SAGE Reference

The SAGE Encyclopedia of Research Design

Coefficient of Concordance

Author: Pierre Legendre
Edited by: Bruce B. Frey
Book Title: The SAGE Encyclopedia of Research Design
Chapter Title: "Coefficient of Concordance"
Pub. Date: 2022
Access Date: February 23, 2022
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks,
Print ISBN: 9781071812129
Online ISBN: 9781071812082
DOI: https://dx.doi.org/10.4135/9781071812082.n89
Print pages: 219-222

Proposed by Maurice G. Kendall and Bernard Babington Smith, Kendall's coefficient of concordance (W) is a measure of the agreement among several (m) quantitative or semiquantitative variables that are assessing a set of n objects of interest. In the social sciences, the variables are often people, called judges, assessing different subjects or situations. In community ecology, they may be species whose abundances are used to assess habitat quality at study sites. In taxonomy, they may be characteristics measured over different species, biological populations, or individuals.

There is a close relationship between Milton Friedman's two-way analysis of variance without replication by ranks and Kendall's coefficient of concordance. They address hypotheses concerning the same data table, and they use the same χ^2 statistic for testing. They differ only in the formulation of their respective null hypothesis. Consider Table 1, which contains illustrative data. In Friedman's test, the null hypothesis is that there is no real difference among the n objects (sites, rows of Table 1) because they pertain to the same statistical population. Under the null hypothesis, they should have received random ranks along the various variables, so that their sums of ranks should be approximately equal. Kendall's test focuses on the m variables. If the null hypothesis of Friedman's test is true, this means that the variables have produced rankings of the objects that are independent of one another. This is the null hypothesis of Kendall's test.

Table 1 Illustrative Example: Ranked Relative Abundances of Four Soil Mite Species (Variables) at 10 Sites (Objects)

	Ranks (column-wise)				Sum of Ranks
	Species 13	Species 14	Species 15	Species 23	Ri
Site 4	5	6	3	5	19.0
Site 9	10	4	8	2	24.0
Site 14	7	8	5	4	24.0
Site 22	8	10	9	2	29.0
Site 31	6	5	7	6	24.0
Site 34	9	7	10	7	33.0
Site 45	3	3	2	8	16.0
Site 53	1.5	2	4	9	16.5
Site 61	1.5	1	1	2	5.5
Site 69	4	9	6	10	29.0

Source: Legendre, P. (2005) Species associations: The Kendall coefficient of concordance revisited. Journal of Agricultural, Biological, & Environmental Statistics, 10, 230. Reprinted with permission from the Journal of Agricultural, Biological, & Environmental Statistics. Copyright 2005 by the American Statistical Association. All rights reserved.

Notes: The ranks are computed columnwise with ties. Right-hand column: sum of the ranks for each site.

Computing Kendall's W

There are two ways of computing Kendall's W statistic (first and second forms of Equations 1 and 2); they lead to the same result. S or S' is computed first from the row-marginal sums of ranks Ri received by the objects:

$$S = \sum_{i=1}^{n} (R_i - \bar{R})^2 \text{ or } S' = \sum_{i=1}^{n} R_i^2 = SSR,$$

where S is a sum-of-squares statistic over the row sums of ranks Ri, and $\bar{R}$ is the mean of the Ri values. Following that, Kendall's W statistic can be obtained from either of the following formulas:

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

where n is the number of objects and m is the number of variables. T is a correction factor for tied ranks:

$$T = \sum_{k = 1}^{g} (t_k^3 - t_k),$$

in which tk is the number of tied ranks in each (k) of g groups of ties. The sum is computed over all groups of ties found in all m variables of the data table. T = 0 when there are no tied values.

Kendall's W is an estimate of the variance of the row sums of ranks Ri divided by the maximum possible value the variance can take; this occurs when all variables are in total agreement. Hence $0 \le W \le 1$, 1 representing perfect concordance. To derive the formulas for W (Equation 2), one has to know that when all variables are in perfect agreement, the sum of all sums of ranks in the data table (right-hand column of Table 1) is $mn(n + 1) / 2$ and that the sum of squares of the sums of all ranks is m2n(n+1)(2n+1) / 6 (without ties).

