TECHNOMETRICS VOL. 6, NO. 3 AUGUST, 1964

Multiple Comparisons Using Rank Sums

OLIVE JEAN DUNN
University of California, Los Angeles

This paper considers the use of rank sums from a combined ranking of k independent samples in order to decide which populations differ. Such a procedure is suggested as a convenient alternative to making separate rankings for each pair of samples, and the two methods are compared. Asymptotic use of the normal tables is given and the treatment of ties is discussed. A numerical example is given.

1. INTRODUCTION

Rank sum methods have been widely used to compare two or more samples and decide whether or not they came from identically distributed populations. See, for example, [12] and [4].

Many research workers, however, are not satisfied with merely knowing that there are some differences among the k populations. They wish to know *which* populations differ. Some of these research workers may be unaware that simple ranking procedures exist for picking out the particular populations which are different.

Such a procedure was given in 1960 by Steel [9]. In his article, he gives tables to use for samples up to and including size six. For cases where it is appropriate to seek confidence intervals for the difference between populations rather than to pick out differences, Nemenyi [6] has described an adaption of Steel's procedure for this purpose.

The present paper suggests a slightly different ranking procedure from Steel's for picking out differences. This will be called Procedure I, and it is compared with Steel's procedure (Procedure II), on the basis of asymptotic theory.

Section 2 and Section 3 of this paper describe Procedures I and II, respectively. Section 4 gives a numerical example. Section 5 gives the mathematical justification of the procedures, and Section 6 discusses the matter of ties and the use of rank sums in contingency tables. Section 7 compares the two procedures from the standpoint of computation, and in Section 8 they are compared from the standpoint of probabilities of correct decisions on specified contrasts under two particular situations involving departure from the null hypothesis.

2. PROCEDURE I

The k samples are combined and then ranked from smallest to largest. When ties exist, they are given the average rank of the tied scores. Let T_i be the sum of the ranks of the ith sample.

Suppose that there are p contrasts among the means which the experimenter wishes to consider. He may wish, for example, to decide whether or not the first population has the same distribution as the second. To do this, he looks at $T_1/n_1 - T_2/n_2$, the difference in mean ranks. Or he might wish to combine his first, third, and fourth samples and decide whether the combined population from which they were drawn is the same as the population from which the fifth sample was drawn; for this he looks at $(T_1 + T_3 + T_4)/(n_1 + n_3 + n_4) - T_5/n_5$. Combining several samples in this manner is appropriate if the k samples were actually one large sample drawn from the combined k populations.

He calculates the value of the contrasts,

$$y_m = \sum_i T_i / \sum_i n_i - \sum_{i'} T_{i'} / \sum_{i'} n_{i'}, \qquad m = 1, \cdots, p, \tag{1}$$

where the summations over i and i' are over distinct subsets of the integers $1, \cdots, k$. A significance level α is selected. Each contrast y_m is then divided by its standard deviation σ_m. If no ties exist, then

$$\sigma_m^2 = [N(N + 1)/12][(\sum_i n_i)^{-1} + (\sum_{i'} n_i')^{-1}] \tag{2}$$

where $N = \sum_{i=1}^k n_i$. The formula for σ_m^2 must be adjusted if there are ties. If there are r groups of tied scores and if the sth group of tied scores has t_s numbers in it, then

$$\sigma_m^2 = \left[\frac{N(N+1)}{12} - \frac{\sum_{s=1}^{r} (t_s^3 - t_s)}{12(N - 1)}\right]\left[\frac{1}{\sum_i n_i} + \frac{1}{\sum_{i'} n_{i'}}\right] \tag{3}$$

The experimenter thus has p values: $y_1/\sigma_1, \cdots, y_p/\sigma_p$. Each of these values y_m/σ_m is then compared with $z_{1-\alpha/2p}$, the $1 - \alpha/2p$ point of the standard normal distribution.

If $y_m/\sigma_m < -z_{1-\alpha/2p}$, he decides that the population contrast is negative.

If $-z_{1-\alpha/2p} < y_m/\sigma_m < z_{1-\alpha/2p}$, he decides that the population contrast may be zero.

