A NEW MEASURE OF RANK CORRELATION

By M. G. KENDALL

1. In psychological work the problem of comparing two different rankings of the same set of individuals may be divided into two types. In the first type the individuals have a given order A which is objectively defined with reference to some quality, and a characteristic question is: if an observer ranks the individuals in an order B, does a comparison of B with A suggest that he possesses a reliable judgment of the quality, or, alternatively, is it probable that B could have arisen by chance? In the second type no objective order is given. Two observers consider the individuals and rank them in orders A and B. The question now is, are these orders sufficiently alike to indicate similarity of taste in the observers, or, on the other hand, are A and B incompatible within assigned limits of probability? An example of the first type occurs in the familiar experiments wherein an observer has to arrange a known set of weights in ascending order of weight; the second type would arise if two observers had to rank a set of musical compositions in order of preference.

The measure of rank correlation proposed in this paper is capable of being applied to both problems, which are, in fact, formally very much the same. For purposes of simplicity in the exposition it has, however, been thought convenient to preserve a distinction between them.

Definition of τ

2. Consider a set of individuals, numbered from 1 to 10, whose objective order is that of the natural sequence 1, 2, 3, ..., 10, and consider an arbitrary ranking such as the following:

$$4 \quad 7 \quad 2 \quad 10 \quad 3 \quad 6 \quad 8 \quad 1 \quad 5 \quad 9$$

Consider the order of the nine pairs of numbers obtained by taking the first number 4, with each succeeding number. The first pair, 4 7, is in the correct order (in the sequence of 1, 2, ..., 10), and we therefore allot it a score $+1$. The second pair, 4 2, is in the wrong order and we score -1. The third pair, 4 10, scores $+1$, and so on, the nine scores being

$$+1-1+1-1+1+1-1+1+1, \text{ totalling } +3.$$

Consider also the scores of the second number, 7, with its eight succeeding numbers. They are

$$-1+1-1-1+1-1-1+1, \text{ totalling } -2.$$

The scores of the third number are

$$+1+1+1+1-1+1+1, \text{ totalling } +5.$$

Proceeding thus with each number, we have 9 scores, as follows

$$+3, \quad -2, \quad +5, \quad -6, \quad +3, \quad 0, \quad -1, \quad +2, \quad +1.$$

The total of these scores is $+5$.

Now the maximum score, obtained if the numbers are all in the objective order (1, 2, ..., 10), is 45. I therefore define a rank correlation coefficient between a variable ranked in the objective order (1, 2, ..., 10) and the variable ranked in the order above as

$$\tau = \frac{\text{actual score}}{\text{maximum possible score}} = \frac{5}{45} = +0{\cdot}11.$$

Generally, if there are n individuals, the maximum score, obtained if and only if they are all in objective order is $(n-1)+(n-2)+\ldots+1 = \frac{n(n-1)}{2}$. Denoting the actual score for any given ranking by Σ, we may calculate a measure of the rank correlation between this ranking and the objective ranking by putting

$$\tau = \frac{2\Sigma}{n(n-1)}. \qquad \ldots\ldots(1)$$

Two short methods for the calculation of τ

3. τ is calculable more easily than might appear at first sight from the above approach. Consider for example the order given above, viz.

$$4 \quad 7 \quad 2 \quad 10 \quad 3 \quad 6 \quad 8 \quad 1 \quad 5 \quad 9$$

We see that the number 1 has two numbers on its right and 7 on its left. We therefore score $+2-7 = -5$, and then strike out the 1, being left with

$$4 \quad 7 \quad 2 \quad 10 \quad 3 \quad 6 \quad 8 \quad 5 \quad 9$$

The number 2 has 6 numbers on its right and two on its left and hence we score $6-2 = +4$. We then strike out the 2 and proceed with the 3 and so on. It will be found that the scores obtained are

$$-5, \quad +4, \quad +1, \quad +6, \quad -3, \quad 0, \quad +3, \quad 0, \quad -1.$$

The total of these scores is $+5$, and is equal to Σ.

