PSYCHOMETRIKA—VOL. 17, NO. 4
DECEMBER, 1952

THE AVERAGE SPEARMAN RANK CORRELATION COEFFICIENT*

SAMUEL B. LYERLY

WASHINGTON, D. C.

A method is derived for finding the average Spearman rank correlation coefficient of N sets of ranks with a single dependent or criterion ranking of n items without computing any of the individual coefficients. Procedures for calculating the exact distribution of ρ_{av} for small values of N and n are described for the null case. The first four moments about zero of this distribution are derived, and it is concluded that for samples as small as $N = 4$ and $n = 4$ the normal distribution can be used safely in testing the hypothesis $\rho_{av} = 0$.

The Spearman rank correlation coefficient ρ (sometimes called the "rank difference" coefficient after one method of calculating it) has been in use among psychologists for about half a century (**8**). Recent researches by Hotelling and Pabst (**2**), Kendall and his co-workers (**4**, **5**, **6**, **7**), and others has stimulated interest in rank correlation methods, largely because of their usefulness as non-parametric procedures—i.e., to provide tests of the null hypothesis in cases where the population distribution of either or both variables is unknown. Kendall's recent book (**7**) contains an excellent summary of rank correlation, including ρ and his own alternative coefficient τ, which in some respects is superior to ρ.

A short-cut method for computing the average of the intercorrelations of N ranked series each consisting of n items has been known for many years (**3**, 218). This method may be expressed in the following formula:

$$\text{Average inter-}\rho = 1 - \frac{N(4n + 2)}{(N - 1)(n - 1)} + \frac{12 \sum S^2}{N(N - 1)(n^3 - n)}, \qquad (1)$$

where

N = number of sets of rankings,
n = number of ranks in each set, and
S = sum of rank numbers for a given object or stimulus.

Kendall and Babington Smith discuss this formula and present the exact distribution of the mean intercorrelation in the null case for several small values of N and n (**6**; see also **7**, Chs. 6 and 7). Kendall's "coefficient of

*This problem first came to the writer's attention in discussions with Dr. Dean J. Clyde.

concordance," W, is a simple function of the average rank intercorrelation so designed that W ranges from 0 to 1 as the degree of agreement among the sets of ranks ranges from no agreement at all to perfect agreement. Approximate tests of significance of W (and hence of the average intercorrelation) based upon the z (or F) distribution or upon χ^2 are suggested for use with larger values of N or n where the computation of the exact probabilities is excessively laborious.

Occasionally there may arise a problem in which we are not interested in the average *inter*correlation of N sets of ranks, but we are concerned with the average correlation of N sets of ranks with a single dependent or criterion ranking. For example, we may ask N individuals to rank independently n objects of art according to their merit, and we may have as a criterion variable the rank-order listing of one or more experts. It is the purpose of this paper to derive a short method of calculating the average of the N ρ's without computing each ρ individually, and to investigate the problem of the significance of the average ρ in the null case.

We shall let y_i ($i = 1, 2, \cdots, n$) be the criterion or dependent set of ranks and x_{ij} ($i = 1, 2, \cdots, n; j = 1, 2, \cdots, N$) be the rank number assigned to the ith stimulus by the jth individual. Then the square of the difference in ranks for a single judgment $(y_i - x_{ij})^2$ is equal to $y_i^2 - 2y_i x_{ij} + x_{ij}^2$. The sum of these squares over a given i is

$$\sum_{j=1}^{N} (y_i - x_{ij})^2 = Ny_i^2 - 2y_i \sum_{j=1}^{N} x_{ij} + \sum_{j=1}^{N} x_{ij}^2 , \tag{2}$$

and the corresponding sum of squares over the whole table is

$$\sum_{i=1}^{n} \sum_{j=1}^{N} (y_i - x_{ij})^2 = \sum_{i=1}^{n} \sum_{j=1}^{N} x_{ij}^2 - 2 \sum_{i=1}^{n} \left(y_i \sum_{j=1}^{N} x_{ij} \right) + N \sum_{i=1}^{n} y_i^2 . \tag{3}$$