If $y_m/\sigma_m > z_{1-\alpha/2p}$, he decides that the population contrast is positive.

When the contrast is of the most usual form, $T_i/n_i - T_{i'}/n_{i'}$, these three decisions mean, respectively, that the mean (or median) of the ith distribution is less than the mean (or median) of the i'th distribution, that the two means may be the same, or that the ith mean is greater than the i'th mean. For several populations combined compared with several other populations combined, the decisions are the same concerning the combined populations.

If no differences exist among the k populations, then the probability of making one or more mistakes in decisions on contrasts is at most equal to α.

3. Procedure II

Procedure II is exactly the same as Procedure I except that now for each of the p comparisons, a separate ranking is made, using only those samples which are used in that particular contrast.

When two samples are ranked together, it is usual to look at the sum of ranks of the smaller sample. Here instead, we will use the difference of the two mean ranks, an equivalent statistic since the two ranks add to a constant. This is done for convenience, in order to be able to use exactly the same formulas for Procedures I and II.

Again one looks at y_m/σ_m , where

$$y'_m = \sum_i T_{im}/\sum_i n_i - \sum_{i'} T_{i'm}/\sum_{i'} n_{i'} , \qquad m = 1, \cdots , p \tag{4}$$

and

$$\sigma'^2_m = \left[\frac{N(N+1)}{12} - \frac{\sum_{s=1}^{r} (t_s^3 - t_s)}{12(N-1)}\right]\left(\frac{1}{\sum_i n_i} + \frac{1}{\sum_{i'} n_{i'}}\right) \tag{5}$$

Here, however, N is defined by $N = \sum_i n_i + \sum_{i'} n_i'$, so equals only the number of observations involved in that particular contrast instead of the number in all k samples, as it did in (2) and (3). The subscript m has been included on the rank totals T_i because they now differ from one contrast to another.

The value of y'_m/σ'_m is compared with $z_{1-\alpha/2p}$, exactly as in the first procedure.

4. A Numerical Example

The author is indebted to Dr. Roger Egeberg and Dr. Ann Elconin for data used to illustrate the methods. The data have been adapted from a study of a group of patients entering the Los Angeles County General Hospital during the years 1959–61. All the patients in the study had been judged medically eligible for being cared for at home under a home care program, and they were then classified into three groups, according to their home situation. The groups were:

Group 1—Eligible: Patients able to be cared for at home.

Group 2—No responsible person: Patients ineligible for home care because they had no person responsible for their care.

Group 3—Responsible person unable: Patients ineligible for home care because the person responsible was unable or unwilling to care for them.

The patients' occupations were recorded for 383 patients on a scale on which increasing numbers represent decreasing prestige levels. The data are given in Table 1.

It should be noted first that there are a great many ties, and that the distributions can hardly be claimed to be continuous. Some discussion of this point will be given in Section 6.

The contrasts which seemed of interest were:

1. Eligible for home care versus ineligible (group 1 versus groups 2 and 3).
2. Responsible person versus no responsible person (groups 1 and 3 versus group 2).
3. Responsible person able versus responsible person unable (group 1 versus group 3).

TABLE 1

Occupation and Home Care Eligibility for 383 Patients Medically Eligible for Home Care

Occupation Level*	(1) Eligible	(2) No Responsible Person	(3) Responsible Person Unable	All
10	3		1	4
20	12	4	2	18
30	10	7	4	21
40	20	10	11	41
50	47	9	10	66
60	74	12	21	107
70	62	26	38	126
Total	228	68	87	383

*Occupational classification:

10	executives, large proprietors, major professionals
20	business managers, medium proprietors, lesser professionals
30	administrators, small owners, semiprofessionals, farm owners
40	clerical, sales, technical, small business, small farm owner
50	skilled manual, small farm
60	semiskilled, tenant farmer
70	unskilled, share cropper

At the outset it was thought that the patients who had a responsible person able and willing to care for them at home (group 1) might differ somewhat in prestige of occupation from the patients who did not (groups 2 and 3). Also, it was anticipated that patients eligible for home care (group 1) might be on the average somewhat lower in prestige of occupation than those who had a responsible person unable or possibly unwilling to keep them at home (group 3).