The above rule is quite general. Its validity will be evident when it is noted that instead of taking the first number with each succeeding number and so on, as in § 2, we consider the pairs contributing to Σ in a different way. Taking the number 1 first, and remembering that all the other numbers are greater than 1, we see that any number on the left must contribute -1 to Σ and any number on the right contributes $+1$. When 1 is struck out the procedure remains valid for 2, and so on.

4. Alternatively, the following procedure may be adopted:

Considering once again the order

$$4 \quad 7 \quad 2 \quad 10 \quad 3 \quad 6 \quad 8 \quad 1 \quad 5 \quad 9$$

we see that the first number, 4, has on its right 6 numbers which are greater. The second number, 7, has on its right 3 numbers which are greater. The third number, 2, has on its right 6 numbers which are greater; and so on. The numbers so obtained are

$$6, \quad 3, \quad 6, \quad 0, \quad 4, \quad 2, \quad 1, \quad 2, \quad 1$$

totalling 25.

There must, therefore, be $45-25 = 20$ numbers lying to the right of successive numbers in the order which are less than those numbers, and hence

$$\begin{aligned} \Sigma &= 25-20 \\ &= +5, \text{ as before.} \end{aligned}$$

Generally, if the number obtained by the above method of counting greater numbers is k

$$\Sigma = 2k - \frac{n(n-1)}{2}.$$

In practice, I find this method convenient and rapid. It has, moreover, the advantage of providing an independent check; for if the process is repeated counting greater numbers which lie *to the left*, giving a total of, say, l,

$$\Sigma = \frac{n(n-1)}{2} - 2l.$$

5. The use of τ can now be extended to the case where no objective order exists. In fact, given two rankings, A and B, of the same set of individuals τ may be defined as the coefficient obtained by regarding one order, A, as an objective order. If, for example, the orders are as follows:

A	6	9	4	3	5	10	2	1	8	7
B	6	5	10	2	3	9	7	4	1	8

τ is given by first rearranging A as an objective order, writing below it the corresponding member in B, thus

A'	1	2	3	4	5	6	7	8	9	10
B'	4	7	2	10	3	6	8	1	5	9

and then calculating τ in the manner of preceding paragraphs. Actually, as will be seen below, it is not necessary in any practical calculations to rewrite the orders in this way.

6. It is a notable fact that the same coefficient τ is reached whichever of the two orders, A and B, is rearranged as an objective order.

Consider again the orders given in the preceding paragraph, namely,

A'	1	2	3	4	5	6	7	8	9	10
B'	4	7	2	10	3	6	8	1	5	9

Rearranging B as an objective order we have

A''	8	3	5	1	9	6	2	7	10	4
B''	1	2	3	4	5	6	7	8	9	10

If we repeat this operation on the A'' and B'' we shall get back to A' and B'. A', B' and A'', B'' are thus reciprocally related and the permutations B' and A'' may be said to be *conjugate*.

We have to show that τ is the same when calculated from B' when A' is the objective ranking as when calculated from A'' when B'' is the objective ranking, i.e. that Σ is the same for two conjugate permutations with regard to an objective order $1, 2, \ldots, n$.

In § 2, the value of Σ for B' was ascertained directly, the various items entering into the sum being

$$+3, \quad -2, \quad +5, \quad -6, \quad +3, \quad 0, \quad -1, \quad +2, \quad +1.$$

Consider now the value of Σ for A'' *obtained by the short method of* § 3.

The sums entering into Σ will be found to be

$$+3, \quad -2, \quad +5, \quad -6, \quad +3, \quad 0, \quad -1, \quad +2, \quad +1,$$

i.e. exactly the same as those for B' obtained by the more direct method; and hence Σ and τ are the same in the two cases.

This result is true in general. If the permutation B' begins with a number a_0 the contribution to $\Sigma_{B'}$ from pairs involving a_0 will be $(n-a_0)-(a_0-1)$. In A'' the a_0th number will be 1 and the contribution to $\Sigma_{A''}$ will also be $(n-a_0)-(a_0-1)$, in the manner of § 3. If the second number in B' is a_1 the contribution to $\Sigma_{B'}$ will be $(n-a_1)-(a_1-1)\pm 1$ according to whether a_1 is greater than a_0 or not. In A'' the a_1th number will be 2, and the contribution to $\Sigma_{A''}$ is also $(n-a_1)-(a_1-1)\pm 1$ according to whether 1 lies to the left or the right of 2 in A'', i.e. whether a_1 is greater than a_0 or not; and so on.

