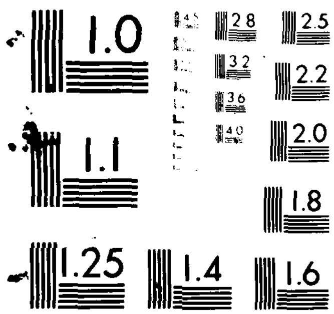
AD-A109 547 SOUTHERN METHODIST UNIV DALLAS TEX DEPT OF STATISTICS F/G 12/1
ON APPROXIMATE CONFIDENCE INTERVALS FOR MEASURES OF CONCORDANCE--ETC(U)
NOV 81 A D PALACHEK, W R SCHUCANY N00014-75-C-0439
UNCLASSIFIED TR-152 NL

1 of 1
AD
A109547

END
DATE
FILMED
2 82
DTIC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL

SOUTHERN METHODIST UNIVERSITY

DEPARTMENT OF STATISTICS

DALLAS, TEXAS 75275

ON APPROXIMATE CONFIDENCE INTERVALS
FOR MEASURES OF CONCORDANCE

by

Albert D. Palachek and William R. Schucany

Technical Report No. 152
Department of Statistics ONR Contract

November, 1981

Research sponsored by the Office of Naval Research
Contract N00014-75-C-0439

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, Texas 75275

ON APPROXIMATE CONFIDENCE INTERVALS

FOR MEASURES OF CONCORDANCE

Abstract

The use of U-statistics based on rank correlation coefficients in estimating the strength of concordance among a group of rankers is examined for cases where the null hypothesis of random rankings is not tenable. The studentized U-statistic is asymptotically distribution-free, and the Student-t approximation is used for small and moderate sized samples. An approximate confidence interval is constructed for the strength of concordance. Monte Carlo results indicate that the Student-t approximation can be improved by estimating the degrees of freedom.

ON APPROXIMATE CONFIDENCE INTERVALS

FOR MEASURES OF CONCORDANCE

I. Introduction

Solutions to the problem of testing for agreement among n sets of rankings of k objects have been proposed by Kendall and Babington-Smith [1939] and Ehrenberg [1952]. Kendall and Babington-Smith proposed the coefficient of concordance W, which is related to Friedman's [1937] χ_r^2 for two-way analysis of variance using ranks. Ehrenberg's statistic is the average of the Kendall rank correlation coefficients τ between the $\binom{n}{2}$ pairs of judges. These statistics have been studied extensively under the usual null hypothesis of random rankings, which implies that for each judge, each of the k! permutations of the ranks 1,...,k is equally likely to be assigned to the k objects.

Very little work has been done in the non-null case. Kraemer [1976] proposed a non-null approximation to the distribution of W, but this approximation is based on an empirical study using data generated from a normal components-of-variance model.

There are many situations, however, in which it is known that there is agreement among the judges in the population, and the investigator would like to estimate the strength of agreement among the judges. For instance, an investigator may know that two populations of judges agree on the preference of k objects and wishes to know which group holds the preference more strongly. What is needed in these situations is a parametric measure of the intensity with which a preference of objects is held by a group of judges.

II. Internal Rank Correlation

Quade [1972] proposed a measure of agreement of rankings that is based on the expected rank correlation between a pair of independent rankings. Denote the rankings of the objects by a sample of n judges as $\underline{X}_i = (X_{i1},\ldots,X_{ik})'$, $i=1,\ldots,n$. Quade's measure of concordance, the internal rank correlation, is given by

$$(2.1) \quad \rho = E[R(\underline{X}_i,\underline{X}_j)] \quad , \quad i \neq j,$$

where $\underline{X}_i$ and $\underline{X}_j$ are independent rankings and $R(\cdot,\cdot)$ is any rank correlation coefficient. Two particular measures are those obtained by using the Spearman [1904] and Kendall [1938] rank correlation coefficients, which will be denoted as R_s and R_k, respectively.

Under the null hypothesis of random rankings one finds that $\rho = 0$. However, ρ is positive if there is agreement among the judges, and $\rho = 1$ when there is perfect agreement. So ρ measures the intensity with which a preference of the objects is held by the judges. An investigator may be interested in estimating ρ to compare with a "norm" or to compare with the estimated internal rank correlations from other populations.

