that the Wilcoxon-Mann-Whitney test is equivalent to a t test performed on ranks replacing scores and that the Kruskal-Wallis test is equivalent to an F test performed on ranks (Conover & Iman, 1981). From that perspective, assignment of ranks to scores is a transformation similar to replacing scores, say, by their logarithms or by their reciprocals. The equivalence implies that "transformation of scores to ranks before performing the t test" is for practical purposes the same thing as "performing the Wilcoxon-Mann-Whitney test in place the t test", and similarly with the Kruskal-Wallis test and the F test.

Researchers usually make a choice between parametric and nonparametric methods after examining the distribution of the data at hand. Current findings suggest that the crucial step in the procedure is ranking itself and not the details of calculations done subsequent to ranking. It also suggests that the well-established power superiority of nonparametric tests for certain non-normal densities is related to the difference between ranks and scores *per se* and not to the details of calculating rank sums. In order to understand the differences in the parametric and nonparametric methods, this approach pays attention to features which the original distribution of scores and the transformed distribution of ranks have in common, as well as to their differences.

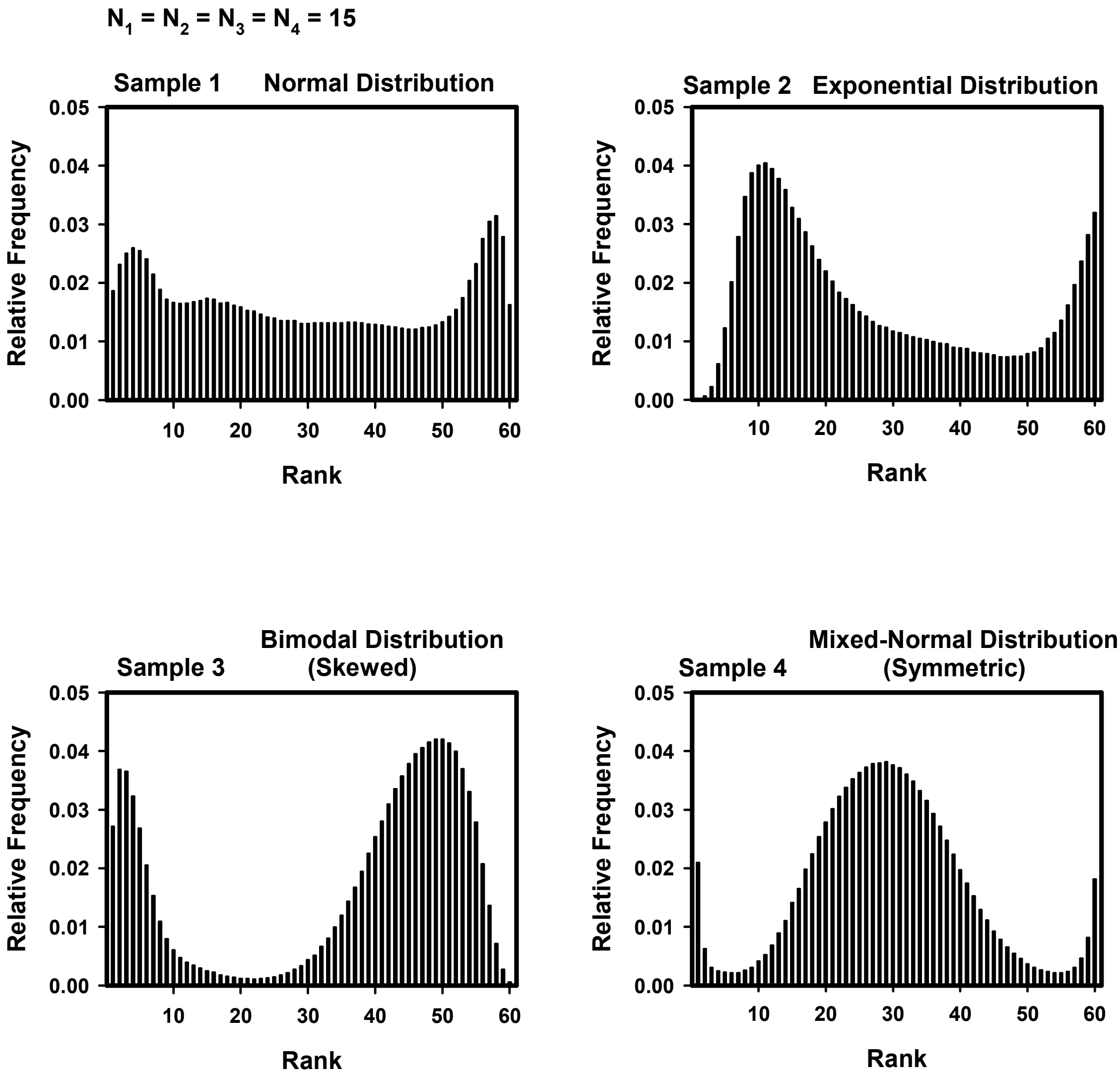

Fig. 11 Relative frequency distributions of ranks for four distributions of scores (two symmetric, 2 skewed).

