SOME RECENT DEVELOPMENTS IN RANKING METHODS [1]

by

M. G. Kendall
Institute of Statistics
North Carolina State College

Institute of Statistics
Mimeograph Series No. 112
July, 1954

1. This research was supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development.

SOME RECENT DEVELOPMENTS IN RANKING METHODS

by M. G. Kendall, Visiting Professor

Institute of Statistics

North Carolina State College

1. Ranking methods received some attention from psychologists and statisticians at the beginning of the present century but were at that time regarded as a minor offshoot of statistical theory. From 1900 to 1935 statisticians were engrossed with the theory of continuous or discontinuous observable variates; and even in psychology, where ranking methods have always been more serviceable, correlations based on ranks were generally considered as approximate representations of some inaccessible correlation based on scales or measurable variates. If one looks back over the history of the subject the few papers which have any current interest stand out more like isolated peaks rising from the plains than a continuous mountain chain. K. Pearson in 1907 discussed the coefficient of rank correlation now knows as Spearman's ρ_s, and returned to the subject from time to time (1914, 1921, 1931, 1932). Greiner in 1909, in a paper which was overlooked by British statisticians for nearly 40 years, considered the estimation of the correlation parameter in bivariate normal populations by what are essentially methods based on order. "Student" in 1921 discussed the standard error of Spearman's ρ_s but did not pursue the topic to the point of further publication. Esscher in 1924 extended Greiner's work but again without exciting any continued interest. It was not until 1936 that the subject began to assume any importance.

2. It would, however, be a mistake to suppose that the ideas underlying ranking theory as it exists today are entirely novel. It is true that earlier writers were mainly concerned to use ranks as makeshifts for variates, and not in their own right. They were, in modern terminology, studying parametric problems by non-parametric methods. But there are signs in several of the earlier papers that the importance of order statistics were appreciated, especially by Karl Pearson, and a series of papers on medians, quantiles and ranges began about 1920 to appear from University College London - and are still appearing with undiminished vigour. The relationships between rank and order statistics, however, have only recently been brought into full view and we may regard the development of order-statistic theory as almost independent of ranking theory until about 1950.

3. In 1936 Hotelling and Pabst published a substantial study of Spearman's ρ_s in which the newer approach is already visible. They determined its first four moments in the null case (i.e. in the population wherein all rankings occur equally frequently), proved that its distribution tended to normality, cleared up some obscurities of notation and re-examined earlier work on the relation between Spearman's ρ_s and the correlation parameter in bivariate normal variation. All this could be regarded as in the line of development of previous work; more significant was the line of thought, explicitly recorded in their title ("Rank correlation and tests of significance involving no assumptions of normality") in which they considered rank correlation, not as a substitute for a variate-correlation, but as a

statistic providing tests of independence which should not depend on assumptions concerning the parent population - what would nowadays be called a distribution-free test. It was about this time that such tests began to be studied in connection with randomization procedures introduced systematically into experimentation by R. A. Fisher; see, for example, the work of Pitman in 1937 and 1938 and of Welch in 1937.

4. In 1938 I discovered the rank correlation coefficient known as τ, although it emerged later that Greiner in 1909 had, in effect, used a similar coefficient to estimate the correlation parameter in bivariate normal variation. Independently Olds (1937) and Kendall and others (1939) studied the distribution of Spearman's ρ in the null case. Friedman (1937) discussed the use of ranks to avoid assumptions of normality in variance-analysis, his work being extended by Kendall and Babington Smith (1939). By this time ranking methods were returning to statistical notice. Woodbury (1940), Sillitto (1947) and Kendall (1945, 1947) discussed the treatment of ties. In 1942 Kendall extended the t-coefficient to the definition of partial rank correlation. Daniels (1944) defined a general coefficient of which τ, Spearman's ρ_s and and product-moment correlation are special cases. Hoeffding (1947) and Daniels and Kendall (1947) discussed the significance of rank correlations in the non-null case.

5. The state of the theory as it existed in 1948 is described in my brochure on Rank Correlation Methods. Some further work was

reviewed in the Symposium on Ranking Methods (Moran, Whitfield and Daniels) held before the Research Section of the Royal Statistical Society in March 1950. Such is the pace of advance when a subject really gets into its stride that four years later it is possible to report a great deal of further progress. That is the object of this memorandum. I divide the treatment under six heads:

(a) Rank correlation coefficients

(b) Sampling from ranked populations

(c) Ranks and variate values

(d) Multivariate ranking theory

(e) Relationship with order-statistics

(f) Paired comparisons.

