1959

The treatment of ties in non-parametric tests.

BOSTON UNIVERSITY

GRADUATE SCHOOL

Thesis

THE TREATMENT OF TIES IN NON-PARAMETRIC TESTS

by

JEANNENE DUENAS

(A.B., Regis College, 1957)

Submitted in partial fulfilment of

the requirements for the degree

of

MASTER OF ARTS

1959

Approved
By

First Reader ______________________
Professor of Mathematics

Second Reader ______________________
Associate Professor of Mathematics

TABLE OF CONTENTS

SECTION 1

INTRODUCTION

Most non-parametric methods assume that the underlying distributions, from which the samples are drawn, are continuous, and hence tied observations (i.e., observations of equal magnitude) occur with probability zero. However, this is not a realistic assumption because, in practice, inability to get precise measurements render the distributions involved discontinuous. Therefore, ties will sometimes occur, and their treatment does affect the result of the particular test being used.

Before investigating the treatment of ties in several non-parametric tests, it may be well to give a brief discussion of the general theory of tests of hypotheses, considering both the parametric and non-parametric cases.

1. Tests of hypotheses.

Suppose that for a certain given experiment we choose a class of distributions, $C = C_o + C_1$, over the space X. A test of a hypothesis is a procedure for choosing one of two decisions at the end of the experiment. One decision is that the distribution belongs to C_o, the other is that it belongs to the complement C_1. These are referred to as the null hypothesis H_o, and the alternative H_1. The two decisions to be made are d_o to accept the hypothesis H_o, and d_1 to accept the alternative H_1.

Each outcome $x \in X$ of the experiment will be associated with either d_o or d_1. Hence, a decision function is represented by a subset of X called the critical region, which consists of all points x that are associated with the decision d_1 to reject the hypothesis (hence to accept the alternative).

Let α be the probability of rejecting the correct null hypothesis, This is called the probability of a Type I error. Let $\beta = \beta\ (C_1)$ be the probability of accepting the incorrect null hypothesis. This is called the probability of a Type II error. The usual procedure of selecting a test is to let α and N (the sample size), be fixed quantities, and then to choose the test that minimizes β in some sense. Actually, it is more customary to talk of the complementary probability $1-\beta$, which is called the power of the test. $1-\beta$ is the probability of rejecting the null hypothesis H_o computed as a function of the true distribution. Thus, in order to determine which test to use in testing hypotheses, we consider only critical regions which have probability of a type I error $\leq \alpha$, and select the test which has optimum power.

Closely related to the concept of power are two other properties, consistency and efficiency.

A test is consistent with respect to a particular alternative if the power of the test with respect to this alternative tends to unity as the sample size tends to infinity.

The concept of efficiency enters when two tests are to be compared with each other. Given two tests of the same size of the same statistical hypothesis, the relative efficiency of the second with respect to the first is given by the ratio N_1/N_2, where N_2 is the sample size of the second test required to achieve the same power for a given alternative ~~as~~ is achieved by the first test with respect to the same alternative when using a sample ~~of size~~ N_1.[1]

Once a test has been chosen to be applied to a particular set of observed data, the following procedure is used in testing the statistical hypothesis:

(1) The null hypothesis is stated.

(2) In practice, α and N are specified, and β is determined from these two.

(3) One determines which resulting values of the chosen statistic, called the critical region, will cause the rejection of the null hypothesis and which

---

[1] G. E. Noether, "On a Theorem of Pitman", Annals of Mathematical Statistics, Vol. 26 (1955), p. 64.

values will cause the acceptance of the null hypothesis.

(4) The value of the statistic is computed from the observed data.

(5) The null hypothesis is either accepted or rejected depending on whether the value obtained in (4) is outside or inside the critical region.[2]

Parametric Case.

In the parametric case, C consists of the class of distributions of a frequency function whose functional form is assumed to be known. The problem we are concerned with is to test hypotheses about the parameters θ, in the set Ω, of this frequency function. C_o corresponds to a subset ω of Ω, and C_1 corresponds to the complement $\Omega - \omega$. We then test the hypothesis $H_o : \theta \in \omega$ against the alternative $H_1 : \theta \in \Omega - \omega$.

In particular, in many of the parametric methods, it is assumed that the populations from which the observations are drawn are normally distributed and that the variances of these populations are equal. In general, these conditions are simply accepted and their truth or falsity determines the meaningfulness of the probability statement arrived at by the parametric test.

When testing a certain set of data, in which we

---

[2] W. J. Dixon and F. J. Massey, Introduction to Statistical Analysis, McGraw-Hill Book Co. Inc., New York, 1951

can safely assume these conditions, the appropriate parametric test is usually the optimum choice for analyzing the data because it will be the most powerful test.