Thus Σ and τ are the same for two conjugate permutations with regard to the objective order $1, 2, \ldots, n$.

7. In practical cases, the value of τ may be found as follows:

Write down above the given rankings the objective ranking. In the example already considered this would give

	1	2	3	4	5	6	7	8	9	10
A	6	9	4	3	5	10	2	1	8	7
B	6	5	10	2	3	9	7	4	1	8

The number 1 in B has an 8 above it in A. In the objective ranking 8 has two numbers to the right and seven to the left. Score, therefore, -5 and strike out the 8 in the objective ranking. The number 2 in B has a 3 above it in A, and 3 in the objective ranking has six numbers to its right (ignoring the number struck out) and two to its left, score $+4$; and so on, the scores being

$$-5, \quad +4, \quad +1, \quad +6, \quad -3, \quad 0, \quad +3, \quad 0, \quad -1,$$

totalling $+5$, which is equal to Σ.

8. τ satisfies certain elementary requirements of a measure of rank correlation. It is $+1$ if and only if correspondence between the rankings of A and B is perfect. It is -1 if and only if the rankings are exactly inverted. For intermediate values it appears to provide a satisfactory measure of the correspondence between the two rankings. A few examples for $n = 10$ will give some idea of the scale of measurement which it provides (an objective order 1, 2, ..., 10 is taken in each case):

Order	τ	ρ*
4 7 2 10 3 6 8 1 5 9	+0·11	+0·14
1 6 2 7 3 8 4 9 5 10	+0·56	+0·64
7 10 4 1 6 8 9 5 2 3	−0·24	−0·37
6 5 4 7 3 8 2 9 10 1	+0·02	+0·03
10 1 2 3 4 5 6 7 8 9	+0·60	+0·45
10 9 8 7 6 1 2 3 4 5	−0·56	−0·76

In the case where no objective ranking exists τ measures the closeness of correspondence between two given rankings in the sense that it measures how accurate either ranking would be if the other were objective. In other words it measures the *compatibility* of two rankings.

9. For the purpose of measuring correlation between ranks, therefore, τ appears to compare favourably with ρ. It is admitted that ρ can take $\frac{n^3-n}{6}$ values between -1 and $+1$, whereas τ can take only $\frac{n^2-n}{2}$ values in the range. This does not, however, appear to constitute a serious disadvantage to the sensitivity of τ.

On the other hand, τ possesses one marked advantage over ρ, in that it is not difficult to find the distribution of values obtained by correlating a given ranking with the members of a universe in which all possible rankings occur equally

* Throughout this paper ρ means the Spearman coefficient of rank correlation defined by

$$\rho = 1 - \frac{6S(d^2)}{n^3 - n},$$

where d is a difference in ranks.

frequently. It is shown below that the distribution of τ tends to normality for large n, resembling ρ in this respect; but in fact τ is surprisingly close to normality even for low values of n, whereas the distribution for ρ has not yet been given, and appears to present peculiar features.*

The sampling distribution of Σ

10. To judge the significance of an observed value of τ or of Σ in the case where an objective order is given, we wish to know whether the value could have arisen by chance from a universe in which all the possible rankings of the n objects occur an equal number of times. It is, therefore, necessary to consider the distribution of Σ in such a universe. The distribution of τ may be found at once from that of Σ by dividing the variate values of Σ by $\frac{n(n-1)}{2}$.

The same distribution may be used to judge the significance of a value of τ expressing the compatibility of two rankings. A significantly negative τ, for example, would mean that if one ranking is taken to be objective the other has not, as judged by the τ-distribution, arisen by chance from the universe in which all possible rankings occur equally frequently; in other words that the two rankings are significantly incompatible.