Rank Correlation Coefficients

6. In his 1944 paper Daniels defined a general coefficient for a set of n observations, each on two variates, say x and y. To any pair of individuals considered as ordered according to the x relationship, the ith and jth, we allot a score a_{ij} subject only to the condition that $a_{ij} = -a_{ji}$. Similarly for y we define a score $b_{ij}(= -b_{ji})$. For each pair of objects that is an a-score and a b-score. Denoting by Σ the summation over all i, j from 1 to n we define

$$\Gamma = \frac{\Sigma\, a_{ij} b_{ij}}{(\Sigma\, a_{ij}^2 \Sigma\, b_{ij}^2)^{\frac{1}{2}}} \quad . \qquad (1)$$

We may either take $a_{ii} = 0$ or ignore items involving $i = j$ in the summation. For various values of the scores this coefficient reduces to τ, Spearman's ρ_s or product-moment correlation as the case may be. It clearly cannot exceed unity in absolute value. It also has a useful property demonstrated by Daniels himself (1948), namely that if two corresponding pairs of members do not agree in the signs of their scores and the members of one pair are interchanged (so that they do agree) the coefficient Γ will increase, provided only that the scores are not zero and do not decrease the further apart the members are. Thus, as one set of ranks is interchanged pair by pair so as to come into closer agreement with a second set, the coefficient increases. This is a useful and, indeed, almost an essential quality of a coefficient which purports to measure the "amount of agreement" between two rankings.

7. The coefficient τ arises when the scores are ± 1 and ρ_s when the score of the ith and jth members is $|i-j|$. The two coefficients are therefore different in the nature and measure rather different properties. There are, however, some inequalities which limit the possible values of one when the other is fixed. Two such are

(a) Daniels (1950)

$$-1 \leq \frac{3n}{n-2} - \frac{2(n+1)}{n-2}\rho_s \leq 1 \qquad (2)$$

where n is the number of objects. For large n this is effectively

$$-1 \leq 3\tau - 2\rho_s \leq 1 \ . \tag{3}$$

For $\tau > 0$ the upper limit may be obtained and not the lower limit; for $\tau < 0$ the opposite is true; when $\tau = 0$ both limits may be attained.

<u>(b) Durbin and Stuart (1951)</u>

$$\rho_s \leq 1 - \frac{1-\tau}{2(n+1)}\{(n+1)(1-\tau) + 4\} \ , \quad t > 0 \ , \tag{4}$$

$$\geq \frac{3n\tau-(n-2)}{2(n+1)} \ , \quad t > 0 \ . \tag{5}$$

For large n these limits are

$$\frac{3}{2}\tau - \frac{1}{2}\rho_s \leq \frac{1}{2} + \tau - \frac{1}{2}\tau^2 \ , \quad t > 0. \tag{6}$$

For $\tau < 0$ we find

$$\frac{1}{2}\tau^2 + \tau - \frac{1}{2} \leq \rho_s \leq \frac{3}{2}\tau + \frac{1}{2} \ . \tag{7}$$

In the case when the n members of the ranking are a random sample from a bivariate normal population David and others (1951) have shown that for large n the correlation between ρ_s and τ is very high. Daniels (1944) had previously proved a conjecture of Kendall and others (1939) that in the null case (i.e. in the population comprising the n! permutations of ranks) the correlation between ρ_s and τ is

$$\frac{2(n+1)}{\{2n(2n+5)\}^{\frac{1}{2}}} \tag{8}$$

8. Apart from questions of interpretation and convenience of calculation there are some new results concerning τ and ρ_s when calculated from samples that have a bearing on their use. In fact, if our n members are regarded as chosen at random from a parent ranking of N members, with sample values t and r_s and parent correlations τ and ρ_s, we have

$$E(t) = \tau \tag{9}$$

(Daniels and Kendall, 1947); so that t is an unbiassed estimator of the parent τ. On the other hand

$$E(r_s) = \frac{1}{(n+1)(N-2)}\{3(N-n)\tau+(n-2)(N+1)\rho_s\} \tag{10}$$

(Durbin and Stuart, 1951; Daniels, 1951); so that r_s is a biased estimator of ρ_s. Even if N is large so that

$$E(r_s) = \frac{n-2}{n+1}\rho_s - \frac{3}{n+1}\tau \tag{11}$$

the estimator is biased. To obtain an unbiassed estimator of ρ_s we should use the function

$$\frac{n+1}{n-2}\left(r_s + \frac{3}{n+1}\right) \tag{12}$$

which depends on the sample values of both r_s and t.

9. A further use of the coefficient τ may be noted. By regarding a dichotomy as a case of heavily tied rankings, when all the members in one part are tied and all the members in the other are also tied Whitfield (1947) developed a measure of correlation in a 2 x q table where the q-variates can be ordered. This is a kind of biserial coefficient. Kendall (1949) extended this to the case of contingency tables where rows and columns have a natural order, and the idea has been further developed by Stuart (1953). The attraction of this approach is that outer bounds can be set to confidence intervals for the coefficients so obtained.

