THE PROBLEM OF m RANKINGS

By M. G. Kendall and B. Babington Smith

1. **Introduction.** If n objects are ranked by m persons according to some quality of the objects there arises the problem: does the set of m rankings of n show any evidence of community of judgment among the m individuals? For example, if a number of pieces of poetry are ranked by students in order of preference, do the rankings support the supposition that the students have poetical tastes in common, and if so is there any strong degree of unanimity or only a faint degree?

The problem in its full generality permits of no assumption about the nature of the quality according to which the objects are ranked, other than that ranking is possible. No hypothesis is made that the quality is measurable, still less that there is some underlying frequency distribution to the quantiles of which the rankings correspond. The quality is to be thought of as linear in the sense that any two objects possessing it are either coincident or may be put in the relation "before and after." A metric may, of course, be imposed on this linear space by convention; but the relationship between objects is invariant under any transformation which stretches the scale of measurement. In particular, it is not a condition of the problem that the ranking shall be based on a distribution according to a normal variate.

It is permissible to denote the rankings by the *ordinal* numbers 1, 2, $\cdots$ n; but it is not permissible, without further discussion, to operate on these numbers as if they were cardinals. This point seems to have been inadequately appreciated. For instance, when $m = 2$ we have the familiar case of rank correlation between a pair of rankings; and this is frequently treated by subtracting corresponding ranks, squaring, and forming the Spearman coefficient

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

To justify this procedure it is necessary to explain what is meant, for example, by such a process as (4th minus 8th), or what the square of this difference of ordinal numbers represents.

It is worth stressing that the necessary transition from ordinals to cardinals can be made without invoking a scale of measurement. When we rank an object as first we mean, in effect, that no member of the set of n is preferred to it; when we rank it as the rth we mean that $(r - 1)$ objects are preferred to it. The ordinals of the ranking are then biunivocally related to the cardinals expressing the number of objects which are preferred. It is thus legitimate to apply the laws of cardinal arithmetic to them. For example, if an object A is ranked r_1 by Brown, r_2 by Jones and r_3 by Robinson we may form the sum $(r_1 + r_2 + r_3)$, which is to be interpreted as meaning that, taking the preferences of the three persons together, there were $(r_1 + r_2 + r_3 - 3)$ cases in which some other object was preferred to A. The point is of some importance, in view of the prevailing practice of replacing ranks by quantiles of the normal distribution—a practice which cannot always be regarded as justifiable and is sometimes little short of desperate.

To fix the ideas, consider the following three rankings of six objects

Object:	A	B	C	D	E	F
	5	4	1	6	3	2
	2	3	1	5	6	4
	4	1	6	3	2	5
Sum of ranks	11	8	8	14	11	11

We may sum the ranks for each object, as shown. These sums (which must add to 63, and in general to $mn(n + 1)/2$) reflect the degree of resemblance among the rankings. If the resemblance were perfect the sums would be 3, 6, 9, 12, 15, 18 (though not necessarily, of course in that order) and in such a case would be as different as possible. On the other hand, when there is little or no resemblance, as in the example given, the sums are approximately equal. It is thus natural to take the variance of these sums as providing some measure of the concordance in the rankings. If S is the observed sum of squares of the deviations of sums of ranks from the mean value $m(n + 1)/2$ (i.e. is n times the variance) we may write

$$W = \frac{12S}{m^2(n^3 - n)} \tag{2}$$

and call W the coefficient of concordance. Here $m^2(n^3 - n)/12$ is the maximum possible value of S, occurring if there is complete unanimity in the rankings, so that W may vary from 0 to 1. In the example given, $S = 25.5$, $W = 0.16$.

The coefficient W has arisen in several ways.

(a) W is simply related to the average of the $\binom{m}{2}$ Spearman rank correlation coefficients between pairs of the m rankings. It is easy to show that the average ρ is given by

$$\rho_{av} = \frac{\dfrac{12S}{n^3 - n} - m}{m^2 - m} \tag{3}$$

$$= \frac{mW - 1}{m - 1} \tag{4}$$

ρ_{av} has been considered by Kelley [3] as a measure of average intercorrelation in rankings, but he gives no results for testing the significance of observed values.