Consider then the universe of values of Σ obtained from an objective order 1, 2, 3, ..., n and the $n!$ possible permutations of the first n integers. Let the number of values of a given Σ be denoted by $u_{n,\Sigma}$. Consider a given ranking of the numbers 1, ..., n, and the effect of inserting an additional number $(n+1)$ in the various possible places in this ranking, from the first place (preceding the first member of the rank of n) to the last place (following the last member of the rank of n).

Inserting the number $(n+1)$ at the beginning will add $-n$ to the value of Σ. Inserting it between the first and second members will add $-(n-2)$ to Σ. Inserting it between the second and third will add $-(n-4)$ to Σ, and so on. Adding the number $(n+1)$ at the end will add $+n$ to Σ.

It follows that

$$\begin{aligned} u_{n+1,\Sigma} = {} & u_{n,\Sigma-n} + u_{n,\Sigma-n+2} + u_{n,\Sigma-n+4} \\ & + \ldots + u_{n,\Sigma+n-4} + u_{n,\Sigma+n-2} + u_{n,\Sigma+n}. \end{aligned} \qquad \ldots\ldots(2)$$

This recursion formula permits of the calculation of the frequency array of Σ.

11. If $n = 2$, there are two values of Σ, $+1$ and -1, i.e. $u_{2,-1} = u_{2,1} = 1$, $u_{2,0} = 0$. From (2) we have

$$u_{3,\Sigma} = u_{2,\Sigma-4} + u_{2,\Sigma-2} + u_{2,\Sigma} + u_{2,\Sigma+2} + u_{2,\Sigma+4},$$

* The fact that ρ tends to normality for large n has recently been proved by Hotelling & Pabst (1936). The remarks above on the behaviour of ρ for low values of n are founded on an expression for the sampling distribution of ρ which will be discussed in a further communication shortly to be published. This communication will also deal with the relation between τ and ρ.

the possible values of Σ ranging from -3 to $+3$. By substituting $\Sigma = -3, \ldots +3$ in the above, we find

$$u_{3,3} = 1, \quad u_{3,2} = 0, \quad u_{3,1} = 2, \quad u_{3,0} = 0,$$

and similar values for the negative values of Σ.

Applying equation (2) again we find

$$u_{4,6} = 1, \quad u_{4,5} = 0, \quad u_{4,4} = 3, \quad u_{4,3} = 0,$$
$$u_{4,2} = 5, \quad u_{4,1} = 0, \quad u_{4,0} = 6, \quad \text{etc.}$$

The successive arrays of Σ may in fact be built up by the following process:

1	1					
	1	1				
		1	1			
1	2	2	1			
	1	2	2	1		
		1	2	2	1	
			1	2	2	1
1	3	5	6	5	3	1
etc.						

At each stage, to find the array for $(n+1)$ we write down the n-array $(n+1)$ times, one under the other and moving one place to the right each time, and then sum the $(n+1)$ arrays. If the total array has a central value, that value is the frequency for $\Sigma = 0$, and all values of Σ must be even. If the total array has two central values, these values are the frequencies for $\Sigma = \pm 1$, and all values of Σ must be odd.

12. The above procedure may be condensed by forming a kind of figurate triangle as follows:

1	1										
2	1	1									
3	1	2	2	1							
4	1	3	5	6	5	3	1				
5	1	4	9	15	20	22	20	15	9	4	1
etc.	etc.										

In this array, a number in the rth row is the sum of the number immediately above it and the $(r-1)$ numbers to the left of that number. The formation of the array is quite simple and several devices shorten the arithmetic. For instance, in part of the array towards the left a number in the rth row is the sum of the number immediately above it and the number immediately to the left. A check is provided by the fact that the total in the rth row is $r!$.

The following table shows the frequency distribution of Σ for values of n from 1 to 10.