Sampling from ranked populations

10. Earlier studies in the sampling theory of rank correlation coefficients were all based on the null-case i.e. the population of all possible rankings; except for a few studies of large sample variances in ranked members of normal samples. Kendall in 1938 gave the distribution of t up to $n = 10$, a general expression for the variance and an approximation for $n > 10$, based on the normal curve, which was good enough for most purposes. Tabulation up to $n = 42$ has recently been carried out by Kaarsemaker and van Wijngaarden (1952). Expressions for the cumulants of the distribution have been given by Moran (1950a) and Silverstone (1950). Finally David and others (1951) developed an expansion in the form of an Edgeworth series which gives four figure accuracy to the significance points of the distribution for $n > 10$. The distribution of t in the null-case of equally frequent permutations is known as completely as practical applications are likely to require.

11. The distribution of Spearman's ρ_s in the null case of equally frequent permutations is much more difficult to ascertain. Olds gives it up to $n = 7$ in 1938, Kendall and others up to and including $n = 8$ in 1939; and David and others added the values for $n = 9$ and 10 in 1951. "Student" knew the variance for general n about 1920; Hotelling and Pabst gave the fourth moment in 1936; and David and others gave the sixth and eithth moments in 1951. Here also David and others used an Edgeworth series to obtain satisfactory significance points for $n > 10$. These results are likely to meet

all practical requirements, at least for some time to come.

12. This is also a convenient place to mention the so-called problem of m rankings, in which m rankings of n, not merely two, are given and the principle problem is to measure their relationship and to test its significance. Kelley in the 1920's had suggested using the Spearman ρ_s between pairs of rankings averaged over all pairs. Friedman (1937) used rank sums to test the significance of departure from homogeneity. Kendall and Babington Smith (1940) proposed a related statistic known as the coefficient of concordance W, approximated to it by a Beta-distribution and worked out the actual distributions for small values of m and n. Friedman (1940) then determined the significance points of W for a range of values.

13. More recently Benard and van Elteren (1953) have considered a more general situation in which not all rankings are of the same number n and have developed expressions for the moments to enable a test to be applied. These are very general results for the null case.

14. When we come to consider the non-null case several different models arise for examination; and the sampling problems associated with these are distinct:

(a) Kendall's model. In this case the two observed rankings of n are supposed to be an arrangement of n values chosen at random from a population of N values, as in section 8 above.

(b) Stuart's model. The m rankings are regarded as a sample drawn, with or without replacement, from a population of M rankings of n.

(c) Daniels' model. One order is determined by an objective criterion, e.g. it might be ordered in time or by reference to some measurable variate. The other is subject to error which causes actual scores to deviate from their means and hence certain orders to appear which differ from the true order (i.e. that based on the means). This is analogous to a regression model.*

15. The first of these models was considered by Daniels and Kendall (1947) and by Hoeffding (1947). With certain restrictions it can be shown that t or r_s both tend to normality with increasing n. It was also shown that their joint distribution tends to bivariate normality. In point of fact, more general asymptotic results are attainable: with certain restrictions any two coefficients of the Γ type tend to joint normality. Hoeffding (1948) defined a class of coefficient even more general and proved the limiting normality.

16. There is one noteworthy feature of this situation, namely that although exact expressions for the variances of t and r_s cannot be obtained without further knowledge of the parent, it is possible to give upper bounds for those variances. In fact Daniels and Kendall show and it follows from a general result of Hoeffding that

---

* In the paired-comparison case a model of this kind has been considered by Thurstone. Cf. Mosteller (1951 a,b,c).

$$\text{var } t \leq \frac{2}{n}(1-\tau^2) \tag{13}$$

This is a very useful result and enables conservative tests to be made of the differences of two coefficients.

17. Sundrum (1953 a), working from the point of view of order relations, has taken matters further. He shows that the first four moments of t depend on 10 parameters which are troublesome but not wholly unmanageable. They may be estimated from the data or determined for specified parental forms either mathematically or by sampling experiments. A frequency distribution can then be fitted by identifying lower moments and the confidence interval for t determined from this distribution.

18. The formula (13) holds when ties are present. An approximate formula of a similar kind for Spearman's r_s is deducible from Hoeffding's general result, namely

$$\text{var } r_s \leq \frac{3}{n}(1-\rho_s^2) \tag{14}$$

but it does not seem to be known how this is affected by ties.

19. Extensions of these results when there are more than two rankings are hampered by a difficulty which is common to much multivariate work, namely the rapid increases in the number of parental correlations; for m rankings there are $\frac{1}{2}m(m-1)$ of these unknown

parental parameters. The difficulties of treating this general topic, though severe, are not to be considered as prohibitive. There is considerable need for a test of the significance of the difference of two concordance coefficients W.

