A COMPARISON OF ALTERNATIVE TESTS OF SIGNIFICANCE FOR THE PROBLEM OF m RANKINGS[1]

BY MILTON FRIEDMAN

A paper published in 1937 [2] suggested that the consilience of a number of sets of ranks can be tested by computing a statistic designated χ_r^2. A mathematical proof by S. S. Wilks demonstrated that the distribution of χ_r^2 approaches the ordinary χ^2 distribution as the number of sets of ranks increases. The rapidity with which this limiting distribution is approached was investigated by obtaining the exact distributions of χ_r^2 for a number of special cases. It was concluded that "when the number of sets of ranks is moderately large (say greater than 5 for four or more ranks) the significance of χ_r^2 can be tested by reference to the available χ^2 tables" [2, p. 695]. The use of the normal distribution was recommended when the number of ranks in each set is large, but the number of sets of ranks is small, although no rigorous justification of this procedure was presented.

Except for the few special cases for which exact distributions were given, the paper did not provide a test of significance for data involving less than six sets of ranks and a small or moderate number of ranks in each set. This important gap has now been filled by M. G. Kendall and B. Babington Smith [1]. In addition, they furnish a somewhat more exact test of significance for tables of ranks for which the earlier article recommended the use of the χ^2 distribution.

Kendall and Smith use a different statistic, W, defined as χ_r^2 divided by its maximum value, $m(n - 1)$, where n is the number of items ranked, and m the number of sets of ranks.[2] The new statistic (independently suggested by W. Allen Wallis [3] who terms it the rank correlation ratio and denotes it by η_r^2) is thus not fundamentally different from χ_r^2. A more radical innovation is the improvement in the test of significance that they suggest. Instead of testing χ_r^2 by reference to the χ^2 distribution for $n - 1$ degrees of freedom, Kendall and Smith, generalizing from the first four moments of W, recommend that the significance of W be tested by reference to the analysis of variance distribution (Fisher's z-distribution) with $z = \frac{1}{2}\log_e\left(\frac{(m-1)W}{1-W}\right)$, $n_1 = (n-1) - \frac{2}{m}$, $n_2 = (m-1)\left[(n-1) - \frac{2}{m}\right]$. For small values of m and n, they introduce con-

[1] The author is indebted to Mr. W. Allen Wallis for valuable criticism and to Miss Edna R. Ehrenberg for computational assistance.

[2] This is Kendall and Smith's notation which will be used in the present paper. The original paper [2] designated the number of items ranked by p, and the number of sets of ranks by n.

tinuity corrections, substituting for $W = \frac{12S}{m^2(n^3 - n)}$, the statistic

$$W_c = \frac{S - 1}{\frac{m^2(n^3 - n)}{12} + 2} = \frac{W - \frac{12}{m^2(n^3 - n)}}{1 + \frac{24}{m^2(n^3 - n)}},$$

where S is the observed sum of squares of the deviations of sums of ranks from the mean value, $m(n + 1)/2$. Comparison with exact distributions of W (or S) for special cases indicates that this test yields very good approximations to the correct probabilities.

In the limit the two tests of significance are identical. Neglecting the correction for continuity, $z = \frac{1}{2} \log_e \left(\frac{(m - 1)\chi_r^2}{m(n - 1) - \chi_r^2} \right) \rightarrow \frac{1}{2} \log_e \left(\frac{\chi_r^2}{n - 1} \right)$, $n_2 = (m - 1)\left[(n - 1) - \frac{2}{m} \right] \rightarrow \infty$, and $n_1 = (n - 1) - \frac{2}{m} \rightarrow (n - 1)$ as $m \rightarrow \infty$. For $n_2 = \infty$, the analysis of variance distribution is identical with the distribution of $\frac{1}{2} \log_e \frac{\chi^2}{n_1}$. The difference between the two tests is thus that one, χ^2, uses a single (limiting) distribution for all values of m, whereas the other, z, adapts the distribution to the value of m.