However, when these conditions are not satisfied, the application of a parametric test may result in an erroneous probability statement. We, therefore, seek a suitable procedure for testing hypotheses in such cases. This is where the non-parametric approach comes in.

Non-parametric Case.

A non-parametric statistical test is one in which no assumptions are made concerning the functional form of the distribution of the population from which the samples in question are drawn, but only general assumptions, like continuity of the cumulative distribution function. Since no assumption about the form of the basic distribution function is required in this significance test, it would hardly be expected to be as efficient as one that permits some such assumption.

For example, the asymptotic relative efficiency of the sign test when compared with the t-test is usually said to be 0.64. According to the definition of relative efficiency, this would imply that if we used the sign test instead of the t-test in analyzing a given set of observations, we are "wasting" more than one-third of all observations. However, this statement can be made only if the condition that the sample is drawn from a normally distributed population is satisfied, this condition being assumed by the t-test. Since, in many cases, one has no way of being certain that this condition is satisfied, then, in these cases, the statement that the sign test has relative efficiency equal to 0.64 has not much meaning. One can be sure, however, that by using a non-parametric test he is performing a test at the stated level of significance, whatever the true population distribution. This statement cannot be said of the t-test.

Of course, using the sign test where the t-test could be correctly applied, would be wasteful of information, unless the greater simplicity of the sign test outweighed the loss of information.

The above applies analogously to other non-parametric tests and their parametric counterparts. Often, very little is lost by using non-parametric tests when the conditions are such that the parametric methods would be the optimum choice. On the other hand, a great deal may be gained by using them when the conditions required by the parametric procedures are not satisfied.[3]

Ranks

In order to insure the non-parametric character

---

[3] G.E. Noether, "Non-parametric Statistics", Boston University Graduate Journal, Vol. V (1957) pp. 110-111.

of a test under the null hypothesis H_o, it is usually necessary in the analysis of data to replace the original observations by ranks. This is done by ordering the N observations according to magnitude and calling the smallest observation 1, the second smallest 2, etc., the largest N.

Tests based on ranks are called "rank tests" and most of the non-parametric tests fall in this class.

In addition to being non-parametric in character, rank tests have the following advantages:

(1) The calculations are often simplified.

(2) Data available only in ordinal form may be used.

(3) When the assumptions of the usual test procedure are too unrealistic, we have not only the problem of distribution theory if the usual test is used, but it is possible that the usual test may not have as good a chance as the rank test of detecting the kinds of differences in which we are interested.[4]

---

[4] W. H. Kruskal and W. A. Wallis, "Use of Ranks in One-Criterion Variance Analysis," American Statistical Association Journal, Vol. 47 (1952), p. 585.

SECTION II

TREATMENT OF TIES

The assumption of continuity is of great importance because most of the non-parametric methods are valid as soon as it can be assumed that the underlying distribution is continuous. However, as we have already stated, this assumption of continuity is not usually satisfied because in actual cases, due to limitations in measurement, the distributions are generally discontinuous. Hence, modifications must be made in the non-parametric test procedures when applying them to cases involving discontinuous observations.

Procedures of treating ties are:

(1) To use mid-ranks, i.e., the average of the ranks that would have been assigned had there been no ties present.

(2) To assign randomly with equal probability the ranks that correspond to a set of tied observations. This procedure has the advantage that under the null hypothesis, H_0, the distribution theory applicable in the case of untied observations usually remains exactly valid. However, there are practical objections to this procedure. It often seems objectionable to base one's decisions on an extraneous random process required for the randomization of the ranks. An extraneous random process, also, as we shall see later, may reduce the power of the test.

Other procedures of treating tied observations are:

(3) To omit the tied observations altogether.

(4) To take the average of the probabilities associated with different values of the statistic obtained by breaking the ties in all possible ways, and to use this average probability in deciding whether to reject or accept the null hypothesis.

In what follows, the treatment of ties in certain specific non-parametric tests will be investigated.

SECTION III

THE TWO-SAMPLE PROBLEM

In the two-sample problem, there are two sets of observations, each being a sample from some probability distribution and we test whether the distributions are the same, i.e., the samples come from the same distribution.

We let X and Y be two random variables with continuous distribution functions F and G respectively. Let N ovservations $x_1, \ldots, x_N$ represent a sample from the X population where the x_i's are assumed independently and identically distributed; and let M observations $y_1, \ldots, y_M$ represent a sample from the Y population where the y_j's are assumed independently and identically distributed. M may be less than, equal to, or greater than N.

We want to test the hypothesis:

$$H_o : F(x) = G(x)$$

against the alternative

$$H_1 : F(x) \neq G(x).$$

The usual parametric technique for testing this particular hypothesis is to apply the Student t-test to the means of the two groups. The t-test assumes that the observations are independent and come from normally distributed populations with equal variances.