20. For Stuart's model - see section 14 above - the situation requiring analysis is more complex. We imagine a population M of individuals (e.g. judges) who rank the same set of n objects. A subset m of M is chosen and we require to discuss the properties of the M individual in relation to their agreement or disagreement about the rankings. The principal problem is to find some concise method of specifying the parent. If any ranking may happen there will be needed n! parameters to describe the relative frequency of all possible rankings. This is far too many and some simplification is essential.

21. Stuart considers the mean and variance of the concordance coefficient in terms of parent k-statistics. The resulting expressions are cumbrous and he was able to make substantial progress only in the case where the mean ranks of all the objects are the same in the population of M. Ehrenberg (1952) has considered the use of the coefficient of agreement u proposed by Kendall and Babington Smith (1940) but was unable to make much headway with the sampling problem for ranks in the non-null case. He points out that in the null case the correlation between u and W is given by

$$\frac{2(n+1)}{(4n^2 + 10n)^{\frac{1}{2}}} \tag{15}$$

which rapidly approaches unity with increasing n. He also rederives a χ^2 approximation to the distribution of u given earlier by Kendall and Babington Smith.

22. The case considered by Daniels (1950) - see section 14 - is again different from the other two models. We now consider a set n of objects ranked by a population of judges, or by the same judge in repeated trials, on a particular attribute <u>whose ranking is known a priori</u>. A model of a similar kind had been considered by Thurstone as early as 1927 (<u>J. Abn. and Soc. Psych.</u> 21, 384). One simple form is given by supposing that there is a continuous scale y and the frequency function of the ratings $y_1 \ldots y_n$ is

$$f(y_1 - m_1)\, f(y_2 - m_2) \ldots f(y_n - m_n) \tag{16}$$

where $m_1 \ldots m_n$ are the expectations in his rankings. By the use of this model Daniels makes considerably more progress than might be expected from its unpromising nature. Effectively the problem becomes one of regression.

Ranks and Variate-Values

23. One of the objects of using ranking methods is to be free from assumptions about the precise form of the parent population. It is, however, of considerable interest to consider the raltionship between rank correlation coefficients and the correlation parameter ρ in a bivariate normal surface; and, indeed to go further by considering relations between ranks and variate-values. Suppose we have a continuous bivariate population, and draw a pair of values (x_i, y_i) and (x_j, y_j). We define the probability of concordance of type 1 as

$$\pi_1 = \text{Prob } (y_i < y_j \mid x_i < x_j) \tag{17}$$

that is to say, the probability that $y_i < y_j$ whenever $x_i < x_j$. If p_1 is the sample estimate of π_1 obtained by counting concordances in the $\frac{1}{2}n(n-1)$ comparisons of pairs of n objects, the rank correlation t (the sample value of τ) is given by

$$t = 2p_1 - 1 \tag{18}$$

24. Similarly, if we have a triad of values (x_i, y_i), (x_j, x_j), (x_k, y_k) the probability of concordance of type 2 is

$$\pi_2 = \text{Prob } (y_i < y_k \mid x_i < x_j) \; . \tag{19}$$

Defining a sample value p_2 as the estimator of π_2 we have for Spearman's r_s

$$r_s = \frac{3t}{n+1} + \frac{6(n-2)}{n+1} \quad (p_2 - \tfrac{1}{2}) \tag{20}$$

showing that r_s depends on both the first and second types of concordance. Since π_1 and π_2 are defined for a continuous population we may regard the limiting forms of (18) and (20) as defining parental values of τ and ρ_s for such a population, i.e.

$$\begin{aligned} \tau &= 2\pi_1 - 1 \\ \rho_s &= 6(\pi_2 - \tfrac{1}{2}) \quad . \end{aligned} \tag{21}$$

This approach to the problem is mainly due to Hoeffding, who has obtained from it some very powerful general methods of attack.

25. An interesting result in this field has also been obtained by Stuart (1954, unpublished). If we draw a sample from a univariate population and rank the members by order of variate-value, what is the expected correlation between the ranks so obtained and the variate-values themselves? Stuart shows that the correlation (product-moment) is surprisingly high. For example, in samples from a rectangular population it is

$$\left(\frac{n-1}{n+1}\right)^{\frac{1}{2}}$$

and in samples from a normal population is $\surd(3/\pi) = 0.9772$ times this amount. It follows that in many problems, if we replace variate-values by ranks we shall not seriously affect conclusions drawn from the data.