The necessity of taking into account the value of m, while it increases the flexibility of the distribution, makes the z test somewhat less convenient in practice than the χ^2 test. Additional computation is required to obtain the values of n_1 and n_2, and to make the continuity corrections. It is also fairly laborious to test the significance of the result, if exact values of z at any level of significance are required. In these instances, two-way interpolation of reciprocals in the analysis of variance tables is necessary since both n_1 and n_2 are always fractional. These difficulties make it desirable to investigate the rapidity with which the significance levels given by the z test approach those given by the χ^2 test, and thus determine the range of values of m and n for which the simpler test can safely be employed. This investigation will yield as a by product the .05 and .01 significance values of χ_r^2 (or W or S) for selected values of m and n as determined by the z test.

Table I presents a summary comparison of the values of χ_r^2 at the .05 and .01 levels of significance as shown by (1) exact distributions, (2) the z test with continuity corrections, (3) the χ^2 test.[3] The significance values are expressed in terms of χ_r^2 rather than W because, for a given number of ranks per set (i.e., a given n), the significance values given by the χ^2 test are the same regardless of the number of sets of ranks (i.e., of the value of m). This would not be so if W were employed, since $W = \chi_r^2/m\,(n - 1)$. The expected value of W depends on

[3] The values of χ_r^2 computed using the z test that are given in Tables I and II were obtained with the aid of Fisher and Yates' Table V [4]. Linear interpolation of reciprocals was employed throughout.

m and approaches zero as $m \to \infty$ while the expected value of χ_r^2 is equal to $n - 1$ for all values of m.

The values given by the z test agree remarkably well with the exact values. With but two exceptions (the .01 values for $n = 3$, $m = 8$ and 10) the exact value differs very much less from the value given by the z test than from the value given by the χ^2 test. In all but three of the 12 comparisons, the z test gives a value below the correct one.[4]

TABLE I

Comparison of Values of χ_r^2 at .05 and .01 Levels of Significance Yielded by Exact Distributions, z Test with Continuity Corrections, and χ^2 Test

n	m	.05 Level of Significance				01 Level of Significance			
		From Exact Distribution		From z test with continuity corrections	From χ^2 test	From Exact Distribution		From z test with continuity corrections	From χ^2 test
		Limits	Interpolated value*			Limits	Interpolated value*		
3	8	5.25–6.25	6.16	6.012	5.991		9.00	8.35	9.21
	9	6.0 –6.22	6.17	6.004	5.991		8.67	8.44	9.21
	10	5.6 –6.2	6.08	5.999	5.991	8.6 – 9.6	9.04	8.51	9.21
	∞			5.991	5.991			9.21	9.21
4	4	7.5 –7.8	7.54	7.43	7.82	9.3 – 9.6	9.42	9.21	11.34
	5	7.32–7.8	7.54	7.52	7.82	9.72– 9.96	9.87	9.66	11.34
	6	7.4 –7.6	7.49	7.57	7.82		10.00	9.95	11.34
	∞			7.82	7.82			11.34	11.34
5	3	8.27–8.53	8.41	8.59	9.49	9.87–10.13	10.05	10.08	13.28
	∞			9.49	9.49			13.28	13.28

* Computed by linear interpolation of probabilities.

Table II gives for a very much larger number of values of m and n the .05 and .01 values of χ_r^2 computed on the basis of the z test with continuity correc-

[4] These comparisons duplicate some of those made by Kendall and Smith and merely serve to confirm their conclusion that the z test with continuity corrections gives exceedingly good results.

The values obtained using the z test without continuity corrections agree less well with the exact values than those obtained with the aid of the continuity corrections. However even if no continuity corrections are made the z test in general yields values closer to the exact values than does the χ^2 test.