TABLE I

Distribution of Σ for values of n from 1 to 10 (only the positive half of the symmetrical distribution shown)

Values of Σ

Values of n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1														
2	0	1													
3	0	2	0	1											
4	6	0	5	0	3	0	1								
5	22	0	20	0	15	0	9	0	4	0	1				
6	0	101	0	90	0	71	0	49	0	29	0	14	0	5	0
7	0	573	0	531	0	455	0	359	0	259	0	169	0	98	0
8	3836	0	3736	0	3450	0	3017	0	2493	0	1940	0	1415	0	961
9	29228	0	28675	0	27073	0	24584	0	21450	0	17957	0	14395	0	11021
10	0	250749	0	243694	0	230131	0	211089	0	187959	0	162337	0	135853	0

Values of Σ

Values of n	15	16	17	18	19	20	21	22	23	24	25	26	27	28
6	1													
7	49	0	20	0	6	0	1							
8	0	602	0	343	0	174	0	76	0	27	0	7	0	1
9	0	8031	0	5545	0	3606	0	2191	0	1230	0	628	0	285
10	110010	0	86054	0	64889	0	47043	0	32683	0	21670	0	13640	0

Values of Σ

Values of n	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
9	0	111	0	35	0	8	0	1									
10	8095	0	4489	0	2298	0	1068	0	440	0	155	0	44	0	9	0	1

The frequency polygon of the distribution is quite close to normality even for $n = 6$. For $n = 10$ the correspondence is very good over the material part of the range, as may be judged roughly by drawing the frequency polygon to Σ and the normal curve with the same area and standard deviation. On an ordinary scale the two curves are hardly distinguishable by the eye above $\Sigma = 5$.

STANDARD ERROR OF τ

13. A little consideration of the above method of obtaining the frequency distribution of Σ will show that the distribution may be arrayed by the function:

$$f = (x^{-1}+x)(x^{-2}+1+x^2)(x^{-3}+x^{-1}+x+x^3)\dots (x^{-(n-1)}+x^{-(n-3)}+\dots+x^{(n-3)}+x^{(n-1)}). \qquad \dots\dots(3)$$

The coefficient of x^{Σ} in f is the frequency of Σ in the distribution.

If we differentiate f with respect to x and then multiply by x the coefficient of x^{Σ} is multiplied by Σ. Writing then θ for the operator $x\frac{\partial}{\partial x}$ we have

$$\mu_0\mu_1 = (\theta f)_{x=1},$$

and generally

$$\mu_0\mu_r = (\theta^r f)_{x=1}. \qquad \dots\dots(4)$$

Applying equation (4) when $r = 2$, I find

$$\mu_2 = \frac{n(n-1)(2n+5)}{18}, \qquad \dots\dots(5)$$

and hence the standard error of τ is given by

$$\sigma_\tau = \frac{1}{3}\sqrt{\frac{2(2n+5)}{n(n-1)}}, \qquad \dots\dots(6)$$

which, as n becomes large, gives

$$\sigma_\tau \sim \frac{2}{3}\cdot\frac{1}{\sqrt{n}}. \qquad \dots\dots(7)$$

Table II shows the proportion of the total frequencies falling outside ranges $\pm\sigma$, $\pm 2\sigma$, $\pm 3\sigma$ for some of the distributions of Table I.

The expected values on the hypothesis of a normal distribution are 0·3173, 0·0455, 0·0027 and it is clear that for most practical purposes in testing the significance of an observed τ for $n = 10$ or greater, the standard error may be used in the ordinary way.

14. Applying equation (4) when $r = 4$, I find

$$\mu_4 = \frac{n(n-1)}{2}\left\{1+\frac{74}{9}(n-2)+\frac{37}{6}(n-2)(n-3)+\frac{32}{25}(n-2)(n-3)(n-4) + \frac{2}{27}(n-2)(n-3)(n-4)(n-5)\right\}. \qquad \dots\dots(8)$$

TABLE II

Proportion of frequencies of the distribution of Σ falling in certain ranges

n	σ_Σ	Proportion falling outside range		
		$\pm\sigma$	$\pm 2\sigma$	$\pm 3\sigma$
6	5·32	0·272	0·056	0·0000
7	6·66	0·381	0·030	0·0004
8	8·08	0·275	0·031	0·0004
9	9·59	0·359	0·045	0·0009
10	11·18	0·291	0·047	0·0009

From this β_2 may be obtained and it is evident that as n becomes large β_2 tends to the value 3. In fact it remains below that value, so that the distribution of Σ and therefore of τ is slightly platykurtic. The following table shows the values of β_2 for some values of n. The corresponding values of β_2 for the distribution of ρ are also given and it will be observed that, as judged by β_2, the approach of τ to normality is appreciably quicker than that of ρ.