26. The relation between ρ (the correlation parameter in bivariate normal variation) and t was, in effect given by Greiner in 1909, and Esscher in 1924 gave the variance of t in normal samples in terms of ρ. An analogous result for r_s was given by Moran (1948)

$$E(r_s) = \frac{6}{\pi(n+1)} \left(\sin^{-1}\rho + (n-2)\sin^{-1}\frac{1}{2}\rho\right) \tag{22}$$

The limiting form of this for large n, namely

$$E(r_s) = \frac{6}{\pi}\sin^{-1}\frac{1}{2}\rho \tag{23}$$

was reached by Karl Pearson in 1907.

In 1949 Kendall examined the effect on non-normality of these formulae and showed in particular that Greiner's formula

$$E(t) = \frac{2}{\pi}\sin^{-1}\rho \tag{24}$$

was not very sensitive to departures from normality as measured by skewness. He also showed that an expression for the variance of r_s corresponding to Esscher's for t was not possible in terms of simple transcendental functions, but derived an expansion for large n as a power series in ρ. Later David and others (1951) extended this result and obtained a similar expansion for the covariance of t and r_s.

27. When higher moments of t and r_s in the normal case are required still more complicated functions make their appearance. Higher types of concordance are also required. Sundrum (1953) has discussed the first four moments of t and considered methods of estimating by sampling experiments the probabilities of concordance required. No similar investigations seem to have been carried out for r_s.

28. As long ago as 1907 K. Pearson suggested using (25) to estimate r from the formula

$$\rho \text{ (est.)} = 2 \sin \frac{\pi r_s}{6} \quad . \tag{25}$$

From (28) it is clear that a better estimator is likely to be

$$\rho \text{ (est.)} = 2 \sin \frac{\pi}{6} \{r_s - \frac{3(t-r_s)}{n+1}\} \quad . \tag{26}$$

In a corresponding way Greiner proposed to estimate ρ from (24) as

$$\rho\ (\text{est.}) = \sin \tfrac{1}{2} \pi\, t \quad . \tag{27}$$

These formulae, however, may still give biased estimators. An examination of the problem of deriving unbiased estimators of ρ from t and r_s has not yet been carried out.

Multivariate ranking theory

29. An analogy, which may prove to be somewhat misleading, between rank correlation and product-moment correlation, has led to inquiries into the possibilities of constructing ranking analogues to such quantities as correlation ratios, partial correlations and multiple correlations. The concordance coefficient W is a kind of ranking correlation ratio. Kendall (1942) defined a partial rank correlation coefficient, but its interpretation is obscure. Hoeffding (1948) showed that the coefficient is asymptotically normal, but practically nothing is known of its small-sample behaviour. Moran (1950) suggested a determinantal expression involving the t-coefficients of several rankings as an analogue of Wilks' criterion in multivariate analysis. Finally Moran (1951) re-examined partial t without reaching any definite conclusion about its distribution; and defined a multiple rank coefficient, suggesting a test of significance based on the F-ratio. The work might, perhaps, usefully be taken further. At the present time multivariate

ranking theory is in a very rudimentary state.

Relationship with order-statistics.

30. The use of ranks to indicate order relationships is convenient in practice and has a kind of historical authority. It seems, however, that order properties among the members of a sample or population are the more basic elements of a situation, and rank numbers are only convenient counts of certain types of order relationship. Although ranks are based on orders they are not, in customary statistical terminology, order statistics. In fact, an order statistic x_j is the variate-value of a member which is defined as being the jth in order of magnitude among a set of a certain size. When the ranks themselves are used we have a rank-order statistic.

31. Statistics based on rank-order form part of the class of statistics which are extremely useful in what is known (improperly) as non-parametric inference or (better) as distribution-free inference. The literature on this subject is now very large and Savage (1954 a) has recently prepared a bibliography of the subject. He has also (1954 b) reviewed the theory of rank-order statistics. Another useful survey is given by van Dantzig and Hemelrijk (1954). These three together relieve me of the necessity of an extended examination of the subject. It will be sufficient to mention the main currents in this broad stream of development.

(a) A number of homogeneity tests in two or more samples (Wilcoxon, 1945; Mann and Whitney, 1947; Terpstra, 1952; Kruskal and Wallis, 1952; Terry, 1952) have been developed using either linear or quadratic functions of ranks;

(b) Some general tests, including tests of independence have been further developed by Hoeffding (1948 a, 1948 b, 1952).

(c) Tests against trend by the use of ranks have been given by Mann (1945) Terpstra (1952) and others.

(d) Various investigations have been made into the power of such tests. General results have been obtained by Hoeffding (1951, 1952) and Lehmann (1951, 1953), and specific results for particular tests or particular alternatives given by various writers. (van Dantzig, 1951; Terry, 1952; van der Vaart, 1950; van der Waerden, 1952, Sundrum, 1953).