When these conditions are not satisfied, the data may be analyzed with one of the non-parametric tests for the two sample case. These will be discussed in the following:

1. <u>The Sign Test</u>

The sign test is applicable if we have paired observations. That is, we have N independent pairs of observations $(X_1, Y_1), \ldots, (X_N, Y_N)$, so that for each X_i $(i = 1, \ldots, N)$ in the X sample, there corresponds a Y_i in the Y sample. The test is based on the differences $X_i - Y_i = Z_i$.

The hypothesis that we test is

$H_0 : P\ (Z_i > 0) = P\ (Z_i < 0)$

Against the alternative

$H_1 : P\ (Z_i > 0) \neq P\ (Z_i < 0)$.

We use as a test criterion, the number of times, denoted by r, that the less frequent signed differences, either positive or negative, occur.

In the continuous case, where $P\ (Z_i = 0) = 0$, r is $B\ (N, \frac{1}{2})$. That is, r possesses a binomial distribution with probability $\frac{1}{2}$. This gives us the cut-off point.

If we should be dealing with discontinuous distributions, where $P(Z_i = 0) > 0$, we shall have to modify the sign test.

Let,

$N+$ = the number of positive Z_i's,

$N-$ = the number of negative Z_i's,

N_o = the number of zero Z_i's,

$$r = \begin{cases} N+, & \text{if } N+ < N- \\ N-, & \text{if } N- < N+ \end{cases}$$

There are three ways of treating these zero differences.

(1) We can count one-half as positive and the other half as negative, as suggested by Dixon and Mood,

(2) We can omit them altogether, thus reducing our N to $N-N_o$, as suggested by Dixon and Massey, or

(3) We can assign the ties at random, either positive or negative values.

The test based on (2) is given by:

(2.1) $r < K(N_o)$, where the cut-off point $K(N_o)$ is the one corresponding to $B(N-N_o, \frac{1}{2})$. Test (2.1) does not coincide with test (1.1) $(r + \frac{1}{2} N_o) < K$, which is based on (1).

The cut-off poing for (1.1) cannot be well defined since the distribution of $r + \frac{1}{2} N_o$, under H_o, depends on the parameter $p_o = P(Z_i = 0)$. The cut-off point, then, is usually taken to be the cut-off point corresponding to $B(N, \frac{1}{2})$. This results in lowering the level of significance of the test and consequently the power of the test is also reduced.[1]

Putter gives a theorem, which, when applied here, shows test (2.1) to be the unique most powerful test based on r and N_o.

Putter further shows that the asymptotic relative efficiency of the randomized test with respect to the (non-randomized) test is $1-p_o$.[2]

Thus, for most applications of the sign test, it would seem that the best way of dealing with ties is to omit them altogether.

2. The Wilcoxon (Mann-Whitney) Test.

In the Wilcoxon test, the two samples may or may not have an equal number of observations. (i.e., $M \lesseqgtr N$).

Here, we pool the samples $x_1, \ldots, x_N, y_1, \ldots, y_M$, and arrange them in ascending order of magnitude giving the smallest observation rank 1, etc.

Under the assumption of continuity, the arrangement is unique with probability 1, since $P(x_i = y_j) = 0$.

---

[1] H. Hemelrijk, "A Theorem on the Sign test When Ties are Present,": Koninkl, Nederl, Akad, Van Wetensch., Vol. 55 (1952), pp. 322-326.

[2] J. Putter, "The Treatment of Ties in Some Non-parametric Tests", Annals of Mathematical Statistics Vol. 26 (1955), pp. 372.

Letting T equal the sum of the ranks of the y's in the ~~ordered~~ sequence, U is defined by

$$U = MN + \frac{M(M+1)}{2} - T.$$

It has been shown that for M, N $>$ 8, the random variable U is approximately normally distributed with the following mean and variance

$$\mu_{NM} = E\ (U) = \tfrac{1}{2}\ NM.$$

$$\sigma_{NM}^2 = \frac{NM\ (N + M + 1)}{12}$$

Hence,

$$\tau = \frac{U - \mu_{NM}}{\sigma_{NM}} = \frac{U - \frac{1}{2}\ NM}{\sqrt{\frac{NM\ (N + M + 1)}{12}}}$$

is asymptotically normal with mean 0 and variance 1.

In the discontinuous case, when P $(x_1 = y_j) > 0$, it may happen that the pooled sample can no longer be uniquely ordered. We, therefore, have the problem of redefining the Wilcoxon test in this case.