The U-statistic estimator of ρ is given by

$$(2.2) \quad \bar{R} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \sum R(\underline{X}_i,\underline{X}_j) \quad ,$$

which Quade refers to as the average internal rank correlation.

The asymptotic ($n\to\infty$) variance of $\sqrt{n}\,\overline{R}$ is given by $4\zeta_R$, where

$$(2.3)\qquad \zeta_R = E[\psi_R^2(\underline{X})]$$

and

$$\psi_R(\underline{X}) = E[R(\underline{X}_1,\underline{X}_2)\,|\,\underline{X}_2 = \underline{X}] - \rho \ .$$

If $\zeta_R > 0$ then Hoeffding's [1948] results show that the limiting distribution of $\sqrt{n}(\overline{R}-\rho)/\sqrt{4\zeta_R}$ is standard normal. Under the null hypothesis of random rankings $\zeta_R = 0$ and the limiting distribution of $\sqrt{n}\,\overline{R}$ is degenerate. Under mild regularity conditions given by Quade [1972], however, $\zeta_R > 0$ when $\rho > 0$.

Since ζ_R is seldom known, one usually estimates this parameter. A consistent estimator of ζ_R can be obtained from a method of Sen [1960] and is given by

$$(2.4)\qquad \hat{\zeta}_R = \frac{1}{n-1}\sum_{i=1}^{n}[V_n(\underline{X}_i)-\overline{R}]^2,$$

where

$$(2.5)\qquad V_n(\underline{X}_i) = \frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n} R(\underline{X}_i,\underline{X}_j), \quad i = 1,\ldots,n,$$

are the sample components of $\overline{R}$.

The asymptotic distribution of $\sqrt{n}(\overline{R}-\rho)/\sqrt{4\hat{\zeta}_R}$ is also standard normal under the regularity condition $\zeta_R > 0$, and this studentized U-statistic can be used to construct approximate tests and confidence intervals.

III. Refinement of Interval Estimation

The distributions of studentized U-statistics are often approximated by the Student-t distribution on n-1 degrees of freedom.

Using this approximation, one can obtain an approximate $100(1-\alpha)\%$ confidence interval for ρ as

$$\bar{R} - t_{\alpha/2}(n-1)\sqrt{\frac{4\hat{\zeta}_R}{n}} < \rho < \bar{R} + t_{\alpha/2}(n-1)\sqrt{\frac{4\hat{\zeta}_R}{n}} \quad ,$$

where $t_\alpha(\nu)$ is the $(1-\alpha)$th quantile of the Student-t distribution on ν degrees of freedom.

The choice of n-1 degrees of freedom seems reasonable since

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} V_n(\underline{X}_i) \quad .$$

However, the sample components are not independent, and the approximation using n-1 degrees of freedom can lead to problems of undercoverage when estimating ρ.

To illustrate this we use a model introduced by Mallows [1957] and later studied by Feigin and Cohen [1978]. Let $\underline{x}_0$ be a fixed vector with one of the k! orderings of the integers 1,...,k, and for every possible ranking $\underline{x}$ let $d(\underline{x}_0,\underline{x})$ denote a "distance" (in a rank correlation sense) between $\underline{x}_0$ and $\underline{x}$. A model which assigns equal probabilities to rankings with the same value of d is then

$$(3.1) \quad P_\theta(\underline{x}) = c(\theta)\theta^{d(\underline{x}_0,\underline{x})} \quad , \quad 0 \le \theta \le 1,$$

where

$$c(\theta) = \left[\sum_{\underline{y}} \theta^{d(\underline{x}_0,\underline{y})}\right]^{-1} \quad ,$$

the summation being over all k! possible rankings. The smaller the rank correlation of $\underline{x}$ with $\underline{x}_0$, the smaller the probability of occurrence of $\underline{x}$. The extreme of $\theta=0$ corresponds to perfect concordance and $\theta=1$ corresponds to random rankings.

Data were generated from this model at various values of θ using distances based on both the Spearman and Kendall rank correlation coefficients. These simulations were performed on the C D C 6600 computer at Southern Methodist University. The possible rankings were denoted by $\underline{x}_i$, $i = 1,\ldots,k!$, and the $\underline{x}_0$ vector used was $\underline{x}_0 = (1,2,\ldots,k)'$. A uniform (0,1) observation u was generated by using the C D C pseudorandom number generator RANF and the generated ranking was then that ranking $\underline{x}_i$ which satisfied

$$F_{i-1} < u \leq F_i ,$$

where $F_0 = 0$ and $F_i = \sum_{j=1}^{i} P_\theta(\underline{x}_j)$ for $i = 1,\ldots,k!$.

