INDIVIDUAL COMPARISONS BY RANKING METHODS

Frank Wilcoxon
American Cyanamid Co.

The comparison of two treatments generally falls into one of the following two categories: (a) we may have a number of replications for each of the two treatments, which are unpaired, or (b) we may have a number of paired comparisons leading to a series of differences, some of which may be positive and some negative. The appropriate methods for testing the significance of the differences of the means in these two cases are described in most of the textbooks on statistical methods.

The object of the present paper is to indicate the possibility of using ranking methods, that is, methods in which scores 1, 2, 3, . . . n are substituted for the actual numerical data, in order to obtain a rapid approximate idea of the significance of the differences in experiments of this kind.

Unpaired Experiments. The following table gives the results of fly spray tests on two preparations in terms of percentage mortality. Eight replications were run on each preparation.

Sample A Percent kill	Rank	Sample B Percent kill	Rank
68	12.5	60	4
68	12.5	67	10
59	3	61	5
72	15	62	6
64	8	67	10
67	10	63	7
70	14	56	1
74	16	58	2
Total 542	91	494	45

Rank numbers have been assigned to the results in order of magnitude. Where the mortality is the same in two or more tests, those tests are assigned the mean rank value. The sum of the ranks for B is 45 while for A the sum is 91. Reference to Table I shows that the probability of a total as low as 45 or lower, lies between 0.0104 and 0.021. The analysis of variance applied to these results gives an F value of 7.72, while 4.60 and 8.86 correspond to probabilities of 0.05 and 0.01 respectively.

Paired Comparisons. An example of this type of experiment is given by Fisher (2, section 17). The experimental figures were the differences in height between cross- and self-fertilized corn plants of the same pair. There were 15 such differences as follows: 6, 8, 14, 16, 23, 24, 28, 29, 41, —48, 49, 56, 60, —67, 75. If we substitute rank numbers for these differences, we arrive at the series 1, 2, 3, 4, 5, 6, 7, 8, 9, —10, 11, 12, 13, —14, 15. The sum of the nagative rank numbers is —24. Table II shows that the probability of a sum of 24 or less is between 0.019 and 0.054 for 15 pairs. Fisher gives 0.0497 for the probability in this experiment by the t test.

The following data were obtained in a seed treatment experiment on wheat. The data are taken from a randomized block experiment with eight replications of treatments A and B. The figures in columns two and three represent the stand of wheat.

Block	*A*	*B*	*A-B*	*Rank*
1	209	151	58	8
2	200	168	32	7
3	177	147	30	6
4	169	164	5	1
5	159	166	—7	—3
6	169	163	6	2
7	187	176	11	5
8	198	188	10	4

The fourth column gives the differences and the fifth column the corresponding rank numbers. The sum of the negative rank numbers is —3. Table II shows that the total 3 indicates a probability between 0.024 and 0.055 that these treatments do not differ. Analysis of variance leads to a least significant difference of 14.2 between the means of two treatments for 19:1 odds, while the difference between the means of A and B was 17.9. Thus it appears that with only 8 pairs this method is capable of giving quite accurate information about the significance of differences of the means.

Discussion. The limitations and advantages of ranking methods have been discussed by Friedman (3), who has described a method for testing whether the means of several groups differ significantly by calculating a statistic χ^2_r from the rank totals. When there are only two groups to be compared, Friedman's method is equivalent to the binomial test of significance based on the number of positive and negative differences in a series of paired comparisons. Such a test has been shown to have an efficiency of 63 percent (1). The present method for comparing the means of two groups utilizes information about the magnitude of the differences as well as the signs, and hence should have higher efficiency, but its value is not known to me.

The method of assigning rank numbers in the unpaired experiments requires little explanation. If there are eight replicates in each group, rank numbers 1 to 16 are assigned to the experimental results in order of magnitude and where tied values exist the mean rank value is used.

TABLE I

For Determining the Significance of Differences in Unpaired Experiments

No. of replicates	Smaller rank total	Probability for this total or less
5	16	.016
5	18	.055
6	23	.0087
6	24	.015
6	26	.041
7	33	.0105
7	34	.017
7	36	.038
8	44	.0104
8	46	.021
8	49	.050
9	57	.0104
9	59	.019
9	63	.050
10	72	.0115
10	74	.0185
10	79	.052

In the case of the paired comparisons, rank numbers are assigned to the differences in order of magnitude neglecting signs, and then those rank numbers which correspond to negative differences receive a negative sign. This is necessary in order that negative differences shall be represented by negative rank numbers, and also in order that the magnitude of the rank assigned shall correspond fairly well with the magnitude of the difference. It will be recalled that in working with paired differences, the null hypothesis is that we are dealing with a sample of positive and negative differences normally distributed about zero.