There is a close relationship between Charles Spearman's correlation coefficient rS and Kendall's W statistic: W can be directly calculated from the mean ($\bar{r}_S$) of the pairwise Spearman correlations rS using the following relationship:

$$W = \frac{(m - 1)\bar{r}_S + 1}{m},$$

where m is the number of variables (judges) among which Spearman correlations are computed. Equation 4 is strictly true for untied observations only; for tied observations, ties are handled in a bivariate way in each Spearman rS coefficient whereas in Kendall's W the correction for ties is computed in a single equation (Equation 3) for all variables. For two variables (judges) only, W is simply a linear transformation of rS: W = (rS+1) / 2. In that case, a permutation test of W for two variables is the exact equivalent of a permutation test of rS for the same variables.

The relationship described by Equation 4 clearly limits the domain of application of the coefficient of concordance to variables that are all meant to estimate the same general property of the objects: variables are considered concordant only if their Spearman correlations are positive. Two variables that give perfectly opposite ranks to a set of objects have a Spearman correlation of -1, hence W = 0 for these two variables (Equation 4); this is the lower bound of the coefficient of concordance. For two variables only, rS = 0 gives W = 0.5. So coefficient W applies well to rankings given by a panel of judges called in to assess overall performance in sports or quality of wines or food in restaurants, to rankings obtained from criteria used in quality tests of appliances or services by consumer organizations, and so forth. It does not apply, however, to variables used in multivariate analysis in which negative as well as positive relationships are informative. Jerrold H. Zar, for example, uses wing length, tail length, and bill length of birds to illustrate the use of the coefficient of concordance. These data are appropriate for W because they are all indirect measures of a common property, the size of the birds.

In ecological applications, one can use the abundances of various species as indicators of the good or bad environmental quality of the study sites. If a group of species is used to produce a global index of the overall quality (good or bad) of the environment at the study sites, only the species that are significantly associated and positively correlated to one another should be included in the index, because different groups of species may be associated to different environmental conditions.

Testing the Significance of W

Friedman's chi-square statistic is obtained from W by the formula

$$\chi^2 = m(n - 1)W.$$

This quantity is asymptotically distributed like chi-square with $v = (n - 1)$ degrees of freedom; it can be used to test W for significance. According to Kendall and Babington Smith, this approach is satisfactory only for moderately large values of m and n.

Sidney Siegel and N. John Castellan Jr. recommend the use of a table of critical values for W when $n \le 7$ and $m \le 20$; otherwise, they recommend testing the chi-square statistic (Equation 5) using the chi-square distribution. Their table of critical values of W for small n and m is derived from a table of critical values of S assembled by Friedman using the z test of Kendall and Babington Smith and reproduced in Kendall's classic monograph, Rank Correlation Methods. Using numerical simulations, Pierre Legendre compared results of the classical chi-square test of the chi-square statistic (Equation 5) to the permutation test that Siegel and Castellan also recommend for small samples (small n). The simulation results showed that the classical chi-square test was too conservative for any sample size (n) when the number of variables m was smaller than 20; the test had rejection rates well below the significance level, so it remained valid. The classical chi-square test had a correct level of Type I error (rejecting a null hypothesis that is true) for 20 variables and more. The permutation test had a correct rate of Type I error for all values of m and n. The power of the permutation test was higher than that of the classical chi-square test because of the differences in rates of Type I error between the two tests. The differences in power disappeared asymptotically as the number of variables increased.

An alternative approach is to compute the following F statistic:

$$F = (m - 1)\, W / (1 - W),$$

which is asymptotically distributed like F with $v_1 = n - 1 - (2/m)$ and $v_2 = v_1(m - 1)$ degrees of freedom. Kendall and Babington Smith described this approach using a Fisher z transformation of the F statistic, z = 0.5 loge(F). They recommended it for testing W for moderate values of m and n. Numerical simulations show, however, that this F statistic has correct levels of Type I error for any value of n and m.