For k = 4, 1000 samples of size n = 25 were generated for several values of θ, and approximate 95% confidence intervals were obtained for the parameters ρ_s and ρ_k, which are the population internal rank correlation measures when using $R_s(\cdot,\cdot)$ and $R_k(\cdot,\cdot)$, respectively.

The empirical coverages that were obtained from these simulations are given in Table 1. The standard error of these proportions at the nominal level is .0069. Most of these coverages are significantly less than .95. Larger samples were also generated for two configurations using the Kendall distance, and the empirical coverages are given in Table 2. These show that the coverages improve when the sample size becomes quite large.

For small and moderate sized samples, however, the intervals are too small. This problem could be due to the choice of n-1 degrees of freedom in the t-approximation. One method of improving the coverage is that of estimating the degrees of freedom. Hinkley [1977] has proposed a method of estimating the degrees of freedom for the studentized jackknife estimator. This method can be used in the present setting to adjust the interval width only, since $\bar{R}$ is invariant to the jackknife procedure.

The pseudovalues for jackknifing $\bar{R}$ can be shown to be

$$(3.2)\quad \hat{\rho}_{n,-j} = 2\,\frac{n-1}{n-2}\,V_n(\underline{X}_{-j}) - \frac{n}{n-2}\,\bar{R}\ ,$$

for $j = 1,\ldots,k$. The jackknife variance estimator of $\bar{R}$ is then given by

$$(3.3)\quad \begin{aligned} V_R &= \frac{1}{n(n-1)} \sum_{j=1}^{n} (\hat{\rho}_{n,-j} - \bar{R})^2 \\ &= \frac{4}{n}\left(\frac{n-1}{n-2}\right)^2 \hat{\zeta}_R\ . \end{aligned}$$

Since $n\,V_R$ is also a consistent estimator of $4\zeta_R$, then $(\bar{R}-\rho)/\sqrt{V_R}$ is asymptotically standard normal if $\zeta_R > 0$.

Hinkley's estimator for the degrees of freedom is given by

$$(3.4)\quad f_n = \frac{2V_R^2}{K_n}\ ,$$

where

$$(3.5)\quad K_n = \frac{\sum_{j=1}^{n} (\hat{\rho}_{n,-j} - \bar{R})^4}{n(n-1)(n-2)^2} - \frac{n-1}{(n-2)^2}\,V_R^2\ .$$

[The expression for K_n given in the Hinkley paper contains a slight error in the coefficient of V_R^2, and the correct expression is given here in (3.5)]. The estimated degrees of freedom can also be expressed in terms of the sample components as

$$(3.6) \qquad f_n = \frac{\frac{2}{n}(n-2)^2\hat{\zeta}_R^2}{\frac{1}{n-1}\sum_{i=1}^{n}[V_n(\underline{X}_i)-\bar{R}]^4 - \frac{n-1}{n}\hat{\zeta}_R^2} \; .$$

The jackknife variance estimator and the degrees of freedom estimator were used on the data that were generated for Table 1. Table 3 gives the empirical coverages that were obtained from these modified approximate 95% confidence intervals. A comparison of these results with those in Table 1 shows that the empirical coverages have improved in almost every configuration. Simulations from samples of size n=10 also show an improvement in coverage when using the estimated degrees of freedom.

Table 4 gives the average lengths of the confidence intervals for ρ_s and ρ_k that were obtained when generating samples from the model using the Kendall distance measure. These average lengths are larger when estimating the degrees of freedom than when using n-1 degrees of freedom. However, this is expected since the empirical coverages have increased.

The minimum estimated degrees of freedom are between 2 and 4 for most of the samples generated, and the minimum among all generated samples is 2.0. However, the average estimated degrees of freedom for some models was greater than n-1, as can be seen in Table 5. So there are many instances where the estimated degrees of freedom are larger than n-1. In these cases the lengths of the confidence intervals are smaller when estimating the degrees of freedom than when using n-1 degrees of freedom. Nevertheless, as Tables 4 and 5 show, even when the average estimated degrees of freedom exceeds n-1, the associated average confidence interval is not shorter.