The method of calculating the probability of occurrence of any given rank total requires some explanation. In the case of the unpaired experiments, with rank numbers 1 to $2q$, the possible totals begin with the sum of the series 1 to q, that is, $q(p+1)/2$; and continue by steps of one up to the highest value possible, $q(3q+1)/2$. The first two and the last two of these totals can be obtained in only one way, but intermediate totals can be obtained in more than one way, and the number of ways in which each total can arise is given by the number of q-part partitions of T, the total in question, no part being repeated, and no part exceeding $2q$. These partitions are equinumerous with another set of partitions, namely the partitions of r, where r is the serial number of T in the possible series of totals beginning with 0, 1, 2, . . . r, and the number of parts of r, as well as the part magnitude, does not exceed q. The latter partitions can easily be enumerated from a table of partitions such as that given by Whitworth (5), and hence serve to enumerate the former. A numerical example may be given by way of illustration. Suppose we have 5 replications of measurements of two quantities, and rank numbers 1 to 10 are to be assigned to the data. The lowest possible rank total is 15. In how many ways can a total of 20 be obtained? In other words, how many unequal 5-part partitions of 20 are there, having no part greater than 10? Here 20 is the sixth in the possible series of totals; therefore $r = 5$ and the number of partitions required is equal to the total number of partitions of 5. The one to one correspondence is shown below:

Unequal 5-*part partitions of* 20	*Partitions of* 5
1-2-3-4-10	5
1-2-3-5-9	1-4
1-2-3-6-8	2-3
1-2-4-5-8	1-1-3
1-2-4-6-7	1-2-2
1-3-4-5-7	1-1-1-2
2-3-4-5-6	1-1-1-1-1

By taking advantage of this correspondence, the number of ways in which each total can be obtained may be calculated, and hence the probability of occurrence of any particular total or a lesser one.

The following formula gives the probability of occurence of any total or a lesser total by chance under the assumption that the group means are drawn from the same population:

$$P=2\left\{1+\sum_{i=1}^{i=r}\sum_{j=1}^{j=q}\sqcap_j^i-\sum_{n=1}^{n=r-q}\left[(r-q-n+1)\sqcap_{q-1}^{q-2+n}\right]\right\}\Big/\frac{\lfloor 2q}{\lfloor q\times\lfloor q}$$

$\sqcap_j^i$ represents the number of j-part partitions of i,
r is the serial number of possible rank totals, 0, 1, 2, $\cdots$ r.
q is the number of replicates, and
n is an integer representing the serial number of the term in the series.

In the case of the paired experiments, it is necessary to deal with the sum of rank numbers of one sign only, + or —, whichever is less, since with a given number of differences the rank total is determined when the sum of + or — ranks is specified. The lowest possible total for negative ranks is zero, which can happen in only one way, namely, when all the rank numbers are positive. The next possible total is —1, which also can happen in only one way, that is, when rank one receives a negative sign. As the total of negative ranks increases, there are more and more ways in which a given total can be formed. These ways for any totals such as —r, are given by the total number of unequal partitions of r. If r is 5, for example, such partitions, are 5, 1-4, 2-3. These partitions may be enumerated, in case they are not immediately apparent, by the aid of another relation among partitions, which may be stated as follows:

The number of unequal j-part partitions of r, with no part greater than i, is equal to the number of j-part partitions of $r-\binom{j}{2}$, parts equal or unequal, and no part greater than $i-j+1$ (4).

For example, if r equals 10, j equals 3, and i equals 7, we have the correspondence shown below:

Unequal 3-part partitions of 10	1-2-7 1-3-6 1-4-5 2-3-5
3-part partitions of 10—3, or 7, no part greater than 5	1-1-5 1-2-4 1-3-3 2-2-3

The formula for the probability of any given total r or a lesser total is:

$$P=2\left[1+\sum_n\left(\sum_{i=n}^{i=r-\binom{n}{2}}\sqcap_n^i\right)\right]\Big/2^q$$

r is the serial number of the total under consideration in the series of possible totals

0, 1, 2, . . ., r,

q is the number of paired differences.