It is to be noted that whereas W may vary from 0 to 1, ρ_{av} may vary from $-1/(m-1)$ to 1, i.e. it is asymmetrical like the coefficient of intraclass correlation, to which it bears some resemblance.[1]

(b) Friedman [1] has considered a quantity χ_r^2 related to W by the equation

$$\chi_r^2 = m(n-1)W. \tag{5}$$

(c) Welch [6] and Pitman [5] have considered the problem of the distribution of variance in an array

$$\begin{matrix} a_1, a_2, \cdots a_n \\ b_1, b_2, \cdots b_n \end{matrix}$$

etc., for permutations of the numbers a, b, etc. in rows.

This is more general than the ranking case, in which $a_1 \cdots a_n$, $b_1 \cdots b_n$ etc. reduce to permutations of the numbers $1 \cdots n$. Since S', the total sum of squares in an array of m rankings of n, is $m^2(n^3 - n)/12$, we have

$$W = \frac{S}{S'} \tag{6}$$

i.e. the ratio of variance between columns to the total variance.

2. **Significance of W.** To test whether an observed value of W is significant it is necessary to consider the distribution of W (or, more conveniently, of S) in the universe observed by permuting the n ranks in all possible ways. No generality is lost by supposing one ranking fixed, and the others will then give rise to $(n!)^{m-1}$ values of S.

The actual distribution of W (or S), as will be seen below, is irregular for low values of m and n, and likely to be quite irregular for moderate values. It may, however, be shown that the first four moments of W are

$$\mu_1' \text{ (about 0)} = \frac{1}{m} \tag{7}$$

$$\mu_2 = \frac{2(m-1)}{m^3(n-1)} \tag{8}$$

$$\mu_3 = \frac{8(m-1)(m-2)}{m^5(n-1)^2} \tag{9}$$

$$\mu_4 = \frac{24(m-1)}{m^7(n-1)^2}\left\{\frac{25n^3 - 38n^2 - 35n + 72}{25(n^3 - n)} + 2(n-1)(m-2) + \frac{n+3}{2}(m-2)(m-3)\right\}. \tag{10}$$

[1] The Spearman rank correlation coefficient is the product-moment coefficient of correlation between the ranks considered as ordinary variate values. ρ_{av} is the intraclass correlation coefficient for the m sets of ranks, also considered as variate values.

Results equivalent to these for the first three moments were given by Friedman [1]; and for the first four moments by Pitman [5].

In a valuable contribution to the subject Friedman showed that the distribution of χ_r^2 tends to that of χ^2 with $(n - 1)$ degrees of freedom as m tends to infinity and suggested the use of χ_r^2 (equation (5)) for an ordinary test of significance in the χ^2 distribution. This is satisfactory for moderately large values, but for small values it is subject to the disadvantage inherent in any attempt to represent a distribution of finite range by one of infinite range—the fit near the tails is not likely to be very good. An improvement is obtained by noting that the first four moments of the Type I distribution,

$$df = \frac{1}{B(p, q)} W^{p-1}(1 - W)^{q-1} \tag{11}$$

are approximately those of W if m and n are moderately large, and

$$p = \frac{n-1}{2} - \frac{1}{m} \tag{12}$$

$$q = (m - 1)\left\{\frac{n-1}{2} - \frac{1}{m}\right\}. \tag{13}$$

For practical purposes it is most convenient to put

$$z = \tfrac{1}{2} \log_e \frac{(m-1)W}{1 - W} \tag{14}$$

so that z can be tested in Fisher's distribution with $(n - 1) - \frac{2}{m}$ $(= n_1)$ and $(m - 1)\left\{(n - 1) - \frac{2}{m}\right\}$ $(= n_2)$ degrees of freedom.

There can be little doubt that this test is quite reliable for moderate values of m and n; but it has hitherto been far from clear how reliable it is for low values of m and n. This point we attempt to clear up in the present paper.

3. **Distribution of S.** For the case $m = 2$ the distribution of S is the same as the distribution of the $S(d^2)$ used in calculating Spearman's rank correlation coefficient. A table showing the distribution up to and including $n = 8$ has already been given (Kendall and others, [4]). Tables giving probabilities that specified values of χ_r^2 would be attained or exceeded were given by Friedman [1] for $n = 3$, $m = 2$–9; and $n = 4$, $m = 2$–4. We have taken this work somewhat further and obtained the distributions for $n = 3$, $m = 2$–10; $n = 4$, $m = 2$–6; and $n = 5$, $m = 3$. Tables 1–4 give the probabilities based on these distributions.

