TI-2302

4D 17 115

WADC TECHNICAL REPORT 52-32

THE STATISTICAL ANALYSIS OF RANKED DATA

PAUL R. RIDER, PH.D.
FLIGHT RESEARCH LABORATORY

AD-A075 994

FEBRUARY 1952

WRIGHT AIR DEVELOPMENT CENTER

20011221164

NOTICES

WADC TECHNICAL REPORT 52-32

THE STATISTICAL ANALYSIS OF RANKED DATA

Paul R. Rider, Ph.D.
Flight Research Laboratory

February 1952

RDO No. 460-51

Wright Air Development Center
Air Research and Development Command
United States Air Force
Wright-Patterson Air Force Base, Ohio

FOREWORD

This report was prepared for the Statistics Research Team, Flight Research Laboratory, Wright Air Development Center by Dr. Paul R. Rider, the project engineer, under RDO No. 460-51, The Statistical Analysis of Ranked Data.

ABSTRACT

Explanations are given of what is meant by ranked data. Questions of rank correlation and concordance are discussed. Coefficients of correlation and concordance are defined and methods of testing them for significance are described and illustrated.

PUBLICATION REVIEW

For the Commanding General:

Leslie B. Williams

LESLIE B. WILLIAMS
Lt. Colonel, USAF
Chief, Flight Research Laboratory
Research Division

TABLE OF CONTENTS

INTRODUCTION

Many of the problems about which the Statistics Research Team is consulted involve ranked data. These problems have to do with such diverse things as photographic film, binoculars used in reconnaissance, coffee tasting, recipe testing, food preferences, and the findings of officers' rating and promotion boards.

The question of how to handle ranked statistical data therefore seems of sufficient importance to warrant the publication of a Technical Report on the subject.

The object of this report is to discuss and explain, in as elementary a manner as is possible with material that is somewhat technical, methods of analysis that are customarily employed in dealing with ranked data. The questions of rank correlation and of concordance of judgment in ranking are discussed. In particular, methods are given for testing whether correlation or concordance that has been found is significant.

SECTION I
RANK CORRELATION

1. Rank. To *rank* a set of objects is to arrange them in order with respect to some characteristic. The set of objects could, for example, be a group of men and the characteristic could be height. When the men are arranged in order of height, the tallest is assigned the rank 1, the next tallest the rank 2, and so on. In this case the characteristic is a measurable one, and the ranking is merely a transformation of variables.

There is usually a distortion in such a transformation. Thus, consider four men whose heights are 6 feet, 2 inches; 6 feet; 5 feet, 11 inches; and 5 feet, 7 inches respectively. The differences in height between consecutive men are 2 inches, 1 inch and 4 inches. Yet the difference in rank between any two consecutive men is 1. (It is assumed for the present that there are no ties in rank.)

It is seen that in the case of a measurable characteristic, rank is a rather rough way of assigning a numerical value to the degree in which the characteristic is possessed. However, there are certain advantages in using ranks. One of these is that the numbers involved in statistical computations and analyses are usually simpler. Another is that sometimes a set of numerical data will be dominated by one or two large items, whereas if the items are ranked the undue influence of these items is eliminated. (See Kendall, Rank Correlation. pp. 14-15.)

It is frequently possible to rank objects according to some characteristic which is difficult or even impossible to measure. Individuals can be ranked according to intelligence or personality, manufactured articles can be ranked according to beauty of design, aircraft can be ranked according to performance or efficiency. Some of the characteristics just mentioned are too vague to allow of measurement, yet they do permit ranking.

2. Measuring Ability to Rank Correctly. We may at times want to measure the ability of an individual to make judgments of a certain type by ranking a set of objects. For example, suppose that there are four objects of the same size and shape but of different weights and that a person attempts to rank them. If his ability to arrange them in the correct order is to be measured, it would seem natural that he should receive the highest possible score if he ranks them in the correct order and the lowest possible score if he arranges them in the reverse order. Any other ordering (that is, ranking) should give him an intermediate score.

In order to develop a measure of ranking ability let us consider a concrete case.