TABLE II

Values of χ_r^2 at .05 and .01 Levels of Significance Computed on the Basis of Kendall and Smith's z test, with Continuity Corrections; .10, .075, .02, .015 Values of χ^2

m	n = 3	4	5	6	7
Values at .05 Level of Significance					
3			8.59	9.90	11.24
4		7.43	8.84	10.24	11.62
5		7.52	8.98	10.42	11.84
6		7.57	9.08	10.54	11.97
8	6.012	7.63	9.18	10.68	12.14
10	5.999	7.67	9.25	10.76	12.23
15	5.985	7.72	9.33	10.87	12.36
20	5.983	7.74	9.37	10.92	12.42
100	5.987	7.80	9.46	11.04	12.56
∞	5.991	7.82	9.49	11.07	12.59
χ^2 (.10)	4.605	6.25	7.78	9.24	10.64
χ^2 (.075)*	5.18	6.90	8.49	10.00	11.45
Values at .01 Level of Significance					
3			10.08	11.69	13.26
4		9.21	10.93	12.59	14.19
5		9.66	11.42	13.11	14.74
6		9.95	11.74	13.45	15.09
8	8.35	10.31	12.13	13.87	15.53
10	8.51	10.52	12.37	14.11	15.79
15	8.74	10.79	12.67	14.44	16.14
20	8.85	10.93	12.82	14.60	16.31
100	9.14	11.26	13.19	14.99	16.71
∞	9.21	11.34	13.28	15.09	16.81
χ^2 (.02)	7.82	9.84	11.67	13.39	15.03
χ^2 (.015)*	8.40	10.46	12.34	14.09	15.77

* Computed from Fisher and Yates' Table IV (4) by linear interpolation between the logarithms of the probabilities.

tions. The values entered for $m = \infty$ are obtained from χ^2 tables for $n - 1$ degrees of freedom and are the significance values by the χ^2 test for all values of m. It is apparent that as m increases the .01 and .05 values of χ_r^2 approach their limiting values very rapidly. For $n = 7$, two-thirds of the difference between the .05 values for $m = 3$ and $m = \infty$, and an even larger proportion of the difference between the .01 values, disappears by the time $m = 10$; and the situation is similar for the other values of n. Except for the .05 values for $n = 3$, the approach to the limit is monotonic from below. The use of the χ^2 test thus tends to lead to the overestimation of the significance values and of the probabilities attached to observed values of χ_r^2. It is clear, however, that for large and even moderate values of m the χ^2 test is, for all practical purposes, equivalent to the z test.

In order to determine more precisely the range of values of m and n for which the approximation given by the χ^2 test is adequate, it is necessary to adopt some convention about the error in estimated significance values of χ_r^2 that is tolerable. Since the conclusion drawn from an observed χ_r^2 depends on the probability that it will be exceeded by chance, this convention clearly should be expressed in terms of the error in the probability.

The structure of published χ^2 tables makes it convenient to accept an estimated probability between .10 and .05 as a tolerable approximation to a correct probability of .05, and an estimated probability between .02 and .01 as a tolerable approximation to a correct probability of .01. These ranges of tolerance are entirely on one side of the correct probability because, as pointed out above, the error in using the χ^2 test is consistent in direction. These ranges are purely arbitrary; of course, and many may think them too broad.

On the basis of this or some similar convention it is possible to make objective statements concerning the range of values of m and n for which the χ^2 test is adequate. The next to the last line in the first section of Table II gives the .10 values of χ^2; the next to the last line in the second section, the .02 values. All the .05 values of χ_r^2 shown in the table exceed the .10 value of χ^2. Using the χ^2 test, all of the values (with two exceptions for $n = 3$) would signify a probability greater than .05 but less than .10. Thus the error made at the .05 level is within the admissible range according to the suggested convention. The χ^2 test is therefore an adequate substitute for the z test at the .05 level for all values of m and n except possibly for a few of the values for which exact distributions are available.

As might be expected, the χ^2 test is less satisfactory at the .01 level. For values of m less than six, the .01 values of χ_r^2 computed using the z test with continuity corrections are less than the .02 value of χ^2. For m greater than 5, the values of χ_r^2 in the table would all be accorded a probability greater than .01 but less than .02 if the χ^2 test were employed. As already noted, this is the range of values of m for which the original paper suggested the χ^2 test could validly be used [2, p. 695].

