J. Statist. Comput. Simul. 1975, Vol. 4, pp. 155–164

Printed in Great Britain

Estimates of the Quantiles of Kendall's Partial Rank Correlation Coefficient

S. MAGHSOODLOO and NORMAN C. PERRY

Auburn University, Auburn, Alabama 36830, U.S.A.

(*Received June 12, 1974*)

The sampling distribution of Kendall's partial rank correlation coefficient, $\tau_{xy\cdot z}$, is not known for $N > 4$, where N is the number of subjects. Moran (1951) used a direct combinatorial method to obtain the distribution of $\tau_{xy\cdot z}$ for $N = 4$; however, ten minor computational errors in his Table 2 apparently resulted in two erroneous entries for his frequency table.

Since the practical limits of the direct combinatorial approach have been reached once $N > 4$, the first main objective of this paper was to obtain the exact distribution of $\tau_{xy\cdot z}$ for $N = 5$, 6, and 7 using an electronic computer. The second was to use the Monte Carlo method to obtain reliable estimates of the quantiles of $\tau_{xy\cdot z}$ for $N = 8, 9, \ldots, 30$.

1. INTRODUCTION

In nonparameteric statistics Kendall's partial rank correlation coefficient

$$\tau_{xy\cdot z} = \frac{\tau_{xy} - \tau_{xz}\tau_{yz}}{\sqrt{(1-\tau_{xz}^2)(1-\tau_{yz}^2)}} \tag{1}$$

is used to measure the extent of agreement between two rankings $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_N)$ of N subjects while keeping the effect of a third such variable $Z = (z_1, z_2, \ldots, z_N)$ constant. Partial τ is an extension of Kendall's τ_{xy}, the rank correlation coefficient between X and Y.

The sampling distribution of τ_{xy} is known and the statistical significance of an observed value of τ_{xy} can be tested (see Conover, 1971, Kendall, 1945 and 1962). The distribution of $\tau_{xy\cdot z}$, except for $N = 3$ and 4†, is not known and

† See Moran (1951) for the distribution of $\tau_{xy\cdot z}$ when $N = 4$.

therefore a test of hypothesis cannot be conducted for $N > 4$. Denoting the αth quantile of $\tau_{xy\cdot z}$ by Q_α, the objective of this paper was to obtain reliable estimates of $Q_{1-\alpha}(\alpha = 0.25, 0.20, 0.10, 0.075, 0.05, 0.025, 0.02, 0.01, 0.005, 0.001)$ for different numbers of subjects (N) in order that the significance of an observed $\tau_{xy\cdot z}$ may be tested.

Once $Q_{1-\alpha}$ is approximated, $\tau_{xy\cdot z}$ would be a useful test statistic in situations where the effect of Z on X and Y cannot be partialled out experimentally.

2. THE EXACT DISTRIBUTION OF $\tau_{xy\cdot z}$ FOR $N = 3, 4, 5, 6$ AND 7

Both τ_{xy} and $\tau_{xy\cdot z}$ have been extensively treated in the literature (Conover, 1971, Kendall, 1945, 1962 and Moran, 1951) of statistics and thus the method of their computations will not be discussed here.

In order to understand the complexity of the distribution of $\tau_{xy\cdot z}$, consider the following rankings of seven subjects ($N = 7$) on three variables:

TABLE I

Variables \ Subjects	*b*	*d*	*c*	*a*	*e*	*g*	*f*
Z	1	2	3	4	5	6	7
X	6	7	5	3	4	1	2
Y	7	6	5	3	4	2	1

where without loss of generality, we have arranged the subjects in such a way that the variable Z will always be the ranking $(1, 2, 3, \ldots, N-2, N-1, N)$. It is apparent from Table I that X and Y are two permutations of the integers $1, 2, \ldots, 7$.

Using Kendall's definition of τ_{xz} (see Kendall, 1962) we have

$$\tau_{xz} = \frac{P-Q}{N_2^C} \tag{2}$$

where P is the number of pairs of X rankings for which the direction of inequality agrees with that for the corresponding Z pair, Q is the number of pairs where X and Z disagree, and $N^C2 = N(N-1)/2$.