These distributions were obtained by two methods. The first consisted of building up the distribution for $(m + 1)$ and n from that of m and n. For

TABLE 1

Probability that a given value of S will be attained or exceeded, for $n = 3$ and values of m from 2 to 10

Values of m

S	2	3	4	5	6	7	8	9	10
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	.866	.944	.931	.954	.956	.964	.967	.971	.974
6	.500	.528	.653	.691	.740	.768	.794	.814	.830
8	.167	.361	.431	.522	.570	.620	.654	.685	.710
14		.194	.273	.367	.430	.486	.531	.569	.601
18		.028	.125	.182	.252	.305	.355	.398	.436
24			.069	.124	.184	.237	.285	.328	.368
26			.042	.093	.142	.192	.236	.278	.316
32			.0046	.039	.072	.112	.149	.187	.222
38				.024	.052	.085	.120	.154	.187
42				.0085	.029	.051	.079	.107	.135
50				$.0^{3}77$	.012	.027	.047	.069	.092
54					.0081	.021	.038	.057	.078
56					.0055	.016	.030	.048	.066
62					.0017	.0084	.018	.031	.046
72					$.0^{3}13$	.0036	.0099	.019	.030
74						.0027	.0080	.016	.026
78						.0012	.0048	.010	.018
86						$.0^{3}32$	.0024	.0060	.012
96						$.0^{3}32$	.0011	.0035	.0075
98						$.0^{4}21$	$.0^{3}86$	.0029	.0063
104							$.0^{3}26$	.0013	.0034
114							$.0^{4}61$	$.0^{3}66$	.0020
122							$.0^{4}61$	$.0^{3}35$	.0013
126							$.0^{4}61$	$.0^{3}20$	$.0^{3}83$
128							$.0^{5}36$	$.0^{4}97$	$.0^{3}51$
134								$.0^{4}54$	$.0^{3}37$
146								$.0^{4}11$	$.0^{3}18$
150								$.0^{4}11$	$.0^{3}11$
152								$.0^{4}11$	$.0^{4}85$
158								$.0^{4}11$	$.0^{4}44$
162								$.0^{6}60$	$.0^{4}20$
168									$.0^{4}11$
182									$.0^{5}21$
200									$.0^{7}99$

TABLE 2

Probability that a given value of S will be attained or exceeded for n = 4 and m = 3 and 5

S	$m = 3$	$m = 5$	S	$m = 5$
1	1.000	1.000	61	.055
3	.958	.975	65	.044
5	.910	.944	67	.034
9	.727	.857	69	.031
11	.608	.771	73	.023
13	.524	.709	75	.020
17	.446	.652	77	.017
19	.342	.561	81	.012
21	.300	.521	83	.0087
25	.207	.445	85	.0067
27	.175	.408	89	.0055
29	.148	.372	91	.0031
33	.075	.298	93	.0023
35	.054	.260	97	.0018
37	.033	.226	99	.0016
41	.017	.210	101	.0014
43	.0017	.162	105	$.0^{3}64$
45	.0017	.141	107	$.0^{3}33$
49		.123	109	$.0^{3}21$
51		.107	113	$.0^{3}14$
53		.093	117	$.0^{4}48$
57		.075	125	$.0^{5}30$
59		.067		

example, with $m = 2$ and $n = 3$ we have the following values of the sums of ranks, measured about their mean:

Type			Frequency
−2	0	2	1
−2	1	1	2
−1	0	1	2
0	0	0	1

Here −2, 1, 1, and 2, −1, −1 are taken to be identical types, for they give the same value of S and will also give similar types when we proceed to $m = 3$ as follows.

In the case $m = 3$ each of the above type will appear added to the six permutations of −1, 0, 1; e.g. the type −2, 0, 2 will give one each of −3, 0, 3; −3, 1, 2;

TABLE 3

Probability that a given value of S will be attained or exceeded for $n = 4$ and $m = 2$, 4 and 6