Here both $\sum \sum x^2$ and $N \sum y^2$ are N times the sum of squares of the first n natural numbers and hence equal to $Nn(n + 1)(2n + 1)/6$. Thus (3) reduces to

$$\sum_{i=1}^{n} \sum_{j=1}^{N} (y_i - x_{ij})^2 = 2 \sum_{i=1}^{n} \sum_{j=1}^{N} x_{ij}^2 - 2 \sum_{i=1}^{n} \left(y_i \sum_{j=1}^{N} x_{ij} \right)$$

$$= \frac{2Nn(n + 1)(2n + 1)}{6} - 2 \sum_{i=1}^{n} \left(y_i \sum_{j=1}^{N} x_{ij} \right), \tag{4}$$

and the average sum of squares of rank differences over the N individuals is

$$\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{N} (y_i - x_{ij})^2 = \frac{2n(n + 1)(2n + 1)}{6} - \frac{2 \sum_{i=1}^{n} \left(y_i \sum_{j=1}^{N} x_{ij} \right)}{N} . \tag{5}$$

Substituting this average value of the sum of squares of differences into the usual formula, $\rho = 1 - 6 \sum d^2/(n^3 - n)$, we have

$$\rho_{av} = 1 - \frac{6\left[2n(n+1)(2n+1)/6 - 2 \sum_{i=1}^{n} \left(y_i \sum_{j=1}^{N} x_{ij}\right)/N\right]}{n^3 - n}$$

$$= 1 - \frac{2(2n+1)}{n-1} + \frac{12 \sum_{i=1}^{n} \left(y_i \sum_{j=1}^{N} x_{ij}\right)}{N(n^3 - n)} . \qquad (6)$$

The following fictitious example will illustrate the computation of ρ_{av} :

y	x				Σx	$y \Sigma x$
	1	2	3	4		
1	3	1	2	2	8	8
2	1	2	3	1	7	14
3	4	4	4	3	15	45
4	2	3	1	4	10	40
			Sums:		40	107

Here $N = 4$ and $n = 4$, and

$$\rho_{av} = 1 - \frac{2(8+1)}{3} + \frac{12 \times 107}{4 \times 60} = .35.$$

It may easily be verified that ρ_{av} is the product-moment correlation coefficient of all Nn pairs of criterion-judgment ranks. This is a consequence of the fact that the distributions of all subsamples (individuals' judgments) are identical and the criterion distribution is likewise identical for all individuals; hence the "total correlation" is equivalent to the mean correlation.

The distribution of ρ_{av} in the null case for any values of N and n can be found exactly, although a considerable amount of arithmetic is involved if either N or n exceeds 7 or 8. Since ρ_{av} is the average of n ρ's, we need only the distribution of ρ in the null case for the appropriate n, and then by ordinary combinatorial methods the distribution of means (or sums) of N samples from this "population" can be found. Perhaps the most difficult part of the process is finding the initial distribution of ρ itself, although Kendall (**5**,**7**) has described methods of finding it and has tabulated the exact distributions of ρ (or of $\sum d^2$) for $n = 3$ to $n = 8$, inclusive. We shall illustrate the method of finding the distribution of ρ_{av} for $n = 3$ and $N = 2$ by starting with the distribution of ρ for $n = 3$ as given by Kendall.

The distribution of ρ in the null case for $n = 3$ is as follows:

ρ	Relative frequency
1.00	1
.50	2
.00	0
− .50	2
−1.00	1

It should be noted that this is a discrete distribution and that only the listed values of ρ are possible, since the sum of squares of differences of 3 pairs of integers can take only a limited number of values. It will also be noted that the distribution is symmetrical about zero. This will obviously be true of the average ρ in the null case.