Paired Comparisons

32. Finally, something may be said of the method of paired comparisons, of which ranking may be regarded as a particular case. Kendall and Babington Smith (1940) discussed the method at some length, derived some tests of significance in the null case, defined coefficients of consistency d in a paired-comparison set and of agreement among judges, and tabulated some of the distributions for low values of m (the number of judges) and n (the number of objects). Moran (1947) proved a conjecture of Kendall and Babington

Smith concerning the moments of the coefficient of consistency d in the null case and showed that the distribution tends to normality for large n; and later (1950 a) he extended the results for the expectation of d in the non-null case, remarking that the variance has not been reduced to an easily calculable form although theoretically it is obtainable.

33. Ehrenberg (1952) also considers the coefficient of agreement u. Kendall and Babington Smith had given the moments and exact distributions for low m and n in the null case and a χ^2 approximation for larger values, which Ehrenberg corrects and rederives. In the non-null case he develops a line of approach suggested independently by Babington Smith (1950, discussion on the paper by A.S.C.Ross, J.R.Stat.Soc., B, 12, 54). If the probability of ranking the ith object higher than the jth is p_{ij}, the expected value of u is

$$E(u) = 1 - \frac{4}{\binom{m}{2}} \sum_{i<j}^{m} p_{ij} (1 - p_{ij}) \tag{28}$$

and the distribution of u tends to normality with either large m or large n, with variance

$$\text{var } u = \frac{8}{\binom{m}{2}\binom{n}{2}^2} \sum_{i<j} \left[(m-1)\{p_{ij}(1-p_{ij})\} - 4p_{ij}^2(1-p_{ij})^2 + 2p_{ij}^2(1-p_{ij})^2 \right] . \tag{29}$$

There is an exception to this rule: if all the p's are equal to $\frac{1}{2}$ the distribution tends for fixed n to a χ^2 distribution. One may also suspect, from what is known about the skewness of the distributions of correlation coefficients, that the tendency of u to normality may be rather slow. An investigation of this point would be useful.

34. An approach of a different kind to the paired comparison problem has been worked out in some detail by Bradley and Terry (1952). They suppose that the n objects have "true" ratings or preferences expressed by the numbers $\pi_1 \ldots \pi_n$ with $\sum_{i=1}^{n} \pi_i = 1$ and that when a pair is compared, the probability that i is preferred to j is $\pi_i/(\pi_i + \pi_j)$. This renders the problem parametric in terms of only n parameters. Bradley and Terry estimate the parameters by maximum likelihood, derive tests of significance and tabulate a number of the functions of ranks required. (Further tables are in preparation.) A scaling approach had been considered by Guttman (1946) and Mosteller (1951 a and b). Bradley (1954) has examined the justification of the model.

35. One of the practical problems of paired comparisons arises from the fact that with n objects $\frac{1}{2}$ n(n-1) comparisons are possible. This may be an inconveniently large number and it is natural to inquire whether the essential object can be achieved by making fewer comparisons in some systematic way. In a separate communication

(1954) I have discussed some of the problems and suggested methods of using incomplete block designs. Durbin (1951) had previously considered a similar problem in ranked data. The same communication discusses the relationship between paired-comparison data and forced rankings based on them.

Bibliography

Benard, A. and van Elteren, P. (1953). A generalization of the method of m rankings. Koninklijke Nederlandse Akademie van Wetenschappen A,56,No. 4 and Indag. Math., 15, No. 4.

Bradley, R. A. (1954). Incomplete Block Rank Analysis. Biometrics.

Bradley, R. A. and Terry, M. E. (1952). The rank analysis of incomplete block designs. Biometrika, 39, 324.

Chambers, E. G. (1946). Statistical techniques in applied psychology. Biometrika, 33, 269.

Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. Biometrika, 33, 129.

Daniels, H. E. (1948). A property of rank correlations. Biometrika, 35, 416.

Daniels, H. E. (1950). Rank correlation and population models. J. Roy. Statist. Soc., B. 12, 171.

Daniels, H. E. (1951). Note on Durbin and Stuart's formula for E (r_s). J. Roy. Statist. Soc., B. 13, 303.

Daniels, H. E. and Kendall, M. G. (1947). The significance of rank correlations where parental correlation exists. Biometrika, 34, 197.

Dantzig, G. B. (1939). On a class of distributions that approach the normal distribution function. Ann. Math. Stat., 10, 247.

David, S. T. and others (1951). Some question of distribution in the theory of rank correlation. Biometrika, 38, 131.

Dixon, W. J. (1953). Power functions of the sign test and power efficiency for normal alternatives. Ann. Math. Stat., 24, 467.