S	$m = 2$	$m = 4$	$m = 6$	S	$m = 6$
0	1.000	1.000	1.000	82	.035
2	.958	.992	.996	84	.032
4	.833	.928	.957	86	.029
6	.792	.900	.940	88	.023
8	.625	.800	.874	90	.022
10	.542	.754	.844	94	.017
12	.458	.677	.789	96	.014
14	.375	.649	.772	98	.013
16	.208	.524	.679	100	.010
18	.167	.508	.668	102	.0096
20	.042	.432	.609	104	.0085
22		.389	.574	106	.0073
24		.355	.541	108	.0061
26		.324	.512	110	.0057
30		.242	.431	114	.0040
32		.200	.386	116	.0033
34		.190	.375	118	.0028
36		.158	.338	120	.0023
38		.141	.317	122	.0020
40		.105	.270	126	.0015
42		.094	.256	128	$.0^{3}90$
44		.077	.230	130	$.0^{3}87$
46		.068	.218	132	$.0^{3}73$
48		.054	.197	134	$.0^{3}65$
50		.052	.194	136	$.0^{3}40$
52		.036	.163	138	$.0^{3}36$
54		.033	.155	140	$.0^{3}28$
56		.019	.127	144	$.0^{3}24$
58		.014	.114	146	$.0^{3}22$
62		.012	.108	148	$.0^{3}12$
64		.0069	.089	150	$.0^{4}95$
66		.0062	.088	152	$.0^{4}62$
68		.0027	.073	154	$.0^{4}46$
70		.0027	.066	158	$.0^{4}24$
72		.0016	.060	160	$.0^{4}16$
74		$.0^{3}94$	.056	162	$.0^{4}12$
76		$.0^{3}94$	.043	164	$.0^{5}80$
78		$.0^{3}94$	.041	170	$.0^{5}24$
80		$.0^{4}72$	.037	180	$.0^{6}13$

TABLE 4

Probability that a given value of S will be attained or exceeded, for n = 5 and m = 3

S	$m = 3$	S	$m = 3$
0	1.000	44	.236
2	1.000	46	.213
4	.988	48	.172
6	.972	50	.163
8	.941	52	.127
10	.914	54	.117
12	.845	56	.096
14	.831	58	.080
16	.768	60	.063
18	.720	62	.056
20	.682	64	.045
22	.649	66	.038
24	.595	68	.028
26	.559	70	.026
28	.493	72	.017
30	.475	74	.015
32	.432	76	.0078
34	.406	78	.0053
36	.347	80	.0040
38	.326	82	.0028
40	.291	86	$.0^3 90$
42	.253	90	$.0^4 69$

$-2, -1, 3$; $-2, 1, 1$; $-1, -1, 2$; and $-1, 0, 1$. These types are then counted for each of the four basic types of $m = 2$ and we get:

Type			Frequency
−3	0	3	1
−3	1	2	6
−2	0	2	6
−2	1	1	6
−1	0	1	15
0	0	0	2

The case $m = 4$ is treated by considering the numbers of types obtained by adding the six permutations of $-1, 0, 1$ to the types for $m = 3$; and so on.

This method is quite convenient for $n = 2$ and $n = 3$. For $n = 4$ it becomes difficult owing to the labour of considering 24 permutations at each stage and to the increase in the number of types. For $n = 5$ there are 120 permutations and the labour becomes excessive.

The second method employed is a generalisation of a procedure we used for the Spearman coefficient. Taking first of all the case $m = 2$, consider the array

$$\begin{matrix} a^2 & a^3 & a^4 & \cdots & a^{(n+1)} \\ a^3 & a^4 & a^5 & \cdots & a^{(n+2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a^{(n+1)} & a^{(n+2)} & a^{(n+3)} & \cdots & a^{2n} \end{matrix}$$

Any permissible set of values of the sums of ranks is obtained by selecting n entries from this array so that no entry appears more than once in the same row or column. If then, subtracting from each index the mean $(n + 1)$ and squaring, we write

$$E = \left\{ \begin{matrix} a^{(n-1)^2} & a^{(n-2)^2} & \cdots & a^1 & a^0 \\ a^{(n-2)^2} & a^{(n-3)^2} & \cdots & a^0 & a^1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a^0 & a^1 & \cdots & a^{(n-2)^2} & a^{(n-1)^2} \end{matrix} \right\} \tag{15}$$

the values of S are the powers of a in E when it is expanded as a sum of $n!$ terms each of which is obtained by multiplying n factors which do not appear in the same row or column. The distribution of S is arrayed by the expansion of E, the number of values of any S being the coefficient of a^S in the expansion.

Similarly, for m rankings, the distribution of S is given by the expansion of an m- dimensional E-function. For example, with $m = 3$ there would be a three-dimensional E-function the bottom plane of which would be

$$\begin{matrix} a^{\left\{3-\frac{3(n+1)}{2}\right\}^2} & a^{\left\{4-\frac{3(n+1)}{2}\right\}^2} & \cdots & a^{\left\{n+2-\frac{3(n+1)}{2}\right\}^2} \\ a^{\left\{4-\frac{3(n+1)}{2}\right\}^2} & a^{\left\{5-\frac{3(n+1)}{2}\right\}^2} & \cdots & a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} \\ \cdots & \cdots & \cdots & \cdots \\ a^{\left\{n+2-\frac{3(n+1)}{2}\right\}^2} & a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} & \cdots & a^{\left\{2n+2-\frac{3(n+1)}{2}\right\}^2} \end{matrix}$$

The plane above this would be

$$\begin{matrix} a^{\left\{4-\frac{3(n+1)}{2}\right\}^2} & \cdots & a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} \\ \cdots & \cdots & \cdots \\ a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} & \cdots & a^{\left\{2n+3-\frac{3(n+1)}{2}\right\}^2} \end{matrix}$$

and so on.