To find the distribution of ρ_{av} for $N = 2$, we merely calculate the distribution of all possible combinations of 2 from the above table. A ρ_{av} of + 1.00 may be obtained in only one way, i.e., when both single ρ's are + 1.00. The next highest possible ρ_{av} is + .75, which is obtained when one ρ is + 1.00 and the other is + .50. The relative frequency of this combination is 4, since the sequence + 1.00, + .50 can occur in two ways and so can the sequence + .50, + 1.00. The complete distribution may be obtained readily by constructing a table as follows:

	1	2	0	2	1				
1	1	2	0	2	1				
2		2	4	0	4	2			
0			0	0	0	0	0		
2				2	4	0	4	2	
1					1	2	0	2	1
Sums:	1	4	4	4	10	4	4	4	1

In this table the entries are all the possible products of pairs of elements of the original distribution. Each row is displaced one space to the right of its predecessor so that the new distribution can be obtained by merely adding columns. The total frequency is $(n!)^N$; and, tabulating the distribution in terms of ρ_{av}, we have:

ρ_{av}	Relative frequency
1.00	1
.75	4
.50	4
.25	4
.00	10
− .25	4
− .50	4
− .75	4
−1.00	1
Total	36

For small values of n and N the above method for finding the distribution of ρ_{av} is quite feasible. For large n and N the method is cumbersome, largely because the initial distribution of ρ itself is difficult to obtain. In such circumstances it is natural to inquire whether a good approximation can be found which will be satisfactory.

It is known that for large n the distribution of ρ itself approaches normality, but it is not known precisely how large n should be in order to use the normal integral for testing purposes. Kendall (7), with some hesitation, suggests 20 as a minimum n for which the normal curve may safely be used, and proposes that $\rho\sqrt{(n-2)/(1-\rho^2)}$, treated as "Student's" t with $n-2$ degrees of freedom, provides a better test in the range $8 < n < 20$. (For $n \leq 8$ exact probabilities have been computed.)

In the present case, where we have N sets of ranks, it is reasonable to suppose that the approach to normality should be fairly rapid. Since ρ_{av} is the mean of N values of ρ, and since ρ is approximately normally distributed, it would follow from the Central Limit Theorem that, for fixed n, the distribution of ρ_{av} would approach the normal as N increases.

The variance of ρ in the null case is $1/(n-1)$. Its fourth moment about zero (2) is

$$\mu_4 = \frac{3(25n^3 - 38n^2 - 35n + 72)}{25n(n+1)(n-1)^3}.$$

Here $\beta_1 = 0$ (all odd moments are zero by virtue of the symmetry of the distribution), and

$$\beta_2 = 3 + \frac{24}{100}\left(\frac{36 - 5n - 19n^2}{n^3 - n}\right),$$

which approaches the normal value of 3 as n increases.

In samples of N from such a population, we calculate from well-known theorems:

$$\sigma^2 = \frac{1}{N(n-1)},$$

$$\mu_4 = \frac{3(25n^3 - 38n^2 - 35n + 72)}{25N^3n(n+1)(n-1)^3} + \frac{3(N-1)}{N^3(n-1)^2},$$

and hence

$$\beta_2 = 3 + \frac{24}{100N}\left(\frac{36 - 5n - 19n^2}{n^3 - n}\right).$$

Thus the distribution of ρ_{av} approaches normality, as judged from its first four moments, very rapidly as n and N increase. Table 1 lists β_2 for certain small values of n and N.

TABLE 1

β_2 for Small Values of n and N

n	N					
	1	2	3	4	5	6
2	1.00	2.00	2.33	2.50	2.60	2.67
3	1.50	2.25	2.50	2.62	2.70	2.75
4	1.85	2.42	2.62	2.71	2.77	2.81
5	2.07	2.54	2.69	2.77	2.81	2.85
6	2.23	2.61	2.74	2.81	2.84	2.87
7	2.35	2.67	2.78	2.84	2.87	2.89

From Table 1 it can readily be seen that the normal value of 3 for β_2 is approximated very nearly for even small values of N and n. As a check on the usefulness of the normal distribution for testing the null hypothesis, the exact distribution of ρ_{av} for $N = 4$ and $n = 4$ was calculated (these values are probably as low as most experimenters would ever need to use), and the .005, .010, .025, and .050 points determined both from the exact relative frequencies and by using the normal approximation based upon the variance $1/N(n-1)$. (The four significance points were chosen because they provide one-tail or two-tail tests of the null hypothesis at the 1% or 5% levels.) Table 2 lists the relative frequencies in the tail of this distribution. From Table 2 we see, incidentally, that the ρ_{av} of .35 obtained in the fictitious example above is not significant at the 5% level, since a value of .50 is required for the one-tail test and .60 for the two-tail test.