Putter considers the randomized treatment of ties with the treatment based on mid-ranks and shows that for small samples the non-randomized treatment presents some practical difficulties, but the asymptotic (large sample) problem can be handled. He shows that although the randomized test is approximately equivalent to the Wilcoxon test where ties are not present, the resulting value of the statistic will be slightly affected since it depends not only upon the observations but also upon the outcome of the randomization procedure. He also shows that this randomization procedure results in reduced efficiency. Hence, the non-randomized procedure, based on mid-ranks is tobe preferred.

The changes required by the mid-rank method will be described in the following. The value of U is affected by the occurrence of ties among the observations involving both samples. The expected value of U, however, remains the same:

$$E\,(U) = \tfrac{1}{2}\, N\, M.$$

The variance of U is changed and is now

$$\sigma^2_{NM}(U) = \left(\frac{NM}{n\,(n-1)}\right)\left(\frac{n^3 - n}{12} - \Sigma T\right)$$

where, $n = N + M$

$$T = \frac{t^3 + t}{12}, \quad t = \text{the number of observations tied}$$

for a given rank.

The summation, Σ , takes place over all groups of ties.

Hence, with correction for ties,

$$\tau = \frac{U - \dfrac{NM}{2}}{\sqrt{\left(\dfrac{NM}{n\,(n-1)}\right)\left(\dfrac{n - n}{12} - \Sigma T\right)}}$$

This correction tends to increase the value of τ slightly, making it more significant.

Siegel recommends that correction for ties should be used only if:

(1) The proportion of ties is quite large,

(2) Some of the t's are quite large,

(3) The value of $\mathcal{T}$ obtained without the correction corresponds to a probability value which is very close to the given critical value.

3. <u>The Wald-Wolfowitz Runs Test</u>

Whereas the sign and the Wilcoxon tests are usually used to test the null hypothesis against the particular alternative of a shift in location, the Wald-Wolfowitz test can be used in testing the null hypothesis against the alternative hypothesis that the two distributions differ in any respect whatsoever. It can, therefore, be used to test a large class of alternatives. However, mathematical investigations have shown that this test is not very powerful against any particular class of alternatives. A further point is that when the null hypothesis is rejected on the basis of the test, it can be asserted that the populations differ, but very little if anything can be said as to how they differ.

The test assumes that the variables under consideration are continuous and that measurements are on at least an ordinal scale.

In applying the test to data from two independent samples of size N and M, we rank the N + M members of the two samples, taken together, in order of increasing size. We then determine the number of runs in this ordered series. A run is defined as any sequence of members from the same sample, either the X sample or the Y sample.

If the null hypothesis is true, the members of both groups will be well mixed. Hence, r, the number of runs, will be relatively large. We, therefore, reject the null hypothesis for small values of r. Tables of significance values for this test are given by Swed and Eisenhart for N, M $\leq$ 20. For large samples when either N or M is larger than 20, r can be taken as approximately normally distributed with

$$\mu_r = \frac{2\ NM}{N + M} + 1,$$

and $$\sigma_r^2 = \frac{2\ N\ M\ (2NM - N - M)}{(N + M)\ (N + M - 1)}.$$

Hence,

$$\tau = \frac{r - \mu_r}{\sigma_r}$$ is normally distributed with mean 0 and variance 1. Tables for the normal distribution are used here to determine the significance of an observed value of r.

In the discontinuous case, the occurrence of ties among members of the same sample does not affect the number of runs, r, and, therefore, the obtained level of significance is unaffected. However, when ties occur between members of the different samples, then the ordered sequence is not unique and, we do not obtain a unique value of r.

Siegel suggests a procedure for treating ties by breaking the ties in all possible ways and to observe the resulting values of r. If all these values are significant with respect to previously set values of α, then ties present no problem, although they do increase the computations. However, if the various possible ways of breaking ties lead to some values of r which are significant and some which are not, the decision becomes difficult. It is suggested that the probability of occurrence associated with each possible value of r be determined and then the average of these probabilities be taken as the probability to be used in deciding to reject or accept the null hypothesis.

If the number of ties between the members in the two different samples is large, r is essentially indeterminate, and in such cases, the Wald-Wolfowitz test is inapplicable.

4. The Moses Test

In the behavioral sciences, it is sometimes expected that some experimental condition will cause some subjects to show extreme behavior in one direction while it causes others to show extreme behavior in the opposite direction.

Suppose one wishes to determine whether the behavior in one group (experimentals) is defensive as contrasted with the behavior of another group (controls). The Moses Test is designed to be used with data (measured in at least an ordinal scale) collected to test hypotheses of this type. Here the null hypothesis that the two groups, experimental and control, come from a common population is tested against the alternative that the experimentals are "extreme" in one or both directions relative to the controls. The Moses Test is most useful if it is believed that the experimental condition will lead to extreme scores in either direction.