The contrasts to be calculated are then:

$$y_1 = T_1/n_1 - (T_2 + T_3)/(n_2 + n_3) \tag{6}$$

$$y_2 = (T_1 + T_3)/(n_1 + n_3) - T_2/n_2 \tag{7}$$

$$y_3 = T_1/n_1 - T_3/n_3 \tag{8}$$

Here, $k = 3$, $p = 3$, $n_1 = 228$, $n_2 = 68$, and $n_3 = 87$.

Hand calculations of the rank totals T_i for Procedure I are given in Table 2. Table 3 gives the same calculation for the third contrast using Procedure II. Since the first two contrasts involve all three groups of patients, Procedure II was exactly the same as Procedure I for these two contrasts. Table 4 calculates the contrasts, their standard deviations with and without adjustment for ties, and gives the ratio of each contrast to its standard deviation.

If $\alpha = .20$ is used, then $z_{1-\alpha/2p} = z_{1-.20/6} = z_{.9667} = 1.834$ from univariate normal tables. Thus for either procedure one decides that μ_1 is less than μ_3. In other words, the patients whose responsible person could and would care for

TABLE 2

Calculation of Rank Totals, Procedure I

	(1)	(2)	(3)	(4)	(5)
Occupation	Cumulative		Rank times Frequency		
Level	Frequency	Rank	Group 1	Group 2	Group 3
10	4	2.5	7.5	0	2.5
20	22	13.5	162	54	27
30	43	33	330	231	132
40	84	64	1,260	640	704
50	150	117.5	5,522.5	1,057.5	1,175
60	257	204	15,096	2,448	4,284
70	383	320.5	19,871	8,333	12,179
		T_i =	42,249	12,763.5	18,503.5

(1) From the last column in Table 1.

(2) Calculated from the figures in column (1) as follows: 2.5 = (4 + 1)/2; 13.5 = 4 + (22 − 4 + 1)/2; 33 = 22 + (43 − 22 + 1)/2, etc.

Check: $\sum_{i=1}^{3} T_i = 73{,}536 = (383)(384)/2$

$T_1/n_1 = 42{,}249/228 = 185.4$
$T_2/n_2 = 12{,}763.5/68 = 187.7$
$T_3/n_3 = 18{,}503.5/87 = 212.7$

TABLE 3

*Calculation of Rank Totals, Procedure II Responsible Person Able vs. Responsible Person Unable**

	(1)	(2)		
	Cumulative			
Occupation	Frequency,		Rank times Frequency	
Level	Groups 1 and 3	Rank	Group 1	Group 3
10	14	2.5	7.5	2.5
20	18	11.5	138	23
30	32	25.5	255	102
40	63	48	960	528
50	120	92	4,324	920
60	215	168	12,432	3,528
70	315	265.5	16,461	10,089
		T_i =	34,577.5	15,192.5

*Other two contrasts include all 3 groups of observations, so calculations are the same as for Procedure I.

(1) From the sums of columns (1) and (3) in Table 2.

(2)Calculated from the figures in column (1) the same way as in Table 2.

Check: $T_{13} + T_{33} = 49,\ 770 = (315)(316)/2$
$T_{13}/n_1 = 34{,}577.5/228 = 151.7$
$T_{33}/n_3 = 15{,}192.5/87 = 174.6$

Table 4

*Calculation of y_m/σ_m by Two Procedures, with and without Adjustment for Ties**