The E-function is difficult to handle in more than three dimensions, but for the two and three dimensional case it is manageable and we used it to obtain the distribution of S for $n = 5$ and $m = 3$.

4. **Adequacy of the z-test.** Tables 1–4 provide exact tests for the values of m and n there given. It remains to be seen how good the ordinary z-test applied to W would be for higher values. It may be presumed that if the test is satis-

factory for any particular value of m and n for which exact results are available, it will be so for higher values of m and n. Since, for ordinary purposes the significance points of z as tabled by Fisher and Yates [2] would be employed, the most useful comparison would seem to be between those tables and the extreme values of tables 1–4.

For $n = 3$, $m = 9$, the 1% level is given approximately by $S = 78$. We have, testing for such a value, $W = 0.4814$, $z = 1.002$, $n_1 = \frac{16}{9}$, $n_2 = \frac{128}{9}$. By linear interpolation of reciprocals in the tables of z we find for the 1% point and these degrees of freedom $z = 0.954$. The correspondence is hardly satisfactory, and the z test might lead to incorrect inferences in practice. Matters improve a good deal, however, if we make continuity corrections, by subtracting unity from S before calculating W, and increasing by two the divisor $m^2(n^3 - n)/12$, so as to allow for the finite range. In this case $z = 0.979$.

For $n = 4$, $m = 6$ the 1% point is approximately $S = 100$. We have $W = 0.5556$, $z = 0.916$; $n_1 = 8/3$, $n_2 = 40/3$. By linear interpolation as before we find $z = 0.888$.

Continuity corrections again materially improve the agreement, giving a value of $z = 0.893$.

For $n = 5$ $m = 3$ there is no very convenient value of S close to the 1% point. For $P = 0.015$ $S = 74$ and for $P = 0.0078$ $S = 76$.

For $S = 74$ (with continuity corrections) $z = 1.020$
$S = 76$ (" " ") $z = 1.089$

By interpolation from the tables $z = 1.075$. The use of the z test would lead to the correct conclusion that a value of S equal to 74 falls below, and that of 76 above, the 1% point.

For values of m and n not included in Tables 1–4 it thus appears that the z-test with continuity corrections will give sufficiently accurate results, if n is greater than 3, at the 1% points. It may be presumed that the results at the 5% points are equally good and probably better. But for finer values of significance, such as 0.1%, it is doubtful whether the test is sound. The tails of the distribution of S for moderate values of m and n are very irregular.

For instance, the following is the tail of the distribution of S for $n = 3, m = 10$ (the total distribution being 10,077,696):

S	Frequency	S	Frequency
96	11,340	146	740
98	30,090	150	252
104	13,830	152	420
114	7,380	158	240
122	4,200	162	90
126	3,240	168	90
128	1,450	182	20
134	1,860	200	1

and the following is the tail for $n = 4$ $m = 6$ (the total being 7,962,624):

S	Frequency	S	Frequency	S	Frequency
100	5536	122	4100	146	810
102	8160	126	4480	148	225
104	10260	128	240	150	264
106	8850	130	1152	152	120
108	3920	132	660	154	180
110	13344	134	1980	158	60
114	5460	136	300	160	36
116	3870	138	600	162	30
118	3900	140	312	164	45
120	2472	144	100	170	18
				180	1

Irregularities of this kind run all through the distributions we have obtained, and frequency diagrams present the same sort of features we have noticed in the case $m = 2$ (Kendall and others, [4]). The representation of such distributions by continuous functions, no matter how close their lower moments, is obviously to be used with some care. Although the B-distribution or the associated z-distribution will give reasonable significance tests at levels of 1% or greater, they will probably be inadequate to represent frequencies occurring in narrow ranges.