The present study indicated that various parameters of the score distributions, including differences in variance, skewness, kurtosis, bimodality, and outlier-proneness, are inherited by ranks. These measures calculated from ranks are somewhat less in magnitude than the original ones calculated from scores. Nevertheless, they are of sufficient magnitude to substantially influence the Type I and Type II errors of well-known significance tests in ways similar to the influence of non-normal data on the *t* test. For example, a ratio of population standard deviations, σ_1/σ_2, of 2.00 or greater is known to severely compromise the Type I error probability of the *t* test. After transformation to ranks, the ratio is likely to be approximately 1.5 instead of 2. Nevertheless, it is still large enough to severely disrupt the Type I error probability of the Wilcoxon-Mann-Whitney test. Similar substantial, although somewhat less, differences in skewness, kurtosis, bimodality, and outlier-proneness also arise.

The shapes of the distributions of ranks found in the present study are consequences of that kind of separation of ranks into two groups corresponding 1-1 to the original treatment groups. Because the ranks are bounded by the total sample size $N_1 + N_2$ of the combined treatment groups, the frequencies of ranks in one group are constrained by the frequencies in the other. For this reason the shapes of the distributions of ranks are modified somewhat and do not precisely reproduce the shapes of the distributions of scores. Moreover, the data in Figures 8, 9, 10, and 11 reveal the same inheritance of properties of scores by ranks in designs involving three and four groups, as well as similar constraints on the shape of the distributions.

It is known that the original version of the Wilcoxon-Mann-Whitney test is sensitive to any difference in distributions, not just differences in means. Unequal variances, especially, can modify the nominal significance level of the test, analogous to the Behrens-Fisher problem in regard to the Student *t* test (Zimmerman, 1996). There have been several proposals for modifying the Wilcoxon-Mann-Whitney test in order to allow for unequal variances (see, for example, Brunner & Munzel, 2000; Cliff, 1996; Neuhauser, Losch, & Jockel, 2007, Wilcox, 2005).

The present results further suggest the desirability of this approach. Depending on characteristics of the original population, which often are not known, sample data may contain various differences in sample statistics of treatment groups greater than might be expected. Therefore, the Type I and Type II errors of significance tests, both parametric and nonparametric, can be altered in unpredictable ways.

REFERENCES

Brunner, E., & munzel, U. (2000). The nonparametric Behrens-Fisher problem: asymptotic theory and small-sample approximation. *Biometrical Journal, 42,* 17-25.

Cliff, N. (1996). *Ordinal methods for behavioral data analysis.* Mahwah, NJ: Erlbaum.

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

Green, R.F. (1976). Outlier-prone and outlier-resistant distributions. *Journal of the American Statistical Association, 71,* 502-505.

Hettmansperger, T.P., & McKean, J.W. (1978). Statistical methods based on ranks. *Psychometrika, 43,* 69-79.

Lehmann, E.L. (1975). *Nonparametrics: Statistical methods based on ranks.* San Francisco: Holden-Day.

Marsaglia, G., & Bray, T.A. (1964). A convenient method for generating normal variables. *SIAM Review, 6,* 260-264.

Marsaglia, G., Zaman, A., & Tang, W.W. (1990). Toward a universal random number generator. *Statistics and Probability Lectures, 8,* 35-39.

Neuhauser, M., Losch, C., & Jockel, K-H. The Chen-Luo test in case of heteroscedasticity. *Computational Statistics and Data Analysis, 51,* 5055-5060.

Neyman, J. & Scott, E.L. (1971). Outlier proneness of phenomena and of related distributions. In J.S. Rustagi (Ed.), *Optimizing methods in statistics* (pp. 413-430). New York: Academic Press.

Pashley, P.J. (1993). On generating random sequences. In G. Keren & C. Lewis (Eds.), *Handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 395-415). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pratt, J.W. (1964). Robustness of some procedures for the two sample location problem. *Journal of the American Statistical Association, 59,* 665-680.

Wilcox, R.R. (2005). *Introduction to robust estimation and hypothesis testing.* New York: Academic Press.

Zaremba, S.K. (1965). Note on the Wilcoxon-Mann-Whitney statistic. *The Annals of Mathematical Statistics, 36,* 1058-1060.

Zimmerman, D.W. (1996). A note on homogeneity of variance of scores and ranks. (1996). *Journal of Experimental Education, 64,* 351-362.

(Manuscript received: 9 December 2009; accepted: 28 January 2010)