Contrast	Coefficients of											
Number	T_1	T_2	T_3	y_m	$\sum_i n_i$	$\sum_{i'} n_{i'}$	N	$N(N+1)/12$	$\frac{1}{\sum n_i} + \frac{1}{\sum n_{i'}}$	σ_m^2	σ_m	y_m/σ_m
						Procedure I						
1	1/228	−1/155	−1/155	−16.3	155	228	383	12, 256	.01084	132.9	11.53	−1.41
								(11, 471)		(124.3)	(11.15)	(−1.46)
2	1/383	−1/68	+1/383	+ 5.2	68	315	383	12, 256	.01788	219.1	14.80	+ .35
								(11, 471)		(205.1)	(14.32)	(+ .36)
3	1/228		−1/87	−27.3	228	87	383	12, 256	.01588	194.6	13.95	−1.96
								(11, 471)		(182.2)	(13.50)	(−2.02)
						Procedure II						
1			Same as for Procedure I, since all groups included in these contrasts									−1.41
												(−1.46)
2												+ .35
												(+ .36)
3	1/228		−1/87	−22.9	228	87	315	8, 295	.01588	131.7	11.48	−1.99
								(7, 744)		(123.0)		(−2.06)

* In parentheses below unadjusted figures are given the adjusted ones. 12,256 − 785 = 11,471, where 785 = $(4^3+18^3+21^3+41^3+66^3+107^3+126^3) - (4 + 18 + 21 + 41 + 66 + 107 + 126)$.

8,295 − 551 = 7,744, where 551 = $(4^3 + 14^3 + 14^3 + 31^3 + 57^3 + 95^3 + 100^3) - (4 + 14 + 14 + 31 + 57 + 95 + 100)$.

them at home were in higher prestige occupations than those whose responsible person could not care for them at home (higher since a low rank indicated high prestige!), a conclusion not anticipated.

5. Mathematical Analysis

The null hypothesis to be tested is that the samples come from populations with identical, continuous distributions. It will be assumed in order to test H_0 that the k populations are identically distributed except for possibly differing in a location parameter.

Considering the samples being ranked as one large sample of size N, the rank assigned to the jth observation, say r_j, is well known to have expected value $E(r_j) = (N + 1)/2$ and variance $\sigma^2_{r_j} = (N^2 - 1)/12$. Further, the covariance between the ranks of the jth and j'th observations is Cov $(r_j, r_{j'}) = -(N + 1)/12$. Then T_i is the sum of n_i of the r_j's.

Thus $T_1, \cdots, T_k$ (or $T_{1m}, \cdots T_{km}$, for Procedure II) have means $n_i(N+1)/2$, where $N = \sum_{i=1}^{k} n_i$ is the size of the combined sample for Procedure I, or $N = \sum_i n_i + \sum_{i'} n_{i'}$ for Procedure II. If there are no ties, the variances of the T_i are equal to $(N + 1)(N - n_i)n_i/12$, and the covariances are equal to $-n_i n_{i'}(N + 1)/12$ (see, for example, Nemenyi [7]).

For samples sufficiently large asymptotic theory may be applied. This tells us that the T_i are approximately jointly normally distributed.

To establish the asymptotic joint normal distribution of the T_i, it is sufficient to establish that all linear combinations of the T_i are normally distributed (Anderson [1], p. 37). This will be shown using a theorem of Wald and Wolfowitz [11].

Let an arbitrary linear combination of the T_i be $z = \sum_{i=1}^{k} c_i T_i$. Since each T_i is the sum of n_i ranks, z is a linear combination of the N r_j, that is,

$$z = \sum_{j=1}^{k} c_j \left(\sum_{i=n_1+\cdots+n_{j-1}}^{n_1+\cdots+n_j} r_i \right). \tag{9}$$

Wald and Wolfowitz's theorem states that z (their L_N) is asymptotically normally distributed provided that two sequences satisfy their condition W. These two sequences (their A_N and D_N) are in this problem the sequences of positive integers $1, 2, \cdots, N$ and the sequence of coefficients consisting of n_1 c_1's, n_2 c_2's, and so forth (in other words, the coefficients of the r_i in z).