5. Some Experimental Distributions. In some previous work we obtained a number of random permutations of the numbers 1–10 and 1–20. These were used to derive some experimental distributions of S which may be worth recording. Table 5 gives the distribution for 200 sets of pentads of 10 and Table 6 that for 100 triads of 20. In the distribution of Table 5, the mean of the grouped distribution is 404. The theoretical mean is 412.5 with a standard error of 12.3. In Table 6 the mean is 1936, the theoretical mean being 1995 with s.e. 53. The distributions accord quite well with expectation.

In conclusion we give two examples to illustrate some points of importance in ranking work. The first is a case in which ranks appear as the primary variate and in which the assumption of normality is clearly illegitimate.

6. Example 1. In some experiments in random series a pack of ordinary playing cards was shuffled and the order of the 13 cards of each suit from the top of the pack was noted. The pack was then reshuffled and again the orders noted. This was done 28 times. The question we wished to discuss was whether the shuffling was good, in the sense that the cards were thoroughly mixed at each shuffle.

Here, for each suit, say diamonds, we have 28 rankings of 13. The sums of ranks were 183, 137, 171, 207, 188, 160, 225, 174, 216, 192, 236, 239, 220. The mean is 196, and $S = 11522$, W (without continuity corrections, which are not

TABLE 5

Experimental Distribution of S in 200 sets (m = 5, n = 10)

S	Frequency
0–	1
50–	2
100–	7
150–	9
200–	21
250–	22
300–	24
350–	26
400–	20
450–	17
500–	12
550–	11
600–	10
650–	4
700–	5
750–	3
800–	3
...	..
1000–	2
...	..
1250–	1
Total	200

TABLE 6

Experimental Distribution of S in 100 sets (m = 3, n = 20)

S	Frequency
800–	4
1000–	8
1200–	8
1400–	6
1600–	12
1800–	15
2000–	20
2200–	12
2400–	6
2600–	5
2800–	0
3000–	3
3200–	0
3400–	1
Total	100

worth making for these values of m and n) = 0.08075, z (equation (14)) = 0.432. This falls just beyond the 1% point.

Similarly for the clubs W was found to be 0.0535; for the hearts, 0.0245; and for spades, 0.0342. None of these values is significant and we conclude that the randomisation introduced by the shuffling was good, at all events, so far as this test was concerned. It may be added that the shuffling was done with much more care than would be taken in an ordinary game of cards.

In psychological work there has sometimes been a confusion between the determination of a measure of agreement between subjects and that of an objective order based on experimental rankings. It may therefore be as well to point out that in its psychological applications the test of W is one of concordance between judgments. There may be quite a high measure of agreement about something which is incorrect.

7. **Example 2.** A number of students were given 12 photographs of persons unknown to them, and asked to rank them in what they judged from the photographs to be their intelligence. For 16 students the sums of ranks were

112, 94, 101, 84, 97, 75, 104, 84, 102, 146, 125, 124

The mean is 104. $S = 4472$, $W = 0.1222$. $z = 0.368$, and is barely significant, being between the 1% and the 5% points.

For 111 students the sums were

818, 670, 908, 410, 706, 526, 780, 485, 596, 1044, 959, 756

$$W = 0.2378, z = 1.768$$

This is highly significant and it is to be inferred that community of judgment exists between students or groups of students. But there was little relationship between the judgments and the intelligence of the photographed subjects as given by the Binet Intelligence Quotient.

Note added in proof:

While this paper was passing through the press Professor W. Allen Wallis, of Stanford University, kindly drew our attention to some unpublished work of his own on this subject. Professor Wallis had also arrived at the coefficient W which, he points out, is the ranking analogue of the correlation ratio. His paper is, we understand, on the point of publication.

REFERENCES

[1] M. Friedman, "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance." *Jour. Am. Stat. Assn.*, Vol. 32(1937), p. 675.

[2] R. A. Fisher, and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 1938, Oliver & Boyd, Edinburgh.

[3] T. L. Kelley, *Statistical Methods*, 1927.

[4] M. G. Kendall, Sheila F. H. Kendall, and Bernard Babington Smith, "The Distribution of Spearman's Coefficient of Rank Correlation in a Universe in which all Rankings Occur an Equal Number of Times," *Biometrika*, Vol. 30(1938), p. 251.

[5] E. J. G. Pitman, "Significance Tests which may be applied to Samples from any Populations: III. The analysis of variance." *Biometrika*, Vol. 29(1938), p. 322.

[6] B. L. Welch, "On the z- test in Randomised Blocks and Latin Squares," *Biometrika*, Vol. 29(1937), p. 21.

LONDON, ENGLAND
AND
UNIVERSITY OF ST. ANDREWS,
SCOTLAND.