IV. Example Application

An investigator is interested in estimating the strength of agreement of male college students on the importance ordering of seven basic human needs. These needs are

A) Self-actualization
B) Cognitive needs
C) Physiological needs
D) Aesthetic needs
E) Esteem needs
F) Belongingness and love, and
G) Safety needs.

A sample of 15 male college students was obtained, and each student ranked the needs based on the criterion of importance. These rankings are given in Table 6.

The Spearman rank correlation coefficients, $R_s(\underline{X}_i, \underline{X}_{i'})$ are found for every pair of rankings, and from these one obtains the sample components $V_n(\underline{X}_i)$ which are also given in Table 6. This leads to

$$\bar{R} = \frac{1}{n} \sum_{i=1}^{n} V_n(\underline{X}_i) = .23979$$

and

$$\hat{\zeta}_R = \frac{1}{n-1} \sum_{i=1}^{n} [V_n(\underline{X}_i) - \bar{R}]^2 = .00910.$$

Then the variance of $\bar{R}$ is estimated by

$$V_R = \frac{4}{n} \left(\frac{n-1}{n-2}\right)^2 \hat{\zeta}_R = .00281.$$

To estimate the degrees of freedom for the t approximation we need

$$\sum_{i=1}^{n} [V_n(\underline{X}_i) - \bar{R}]^4 = .00286.$$

Using (3.6), the degrees of freedom are estimated by

$$f_n = \frac{\frac{2}{15}(13)^2(.00910)^2}{\frac{1}{14}(.00286) - \frac{14}{15}(.00910)^2} = 14.69 \quad .$$

Since $t_{.025}(14.69) \approx 2.135$, an approximate 95% confidence interval for ρ is

$$.23979 - (2.135)\sqrt{.00281} < \rho < .23979 + (2.135)\sqrt{.00281} ,$$

i.e. $.1266 < \rho < .3530$.

This gives an approximate 95% confidence interval for the strength of agreement among male college students on the ordering of the seven basic human needs. Notice that for this data set the estimated degrees of freedom statistic is quite close to n-1. As the Monte Carlo results suggest, other samples from this same population may be expected to yield values of f_n that differ considerably from n-1 in either direction.

V. Conclusions

Knowledge of the parameter ρ can be very useful to an investigator who wants to determine the strength of agreement among a population on the rank-order preference of k objects. This parameter can be estimated without putting model constraints on the rankings since the U-statistic estimator is asymptotically distribution-free. The estimation of ρ can also be improved by using Hinkley's method of estimating the degrees of freedom for the Student-t approximation to the distribution of the studentized U-statistic. This method provides better coverage by increasing the lengths of intervals that are too short, and it can lead to more accurate estimation in many cases by decreasing the lengths of some intervals that are too long.

TABLE 1

Empirical Coverage of Approximate 95% Confidence Intervals for ρ_s and ρ_k with k=4 and n=25 (1,000 Simulations)

	Kendall Distance		Spearman distance	
θ	ρ_s	ρ_k	ρ_s	ρ_k
.2	.923	.936	.938	.936
.3	.906	.928	.921	.929
.4	.938	.937	.919	.925
.5	.929	.932	.924	.940
.6	.917	.915	.929	.951
.7	.929	.930	.924	.927
.8	.921	.922	.934	.939
.9	.934	.952	.916	.916

TABLE 2

Empirical Coverage of Approximate 95% Confidence Intervals for ρ_s and ρ_k with k=4

(1,000 Simulations)

	n=50		n=75		n=100	
θ	ρ_s	ρ_k	ρ_s	ρ_k	ρ_s	ρ_k
.3	.914	.922	.942	.946	.942	.945
.6	.926	.927	.944	.948	.936	.938

TABLE 3

Empirical Coverage of Approximate 95% Confidence Intervals for ρ_s and ρ_k using Estimated Degrees of Freedom (1,000 Simulations)

θ	Kendall Distance ρ_s	Kendall Distance ρ_k	Spearman distance ρ_s	Spearman distance ρ_k
.2	.946	.951	.939	.938
.3	.927	.941	.970	.980
.4	.953	.951	.938	.942
.5	.938	.937	.937	.948
.6	.927	.926	.951	.958
.7	.935	.938	.939	.941
.8	.924	.928	.941	.947
.9	.934	.953	.926	.926