Condition W for a sequence H_N of real numbers $h_1, \cdots h_N$ is that for all integral $r > 2$,

$$\frac{\mu_r(H_N)}{[\mu_2(H_N)]^{r/2}} = 0(1), \tag{10}$$

where

$$\mu_r(H_N) = N^{-1} \sum_{\alpha=1}^{N} \left(h_\alpha - N^{-1} \sum_{\beta=1}^{N} h_\beta \right)^r. \tag{11}$$

Condition W is easily seen to be satisfied by the first N positive integers. For the sequence of coefficients $c_1, \cdots, c_1, c_2, \cdots, c_2, \cdots, c_k, \cdots, c_k$, it can be shown that condition W is satisfied provided as N approaches infinity, each n_i approaches $b_i N$, $i = 1, \cdots, k$, where b_i are non-negative constants which add to one.

Thus, the asymptotic distribution of the T_i is a multivariate normal distribution with means equal to $n_i(N + 1)/2$, variances equal to $(N + 1)(N - n_i)n_i/12$, and covariances equal to $-n_i n_{i'}(N + 1)/12$. Therefore, the usual methods for picking out individual contrasts whose means differ from zero apply, since one simply has k normally distributed variates. Tukey's method which uses the distribution of the range applies directly only when the n_i are all equal [10]. Scheffé's method [8] using the F distribution to obtain confidence intervals usually gives unduly long intervals, so that the choice in general falls to the use of the univariate normal distribution, with the level adjusted to give an overall test level of $1 - \alpha$. These methods, Tukey's and Scheffé's, and that using the univariate t distribution have been compared by Dunn [2] for confidence intervals, and the same considerations apply here in choosing among them.

As indicated earlier, p contrasts y_m are selected with a view to deciding which populations differ in the location parameter. If y_m is any linear combination of the T_i, say $y_m = \sum_i a_{mi} T_i$, then its variance σ_m^2 can be calculated from the variances and covariances of the T_i's. It is found to be

$$\sigma_m^2 = \frac{N(N+1)}{12}\left[\sum_{i=1}^{k} n_i a_{mi}^2 - \left(\sum_{i=1}^{k} n_i a_{mi}\right)^2\right]. \tag{12}$$

For $y_m = \sum_i T_i / \sum_i n_i - \sum_{i'} T_{i'} / \sum_{i'} n_{i'}$, where the i is summed over a subset of the integers $1, \cdots, k$ and i' is summed over a second subset, (12) reduces to

$$\sigma_m^2 = \frac{N(N+1)}{12}\left[\frac{1}{\sum_i n_i} + \frac{1}{\sum_{i'} n_{i'}}\right]. \tag{13}$$

This, in conjunction with the fact that y_m is approximately normally distributed, justifies the use of the $1 - \alpha/2p$ point from the normal tables on each of the p y_m/σ_m values; if H_0 is true and all the populations are identical, then the probability of making all p decisions correctly is at least $1 - \alpha$.

It should be mentioned that choosing the $1 - \alpha/2p$ point of the univariate normal is a conservative choice in order to obtain a $1 - \alpha$ overall level. Probably the $\frac{1}{2} + \frac{1}{2}(1 - \alpha)^{1/p}$ point could also be used, though proof of the necessary inequality has not been given in full generality for p larger than 3. For $\alpha = .05$, it makes little difference. However, for higher values of α, say $\alpha = .20$ or $.25$, there is some difference, and it would be helpful if the use of the $\frac{1}{2} + \frac{1}{2}(1 - \alpha)^{1/p}$ point could be substituted for the $1 - \alpha/2p$ point. See Dunn [3].

On the general subject of choice of α, I believe that in making multiple tests and comparisons, one might tend to use a value of α considerably larger than the traditional .05. The advantage of using the overall level rather than making p tests each at a .05 level, say, lies in being able to communicate one's results better with an overall level. And so it seems that there is usually no reason to choose the level so high that substantial differences become exceedingly difficult to establish.

6. Mathematical Analysis: Treatment of Ties

The treatment of ties seems of particular importance since the choice of rank sum methods often means that the distribution from which one is sampling is not continuous but instead involves a number, perhaps small, of ranked categories.