TABLE III

Values of β_2 in the distribution of Σ and of ρ for certain values of n

n	$\beta_2(\Sigma)$	$\beta_2(\rho)$
5	2·53	2·07
10	2·78	2·54
20	2·89	2·77
30	2·93	2·85

In general, as will be seen below, the moment of order $2s$ is a polynomial of degree n^{3s}.

Proof of the normality of τ for large n

15. We shall prove that as $n \to \infty$

$$\mu_{2s} \sim \frac{(2s)!}{2^s s!}(\mu_2)^s,$$

where μ_{2s} is the $2s$th moment of the distribution of Σ. In virtue of the symmetry of the distribution moments of odd order vanish and it follows from the Second Limit Theorem of Probability (see Fréchet & Shohat, 1931) that the distribution of Σ, and hence that of τ, tends to normality in the sense that the frequency between τ_1 and τ_2 tends to

$$\frac{1}{\sigma\sqrt{(2\pi)}}\int_{\tau_1}^{\tau_2} e^{-\frac{x^2}{2\sigma^2}}\,dx.$$

16. Consider the effect of operating on the product f (equation (3)) by $\theta \equiv x\dfrac{\partial}{\partial x}$. The first operation will result in a sum of terms of type

$$\{-rx^{-r}-(r-2)\,x^{-(r-2)}-\ldots+(r-2)\,x^{(r-2)}+rx^r\}$$

multiplied by the remaining terms of f unchanged. When x is put equal to unity we may write this as the sum of terms

$$\frac{-r-(r-2)-\ldots+(r-2)+r}{1+1+\ldots+1+1}\,n! = \frac{-r-(r-2)-\ldots+(r-2)+r}{r}\,n!.$$

Similarly the second operation will bring out terms like

$$\frac{r^2+(r-2)^2+\ldots+(r-2)^2+r^2}{r}\,n!$$

and $$n!\left\{\frac{-r-(r-2)-\ldots+(r-2)+r}{r}\right\}\left\{\frac{-t-(t-2)-\ldots+(t-2)+t}{t}\right\}.$$

Generally, operating $2s$ times will bring out terms like

$$n!\left\{\frac{r^{2s}+(r-2)^{2s}+\ldots+(r-2)^{2s}+r^{2s}}{r}\right\},$$

$$n!\left\{\frac{-r^{2s-1}-(r-2)^{2s-1}-\ldots+(r-2)^{2s-1}+r^{2s-1}}{r}\right\} \times\left\{\frac{-t-(t-2)-\ldots+(t-2)+t}{t}\right\},$$

etc.

When x is put equal to unity any term beginning with an odd superscript in the powers will vanish. Consider now the sum of terms like

$$n!\left\{\frac{r^2+\ldots+r^2}{r}\right\}\left\{\frac{t^2+\ldots+t^2}{t}\right\}\ldots\left\{\frac{u^2+\ldots+u^2}{u}\right\}, \qquad \ldots\ldots(9)$$

containing s factors.

It will be proved below that this term contributes the greatest power of n to the total sum giving $\mu_0\mu_{2s}$.

Further, in virtue of the multinomial form of Leibniz' theorem, the factor by which this term is multiplied in the expansion of $(\theta^{2s}f)$ is

$$\frac{(2s)!}{2!\,2!\ldots 2!} = \frac{(2s)!}{2^s}.$$

Hence, since $\mu_0 = n!$ we have

$$\mu_{2s} \sim \frac{(2s)!}{2^s} \{\text{sum of terms like (9)}\}. \qquad \ldots\ldots(10)$$

Each term in (9) is of type

$$\frac{1}{r}\{r^2 + (r-2)^2 + \ldots + (r-2)^2 + r^2\},$$

i.e. is of order $\frac{r^2}{3}$. The summation will therefore tend to the sum of terms like $\frac{1}{3^s}\{1^2 . 2^2 . \ldots . s^2\}$, each term containing s squares of the numbers 1, 2, ..., $(n-1)$. Call this Π_s.