Suppose that a set of four objects has been placed in the order 2314 instead of the correct or natural order 1234. We consider the number of pairs of ranks in the actual ranking which are in natural order and the number of pairs of ranks which are in inverted order. Let us score the pairs as in Table 1. When a pair of ranks is in the natural order, we place a 1 in the natural order column, labeled P; when a pair of ranks is in the inverted order, we place a 1 in the inverted order column labeled Q. In the score column, labeled S_τ (the subscript is used to avoid confusion with another S which will be used later), we place a + 1 for each 1 in the P column and a − 1 for each 1 in the Q column. It follows that

$$S_\tau = P - Q. \tag{1}$$

It is not necessary to construct a table. It has been constructed here for the purpose of explaining the method of scoring. When this method is understood, the value of P or Q or S_τ can be found quite quickly after a slight amount of practice.

TABLE 1

Pair of ranks	Natural order P	Inverted order Q	Score S_τ
23	1		+ 1
21		1	− 1
24	1		+ 1
31		1	− 1
34	1		+ 1
14	1		+ 1
Total	4	2	2

The following fact should be noted, as it is very helpful in calculating the score. If n is the number of objects ranked, then the number of pairs of ranks is $\frac{1}{2}n(n-1)$. Consequently,

$$P + Q = \tfrac{1}{2}n(n-1) \tag{2}$$

and if either P or Q has been found then the other can be found at once as can S_τ.

Often the easiest procedure is first to find Q by counting the number of inverted pairs.

3. Measuring Agreement between Two Rankings. The preceding method can be used to measure the agreement of two rankings. Suppose for example that two persons, X and Y, rank the same set of five objects, a, b, c, d, e. Suppose that X ranks them in the order c, e, a, d, b and that Y ranks them in the order a, c, e, b, d. We regard one of the rankings (it does not matter which) as standard and compare the other ranking with it. If we take X's ranking as the standard, the situation can be exhibited as follows:

Object:	c	e	a	d	b
X's ranking:	1	2	3	4	5
Y's ranking:	2	3	1	5	4

The rest of the procedure is as before. Here let us list and count the number of inverted pairs in Y's ranking. They are 21, 31, 54; hence $Q = 3$. Since the number of objects ranked is $n = 5$, we find from (2) that $P + 3 = \frac{1}{2} \times 5 \times 4 = 10$, or $P = 7$, from which it follows from (1) that $S_\tau = 7 - 3 = 4$.

4. Kendall's Coefficient of Rank Correlation (τ). It is easily seen that the score S_τ, discussed above, is dependent upon n. For this reason it is a somewhat vague measure of the agreement of one ranking with another. Thus, for example, if four objects are being ranked, the maximum value that S_τ can have is 6. This would indicate perfect agreement in ranking. On the other hand, a value of 6 for S_τ might indicate very poor agreement if a larger number of objects were being ranked.

Consequently it is desirable to have a coefficient which will have the value $+1$ when two rankings are in perfect agreement and the value -1 when one ranking is exactly the reverse of the other. Such a coefficient is *Kendall's coefficient of rank correlation*,

$$\tau = \frac{S_\tau}{\frac{1}{2}n(n-1)}. \tag{3}$$

(The denominator is the maximum value that S_τ can have in the case of n ranks.) Equivalent expressions for τ are the following:

$$\tau = \frac{2P}{\frac{1}{2}n(n-1)} - 1, \tag{4}$$

$$\tau = 1 - \frac{2Q}{\frac{1}{2}n(n-1)}, \tag{5}$$

where P and Q have been defined earlier. In the above example τ has the value $2/5 = 0.4$.

The coefficient τ is a measure of the correlation or agreement of any ranking with a standard ranking or of two rankings with each other. It may also be used to measure the correlation between two characteristics when the same set of objects has been ranked with respect to both of these characteristics. For example, suppose that a group of ten men have been ranked with respect to initiative and also with respect to reliability. The value of τ can of course be calculated; it measures the correlation between the two traits or characteristics in this group of ten men.

5. Spearman's Coefficient of Rank Correlation (ρ). Since it is frequently referred to, we shall discuss briefly at this point another coefficient. This is *Spearman's coefficient of rank correlation*. It is designated by ρ and is calculated as in the following example.