TABLE 2

A Portion of the Distribution of ρ_{av} for $N = 4$ and $n = 4$

ρ_{av}	Relative frequency	Cumulative frequency	% cumulative frequency
1.00	1	1	.000003
.95	12	13	.00004
.90	58	71	.0002
.85	160	231	.001
.80	347	578	.002
.75	704	1282	.004
.70	1194	2476	.007
.65	1852	4328	.013
.60	2885	7213	.022
.55	3968	11181	.034
.50	5544	16725	.050
.45	7196	23921	.072
....			

Total Frequency = 331776

Table 3 lists the significant values of ρ_{av} as calculated from the exact values in Table 2 and the significant values as estimated from the normal approximation.

TABLE 3

Significant Points of the Distribution of ρ_{av} for $N = 4$ and $n = 4$

Significance level	Normal Approximation: By formula	Normal Approximation: Next higher possible value	Exact value
.005	.744	.75	.75
.010	.671	.70	.70
.025	.566	.60	.60
.050	.475	.50	.50

From Table 3 we see that for N and n as low as 4, the use of the normal integral in testing the significance of ρ_{av} at the 1% or 5% levels, using either the one-tail or the two-tail test, results in the same decision with respect to accepting or rejecting the null hypothesis as the use of the exact distribution. Since ρ_{av} approaches normality even more closely with larger N and n, and since most experimental problems would involve larger N or n or both, we are on safe ground in concluding that the normal curve approximation is appropriate in testing the null hypothesis of a zero correlation in the population (or, more explicitly, of an average ρ of zero with the criterion in the population of sets of ranks from which the sample of sets was drawn).

In the non-null case, i.e., when the population ρ_{av} is not zero, no exact significance test is known, since the distribution of ρ itself is unknown for such populations. Thus we cannot test the hypothesis that an observed ρ_{av} is a sample from a hypothetical population in which the average ρ is some value other than zero, nor can we make an exact test of the difference between two sample values of ρ_{av}. Such tests must await the development of feasible methods of calculating or approximating the distribution of ρ for non-zero population values.*

REFERENCES

1. Friedman, Milton. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. statist. Ass.*, 1937, **32**, 675–701.
2. Hotelling, H., and Pabst, M. R. Rank correlation and tests of significance. *Ann. math. Statist.*, 1936, **7**, 29–43.
3. Kelley, T. L. Statistical method. New York: The Macmillan Co., 1923.
4. Kendall, M. G. The advanced theory of statistics, Vol. I, 4th Ed. London: Charles Griffin and Co., 1948.
5. Kendall, M. G., Kendall, S. F. H., and Babington Smith, B. The distribution of Spearman's coefficient of rank correlation. *Biometrika*, 1938, **30**, 251–273.
6. Kendall, M. G., and Babington Smith, B. The problem of *m* rankings. *Ann. math. Statist.*, 1939, **10**, 275–287.
7. Kendall, M. G. Rank correlation methods. London: Charles Griffin and Co., 1948.
8. Spearman, Charles. 'Footrule' for measuring correlation. *Brit. J. Psychol.*, 1906, **2**, 89–108.

Manuscript received 1/8/52

Revised manuscript received 6/13/52

*One of the pre-publication reviewers of this paper has pointed out that the normal deviate test of ρ_{av} is equivalent to a χ^2 test of a linear relationship among the sums of ranks. Similar to Friedman's χ_r^2 (1), which he developed as a test of differences among the sums of ranks (and hence as a test of "concordance"), a χ_r with one degree of freedom as a test of linearity reduces to $\rho_{av}/\sigma_{\rho_{av}}$.