The Moses Test focuses on the span or spread of the control cases. If there are n_c control cases and n_e experimental cases, and the $n_c + n_e$ scores are arranged in the order of magnitude then, under the null hypothesis (that the E's and C's aome from the same population), the E's and C's are expected to be well mixed in the ordered series. However, if the alternative is true, then one of the following situations will hold:

(1) The C's will be congested at the high end of the series,
(2) The C's will be congested at the low end of the ordered series,
(3) The C's will be congested in the middle of the ordered series.

The Moses test determines whether the C scores are so closely congested relative to the $n_c + n_e$ scores that the null hypothesis should be rejected.

In applying the Moses test, the members of both groups are arranged in a single ordered series, retaining the group identity of each member. The span S', of the C scores is then determined by noting the lowest and highest C scores and counting the number of observations between them, including both extremes. The span S', is then the smallest number of consecutive scores in an ordered series needed to include all C scores. For computational simplicity, each score is ranked and S' is determined from the ordered series of the ranks assigned to the $n_e + n_c$ cases. The hypothesis is rejected for values of S' that are too small.

When n_c is large, a modification is necessary since the span of C's is an inefficient index to the spread of the group, due to possible sampling fluctuations. In this case, Moses suggests that the researcher, in advance of collecting his data, arbitrarily select some small number h. Then h control scores are subtracted from both extremes of the span or range of control scores. The span is then found from the remaining scores, and this span, called the truncated span, is denoted by S_h.

Tables are given by Moses for the probability of the occurrence of the observed value of S_h or less under the null hypothesis. The null hypothesis is rejected when the probability of occurrence is $\leq \alpha$, where α is the chosen level of significance.

The occurrence of ties between two or more members of the same sample does not affect the value of S_h. However, when there are tied observations between members of the two different samples, there may be more than on value of S_h, depending on how the tie is broken. Hence, the ties should be broken in all possible ways, and the corresponding probabilities under the null hypothesis should be found. Then the average of these probabilities should be taken for use in deciding whether to accept or reject the null hypothesis. If the number of ties between the two samples is very large, the Moses test is inapplicable.

SECTION IV

THE C-SAMPLE PROBLEM

The C-sample problem is an extension of the two-sample problem to a consideration of C samples. The problem is to test whether C independent samples can be regarded as coming from the same population. That is, the differences among the various samples are to be regarded simply as chance variations which usually occur in drawing random samples from the same population. The alternative usually assumes that the populations are approximately of the same form, the difference being only a shift or translation.

The usual parametric technique for testing whether several independent samples come from the same population is the F-test, for the one-way analysis of variance. The assumptions associated with the statistical model underlying the F-test are that the observations are independently drawn from normally distributed populations, all of which have the same variance. The hypothesis tested here is that the samples are from populations with the same mean, that is,

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_c .$$

The alternative is that at least one μ_i, $i = 1, \cdots, C$, differs from the others.

If the assumptions of the F-test do not hold for a particular set of observations, a non-parametric test may be preferable in analyzing the data. Such a test is the Kruskal-Wallis Test.

1. The Kruskal-Wallis Test

This test is based on ranks. It is required that the observations in all C samples be ranked together, and the sum of the ranks for each sample be obtained. If no ties occur, the following test statistic is computed:

$$(11) \quad H = \frac{12}{N(N+1)} \sum_{i=1}^{C} \frac{R_i^2}{n_i} - 3(N+1), \text{ where}$$

C = the number of samples,
n_i = the number of observations in the i th sample,
$N = \sum n_i$, the number of observations in all samples combined
R_i = the sum of the ranks in the i th sample.

The null hypothesis, i.e., the hypothesis that the various samples are drawn from the same population, will be rejected for large values of H. Under the null hypothesis when the n_i are not too small, H is distributed as χ^2 with C-1 degrees of freedom. Thus, we can use the tables of the χ^2 distribution which are available. However, when the n_i are small and C = 2, tables are available from Wilcoxon, Festinger and White. For C = 3 and all $n_i \leq 5$, tables are given by Kruskal and Wallis. For other cases where the χ^2 approximation is not adequate, two approximations, the Γ approximation and the B approximation, are described by Kruskal and Wallis.

If ties occur, each observation in the tied group is given the mean of the ranks for which it is tied. The H which is computed from above is divided by.

$$(2.1) \qquad 1 - \frac{\sum T}{N^3 - N}$$

where summation takes place over all groups of ties and $T = (t-1)\, t\, (t+1) = t^3 - t$ for each group of ties, t being the number of tied observations in the group.