Consider the rankings used in a previous illustration:

Object:	a	b	c	d	e
X's ranking:	3	5	1	4	2
Y's ranking:	1	4	2	5	3
Difference, d	2	1	-1	-1	-1
d^2	4	1	1	1	1

If the sum of squares of the differences is denoted by $\Sigma\, d^2$, then Spearman's coefficient is defined by the equation

$$\rho = 1 - \frac{6\,\Sigma\, d^2}{n(n+1)(n-1)}, \qquad (6)$$

in which n, as usual, denotes the number of objects ranked. In the present example,

$$\rho = 1 - \frac{6 \times 8}{5 \times 6 \times 4} = \frac{3}{5} = 0.6$$

For the same set of rankings the value of τ was calculated to be 0.4. In general the values of τ and ρ will not be the same. Spearman's coefficient is merely the ordinary Pearsonian coefficient of correlation between two rankings. Like Kendall's coefficient, it has the value $+1$ when the rankings are in perfect agreement and the value -1 when the rankings are in perfect disagreement, that is, when one ranking is exactly the reverse of the other. On the whole, Kendall's coefficient is to be preferred.

6. Ties in Ranks. It is readily realized that ties in rank may sometimes occur, since two or more objects may possess a certain characteristic in exactly the same degree or in indistinguishable degrees. In the case of ties some adjustments are necessary in dealing with correlation.

In the first place, the convention which we shall adopt in assigning ranks in the case of a tie is an averaging process. For example, if it is impossible to distinguish between the objects which would be ranked fourth and fifth, the average rank of 4½ will be assigned to each. If four objects are tied for ranks 2, 3, 4, 5, the rank assigned to each will be the average $(2 + 3 + 4 + 5)/4 = 3\frac{1}{2}$ (The same result is obtained if the first and last ranks, namely 2 and 5, are averaged.)

7. Calculation of S_τ and τ for Tied Ranks. When there are ties in ranks it becomes impossible to use the formulas previously given for S_τ and τ. Consequently, new definitions for these two functions will be given. The new definitions will yield the same results as the old ones for the case in which there are no ties.

In order that the meanings of the definitions may be clear let us consider an example of tied rankings. Suppose that six objects, *a, b, c, d, e, f,* are ranked by *X* and *Y* as follows:

X's ranking: (*c, d* tied), *b,* (*a, e, f* tied)

Y's ranking: *d, c,* (*b, f,* tied), *a, e*

Thus we have

Object:	*a*	*b*	*c*	*d*	*e*	*f*
i and *j*:	1	2	3	4	5	6
X's ranking:	5	3	1½	1½	5	5
Y's ranking:	5	3½	2	1	6	3½

The row labeled '*i* and *j*' gives the number of each object. It is placed in the foregoing scheme for later use in explanation and calculation. Incidentally, these objects may be arranged in any order. Here, since they are represented by letters of the alphabet, it seemed natural to arrange them in alphabetical order.

We now define the quantity x_{ij} $(i > j)$ to be $+1$ if *X*'s ranking ot the *i*th object is less than his ranking of the *j*th object, -1 if his ranking of the *i*th object is greater than his ranking of the *j*th object, and 0 if his ranking of the *i*th object is the same as his ranking of the *j*th object. For illustration $x_{12} = -1$, since his ranking of object number 1 (namely 5) is greater than his ranking of object number 2 (namely 3); $x_{34} = 0$, since *X*'s rankings of the 3rd and 4th objects are the same (namely 1½); $x_{36} = +1$, since the rank 1½ is less than the rank 5.

The quantity y_{ij} is similarly defined. Then

$$S_\tau = \Sigma\, x_{ij}\, y_{ij} \tag{7}$$

and

$$\tau = \frac{\Sigma\, x_{ij}\, y_{ij}}{\sqrt{\Sigma\, x_{ij}^2}\ \sqrt{\Sigma\, y_{ij}^2}} \tag{8}$$

It will be noted that the numerator of τ is S_τ.

The calculation of τ for the foregoing example is shown in Table 2.