Then Π_s is $\frac{1}{s!}$ times the sum of terms in

$$\frac{1}{3^s}\{1^2 + 2^2 + \ldots + (n-1)^2\}^s, \qquad \ldots\ldots(11)$$

which contain s different factors.

Now (11) is of order $\frac{n^{3s}}{9^s} \sim (\mu_2)^s$. Hence if the product term Π_s tends to the sum (11)

$$\Pi_s \sim \frac{(\mu_2)^s}{s!},$$

and in virtue of (10)

$$\mu_{2s} \sim \frac{(2s)!}{2^s} \cdot \frac{(\mu_2)^s}{s!}.$$

To complete the demonstration, we have therefore to show that (11) tends asymptotically to the sum of its terms $s!\,\Pi_s$, i.e. that sums of terms like

$$1^4 . 2^2 . \ldots . (s-1)^2, \quad 1^6 . 2^2 . \ldots . (s-2)^2$$

tend in comparison to zero.

This may be shown inductively.

Consider first of all

$$\{1^2 + 2^2 + \ldots + (n-1)^2\}^2 = 2\Pi_2 + 1^4 + 2^4 + \ldots + (n-1)^4.$$

The expression on the left $\sim \frac{n^6}{9}$. But the sum of fourth powers on the right $\sim \frac{n^5}{5}$, which is of lower order. Hence the sum on the right $\sim 2\Pi_2$. Multiplying by $\{1^2 + 2^2 + \ldots + (n-1)^2\}$ we have

$$\{1^2 + 2^2 + \ldots + (n-1)^2\}^3 \sim 2\Pi_2\{1^2 + 2^2 + \ldots + (n-1)^2\}$$
$$\sim 6\Pi_3 + \text{terms of type } 1^4 2^2.$$

These terms will be less in sum than

$$2\{1^2 + 2^2 + \ldots + (n-1)^2\}\{1^4 + 2^4 + \ldots + (n-1)^4\},$$

which $\sim 2 \cdot \frac{n^3}{3} \cdot \frac{n^5}{5}$, of order 8. But the expression on the left is of order 9. Hence $\{1^2 + 2^2 + \ldots + (n-1)^2\}^3 \sim 6\Pi_3$ and so on.

We can now justify the assertion that the maximum power of n arises from terms like $(1^2 \cdot 2^2 \cdot \ldots \cdot s^2)$. In fact, by a similar line of reasoning to that just given it will be seen that sums of terms of type $\{1^4 \cdot 2^2 \cdot \ldots \cdot (s-1)^2\}$, etc. are of lower order.

The demonstration is complete.

17. It appears therefore that the coefficient τ has a good claim to serious consideration as a measure of rank correlation. It is easily calculable. In the important case of the distribution wherein all possible rankings occur equally frequently its standard error is known; for the values of n likely to be required in practice it may be taken to be normally distributed; and where there is doubt the distribution can be obtained in an exact form.

It should also be remarked that τ has a natural significance. An observer who is given a set of objects (such as coloured discs) to rank appears to follow a process something like this: First of all he searches for the beginning of the series, say the disc of lightest shade. Having selected a disc, he compares it with each of the remainder to verify the propriety of his choice. The coefficient τ gives him one mark for each comparison which is made correctly, and subtracts a mark for each error.* When the first disc is selected, he proceeds as before with a second; and so on. τ follows this process exactly. It appears to be a logical measure of ranking carried out by the process and should therefore prove useful in psychological work.

REFERENCES

FRÉCHET, M. & SHOHAT, J. (1931). "A proof of the generalised second limit theorem in the theory of probability." *Trans. Amer. Math. Soc.* **33**, 533.

HOTELLING & PABST (1936). "Rank correlation and tests of significance involving no assumptions of normality." *Ann. Math. Statist.* **7**, 29.

* Inasmuch as comparisons between extremes in the series will generally be easier than comparisons between neighbouring members it might in some cases be preferable to weight the marking given to different comparisons according to some selected scale. The determination of such a scale, however, would depend to some extent on the circumstances of individual cases and would present considerable difficulty where no objective order is known to exist, apart from adding greatly to the complexity of the distribution of the measure so obtained.