For the X permuation of Table I, $P = 3$, $Q = 18$, and using (2) $\tau_{xz} = -0.7143$. Similarly $\tau_{yz} = -0.9048$ and $\tau_{xy} = 0.8095$. Equation (1) then

yields

$$\tau_{xy\cdot z} = \frac{0.8095-0.6463}{\sqrt{(1-0.5102)(1-0.8187)}} = 0.548.$$

Keeping the variable Z fixed, we ask the question: Is the sample point 0.548 significant at a preassigned level α so as to warrant the rejection of the null hypothesis H_0: "X and Y are independent?" To answer this, we need to obtain the approximate quantiles of $\tau_{xy\cdot z}$.

For each N there are $N!$ permuations of X and of Y and thus $(N!)^2$ possible values for $\tau_{xy\cdot z}$; however it can easily be argued that of this number exactly $4N!-4$ sets of ranks result in τ_{xz} or τ_{yz} equal to ± 1, yielding in equation (1) and undefined expression for $\tau_{xy\cdot z}$. Therefore defining $S(N)$ as the total number of elements in the sample space of partial τ, we have

$$S(N) = (N!)^2-(4N!-4) \qquad (3)$$
$$= (N!-2)^2$$

It should be noted that Moran (1951) gives relation (3) for $N = 4$, and he notes that the symmetry properties of $\tau_{xy\cdot z}$ further decrease the number of necessary computations from $S(N)$ to about $S(N)/8$.

For $N = 3$ the distribution of $\tau_{xy\cdot z}$ can easily be obtained by combinatorial methods. The result of this procedure gives a frequency of 4 for $\tau_{xy\cdot z} = -1, -0.50, 0.50$, and 1. Note that of the $(3!)^2 = 36$ sets of ranks, 20 result in division by zero in (1), leaving only 16 cases where $\tau_{xy\cdot z}$ is defined (which agrees with equation (3)).

Moran (1951) used combinatorial means to obtain the distribution of partial τ for $N = 4$; however, two of the frequencies in his Table 3 for $\tau_{xy\cdot z} = 0.25$ and $\tau_{xy\cdot z} = 0.50$ are incorrect apparently due to the fact that 10 minor computational errors were made in Table 2 of Moran (1951). The correct frequency of $\tau_{xy\cdot z}$ at 0.2500 is 10 (not 8 as reported by Moran) and the frequency for $\tau_{xy\cdot z} = 0.5000$ is 10 (not 12).

If we were to use combinatorial means to obtain the distribution of $\tau_{xy\cdot z}$ for $N = 5$, approximately $\frac{1}{8}(5!-2)^2 = 1891$ calculations would be necessary; similarly about 64440 calculations are required for $N = 6$. So it seems obvious that the practical limits of the direct combinatorial approach have been reached once $N > 4$. Therefore a FORTRAN IV program was written to obtain the exact distribution of $\tau_{xy\cdot z}$ for general N subject to limitations of computer time.

The distributions of partial τ for $N = 4, 5, 6$, and 7 are given in Table III. They were obtained by a computer program† which determines all permuta-

† The computer programs are available upon request.

tions of the integers 1 through N and calculates $\tau_{xy\cdot z}$ for the appropriate $(N!-2)^2$ pairwise combinations of these permutations. The results in Table III were then used to obtain the quantiles of $\tau_{xy\cdot z}$ as given in Table II. Note that the intervals whose frequencies for $N = 4, 5, 6$, and 7 are zero are not reported, and since the distribution of partial τ is symmetric about $\tau_{xy\cdot z} = 0$, only the left half of the distributions is given in Table III.

Since the computer time on IBM 370–I 155 for obtaining the distribution of $\tau_{xy\cdot z}$ (at $N = 7$) was about 165 minutes, Monte Carlo sampling was used to estimate $Q_{1-\alpha}$ for $N > 7$.

3. RESULTS AND RECOMMENDATIONS

The program that gave the frequency distributions of Table III was extensively modified to generate arrays containing randomly assigned integers from 1 to N corresponding to the rankings on the variables X and Y. The modified program, which is available on request, has several subroutines one of which, called ADD, uses the symmetry of $\tau_{xy\cdot z}$ about zero, to effectively double the sample sizes (see Maghsoodloo, 1971) used in Monte Carlo sampling. In order to obtain reliable estimates of $Q_{1-\alpha}$ for $N > 7$ and hold computer time to an acceptable level, a sample of thirty to fifty thousand members (starting with $N = 7$) was generated; then the sample size n was increased for each successive run and when two consecutive runs yielded approximately the same $\hat{Q}_{1-\alpha}$, the corresponding results were accepted as the estimates. The effective sample sizes along with the corresponding computer times, to the nearest minutes, are summarized in Table IV.