TABLE 2

$i\ j$	x_{ij}	y_{ij}	$x_{ij}\ y_{ij}$	$i\ j$	x_{ij}	y_{ij}	$x_{ij}\ y_{ij}$
12	− 1	− 1	1	26	1	0	0
13	− 1	− 1	1	34	0	− 1	0
14	− 1	− 1	1	35	1	1	1
15	0	1	0	36	1	1	1
16	0	− 1	0	45	1	1	1
23	− 1	− 1	1	46	1	1	1
24	− 1	− 1	1	56	0	− 1	0
25	1	1	1	Total (entire table)			10

It is readily seen that

$$\Sigma\, x_{ij}^2 = 11,\ \Sigma\, y_{ij}^2 = 14,\ \Sigma\, x_{ij}\, y_{ij} = 10\ (= S_\tau)$$

and, from (8),

$$\tau = \frac{10}{\sqrt{11}\ \sqrt{14}} = 0.81.$$

As was stated earlier, if there are no ties in ranks the value of S_τ calculated according to the new definition will be identical with that calculated according to the first definition. Moreover, when there are no ties, both $\Sigma\, x_{ij}^2$ and $\Sigma\, x_{ij}^2$ reduce to $\frac{1}{2}n\,(n - 1)$, so that the product of the square roots of these quantities has this value, and we are led to formula (3). The reader will find it instructive to calculate S_τ and τ for the example of section 3, employing the new method.

8. **Alternative Formula for τ.** As was stated in the preceding section, when there are no ties, both $\Sigma\, x_{ij}^2$ and $\Sigma\, y_{ij}^2$ $(i > j)$ reduce to $\frac{1}{2}n\,(n - 1)$. This is readily seen, since in this case each x_{ij} and y_{ij} will be either + 1 or − 1. However, if a tie of t objects exists in X's rankings, then each corresponding x_{ij} will be 0. Now for t objects there are (since we are taking $i > j$) exactly $\frac{1}{2}\,t\,(t - 1)$ such values of x_{ij}. Thus, the total number of zero values for x_{ij} is $\frac{1}{2}\,\Sigma\ t\ (t - 1)$. For instance, if 2 objects are tied in rank, also 3 others, then

$$\tfrac{1}{2} \Sigma\, t\, (t - 1) = \tfrac{1}{2} \times 2 \times 1 + \tfrac{1}{2} \times 3 \times 2 = 4$$

We may therefore write

$$\Sigma\, x_{ij}^2 = \tfrac{1}{2} n\, (n - 1) - \tfrac{1}{2} \Sigma\, t\, (t - 1) \qquad (9)$$

Similarly, if we use u to denote ties in Y's ranking, we have the corresponding formula

$$\Sigma\, y_{ij}^2 = \tfrac{1}{2} n\, (n - 1) - \tfrac{1}{2} \Sigma\, u\, (u - 1). \qquad (10)$$

Consequently the denominator in (8) may be replaced by the product of the square roots of the right-hand members of (9) and (10).

SECTION II
SIGNIFICANCE TESTS FOR RANK CORRELATION

9. Significance Tests for Rank Correlation. Let us consider a set of objects which possess a certain characteristic or quality in different degrees, so that they have an inherent ranking. Suppose that a person attempts to rank them. He will make a certain score S_τ. What information does this score yield concerning the ability of the person to judge this particular quality? Might he not have achieved a score this high or higher simply by ranking the objects at random?

If two persons are ranking the same set of objects, does a certain score S_τ really indicate that they are in substantial agreement or might not the score, or a higher one, have occurred purely as a matter of chance?

If a certain value has been found for the correlation coefficient τ, can we conclude that there is actually some correlation between the two characteristics being investigated? Perhaps the value is so large that correlation is unmistakably indicated *for this given set of objects*. However, this set may be regarded as a sample from a larger 'population' of similar objects. Another sample would doubtless yield a different value of τ. Therefore, can the value actually found be interpreted as indicating the existence of correlation in the population, or might not a value this large or larger happen fairly often in samples?

Questions such as the foregoing suggest the desirability of having some means of testing how unusual an observed score or coefficient is. If a certain score is unusual, we may say that it is *significant*, meaning that its value is decidedly different from what is to be expected as a matter of chance. A test which tells how unusual an observed value is, is called a *significance test*.

The following considerations may throw some light on the meaning of a significance test as well as showing how such a test may sometimes be devised.

For four objects, a, b, c, d, there are 24 possible rankings. (For n objects there are $n!$ possible rankings.) These 24 possibilities, together with the corresponding scores, are shown in Table 3.