In view of the arbitrary nature of the convention as to the permissible error in the probability attached to an observed value of χ_r^2, it is interesting to investigate the effect of an alternative and stricter convention, namely, that only probabilities from .075 to .05 and from .015 to .01 be accepted as approximations to correct probabilities of .05 and .01 respectively. The .075 and .015 values of χ^2 are given in the last lines of the two sections of Table II. On the basis of this convention the χ^2 test is adequate at the .05 level for m greater than three, and

TABLE III

Values of S at .05 and .01 Levels of Significance Computed on the Basis of Kendall and Smith's z test, with Continuity Corrections

m	n					Additional values for $n = 3$	
	3	4	5	6	7	m	S
			Values at .05 Level of Significance				
3			64.4	103.9	157.3	9	54.0
4		49.5	88.4	143.3	217.0	12	71.9
5		62.6	112.3	182.4	276.2	14	83.8
6		75.7	136.1	221.4	335.2	16	95.8
8	48.1	101.7	183.7	299.0	453.1	18	107.7
10	60.0	127.8	231.2	376.7	571.0		
15	89.8	192.9	349.8	570.5	864.9		
20	119.7	258.0	468.5	764.4	1158.7		
			Values at .01 Level of Significance				
3			75.6	122.8	185.6	9	75.9
4		61.4	109.3	176.2	265.0	12	103.5
5		80.5	142.8	229.4	343.8	14	121.9
6		99.5	176.1	282.4	422.6	16	140.2
8	66.8	137.4	242.7	388.3	579.9	18	158.6
10	85.1	175.3	309.1	494.0	737.0		
15	131.0	269.8	475.2	758.2	1129.5		
20	177.0	364.2	641.2	1022.2	1521.9		

at the .01 level for m greater than nine, except possibly for a few of the values for which exact distributions are available. Thus even so drastic a lowering of the permissible margin of error as halving it limits only slightly the range of values of m for which the χ^2 test is adequate.

Table II provides, of course, a direct means of testing the significance of observed values of χ_r^2 for the tabled values of m and n. For this purpose, however, Table III, giving the significance values of S is more useful, since it obviates the necessity of converting S into χ_r^2. For $n = 3$ Table III includes a few values of m in addition to those in Table II.

SUMMARY

The preceding analysis suggests that the χ^2 test of the significance of χ_r^2 (or W or η_r^2), while less accurate than the z test proposed by Kendall and Smith, is adequate for practical purposes at the .01 level of significance if the number of sets of ranks (m) is greater than 5; and at the .05 level for any number of sets of ranks, provided the number of ranks in each set (n) is more than 3. Exact distributions are now available for $n = 3$, $m = 3$ to 10; $n = 4$, $m = 3$ to 6; $n = 5$, $m = 3$ [1]. The .05 and .01 values of χ_r^2 and S, computed using the Kendall and Smith z test with continuity corrections, are given in Tables II and III of the present note for $n = 3$ to 7 and selected values of m from 3 to 100. For n greater than 7 and m less than 6, the z test with continuity corrections should be employed. For all other combinations of n and m not covered by the exact distributions or by Tables II and III, the χ^2 test is adequate.

REFERENCES

[1] M. G. Kendall and B. Babington Smith, "The Problem of m Rankings," *Annals of Math. Stat.*, Vol. X (1939), pp. 275–87.

[2] Milton Friedman, "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," *Jour. Am. Stat. Assn.*, Vol. XXXII (1937), pp. 675–701.

[3] W. Allen Wallis, "The Correlation Ratio for Ranked Data," *Jour. Am. Stat. Assn.*, Vol. XXXIV (1939), pp. 533–8.

[4] R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (London: Oliver and Boyd, 1938).

National Bureau of Economic Research,
New York.