Dubois, P. (1939). Formulas and tables for rank correlation. Psych. Rec., 3, 46.

Durbin, J. (1951). Incomplete blocks in ranking experiments. Brit. J. Psych. (Stat. Section) 4, 85.

Durbin, J. and Stuart, A. (1951). Inversions and rank correlation coefficients. J. Roy. Statist. Soc., B, 13, 303.

Eells, W. C. (1929). Formulas for probable errors of coefficients of correlation. Jour. Am. Stat. Ass., 24, 170.

Ehrenberg, A. S. C. (1952). On sampling from a population of rankers. Biometrika, 39, 82.

Esscher, F. (1924). On a method of determining correlation from the ranks of variates. Skand. Akt., 7, 201.

Feller, W. (1945). Limit theorems in probability. Bull. Am. Math. Soc., 51, 800.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Jour. Am. Stat. Ass., 32, 675.

Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. Ann. Math. Stat., 11, 86.

Greiner, R. (1909). Uber das Fehlersystem der Kollektivmasslehre. Zeitschrift fur Mathematik and Physik, 57, 121, 225 and 337.

Guttman, L. (1946). An approach for quantifying paired comparisons and rank order. Ann. Math. Stat., 17, 144.

Haden, H. G. (1947). A note on the distribution of the different orderings of n objects. Proc. Camb. Phil. Soc., 43, 1.

Hoeffding, W. (1947). On the distribution of the rank correlation coefficient τ when the variates are not independent. Biometrika, 34, 183.

Hoeffding, W. (1948 a). A class of statistics with asymptotically normal distribution. Ann. Math. Stat., 19, 293.

Hoeffding, W. (1948 b). A non-parametric test of independence. Ann. Math. Stat., 19, 546.

Hoeffding, W. (1951). Optimum non-parametric tests. Second Berkeley Symposium, 83.

Hoeffding, W. (1952). The large sample power of tests based on permutations of observations. Ann. Math. Stat., 23, 169.

Hotelling, H., and Pabst, M. R. (1936). Rank correlations and tests of significance involving no assumptions of normality. Ann. Math. Stat., 7, 29.

Irwin, J. O. (1925). The further theory of Francis Galton's individual difference problem. Biometrika, 17, 100.

Kaarsemaker, L., and van Wijngaarden, A. (1952). Tables for use in rank correlation. Report R 73, Computation Department of the Mathematical Centre, Amsterdam.

Kendall, M. G. (1938). A new measure of rank correlation. Biometrika, 30, 81.

Kendall, M. G. (1942 a). Partial rank correlation. Biometrika, 32, 277.

Kendall, M. G. (1942 b). Note on the estimation of a ranking. J. R. Stat. Soc., 105, 119.

Kendall, M. G. (1943). The advanced Theory of Statistics, Vol. 1 (5th edition, 1952).

Kendall, M. G. (1945). The treatment of ties in ranking problems. Biometrika, 33, 239.

Kendall, M. G. (1947). The variance of τ when both rankings contain ties. Biometrika, 34, 297.

Kendall, M. G. (1949). Rank and product-moment correlation. Biometrika, 36, 177.

Kendall, M. G. (1954). Further contributions to paired comparisons. In press.

Kendall, M. G., and others (1939). The distribution of Spearmen's coefficient of rank correlation, etc. Biometrika, 30, 251.

Kendall, M. G. and Babington Smith, B. (1939). The problem of m rankings. Ann. Math. Stat., 10, 275.

Kendall, M. G., and Babington Smith, B. (1940). On the method of paired comparisons. Biometrika, 31, 324.

Kruskal, W. H. (1952). A non-parametric test for the several sample problems. Ann. Math. Stat., 23, 525.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. J. Am. Stat. Ass., 47, 583.

Lehmann, E. L. (1951). Consistency and unbiasedness of certain non-parametric tests. Ann. Math. Stat., 22, 155.

Lehmann, E. L. (1953). The power of rank tests. Ann. Math. Stat., 24, 23.

Mann, H. B. (1945). Non-parametric tests against trend. Econometrica, 13, 245.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is larger than the other. Ann. Math. Stat., 15, 50.

Moran, P. A. P. (1947). On the method of paired comparisons. Biometrika, 34, 363.

Moran, P. A. P. (1948 a). Rank correlation and permutation distributions. Proc. Camb. Phil. Soc., 44, 142.

Moran, P. A. P. (1948 b). Rank correlation and product-moment correlation. Biometrika, 35, 203.

Moran, P. A. P. (1950 a). Recent developments in ranking theory. J. Roy. Stat. Soc. (B) 12, 153.

Moran, P. A. P. (1950 b). A curvilinear ranking test. J. Roy. Stat. Soc., 12, 292.