In this way probability tables may be readily

TABLE II

For Determining the Significance of Differences in Paired Experiments

Number of Paired Comparisons	Sum of rank numbers, + or —, whichever is less	Probability of this total or less
7	0	0.016
7	2	0.047
8	0	0.0078
8	2	0.024
8	4	0.055
9	2	0.0092
9	3	0.019
9	6	0.054
10	3	0.0098
10	5	0.019
10	8	0.049
11	5	0.0093
11	7	0.018
11	11	0.053
12	7	0.0093
12	10	0.021
12	14	0.054
13	10	0.0105
13	13	0.021
13	17	0.050
14	13	0.0107
14	16	0.021
14	21	0.054
15	16	0.0103
15	19	0.019
15	25	0.054
16	19	0.0094
16	23	0.020
16	29	0.053

prepared for the 1 percent level or 5 percent level of significance or any other level desired.

Literature Cited

1. Cochran, W. G., The efficiencies of the binomial series tests of significance of a mean and of a correlation coefficient. Jour. Roy. Stat. Soc., 100:69-73, 1937.
2. Fisher, R. A. The design of experiments. Third ed., Oliver & Boyd, Ltd., London, 1942.
3. Friedman, Milton. The use ranks to avoid the assumption of normality. Jour. Amer. Stat. Assn. 32:675-701. 1937.
4. MacMahon, P. A. Combinatory analysis, Vol. II, Cambridge University Press. 1916.
5. Whitworth, W. A. Choice and chance, G. E. Stechert & Co., New York, 1942.

TEACHING AND RESEARCH AT THE STATISTICAL LABORATORY, UNIVERSITY OF CALIFORNIA

The Statistical Laboratory of the University of California, Berkeley, was established in 1939 as an agency of the Department of Mathematics. The functions of the Laboratory include its own research, help in the research carried on in other institutions, and a cycle of courses and of exercises for students.

The outbreak of the war soon after the establishment of the Statistical Laboratory influenced the direction of its research to a considerable extent. War problems were studied unofficially first and then under a contract with the National Defense Research Committee. These activities trained a considerable number of persons who were later absorbed by the Services. Also, the Laboratory acquired an efficient set of computing machines and other equipment.

In the future, the Laboratory's own research will be concerned with developing statistical techniques on one hand and with analyses of applied problems on the other.

Cooperation with other institutions is based on the principle of free choice and, therefore, care is taken to avoid anything suggesting a tendency to centralize statistical research in the Laboratory of the Department of Mathematics to the exclusion or the restraint of such research in other Departments of the University. The Statistical Laboratory has a hand in pieces of research for which the experimenter requests statistical help. The help offered consists primarily of advice. However, in cases where the numerical treatment of the problem is complex, computations are performed in the Laboratory.

Because of the voluntary basis of cooperation, contacts with institutions of experimental research are not systematic. The closest and most fruitful contacts thus far have been with the California Forest and Range Experiment Station located on the Berkeley Campus. Perhaps unexpectedly, in addition to practical results, these contacts originated a purely theoretical paper in the Annals of Mathematical Statistics. Other theoretical papers are expected because it appears that forest and range experimentation involves specific and difficult problems which require new statistical methods. Also, in contact with the Department of Entomology, some very interesting problems were found. The work done on them is also reflected in some publications.

The teaching in the Statistical Laboratory is geared to train research workers and teachers of statistics. The cycle of courses and of laboratory work now offered is essentially that planned in 1939, but some changes are under consideration. Both the original set-up and the reforms considered are based on the following premises.

The first and generally admitted premise is that a university teacher must be a research worker, that is to say, must be effectively capable of inventive work. As applied to statistics such inventive work may be of two kinds. First, inventiveness may express itself in developing new sections of the mathematical theory of statistics. In the present state of our science this requires not only the knowledge of the existing theory of statistics but also a considerable mastery of the theory of functions and other branches of mathematics. Next, the inventiveness may concern the techniques in statistical design of experiments or of observation in relation to the already existing statistical tools. Here the success of the research depends on a thorough knowledge of the tools and also of the particular domain in which they have to be used.

It is obvious that proficiency in the first of these items is especially desirable for a teacher