The estimates of the upper quantiles of $\tau_{xy\cdot z}$ are given in Table V. The third decimal of the values reported were estimated in such a way that a test of significance would be conservative, and the quantile estimates in Table V were obtained from the sampling distributions whose effective sample sizes are given in Table IV. It should be noted that the output of the corresponding computer program gives only the sampling frequency distributions, which were subsequently used to compute $\hat{Q}_{1-\alpha}$ using a desk calculator.

Due to the symmetry of $\tau_{xy\cdot z}$ about zero, the estimates of the lower quantiles may be obtained from

$$\hat{Q}_{\alpha} = -\hat{Q}_{1-\alpha}$$

where $\alpha = 0.001, 0.005, 0.01, 0.02, 0.025, 0.05, 0.075, 0.10, 0.20$, and 0.25. For example, the estimate of the 5% quantile (fifth percentile) of $\tau_{xy\cdot z}$ for $N = 7$ is $\hat{Q}_{0.05} = -\hat{Q}_{0.95} = -0.527$.

To illustrate the use of Table V let us refer to the example cited earlier in Section 2 where $\tau_{xy\cdot z}$ was computed to be 0.548 for $N = 7$.

TABLE II

Approximate† quantiles of $\tau_{xy\cdot z}$ for $N = 3, 4, 5, 6, 7$

N \ Quantiles	$Q_{0.75}$	$Q_{0.80}$	$Q_{0.90}$	$Q_{0.925}$	$Q_{0.950}$	$Q_{0.975}$	$Q_{0.98}$	$Q_{0.99}$	$Q_{0.995}$	$Q_{0.999}$
3	0.50	1	1	1	1	1	1	1	1	1
4	0.4472	0.5000	0.7071	0.7071	0.7071	1	1	1	1	1
5	0.338	0.403	0.539	0.616	0.661	0.807	0.811	0.819	1	1
6	0.277	0.328	0.476	0.533	0.591	0.670	0.726	0.765	0.866	1
7	0.233	0.282	0.421	0.475	0.527	0.617	0.632	0.712	0.761	0.901

† The word approximate is used since the exact values of $\tau_{xy\cdot z}$ are not known for $N = 5, 6, 7$ and therefore the inequalities that define $Q_{1-\alpha}$ are approximately satisfied. Furthermore the third decimals for $N = 5$, 6, and 7 were selected so that a test of significance will be slightly conservative.

TABLE III
The frequency distribution of $r_{xy \cdot z}$ for $N = 4, 5, 6, 7$

Class Interval	$N = 4$	$N = 5$	$N = 6$	$N = 7$	Class Interval	$N = 4$	$N = 5$	$N = 6$	$N = 7$
[−1.00, −0.99)	22	118	718	5038	[−0.40, −0.39)	0	0	1944	191392
[−0.91, −0.90)	0	0	0	21168	[−0.39, −0.38)	0	0	3028	55984
[−0.90, −0.89)	0	0	0	3696	[−0.38, −0.37)	0	60	6184	112752
−0.89, −0.88	0	0	0	2520	−0.37, −0.36	0	0	2984	174000
−0.88, −0.87	0	0	1500	0	−0.36, −0.35	0	336	4420	241644
−0.87, −0.86	0	0	820	1536	−0.35, −0.34	0	0	4416	140720
−0.86, −0.85	0	0	600	0	−0.34, −0.33	12	176	3864	322104
−0.85, −0.84	0	0	0	816	−0.33. −0.32	0	228	7956	175144
−0.83, −0.82	0	0	380	26320	−0.32, −0.31	36	0	3000	278808
−0.82, −0.81	0	176	0	10760	−0.31, −0.30	0	0	1864	206568
−0.81, −0.80	0	144	0	27180	−0.30, −0.29	0	0	624	43624
−0.80, −0.79	0	0	0	8040	−0.29, −0.28	0	0	12700	267880
−0.79, −0.78	0	0	200	11960	−0.28, −0.27	0	176	4560	224616
−0.77, −0.76	0	96	1744	8060	−0.27, −0.26	0	0	5092	198344
−0.76, −0.75	0	0	1544	0	−0.26, −0.25	0	0	0	179412
−0.75, −0.74	0	0	0	50092	−0.25, −0.24	20	324	1704	326284
−0.74, −0.73	0	0	2384	2720	−0.24, −0.23	0	0	1200	337868
−0.73, −0.72	0	0	1048	17052	−0.23, −0.22	0	0	0	315120
−0.72, −0.71	0	0	0	65420	−0.22, −0.21	0	792	13356	160440
−0.71, −0.70	36	0	800	31656	−0.21, −0.20	0	0	1088	385392
−0.70, −0.69	0	0	816	27224	−0.20, −0.19	12	136	11884	216868
−0.69, −0.68	0	0	80	7424	−0.19, −0.18	0	0	13048	135168
−0.68, −0.67	0	0	0	56376	−0.18, −0.17	0	0	0	101984
−0.67, −0.66	0	180	3492	42664	−0.17, −0.16	0	340	5156	146920
−0.66, −0.65	0	228	544	0	−0.16, −0.15	0	0	228	432892
−0.65, −0.64	0	0	2320	21624	−0.15, −0.14	0	0	0	245428