In permutation tests of Kendall's W, the objects are the permutable units under the null hypothesis (the objects are sites in Table 1). For the global test of significance, the rank values in all variables are permuted at random, independently from variable to variable because the null hypothesis is the independence of the rankings produced by all variables. The alternative hypothesis is that at least one of the variables is concordant with one, or with some, of the other variables. Actually, for permutation testing, the four statistics SSR (Equation 1), W (Equation 2), χ^2 (Equation 5), and F (Equation 6) are monotonic to one another since n and m, as well as T, are constant within a given permutation test; thus they are equivalent statistics for testing, producing the same permutational probabilities. The test is one-tailed because it recognizes only positive associations between vectors of ranks. This may be seen if one considers two vectors with exactly opposite rankings: They produce a Spearman statistic of 1, hence a value of zero for W (Equation 4).

Many of the problems subjected to Kendall's concordance analysis involve fewer than 20 variables. The chi-square test should be avoided in these cases. The F test (Equation 6), as well as the permutation test, can safely be used with all values of m and n.

Contributions of Individual Variables to Kendall's Concordance

The overall permutation test of W suggests a way of testing a posteriori the significance of the contributions of individual variables to the overall concordance to determine which of the individual variables are concordant with one or several other variables in the group. There is interest in several fields in identifying discordant variables or judges. This includes all fields that use panels of judges to assess the overall quality of the objects or subjects under study (sports, law, consumer protection, etc.). In other types of studies, scientists are interested in identifying variables that agree in their estimation of a common property of the objects. This is the case in environmental studies in which scientists are interested in identifying groups of concordant species that are indicators of some property of the environment and can be combined into indices of its quality, in particular in situations of pollution or contamination.

The contribution of individual variables to the W statistic can be assessed by a permutation test proposed by Legendre. The null hypothesis is the monotonic independence of the variable subjected to the test, with respect to all the other variables in the group under study. The alternative hypothesis is that this variable is concordant with other variables in the set under study, having similar rankings of values (one-tailed test). The statistic W can be used directly in a posteriori tests. Contrary to the global test, only the variable under test is permuted here. If that variable has values that are monotonically independent of the other variables, permuting its values at random should have little influence on the W statistic. If, on the contrary, it is concordant with one or several other variables, permuting its values at random should break the concordance and induce a noticeable decrease on W.

Two specific partial concordance statistics can also be used in a posteriori tests. The first one is the mean, $\bar{r}_j$, of the pairwise Spearman correlations between variable j under test and all the other variables. The second statistic, W_j, is obtained by applying Equation 4 to $\bar{r}_j$ instead of $\bar{r}$, with m the number of variables in the group. These two statistics are shown in Table 2 for the example data; $\bar{r}_j$ and W_j are monotonic to each other because m is constant in a given permutation test. Within a given a posteriori test, W is also monotonic to W_j because only the values related to variable j are permuted when testing variable j. These three statistics are thus equivalent for a posteriori permutation tests, producing the same permutational probabilities. Like $\bar{r}_j$, Wj can take negative values; this is not the case of W.

Table 2 Results of (a) the Overall and (b) the A Posteriori Tests of Concordance Among the Four Species of Table 1; (c) Overall and (d) A Posteriori Tests of Concordance Among Three Species

(a) Overall test of W statistic, four species. H0: The four species are not concordant with one another.

Kendall's W =	0.44160	Permutational p value = .0448*
F statistic =	2.37252	F distribution p value = .0440*
Friedman's chi-square =	15.89771	Chi-square distribution p value = .0690

(b) A posteriori tests, four species. H0: This species is not concordant with the other three.

	$\bar{r}_j$	Wj	p Value	Corrected p	Decision at α = 5%
Species 13	0.32657	0.49493	.0766	.1532	Do not reject H0
Species 14	0.39655	0.54741	.0240	.0720	Do not reject H0

Species 15	0.45704	0.59278	.0051	.0204*	Reject H0
Species 23	−0.16813	0.12391	.7070	.7070	Do not reject H0

(c) Overall test of W statistic, three species. H0: The three species are not concordant with one another.

Kendall's W =	0.78273	Permutational p value =:0005* asdetsfaf adfa fa
F statistic =	7.20497	F distribution p value =:0003*
Friedman's chi-square =	21.13360	Chi-square distribution p value =.0121*

(d) A posteriori tests, three species. H0: This species is not concordant with the other two.