TABLE 4

Average Lengths of Approximate 95% Confidence Intervals for ρ_s and ρ_k using n-1 and Estimated Degrees of Freedom with k=4 and n=25

(1,000 Simulations)

	n-1 d.f.		Est.d.f.	
θ	ρ_s	ρ_k	ρ_s	ρ_k
.2	.3151	.3216	.3688	.3532
.3	.3943	.3612	.4451	.3932
.4	.4193	.3651	.4556	.3904
.5	.4030	.3438	.4301	.3646
.6	.3542	.2991	.3710	.3130
.7	.2986	.2517	.3093	.2610
.8	.2327	.1968	.2392	.2026
.9	.1919	.1637	.1964	.1678

TABLE 5

Average Estimated Degrees of Freedom
for Confidence Intervals for ρ_s and ρ_k
Using Kendall Distance with k=4 and n=25
(1,000 Simulations)

θ	ρ_s	ρ_k
.2	13.4	19.0
.3	15.1	19.7
.4	19.3	23.0
.5	23.5	25.4
.6	29.5	29.8
.7	36.2	34.6
.8	42.3	39.2
.9	46.6	42.1

TABLE 6

Rankings of Basic Human Needs by Male College Students

Student	Basic Human Needs							$V_n(\underline{X}_i)$
	A	B	C	D	E	F	G	
1	4	7	3	2	5	1	6	.20918
2	1	3	4	7	6	2	5	.28571
3	4	7	1	5	6	3	2	.25510
4	1	4	6	7	3	2	5	.17347
5	7	6	3	5	4	2	1	.06122
6	2	5	4	6	3	1	7	.36480
7	3	1	2	6	5	4	7	.18878
8	6	3	2	7	4	1	5	.34439
9	1	4	2	5	6	3	7	.32653
10	2	3	1	4	7	6	5	.05357
11	4	7	3	2	5	1	6	.20918
12	5	6	4	7	3	1	2	.23214
13	3	6	1	7	2	4	5	.35714
14	7	6	1	5	3	2	4	.26275
15	3	4	1	7	2	5	6	.27296

REFERENCES

Ehrenberg, A.S.C. On sampling from a population of rankers. Biometrika, 1952, 39, 82-87.

Feigin, P.D. and Cohen, A. On a model for concordance between judges. Journal of the Royal Statistical Society, Series B, 1978, 12, 203-213.

Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association, 1937, 32, 675-701.

Hinkley, D.V. Jackknife confidence limits using Student t approximations. Biometrika, 1977, 64, 21-28.

Hoeffding, W. A class of statistics with asymptotically normal distribution. Annals of Mathematical Statistics, 1948, 19, 293-325.

Kendall, M.G. A new measure of rank correlation. Biometrika, 1938, 30, 81-93.

Kendall, M.G. and Babington-Smith, B. The problem of m rankings. Annals of Mathematical Statistics, 1939, 10, 275-287.

Kraemer, H.C. The small sample nonnull properties of Kendall's coefficient of concordance for normal population. Journal of the American Statistical Association, 1976, 71, 608-613.

Mallows, C. L. Non-null ranking models I. Biometrika, 1957, 44, 114-130.

Quade, D. Average internal rank correlation. Technical Report, Mathematical Centre, University of Amsterdam, 1972.

Sen, P.K. On some convergence properties of U-statistics. Calcutta Statistical Association Bulletin, 1960, 10, 1-18.

Spearman, C. The proof and measurement of association between two things. American Journal of Psychology, 1904, 15, 72-101.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 152	2. GOVT ACCESSION NO. AD-A109547	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On Approximate Confidence Intervals for Measures of Concordance		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER 152
7. AUTHOR(s) Albert D. Palachek and William R. Schucany		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0439
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Dallas, Texas 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE November 1981
		13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purposes of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The use of U-statistics based on rank correlation coefficients in estimating the strength of concordance among a group of rankers is examined for cases where the null hypothesis of random rankings is not tenable. The studentized U-statistic is asymptotically distribution-free, and the Student-t approximation is used for small and moderate sized samples. An approximate confidence interval is constructed for the strength of concordance. Monte Carlo results indicate that the Student-t approximation can be improved by estimating the degrees of freedom.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)