TABLE 3

Ranking	S_τ	Ranking	S_τ	Ranking	S_τ
1234	6	2314	2	3412	−2
1243	4	2341	0	3421	−4
1324	4	2413	0	4123	0
1342	2	2431	−2	4132	−2
1423	2	3124	2	4213	−2
1432	0	3142	0	4231	−4
2134	4	3214	0	4312	−4
2143	2	3241	−2	4321	−6

The information given in Table 3 is summarized in Table 4. In the first column of the latter are listed the various values of S_τ. In the second column are listed the corresponding values of τ. We find from (3), since here $n = 4$, that each value of τ is 1/6 of the corresponding value of S_τ.

The third column gives the frequency of occurrence of the values of S_τ and τ. In the fourth column these frequencies have been converted to probabilities by division by 24. Thus, for example, the 0.125 on the same line with the −4 in the S_τ column means that there are 125 chances in 1,000 (that is, 1 chance in 8) of obtaining a score of −4 (or a value of τ equal to −0.67) if four objects are ranked by some purely random process.

The final column is the cumulative probability, that is, the probability of obtaining a value of S_τ or τ *as large as or larger than* that shown in the same line. For example, the probability of obtaining a value of S_τ equal to or greater than 2 (or a value of τ

TABLE 4

S_τ	τ	Frequency	Probability	Cum. Prob.
− 6	− 1	1	0.042	1.000
− 4	− 0.67	3	0.125	0.958
− 2	− 0.33	5	0.208	0.833
0	0	6	0.250	0.625
2	0.33	5	0.208	0.375
4	0.67	3	0.125	0.167
6	1	1	0.042	0.042
Total		24	1.000	

equal to or greater than 0.33) is 0.375. The probability of a value of S_τ equal to or greater than 2 in absolute value (or a value of τ equal to or greater than 0.33 in absolute value) is 0.375 + 0.375, or 0.750.

10. Meaning of a Significance Test. The average score and the average value of τ in Table 4 are 0. This is to be expected, since a person possessing no ability to judge a certain characteristic would obtain positive and negative scores having the same numerical value with about equal frequency. Likewise, if two persons are ranking the same set of things and each is performing the ranking by some random process, they will be in disagreement just about as often as they are in agreement. Furthermore, if samples are taken from a population in which no correlation exists between two characteristics, then it would seem rather reasonable to find that the value of either S_τ or τ turns out to be 0 on the average.

The more that a value of S_τ or τ deviates from 0, the less likely is this value to occur. If the probability of obtaining a score equal to or greater than a specified value S_τ is p, then S_τ is said to be *significant at the level* p. Similarly, if the probability of obtaining a correlation coefficient equal to or greater than τ is p, then τ is said to be significant at the level p. Since the values of S_τ and τ are symmetrically distributed, when a value of either is significant at the level p, then the corresponding absolute value is significant at the level $2p$. Thus, from Table 4, it is seen that the significance level of the value 6 for S_τ (or of 1 for τ) is 0.042, or 4.2%. The significance level of $|S_\tau| = 6$ (or of $|\tau| = 1$) is twice 4.2%, or 8.4%.

A significance test involving absolute values is often called a *two-sided test*, one involving algebraic values is called a *one-sided test*. Care must be taken to note which kind of test is being used.

The significance levels most frequently used in practice are 5% and 1%. It is customary to say that a value which is significant at the 5% level is *significant*, and that a value which is significant at the 1% level is *highly significant*. These levels and these terms are entirely arbitrary, however.

11. Tables for Testing Significance of Rank Correlation. Tables such as Table 4, showing the frequency or probability distribution of S_τ and τ are easily constructed. However, it is possible to make use of existing tables such as those found in [1]* vol. 1, pages 404-405; [2], page 141; and [3], pages 620-621. The last reference gives the distribution of Q, the number of inversions in rank, which is quite equivalent to the distribution of S_τ.

The tables referred to above extend only as far as $n = 10$, that is, they can be used only in the case of 10 or fewer rankings. It can be shown that, as n increases, the distribution of S_τ approaches a normal distribution with mean 0 and variance

$$n\ (n - 1)(2n + 5)/18. \tag{11}$$

When n is greater than 10 we assume that S_τ is distributed in this way and make use of a table of the normal probability integral. When the normal distribution is used in testing a value of S_τ we make a so-called *correction for continuity*, which consists in subtracting 1 from S_τ.