Kruskal and Wallis suggested in 1952 [4] that for any group of t_s tied observations, the s ranks be replaced by their mean. They note that the variance of an individual rank is decreased from $(N^3 - N)/12N$ to

$$[(N^3 - N) - \sum (t_s^3 - t_s)]/12N.$$

It can be shown directly in a similar manner that the covariance of r_i and $r_{i'}$ is increased from $-(N+1)/12$ to $-[(N^3 - N) - \sum_{s=1}^{r} (t_s^3 - t_s)]/12N(N-1)$. Then the variance of $y_m = \sum a_{mi} T_i$ can be calculated directly and is found to be

$$\sigma_m^2 = \left[\frac{N(N+1)}{12} - \frac{\sum_{s=1}^{r} (t_s^3 - t_s)}{12(N-1)}\right]\left[\sum_{i=1}^{k} n_i a_{mi}^2 - \left(\sum_{i=1}^{k} n_i a_{mi}\right)^2\right] \tag{14}$$

For $y_m = \sum_i T_i / \sum_i n_i - \sum_{i'} T_{i'} / \sum_{i'} n_{i'}$,

$$\sigma_m^2 = \left[\frac{N(N+1)}{12} - \frac{\sum_{s=1}^{r} (t_s^3 - t_s)}{12(N-1)}\right]\left[\frac{1}{\sum_i n_i} + \frac{1}{\sum_{i'} n_{i'}}\right] \tag{15}$$

This variance can also be written down as a particular case of the variance of the difference of two means from a finite population of size N, given in Nemenyi [7].

When as in the numerical example ties are the rule rather than the exception, one actually has k samples from k multinomial distributions, or else one large sample of size N from a multinomial distribution. The distribution of the rank sums T_i is then complicated. However, in the conditional distribution of the counts in all the cells with the number of the observations in each category fixed (that is, the number of "ties" fixed), the various probabilities in the multinomial distribution do not appear, under the null hypothesis. In the conditional distribution of counts, the rank sums may be calculated and used just as in the numerical example. The rank sum methods considered here are appropriate for use with contingency tables or with k samples from multinomial distributions. They are simply methods based on conditional distributions with fixed marginals.

It is interesting in the numerical example to notice how slight the adjustment is for ties. This is in line with the findings of Lehman [5] whose work was with very small samples. Using the normal approximation for large samples, the adjustment for ties, however, is easy to make. Without it, one is always being more conservative than is necessary.

7. Comparison of Methods: Computation

Procedure I can be justified on the basis of being a convenient approximation to Procedure II, as the most obvious comparison between the two methods of ranking is that the first is easier to do than the second. For machine users, either can be programmed, and the difference in the amount of work then becomes unimportant.

Frequently the comparisons are simply all $k(k - 1)/2$ of the form $T_i/n_i - T_{i'}/n_{i'}$. For either procedure, it has been found convenient in programming the procedures to have an option between a) using all $k(k - 1)/2$ contrasts of the form $T_i/n_i - T_{i'}/n_{i'}$ and b) reading in the desired contrasts. We have chosen as output to have, in either case, a table giving, for each contrast y_m , the coefficients of each rank total, the sample value of the contrast, its standard deviation, and the sample value divided by its standard deviation. If $y_m = \sum a_{mi}T_i$, the headings for the columns in the output are:

$$a_{m1} \quad a_{m2} \cdots \quad a_{mk} \quad y_m \quad \sigma_m \quad y_m/\sigma_m$$

The values in the last column may then easily be compared to $\pm z_{1-\alpha/2p}$.

8. Comparison of Methods: Probabilities of Correct Decisions when the Null Hypothesis is False

In both procedures the actual level of significance is unknown and is somewhat lower than the specified level, α. That is, the procedures are both conservative. This is for practical purposes unavoidable in considering multiple contrasts among means, since the actual level varies for different sets of contrasts. In this paper, the attitude is adopted that the actual level is of little importance; what really matters to the investigator are 1) knowing that the level of significance is bounded below by a certain set α and 2) the probabilities of making correct decisions on various types of contrast under various departures from the null hypothesis.

To compare the two procedures, one may look at easily handled departures from the null hypothesis and calculate the probabilities of making correct decisions by the two methods. In this section it will be assumed that the distributions are continuous, so that the probability of a tie is zero.