The following table (Table 2.1), taken from Kruskal and Wallis, gives values of T for t = 1, ..., 10.

Table 2.1

t	1	2	3	4	5	6	7	8	9	10
T	0	6	24	60	120	210	336	504	720	990

Since $0 \leq \left[1 - \frac{\sum T}{N^3 - N}\right] \leq 1$, (2.1) increases H.

If all N observations are equal (2.1) reduces H to the indeterminate form 0/0. If no ties occur, each value of t = 1, so that T = 0 and (1.1) remains unchanged by (2.1). Hence, the general expression, which holds whether or not ties are present, and assuming that such ties as occur are given mean ranks, is given by

$$(2.2) \quad H = \frac{\frac{12}{N(N+1)} \sum_{i=1}^{c} \frac{R_i^2}{n_i} - 3(N+1)}{1 - \frac{\sum T}{N^3 - N}}$$

The difference between (2.2) and (1.1) is negligable in many instances. For example, with $C \leq 10$, a χ^2 probability of 0.01 or more obtained from (1.1) will not be changed by more than ten percent by using (2.2), provided that not more than one-fourth of the observations are involved in ties. For large samples, H is still distributed as χ^2 (C-1) when ties are handled by mean ranks. However, the tables for small samples, although still useful, are no longer exact.

Another method of treating ties is the randomization procedure in which the ranks within a group of tied observations are assigned at random. Since the null hypothesis is that the ranks are distributed at random, the distribution of H, under the null hypothesis is the same as if no ties were present. However, complications in making and verifying computations are introduced in order to provide the adequate randomization which is necessary in using this procedure. It further seems that the introduction of extraneous random variability results in a reduced power of the test. In the case of the H test, we do not know whether mean ranks gives more or less power than random ranking of ties. The answer may vary with different alternative hypotheses and different levels of significance. A few computations for small samples and simple distributions, some carried out by Kruskal and Wallis and some by Howard L. Jones, showed mean ranks superior sometimes and random ranks others.

For theoretical purposes, random ranking of ties is easier to handle. However, for computational purposes, Kruskal and Wallis suggest the mean-rank method. The difference between the two methods will ordinarily be small.

SECTION V

RANK CORRELATION TESTS

In some problems, several variables are studied simultaneously to see how they are associated. One measure of the degree of association between the variables is their correlation.

In the parametric case, the usual measure of correlation is the Pearson product-moment correlation coefficient, which is based upon the assumption that the underlying distribution is bivariate normal.

If, with a given set of observations, the normality assumption is unrealistic, non-parametric methods may be used. Moreover, it is found that, especially with small samples, the computation of non-parametric measures of correlation are easier than those for the Pearson product-moment correlation coefficient. We will discuss two of the non-parametric measures of correlation and their tests of significance.

1. Kendall's Rank Correlation Coefficient

The method to be used is the following: Consider a set of N individuals to be ranked according to two variables X and Y. The observations an the X variable are ranked from 1 to N. Likewise, for the observations on the Y variable. The list of N individuals are arranged so that the X ranks of the individuals are in their natural order, i.e., 1, 2, ..., N. Observe the Y ranks in the order in which they occur when the X ranks are in natural order. The value of S for this order of the Y ranks is determined. The method for finding S will be discussed in the following example: Consider the following rankings on X and Y.

X:	1	2	3	4	5	6	7	8	9	10
Y:	4	7	2	10	3	6	8	1	5	9

Looking at the Y rankings, consider the first number on the left which is 4. Count the number of ranks to its right which are larger, and subtract from this the number of ranks to its right which are smaller. We get 6-3 = 3. The same thing is done with the next number 7 and so on. The differences which we get are:

$$3,-2,5,-6,3,0,-1,2,1$$

adding these differences we get

$$S = 3-2 + 5-6 + 3 + 0 -1 + 2 + 1 = 5$$

Now, the maximum score obtained if the rankings are in the objective order, 1, 2, ..., 10, is 45.

The rank correlation coefficient τ is defined as the ratio of the actual score, S, to the maximum score, S^1.

In our example:

$$\tau = \frac{5}{45} = 0.111.$$

Generally, if there are N individuals, the maximum score obtained, if they are in the order 1,2,..., N, is $(N-1) + (N-2) + \ldots + 1 = \frac{N(N-1)}{2} = S^1$.