	$\bar{r}_j$	Wj	p Value	Corrected p	Decision at α = 5%
Species 13	0.69909	0.79939	.0040	.0120*	Reject H0
Species 14	0.59176	0.72784	.0290	.0290*	Reject H0
Species 15	0.73158	0.82105	.0050	.0120*	Reject H0

Source: (a) and (b): Adapted from Legendre, P. (2005). Species associations: The Kendall coefficient of concordance revisited. Journal of Agricultural, Biological, and Environmental Statistics, 10, 233. Reprinted with permission from the Journal of Agricultural, Biological and Environmental Statistics. Copyright 2005 by the American Statistical Association. All rights reserved.

Notes: $\bar{r}_j$ = mean of the Spearman correlations with the other species; Wj = partial concordance per species; p value = permutational probability (9,999 random permutations); corrected p = Holm-corrected p value. * = Reject H0 at =:05:

There are advantages to performing a single a posteriori test for variable j instead of (m − 1) tests of the Spearman correlation coefficients between variable j and all the other variables: The tests of the (m − 1) correlation coefficients would have to be corrected for multiple testing, and they could provide discordant information; a single test of the contribution of variable j to the W statistic has greater power and provides a single, clearer answer. In order to preserve a correct or approximately correct experimentwise error rate, the probabilities of the a posteriori tests computed for all species in a group should be adjusted for multiple testing.

A posteriori tests are useful for identifying the variables that are not concordant with the others, as in the examples, but they do not tell us whether there are one or several groups of congruent variables among those for which the null hypothesis of independence is rejected. This information can be obtained by computing

Spearman correlations among the variables and clustering them into groups of variables that are significantly and positively correlated.

The example data are analyzed in Table 2. The overall permutational test of the W statistic is significant at α = 5%, but marginally (Table 2a). The cause appears when examining the a posteriori tests in Table 2b: Species 23 has a negative mean correlation with the three other species in the group ($\bar{r}_j$ = .168). This indicates that Species 23 does not belong in that group. Were we analyzing a large group of variables, we could look at the next partition in an agglomerative clustering dendrogram, or the next K-means partition, and proceed to the overall and a posteriori tests for the members of these new groups. In the present illustrative example, Species 23 clearly differs from the other three species. We can now test Species 13, 14, and 15 as a group. Table 2c shows that this group has a highly significant concordance, and all individual species contribute significantly to the overall concordance of their group (Table 2d).

In Table 2a and 2c, the F test results are concordant with the permutation test results, but due to small m and n, the chi-square test lacks power.

Discussion

The Kendall coefficient of concordance can be used to assess the degree to which a group of variables provides a common ranking for a set of objects. It should be used only to obtain a statement about variables that are all meant to measure the same general property of the objects. It should not be used to analyze sets of variables in which the negative and positive correlations have equal importance for interpretation. When the null hypothesis is rejected, one cannot conclude that all variables are concordant with one another, as shown in Table 2 (a) and (b); only that at least one variable is concordant with one or some of the others.

The partial concordance coefficients and a posteriori tests of significance are essential complements of the overall test of concordance. In several fields, there is interest in identifying discordant variables; this is the case in all fields that use panels of judges to assess the overall quality of the objects under study (e.g., sports, law, consumer protection). In other applications, one is interested in using the sum of ranks, or the sum of values, provided by several variables or judges, to create an overall indicator of the response of the objects under study. It is advisable to look for one or several groups of variables that rank the objects broadly in the same way, using clustering, and then carry out a posteriori tests on the putative members of each group. Only then can their values or ranks be pooled into an overall index.

See also Friedman Test; Holm's Sequential Bonferroni Procedure; Spearman Rank Order Correlation

Pierre Legendre
http://dx.doi.org/10.4135/9781071812082.n89
10.4135/9781071812082.n89

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association, 32, 675–701.
Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. Annals of Mathematical Statistics, 11, 86–92.
Kendall, M. G. (1948). Rank correlation methods (http://dx.doi.org/1st ed.). London UK: Charles Griffith.
Kendall, M. G., & Babington Smith, B. (1939). The problem of m rankings. Annals of Mathematical Statistics, 10, 275–287.
Legendre, P. (2005). Species associations: The Kendall coefficient of concordance revisited. Journal of Agricultural, Biological, & Environmental Statistics, 10, 226–245.
Zar, J. H. (1999). Biostatistical analysis (4th ed.). Upper Saddle River, NJ: Prentice Hall.