First, consider the possibility that the k populations have continuous distributions which do not overlap at all, and, to be specific, that $\mu_1 < \mu_2 < \cdots \mu_k$. Suppose that all $k(k - 1)/2$ contrasts of the form $\mu_i - \mu_{i'}$, $i > i'$, are being considered, and that the n_i are all equal, say, to n, so that $N = kn$.

Under Procedure I, $T_i = (i-1)n^2+n(n+1)/2$, so that $T_i/n-T_{i'}/n = (i-i')n$. The correct decision (that $\mu_i > \mu_{i'}$) is made provided

$$n > (i - i')^{-1} z_{1-\alpha/2p} \sqrt{k(kn + 1)/6} \tag{16}$$

With the second procedure, $T_{im}/n - T_{i'm}/n = n$, and the corret decision is reached provided

$$n > z_{1-\alpha/2p} \sqrt{2(2n + 1)/6} \tag{17}$$

The right hand side of (16) is larger than the right hand side of (17) provided

$$i - i' < \sqrt{k(kn + 1)/2(2n + 1)} \tag{18}$$

Thus, for any particular values of n and k, and $i-i' < \sqrt{k(kn + 1)/2(2n + 1)}$, if n is large enough so that a correct decision is always reached using the first procedure, then certainly a correct decision is reached using the second procedure, so that the second procedure is at least as good as the first. In the same way, for $i - i' > \sqrt{k(kn + 1)/2(2n + 1)}$, the first procedure is at least as good as the second.

In this extreme example, any reasonable person would naturally reject the null hypothesis with samples of even moderate size. The example may, nevertheless, yield a clue as to the difference between the two tests. As might be expected, using all k populations seems to be a help in detecting differences for those which are "far apart;" it is a hindrance for the "closer" populations. Here "far apart" is used to indicate that two population means are separated by relatively many other population means; "closer" indicates that two population means are separated by relatively few other population means.

Note also that in this example, the original populations were *not* assumed to be normally distributed (since they were nonoverlapping). The normal distribution has been used in evaluating probabilities of correct decisions because the T_i are asymptotically normal.

A second situation might be that $k - 1$ populations are identical, but that $\mu_k > \mu_i$, for $i = 1, \cdots, k - 1$, and that there is no overlap between the kth population and the first $k - 1$ populations. The distributions are continuous, but otherwise arbitrary. Again taking $N = kn$, under the null hypothesis $T_i/n_i - T_{i'}/n_{i'}$ would be approximately normally distributed, with mean zero and variance (using $N = kn$ in (2)) equal to $k(kn + 1)/6$. The correct decision is reached (that μ_i may equal $\mu_{i'}$) provided $T_i/n_i - T_{i'}/n_{i'}$ is less in absolute value than $z_{1-\alpha/2p} \sqrt{k(kn + 1)/6}$. Actually, however, the null hypothesis is false, and for $i, i' \neq k$, $T_i/n_i - T_{i'}/n_{i'}$ is approximately normally distributed with mean zero and variance equal to $(k - 1)[(k - 1)n + 1]/6$, so that the probability of a correct decision is seen to be

$$2\Phi[z_{1-\alpha/2p}\sqrt{k(kn + 1)/(k - 1)[(k - 1)n + 1]}] - 1,$$

where Φ is the c.d.f. of the standard univariate normal distribution.

With the second procedure, $T_{im}/n - T_{i'm}/n$ has mean zero and variance $\sqrt{2(2n + 1)/6}$, just as it does under the null hypothesis, so that the probability of a correct decision on $\mu_i - \mu_{i'}$ is $2\Phi(z_{1-\alpha/2p}) - 1$.

Since $\sqrt{k(kn + 1)/(k - 1)[(k - 1)n + 1]} > 1$ for all values of k and n under consideration, the probability of a correct decision is improved by using the combined ranking.