Moran, P. A. P. (1951). Partial and multiple rank correlation. Biometrika, 38, 26.

Mosteller, F. (1951 a,b,c). Remarks on the method of paired comparisons. I. Psychometrika, 15, 3; II ibid. 16, 203; III ibid. 16, 207.

Olds, E. G. (1938). Distribution of sums of squares of rank differences for small numbers of individuals. Ann. Math. Stat., 9, 133.

Pearson, K. (1907). Mathematical contributions to the theory of evolution. XVI. On further methods of determining correlation. Drapers' Co. Res. Mem., Biometric Series IV. Cambridge University Press.

Pearson, K. (1914, 1921). On an extension of the method of correlation by grades or ranks. Biometrika, 10, 416, and Second note, ibid., 13, 302.

Pearson, K., and Pearson, M. V. (1931, 1932). On the mean character and variance of an individual, and on the mean and variance of the intervals between ranked individuals. Biometrika, 23, 364, and 23, 203.

Pitman, E. J. G. (1937, 1938). Significance tests which may be applied to samples from any populations. Supp. J. R. Stat. Soc., 4, 119. II. The correlation coefficient test. Supp. J. R. Stat. Soc., 4, 225. III. The analysis of variance test. Biometrika, 29, 332.

Savage, I. R. (1952). Bibliography of non-parametric statistics and related topics. Nat. Bureau Standards. Rep. 182.

Savage, I. R. (1954). Rank order statistics. Memorandum for private circulation.

Sillitto, G. P. (1947). The distribution of Kendall's τ coefficient of rank correlation in ranking containing ties. Biometrika, 34, 36.

Silverstone, H. (1950). A note on the cumulants of Kendall's S-distribution. Biometrika, 37, 231.

Smith, B. Babington, (1950). Discussion of Professor Ross's paper. J. Roy. Stat. Soc., B, 12, 54.

Spearman, C. (1904). The proof and measurement of association between two things. Am. Jour. Psych., 15, 88.

Spearman, C. (1906). A footrule for measuring correlation. Brit. Jour. Psych., 2, 89.

Spearman, C. (1910). Correlation calculated from faulty data. Brit. J. Psych., 3, 271.

Stuart, A. (1953). The estimation and comparison of strengths of association in contingency tables. Biometrika, 40, 105.

Stuart, A. (1954). The correlation between variate-values and ranks in samples from a continuous distribution. Brit. J. Psych. Stat. Section.

"Student" (1921). An experimental determination of the probable error of Dr. Spearman's correlation coefficient. Biometrika, 13, 263.

Sundrum, R. M. (1953). Theory and Applications of Distribution-Free Methods in Statistics. Ph.D. Thesis. London University.

Sundrum, R. M. (1953 a). Moments of the rank correlation coefficient τ in the general case. Biometrika, 40, 409.

Sundrum, R. M. (1953 b). A method of systematic sampling based on order properties. Biometrika, 40, 452.

Sundrum, R. M. (1953 c). The power of Wilcoxon's 2-sample test. J. Roy. Stat. Soc. B, 15, 246.

Terpstra, T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend when ties are present in one ranking. Proc. Kon. Ned. Ak. Wet. A. 55, Indag. Math. 14, 327.

Terry, M. E. (1952). Some rank order tests which are most powerful against specific parametric alternatives. Ann. Math. Stat. 23, 346.

Van Dantzig, D. (1951). On the consistency and the power of Wilcoxon's two sample test. Proc. Kon. Ned. Ak. Wet. A, 54.

Van Dantzig, D. and Hemelrikj, J. (1954). Statistical methods based on few assumptions. Proc. Int. Inst. Stat. Rome.

Van der Vaart, H. R. (1950). Some remarks on the power function of Wilcoxon's test for the problem of two samples. I and II. Proc. Kon. Ned. Ak. Wet.,53, 494 and 507.

Van der Waerden, B. L. (1952). Order tests for the two-sample problem and their power. Proc. Kon. Ned. Ak. Wet., 55.

Wallis, W. A. (1939). The correlation ratio for ranked data. Jour. Am. Stat. Ass., 34, 533.

Watkins, G. P. (1933). An ordinal index of correlation. Jour. Am. Stat. Ass., 28, 139.

Welch, B. L. (1937). On the z-test in randomized blocks and Latin Squares. Biometrika, 29, 21.

Whitfield, J. W. (1947). Rank correlation between two variables, one of which is ranked, the other dichotomous. Biometrika, 34, 292.

Whitfield, J. W. (1950). Uses of the ranking method in psychology. J. Roy. Stat. Soc., 12, 163.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics, 1, 80.

Woodbury, M. A. (1940). Rank correlation when there are equal variates. Ann. Math. Stat., 11, 358.