−0.64, −0.63	24	0	0	68816	−0.14, −0.13	0	0	5924	660272
−0.63, −0.62	0	0	0	51696	−0.13, −0.12	0	0	5152	183424
−0.62, −0.61	0	168	1328	173804	−0.12, −0.11	0	24	4244	303424
−0.61, −0.60	0	0	1956	30588	−0.11, −0.10	0	228	4616	52580
−0.60, −0.59	0	84	3664	26004	−0.10, −0.09	0	0	10560	143808
−0.59, −0.58	0	144	312	64928	−0.09, −0.08	0	432	0	352020
−0.58, −0.57	0	0	5884	29320	−0.08, −0.07	0	0	9356	188340
−0.57, −0.56	0	0	2480	0	−0.07, −0.06	0	0	3664	550048
−0.56, −0.55	0	0	800	191892	−0.06, −0.05	0	0	9256	255632
−0.55, −0.54	0	0	0	37512	−0.05, −0.04	0	168	0	191636
−0.54, −0.53	0	288	4208	108152	−0.04, −0.03	0	0	5944	386200
−0.53, −0.52	0	108	0	146004	−0.03, −0.02	0	0	4032	414124
−0.52, −0.51	0	0	0	65332	−0.02, −0.01	0	0	0	42576
−0.51, −0.50	0	96	0	106976	−0.01, −0.00	0	0	0	93388
−0.50, −0.49	20	228	4424	80108	$\tau = 0.0$	48	880	20304	596672
−0 49, −0.48	0	0	5092	228500					
−0.48, −0.47	0	0	5616	105836					
−0.47, −0.46	0	0	5248	27824					
−0.46, −0.45	0	0	1296	73952					
−0.45, −0.44	36	0	2956	89200					
−0.44, −0.43	0	0	0	165744					
−0.43, −0.42	0	144	6332	230808					
−0.42, −0.41	0	0	1544	204040					
−0.41, −0.40	0	840	2432	201176					

TABLE IV

Sample sizes and computer times of the simulation runs for $N = 7, 8, \ldots, 30$

N	Effective Sample Size 1st Run	2nd Run	3rd Run	Total Computer Time (Minutes)
7	60,000	100,000	150,000	13
8	100,000	200,000	400,000	31
9	200,000	300,000	400,000	46
10	500,000	600,000	700,000	101
11	600,000	700,000	800,000	140
12†	1,200,000			27†
13	1,600,000			37
14	2,000,000			52
15	2,400,000			68
16	2,800,000			84
17	3.200,000			103
18	3,600,000			123
19	4,000,000			152
20	4,400,000			180
25‡	4,000,000‡			213
30‡	4,000,000‡			277

† At this point the computer time was becoming too excessive to be acceptable and therefore not only was just one run used, but also the corresponding program was modified by Mr. Don Hudson of Auburn's computer centre to reduce the computer time to about 1/4 of its original size. The modified program is available on request.

‡ Since the computer time was too large, only one effective sample of size 4,000,000 was used.