As an illustration, suppose that a value of $S_\tau = 55$ has been obtained in the ranking of 14 objects. According to (11) the variance of S_τ is $14 \times 13 \times 33/18 = 333.67$. The standard deviation of S_τ is the square root of this value, namely 18.3. We make the correction for continuity and calculate the quantity $x = (55 - 1)/18.3 = 2.95$. Considering this value as a normal deviate with unit standard deviation, we find, from tables of the normal probability integral, that the probability of an absolute value this large or larger is 0.0032. That is, such values would happen, as a matter of pure chance, only about 32 times out of 10,000. The value therefore is very significant. Stating the matter in a slightly different form, if we hypothesize that the ranking of these 14 objects was done by a purely random process, then a value of S_τ as large as or larger than the one observed, namely 55, would cause us to reject the hypothesis.

It may be noted at this point that the distribution of τ also tends to normality with increasing n.

* Numbers in square brackets refer to the corresponding numbers in the Bibliography at the end of this report.

12. Significance Tests for Rank Correlation When There Are Tied Ranks. According to Kendall ([2], page 43), the distribution of τ for any fixed number of ties approaches normality with increasing n, and it is usually permissible to employ the normal approximation when $n \geqq 10$, although when many ties exist a special consideration may be necessary. For $n \leqq 10$, the tables of Sillito [4] will be found useful.

13. Tests When Correlation Exists in the Population. It has been pointed out (section 9) that when a measure of the rank correlation between two characteristics has been calculated for a given set of objects, this set may be regarded as a sample from a 'population' of similar objects. This population will have a value of the rank correlation coefficient, let us call it T, which may or may not be zero. We may wish to test whether the value of τ observed in a sample deviates significantly from the population value T.

For definiteness let us consider a population of five objects having inherent rankings according to two different characteristics. Suppose that when the objects are arranged according to these characteristics the situation is as follows.

Rank according to 1st characteristic:	1	2	3	4	5
Rank according to 2nd characteristic:	1	5	3	2	4

Suppose that we take a sample of three from this population, for example,

2	3	5
5	3	4

In this sample the value of S_τ is found to be -1 and the value of τ is $-1/3$. Now the algebra of combinations tells us that the number of possible samples of three from this population of five is 10. Since, in any sample, the ranking according to the first characteristic will always be in natural order, it is necessary to consider the ranking according to the second characteristic only. In Table 5 the 10 possible samples are listed (second ranking only), together with the corresponding values of S_τ and τ.

The mean value of τ is readily found to be $\bar{\tau} = 1/5$. The value of the population correlation coefficient, T, is also 1/5, and it can be shown that in general the mean value of τ in samples is always equal to T. A similar statement cannot be made above the average value of S_τ in samples, however. Nor can much be said about the variance of τ in samples except that it can never exceed

$$2\,(1 - T^2)/n \tag{12}$$

where n is the number of rankings in the sample. Practically all that can be done is to assume that τ is normally distributed with mean T and with standard deviation equal

TABLE 5

Sample	S_τ	τ	Sample	S_τ	τ
153	1	1/3	124	3	1
152	1	1/3	532	- 3	- 1
154	1	1/3	534	- 1	- 1/3
132	1	1/3	524	- 1	- 1/3
134	3	1	324	1	1/3

to the square root of the expression (10). This gives a conservative test in the sense that will be explained after an example. If n is less than 10 there seems to be no good test available.

As an illustration of testing the significance of an observed value of Kendall's coefficient, let us suppose that the value $\tau = 0.82$ has been calculated from a sample of 15. Can the sample be regarded as having come from a population in which the correlation coefficient is $T = 0.50$?

Using (12) we calculate the maximum value of the variance of τ to be $2\ (1 - 0.25)/15 = 0.10$. The corresponding maximum value of the standard deviation of τ is $\sigma_\tau = \sqrt{0.10} = 0.316$. Next we calculate the quantity

$$x = \frac{\tau - T}{\sigma_\tau} = \frac{0.82 - 0.50}{0.316} = 1.01$$

Regarding this value as a normal deviate, we find, upon consulting a table of the normal probability integral, that the probability of a deviation this large or larger numerically is 0.312. Thus we cannot reject the hypothesis that the sample came from a population having $T = 0.50$.