Denoting the actual score by S, the coefficient of rank correlation becomes:

$$\tau = \frac{2S}{N(N-1)} .$$

Now, if the N individuals constitute a random sample from some population, we test the null hypothesis that the observed value of τ indicates the existence of an association between the X and Y variables in that population. For $N \leq 10$, a table is given by Kendall which shows the associated probability of a value as large as an observed S. If this probability is equal to or less than the chosen level of significance, α, then the null hypothesis is rejected. For $N > 10$, τ is considered approximately normally distributed with

$$\mu_\tau = 0.$$

and

$$\sigma_\tau^2 = \frac{2\ (2N + 5)}{9N\ (N-1)}$$

Therefore, the statistic

$$Z = \frac{\tau - \mu_\tau}{\sigma_\tau} \quad , \text{ where } Z = N\ [0,1]$$

may be used and the probability be found in the tables for the normal distribution. Here again a probability that is less than or equal to α will result in rejection of the null hypothesis.

If tied observations on either the X or Y variable should occur, the mid-rank method is used. The tied observations are given the average of the ranks they would have had if no ties had occurred.

The presence of ties requires a change in the denominator of the formula for τ, which now becomes

$$\tau = \frac{S}{\sqrt{\frac{1}{2}N\,(N-1) - T_x}\;\sqrt{\frac{1}{2}\,N\,(N-1) - T_y}}\,.$$

Where, $T_x = \frac{1}{2}\Sigma\, t\,(t-1)$, t being the number of tied observations in each group of ties on the X variable. $T_y = \frac{1}{2}\ t\,(t-1)$, t being the number of tied observations in each group of ties on the Y variable.

G. P. Sillitto gives still another formula for the correction of ties in this case. He states that the maximum score possible, S^1, is reduced by the presence of ties, and the presence of each tied pair reduces the maximum possible score by unity, so that

$S^1 = \frac{1}{2}\,N\,(N-1) - P_2$ for the case of a ranking of N individuals containing P_2 pairs.

Generally, each r-tuplet tie reduces the maximum possible score, by $\frac{1}{2}r\,(r-1)$ so that for a ranking of N individuals containing P_2 pairs, P_3 triplets, ..., P_r r-tuplets.

$$\tau = \frac{2S}{N\,(N-1) - 2\,p_2 - 6\,p_3 - \ldots - r(r-1)\,p_r}.$$

2. <u>Kendall's Coefficient of Concordance.</u>

In the last test, we were concerned with the test of the correlation between two sets of rankings of N individuals. Here, we consider the test of the relation among m rankings of N individuals. The degree of association among these m variables can be measured by Kendall's coefficient of concordance.

In order to illustrate the method, consider the following array of m rows, where the m rows stand for rankings of the N individuals:

$$\begin{matrix} a_1 & a_2 & \cdots & a_N \\ b_1 & b_2 & \cdots & b_N \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ m_1 & m_2 & \cdots & m_N \end{matrix}$$

where the a's, b's, etc., are permutations of the natural numbers 1 to N.

Letting S = the sum of squares of deviation of the column totals about their means, Kendall's coefficient of concordance is given by

$$W = \frac{S}{\frac{1}{12} m^2 (N^3 - N)}.$$

To compute S, we find the sum of the ranks, R_j, in each column of the m x N array. We then sum all the R_j and divide this sum by N to obtain the mean value of R_j.

Hence, $S = \sum \left(R_j - \frac{\sum R_j}{N}\right)^2$

and $$W = \frac{\sum \left(R_j - \frac{\sum R_j}{N}\right)^2}{\frac{1}{12} m^2 (N^3 - N)}.$$

To determine if W is significantly different from zero, Milton Friedman has constructed a table for $N \leq 7$, which is adapted by Siegel, and which gives critical values of S associated with W's significant at the 0.05 and 0.01 levels.

For $N > 7$, the following statistic is used:

$$\chi^2 (N-1) = \frac{S}{\frac{1}{12} mN (N + 1)},$$

which is approximately chi square distributed with (N-1) degress of freedom. To determine whether W is significantly different from zero, we fefer to χ^2 tables which give critical values of χ^2 (N-1).

When tied observations occur, the mid-rank method again is used. The effect of tied ranks is to decrease the value of W. If the number of ties is small, the effect is negligable and no correction is made, However, if the proportion of ties is large, a correction factor is used, namely;

$$T = \frac{\Sigma(t^3 - t)}{12},$$

where t = the number of observations in a group tied for a given rank and the summation takes place over all group of ties within any one of the m rankings.

With correction for ties,

$$W = \frac{\sum \left(R_j - \frac{\Sigma R_j}{N} \right)^2}{\frac{1}{12} m^2 (N^3 - N) - m \sum_T T},$$ where

summation over T refers to summation of all values of T for the m rankings.

SECTION VI

CONCLUSION

Various suggestions of how to treat tied observations in non-parametric tests have been presented. Some of these methods can be justified by considerations of power and/or efficiency. In other cases, the suggested method is usually one of convenience and its influence on the behavior of the test still requires investigation.