To test the two-sided hypothesis H_0: "X and Y are independent with Z fixed", $\hat{Q}_{0.975}$ is read from Table 9 and the acceptance region becomes approximately $(-0.617, 0.617)$. Since the sample value of $\tau_{xy\cdot z} = 0.548$ lies in this interval, H_0 cannot be rejected at the 5% level. The critical level is $\hat{\alpha} \doteq 0.08$.

The author's future objective is to find a limiting density function that approximates the frequency function of $\tau_{xy\cdot z}$ for large N, say $N \geqq N_0$. Unfortunately, no general expression exists for any of the first four moments of partial τ; Hoeffding (1948, p. 324) states that the frequency function of $\sqrt{N}(\tau_{xy\cdot z} - \rho_{xy\cdot z})$, where $\rho_{xy\cdot z}$ is the population partial correlation, tends to normality with zero mean and a variance whose complicated expression he evaluates.

TABLE V

Estimates of quantiles of $\tau_{xy\cdot z}$, $Q_{1-\alpha}$, for $N = 7, 8, \ldots, 30$

N \ $1-\alpha$	0.75	0.80	0.90	0.925	0.950	0.975	0.98	0.990	0.995	0.999
7	0.233	0.282	0.421	0.475	0.527	0.617	0.632	0.712	0.761	0.901
8	0.206	0.254	0.382	0.430	0.484	0.565	0.580	0.648	0.713	0.807
9	0.187	0.230	0.347	0.391	0.443	0.515	0.542	0.602	0.660	0.757
10	0.170	0.215	0.325	0.365	0.413	0.480	0.504	0.562	0.614	0.718
11	0.162	0.202	0.305	0.343	0.387	0.453	0.475	0.530	0.581	0.677
12	0.153	0.190	0.288	0.322	0.465	0.430	0.451	0.505	0.548	0.643
13	0.145	0.180	0.273	0.305	0.347	0.410	0.428	0.481	0.527	0.616
14	0.137	0.172	0.260	0.293	0.331	0.391	0.408	0.458	0.503	0.590
15	0.131	0.164	0.249	0.278	0.317	0.375	0.391	0.439	0.482	0.567
16	0.125	0.157	0.240	0.267	0.305	0.361	0.377	0.423	0.466	0.549
17	0.121	0.151	0.231	0.258	0.294	0.348	0.363	0.410	0.450	0.532
18	0.117	0.147	0.222	0.250	0.284	0.336	0.351	0.395	0.434	0.514
19	0.114	0.141	0.215	0.241	0.275	0.326	0.340	0.382	0.421	0.498
20	0.110	0.137	0.208	0.234	0.267	0.317	0.331	0.372	0.410	0.484
25	0.097	0.120	0.183	0.206	0.235	0.278	0.291	0.328	0.362	0.429
30	0.087	0.108	0.165	0.185	0.211	0.251	0.263	0.297	0.328	0.390

0.365

Moran (1951) computes the variance of $\tau_{xy\cdot z}$ for $N = 4$ to be 0.2817, but using the frequencies given in Table III the variance of partial τ for $N = 4$ is 0.2829. The programs of Tables II and V could have been easily supplemented to approximate the variance of $\tau_{xy\cdot z}$ for $N = 5, 6, 7$ and $N = 8, 9, \ldots, 30$, respectively. The next avenue of research should be to investigate how rapidly the pdf of $\sqrt{N}\,\tau_{xy\cdot z}$ tends to normality under the null hypothesis of independence and to determine the value of N_0.

References

Conover, W. J. (1971), *Practical Nonparametric Statistics*, John Wiley and Sons, Inc.

Hoeffding, W. (1948), A class of statistics with asympototically normal distribution. *Ann. Math. Statis.* **19**, 293.

Kendall, M. G. (1975), *The Advanced Theory of Statistics*, Vol. 1, 2nd Edition. Charles Griffin and Company Limited.

Kendall, M. G. (1962), *Rank Correlation Methods*, 3rd Edition, Hafner Publishing Company.

Maghsoodloo, S. (1971), An investigation by high speed sampling of the frequency distribution of rank correlation. *Computing* **8**, 1–12, by Springer-Verlag ¹1971: Paul Gerin, A-1021 Wren.

Moran, P. A. P. (1951), *Partial and Multiple Rank Correlation*, Biometrika, Vol. 38, p. 28.

and Norman C. Perry