Now the value which we have used for σ_τ, namely 0.316, is the maximum value which σ_τ can have. It might be smaller than this, in which case our value of x would have been larger. Conceivably it could have been large enough to have caused us to reject the hypothesis that $T = 0.50$. Thus, using the maximum value of σ_τ will never cause significance to be indicated oftener than it should be, and in this sense it is conservative.

SECTION III
RANK CONCORDANCE

14. Concordance. Up to this point we have considered the case in which just two rankings are involved. It is, however, desirable to have some measure of the agreement in rankings when there are several persons making the rankings.

Suppose then that there are m rankings of n objects. To make the matter more concrete let us consider the following case of 3 judges, X, Y, and Z, who have ranked 5 objects, a, b, c, d, e.

Object:	a	b	c	d	e
X's ranking:	4	1	2	3	5
Y's ranking:	3	4	1	2	5
Z's ranking:	1	4	2	5	3
Sum:	8	9	5	10	13

We have given not only the rankings but the sum of ranks. It can be shown that the 'best' estimate of the ranks, in a certain least-squares sense, is that obtained by ranking the objects according to the sum of the ranks assigned to them by the judges. In the present case the object c would be given rank 1, a would be given rank 2, b rank 3, d rank 4, and e rank 5.

The grand total of the sum of ranks is $8 + 9 + 5 + 10 + 13 = 45$, and the mean sum is $45/5 = 9$. In general the grand total is $\frac{1}{2} mn\,(n + 1)$ and the mean sum is $\frac{1}{2} m\,(n + 1)$.

If there were complete agreement among the 3 rankings the sums would be 3, 6, 9, 12, 15 (although not necessarily in this order). In general the sums would be m, $2m$, ..., nm.

Let us designate by S_W the sum of squares of deviations from the mean sum. This may be taken as a measure of the agreement, or *concordance,* among the rankings. In the example under consideration,

$$S_W = (8 - 9)^2 + (9 - 9)^2 + (5 - 9)^2 + (10 - 9)^2 + (13 - 9)^2 = 34$$

In the general case the maximum value that S_W can have is $m^2\, n\, (n + 1)(n - 1)/12$, which is the value it assumes when the agreement is perfect.

15. Coefficient of Concordance (W). We now define W, the *coefficient of concordance*, by means of the following equation:

$$W = \frac{12\, S_W}{m^2\, n\, (n + 1)(n - 1)} \quad . \tag{13}$$

This coefficient can vary in value between 0 and 1, assuming the latter value when the concordance is perfect. In the present example,

$$W = \frac{12 \times 34}{3^2 \times 5 \times 6 \times 4} = 0.38.$$

Unlike τ, W cannot assume negative values. The value -1 is assumed by τ when there is complete disagreement between 2 rankings, that is, when one ranking is exactly the reverse of the other. Complete disagreement is impossible, however, in the case of more than 2 rankings. Thus, if X and Y are in complete disagreement, Z cannot be in complete disagreement with both of them.

16. Relation between W and Spearman's Coefficient. Suppose that Spearman's coefficient is calculated for each pair of rankings and the average of the values obtained is denoted by $\bar{\rho}$, then it can be shown that

$$\bar{\rho} = \frac{mW - 1}{m - 1}, \tag{14}$$

or

$$W = \frac{(m-1)\bar{\rho} + 1}{m} \tag{15}$$

SECTION IV
SIGNIFICANCE TESTS FOR CONCORDANCE

17. Significance Tests for Concordance. Tests of significance similar to those applied to S_τ can also be applied to S_W. Tables for this purpose are to be found in [1] vol. 1 and [2].

As an example, suppose that a value $S_W = 70$ has been obtained from 3 rankings of 5 objects. In [1], vol. 1, page 415, Table 16.8 or [2], page 149, Appendix, Table 5D, it is found that a value of S_W this large or larger has a probability of 0.026. This value may therefore be regarded as significant.