BIBLIOGRAPHY

Dixon, W. J. and Massey, F. J., An Introduction to Statistical Analysis, McGraw-Hill Book Co., 1951.

Dixon, W. J. and Mood, A. M., "The Statistical Sign Test", Journal of the American Statistical Association., Vol. 41 (1946) pp. 557-566.

Fraser, D. A. S., Non-parametric Methods in Statistics, John Wiley & Sons, Inc., 1957, London.

Friedman, M., "The Use of Ranks to Avoid the Assumption of Normality Implicit in Analysis of Variance", Journal of the American Statistical Association, Vol. 32 (1937) pp. 675-701.

Hoel, P. G., Introduction to Mathematical Statistics, Wiley & Sons, Inc., New York, 1955.

Hemelrijk, J., "A Theorem on the Sign Test When Ties are Present", Koninkl, Nederl, Akad, Van Wetensch, Vol. 55 (1952) pp. 322-326.

Hemelrijk, J., "Note on Wilcoxon's Two Sample Test When Ties are Present", Annals of Mathematical Statistics, Vol. 23 (1952), pp. 133-135.

Kendall, M. G., "The Treatment of Ties in Ranking Problems", Biometrika, Vol. 33 (1943-46), pp. 239-251.

Kendall, M. G., "A New Measure of Rank Correlation", Biometrika, Vol. 30 (1938), pp. 81-93.

Kruskal, W. H. and Wallis, W. A., "Use of Ranks in One-Criterion Variance Analysis", American Statistical Association Journal, Vol. 47 (1952), pp. 583-621.

Mann, H. B. and Whitney, D. R., "On a Test of Whether One of Two Random Variables is Stochastically Larger Than the Other", Annals of Mathematical Statistics Vol. 18 (1947) pp. 50-60.

Moses, L. E., "A Two-Sample Test", Psychometrika, Vol. 17 (1952), pp. 239-247.

Noether, G. E., "Non-parametric Statistics", Boston University Graduate Journal, Vol. V, #7, (1957), pp. 109-111.

Noether, G. E., "On a Theorem of Pitman", Annals of Mathematical Statistics, Vol. 26 (1955), pp. 64-68.

Putter, J., "The Treatment of Ties in Some Non-parametric Tests", Annals of Mathematical Statistics, Vol. 26 (1955), pp. 368-386.

Siegel, S., Non-parametric Statistics, McGraw-Hill Book Co., 1956, New York, Toronto, London.

Sillitto, G. P., "The Distribution of Kendall's Test Coefficient of Rank Correlation in Rankings Containing Ties", Biometrika, Vol. 34 (1947) pp. 36-40.

Terpstra, T. J., The Asymptotic Normality and Consistency of Kendall's Test Against Trend, When Ties are Present in One Ranking", Koninkl, Nederl, Akad, Van Wetensch, Vol. 55 (1952) pp. 327-333.

Wilcoxon, F., "Probability tables for individual comparisons by ranking methods," Biometrics, Vol. 3 (1947), pp. 119-122.

ABSTRACT

In most of the non-parametric methods, only a few general assumptions are made concerning the underlying distribution of the population from which a certain sample is drawn. One of the most frequent of these assumptions is that of continuity, i.e., that the population from which the sample is drawn possesses a continuous distribution, and, therefore, the probability of two or more equal observations is zero. Actually, however, due to limitation of measurement, experimental data are such that they must usually be regarded as coming from discontinuous distributions and equal observations will occur. When this is the case, one speaks of the occurrence of "tied" observations, or simply "ties", in the data.

In applying a non-parametric test to data of this type, where ties occur, several problems arise because here the assumption of continuity no longer holds. Therefore, modifications are necessary in order to apply the non-parametric test.

Almost all of the non-parametric methods are based on ranks, i.e., arranging the observations in increasing order of magnitude and giving the smallest observation the rank 1, the next smallest 2, etc. When ranking in the case of two or more observations with

equal magnitude, one encounters the problem of how to rank these tied observations in order to secure optimum performance of the test being applied. There are two main procedures of treating ties of this kind:

(1) To use the average of the ranks that would have been assigned had there been no ties present, called the mid-rank method, or

(2) To assign randomly and with equal probability the ranks that correspond to a set of tied observations, called the randomization method.

In this paper are discussed the different solutions that various statisticians have suggested in treating tied observations when applying the following non-parametric tests: (1) the sign test, (2) the Wilcoxon (Mann-Whitney) test, (3) the Wald-Wolfowitz Runs test, (4) The Moses test, (5) the Kruskal-Wallis test, (6) Kendall's rank correlation coefficient test, and (7) Kendall's coefficient of concordance test.

It is found that most of the statisticians recommend the mid-rank method of treating ties.