For $\mu_k - \mu_i$, $i \neq k$, for the first procedure one obtains the means and variances as follows. The quantity T_k is always equal to $n[(k - 1)n + (n + 1)/2]$, and so its variance is zero. For T_i, $i \neq k$, $n_i = n$, $N = (k - 1)n$ are substituted in $E(T_i) = n_i(N + 1)/2$ and in Var $T_i = (N + 1)(N - n_i)n_i/12$, to obtain $E(T_i) = [(k - 1)n + 1]n/2$ and Var $T_i = [(k - 1)n + 1][(k - 1)n - n]n/12$. The mean of $T_k/n - T_i/n$ reduces to $kn/2$, and its variance equals $[(k - 1)n + 1](k - 2)/12$. The probability of a correct decision ($\mu_k > \mu_i$) is then

$$1 - \Phi[(z_{1-\alpha/2p}\sqrt{k(kn + 1)/6} - kn/2)/\sqrt{[(k - 1)n + 1][k - 2]/12}] \qquad (19)$$

With the second procedure, $T_{km}/n - T_{im}/n$ has mean n and variance

$(2n + 1)/12$, so that the corresponding probability of a correct decision is

$$1 - \Phi[(z_{1-\alpha/2p}\sqrt{2(2n + 1)/6} - n)/\sqrt{(2n + 1)/12}]. \qquad (20)$$

It can be seen from (19) and (20) that for a given k, if n is large enough, the probability of a correct decision is higher with Procedure I than with Procedure II.

9. Conclusions

In the numerical example, the one contrast which differed in the two procedures had a slightly higher value of $|y_m/\sigma_m|$ under Procedure I; the difference was slight.

Tables for Procedure I for small samples would involve an additional parameter $(N - \sum_i n_i - \sum_{i'} n_{i'})$ so would be more difficult to produce except for equal sample sizes.

For hand computations, Procedure I is of course more convenient than Procedure II.

The existence of many ties does not invalidate the use of either procedure.

The special cases considered in Section 8 indicate that Procedure I may be somewhat better at picking out differences between populations which are close together, while Procedure II may be better at picking out differences between the more distant populations. Comparisons made between the methods under more usual situations might be illuminating.

In general, either procedure seems good where ranking methods are appropriate. Perhaps until further work is done comparing the two procedures, II may be preferred for small samples because of lack of available tables for I, and I may be recommended for larger samples when calculations are to be done by hand.

References

1. Anderson, T. W., 1958 *An Introduction to Multivariate Statistical Analysis,* John Wiley and Sons, Inc.
2. Dunn, Olive Jean, 1959. Confidence Intervals for the Means of Dependent, Normally Distributed Variables, *Jour. of the Amer. Stat. Assoc. 54,* 613–621.
3. Dunn, Olive Jean, 1958. Estimation of the Means of Dependent Variables, *Annals of Math Stat. 29,* 1095–1111.
4. Kruskal, Walter and Wallis, W. A., 1952. Use of Ranks in One-Criterion Variance Analysis, *Jour. of the Amer. Stat. Assoc. 47,* 583–621.
5. Lehman, Shirley Young, 1961. Exact and Approximate Distributions for the Wilcoxon Statistic with Ties, *Jour. of the Amer. Stat. Assoc. 56,* 293–298.
6. Nemenyi, Peter, 1962. Confidence Regions Bounded by Generalized Order Statistics, Mimeographed paper.
7. Nemenyi, Peter, 1961. Some Simple Approaches to Distribution-Free Multiple Comparisons, Mimeographed paper.
8. Scheffé, H., 1953. A Method for Judging all Contrasts in the Analysis of Variance, *Biometrika 40,* 87–104.
9. Steel, R. G. D., 1960. A Rank Sum Test for Comparing all Pairs of Treatments, *Technometrics 2,* 197–208.
10. Tukey, J. W., 1953. The Problem of Multiple Comparisons, Unpublished Dittoed Notes, Princeton University, 396 pp.
11. Wald, Abraham and Wolfowitz, Jacob, 1944. Statistical Tests based on Permutations of the Observations, *Annals of Math. Stat. 15,* 358–372.
12. Wilcoxon, F., 1949. Some Rapid Approximate Statistical Procedures, American Cyanamid Company.