Since the tables referred to are not extensive it is useful to have methods of testing the significance of concordance for larger values of m and n. One such method consists in making the transformation

$$F = \frac{(m - 1)\, W}{1 - W}, \quad \nu_1 = n - 1 - \frac{2}{m}, \quad \nu_2 = (m - 1)\, \nu_1 , \qquad (16)$$

and using tables of Snedecor's F (for example, those to be found in [3]) with ν_1 and ν_2 degrees of freedom. In this case a *correction for continuity* should be made. This consists in subtracting 1 from the numerator and adding 2 to the denominator of the fraction in

$$W = \frac{S_W}{m^2\ (n^3 - n)/12} \quad .$$

To illustrate the method we shall use the above example, although naturally if the values of m and n fall within the range covered by probability tables of W, these tables should be used. We find

$$W = \frac{70 - 1}{9 \times 120/12 + 2} = 0.75,$$

$$F = \frac{(3 - 1) \times 0.75}{1 - 0.75} = 6.00,$$

$$\nu_1 = 5 - 1 - \frac{2}{3} = 3\frac{1}{3}, \quad \nu_2 = (3 - 1) \times 3\frac{1}{3} = 6\frac{2}{3} .$$

Since the degrees of freedom, ν_1 and ν_2 are fractional we must use two-way interpolation in a table of F:

Interpolation for 5% Point of F

	$\nu_1 = 3$	$\nu_1 = 3\frac{1}{3}$	$\nu_1 = 4$
$\nu_2 = 6$	4.76	4.683	4.53
$\nu_2 = 6\frac{2}{3}$		4.410	
$\nu_2 = 7$	4.35	4.273	4.12

Alternative method (for check)

	$\nu_1 = 3$	$\nu_1 = 3\frac{1}{3}$	$\nu_1 = 4$
$\nu_2 = 6$	4.76		4.53
$\nu_2 = 6\frac{2}{3}$	4.487	4.410	4.257
$\nu_2 = 7$	4.35		4.12

The 5% point is 4.41. By the same method, the 1% point is determined to be 8.69. The value $F = 6.00$ (or $W = 0.75$) is therefore significant at the 5% but not at the 1% level.

If we wish to interpolate between these levels we have the following set up:

F	P
4.41	0.05
6	
8.69	0.01

For the value of P we find 0.035 as against 0.026 given by the exact method.

For $n > 7$ we may use the chi-square distribution as follows. Set

$$\chi^2 = m\ (n - 1)\ W, \tag{17}$$

where W is to be corrected for continuity. The expression (17) has a chi-square distribution with $n - 1$ degrees of freedom.

Although in the preceding example n is only 5 and the use of χ^2 is not justified we shall use it here for purposes of illustration.

We find

$$\chi^2 = 3 \times (5 - 1) \times 0.75 = 9.00$$

Using a table of χ^2 (for example, that to be found in [3]) we find for 4 degrees of freedom that this value is not significant at the 5% level, which illustrates the use of the method and also shows that it is unsafe to use it when n is not greater than 7.

18. Tied Ranks. When ties occur the F-test and the chi-square test require no modification unless the number of ties is large. In this situation the test is complicated and no attempt will be made to discuss it here. The interested reader is referred to [1] and [2].

19. Relation between Concordance and Correlation. The score S_τ and Kendall's coefficient τ are meaningless when more than two rankings are concerned. However, the score S_W and the coefficient of concordance W can be calculated for two rankings, that is for $m = 2$, just as well as for any other values of m. In fact, if when $m = 2$ the values of S_W S_τ are calculated for two given rankings, these values will have exactly the same probability levels. A similar statement may be made for W and τ.

BIBLIOGRAPHY

[1] Kendall, Maurice G. *The Advanced Theory of Statistics.* vol. 1 (4th edition, 1948), vol. 2 (2nd edition, 1948). London: Charles Griffin & Company.

[2] Kendall, Maurice G. *Rank Correlation Methods.* (1948) London: Charles Griffin & Company.

[3] Rosander, A. C. *Elementary Principles of Statistics.* (1951) New York: D. Van Nostrand Company, Inc.

[4] Sillitto, G. P. The distribution of Kendall's τ coefficient of rank correlation in rankings containing ties. *Biometrika.* vol. 34 (1947).

McGregor & Werner, Inc.
Dayton, O. 100, 29 Sept. 52