

Simulation Study to Compare the Performance of Signed Klotz and the Signed Mood Generalized Weighted Coefficients

Sandra M. Aleixo[1(✉)] and Júlia Teles[2]

[1] CEAUL and Department of Mathematics, ISEL – Instituto Superior de Engenharia de Lisboa, IPL – Instituto Politécnico de Lisbon, Rua Conselheiro Em ídio Navarro, 1, 1959-007 Lisbon, Portugal
sandra.aleixo@isel.pt

[2] CIPER and Mathematics Unit, Faculdade de Motricidade Humana, Universidade de Lisbon, Estrada da Costa, 1499-002 Cruz Quebrada - Dafundo, Portugal

Abstract. In this work a Monte Carlo simulation study was carried out to compare the performance of three weighted coefficients that emphasized the top and bottom ranks at the same time, namely the signed Klotz and the signed Mood weighted coefficients previously proposed by the authors [2] and the van der Waerden weighted coefficient [5], with the Kendall's coefficient that assigns equal weights to all rankings. The coefficients for m judges studied in this paper generalize the coefficients previously proposed by the authors for two judges in a preliminary work [1] where a simulation study is also carried out. As the main result of the simulation study, we highlight the best performance of Klotz coefficient in detecting concordance in situations where the agreement is located in a lower proportion of extreme ranks, contrary to the case where the agreement is located in a higher proportion of extreme ranks, in which the Signed Mood and van der Waerden coefficients have the best performance.

Keywords: Monte Carlo simulation · Weighted coefficients · Concordance measures · Signed Klotz coefficient · Signed Mood coefficient · van der Waerden coefficient

1 Introduction

Two weighted correlation coefficients R_S, that allow to give more weight to the lower and upper ranks simultaneously, were proposed by the authors in a previous work [1]. These indexes were obtained computing the Pearson correlation coefficient with a modified Klotz and modified Mood scores. In the sequel of that work, two new generalized weighted coefficients A_S, the signed Klotz A_K and

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12251, pp. 157–172, 2020.
https://doi.org/10.1007/978-3-030-58808-3_13

the signed Mood A_M, to measure the concordance among several sets of ranks, putting emphasis on the extreme ranks, were deduced [2]. The van der Waerden weighted coefficient A_W [5] could also be included in these generalized coefficients. In that work, a relationship between the generalized A_S and the average of pairwise R_S was derived. Besides that, the distribution of the generalized weighted agreement coefficients is derived, and an illustrative example is shown. The present paper aims to extend this last work carried out by the authors [2], performing a simulation study to assess the behavior of the new coefficients. So, the goal of this paper is the comparison of the performance of four coefficients: three weighted coefficients to measure agreement among several sets of ranks emphasizing top and bottom ranks at the same time, the signed Klotz, the signed Mood and the van der Waerden coefficients, with the Kendall's coefficient that attributes equal weights to all rankings. This comparison is made through the implementation of a simulation study using the Monte Carlo method, as usual in this kind of studies [3,7,9].

The remainder of paper is organized in five sections. In Sect. 2, the four concordance coefficients compared in the simulation study performed in this work, the Kendall's, the van der Waerden, the Klotz and the Mood coefficient are presented, and treating these coefficients as test statistics, their asymptotic distributions, under the null hypotheses of no agreement among the rankings, are stated. The Monte Carlo simulation study design used in this paper is outlined in Sect. 3. In Sect. 4 the results of the mentioned simulation study are shown and finally, in Sect. 5, conclusions are drawn.

2 Coefficients to Measure Agreement Among Several Sets of Ranks

In the simulation study presented in this paper, it will be used four coefficients to measure agreement among several sets of ranks: the Kendall's coefficient that assigns equal weights to all rankings and three weighted coefficients that emphasized the top and bottom ranks at the same time, the van der Waerden weighted coefficient [5] and the two coefficients previously proposed by the authors, the signed Klotz and the signed Mood [2]. Therefore, this section will briefly present these coefficients, namely the general expression and asymptotic distributions, since they are necessary to carry out the Monte Carlo simulation study.

To present the coefficients described below, it is considered that there are n subjects to be ranked by $m > 2$ independent observers or judges or in m moments, giving rise to m sets of ranks, and also that R_{ij} represents the rank assigned to the jth subject by the ith observer, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

2.1 Kendall's Coefficient of Concordance

The Kendall's coefficient of concordance W [6,11] is the ratio between the variance of the sum of the ranks assigned to subjects, represented by S, and the

maximum possible value reached for the variance of the sum of the ranks taking into account the values of n and m,

$$W = \frac{S}{\max(S)},$$

where $\max(S) = m^2 n(n^2-1)/12$ is the value of S when there is a total agreement among the ranks. Representing by $R_j = \sum_{i=1}^{m} R_{ij}$ the sum of the ranks assigned to the jth subject, for $j = 1, \ldots, n$, the Kendall's coefficient of concordance W is given by

$$W = \frac{n\sum_{j=1}^{n} R_j^2 - \left(\sum_{j=1}^{n} R_j\right)^2}{\frac{m^2 n(n^2-1)}{12}}.$$

The Kendall's coefficient can take values in the interval $[0, 1]$ [6]. This coefficient attains: the maximum value 1 when there is perfect agreement among the m sets of ranks, only if there are no tied observations; the minimum value zero when there is no agreement, only if $m(n+1)$ is even.

Although the coefficient W is normally used as a measure of agreement among rankings, it could also be used as a test statistic. There is a relationship between W and the Friedman χ_r^2 statistic, $\chi_r^2 = m(n-1)W$, which has an approximate chi-squared distribution with $n-1$ degrees of freedom [4].

2.2 Weighted Coefficients to Measure the Concordance Among Several Sets of Ranks Emphasizing Extreme Ranks at the Same Time

Consider a generic score S_{ij} which is computed based on the rank R_{ij}, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. The score S_{ij} can be:

- the van der Waerden score [10]

$$W_{ij} = \Phi^{-1}\left(\frac{R_{ij}}{n+1}\right),$$

- the signed Klotz score [1]

$$SK_{ij} = sign\left(R_{ij} - \frac{n+1}{2}\right)\left(\Phi^{-1}\left(\frac{R_{ij}}{n+1}\right)\right)^2,$$

- or the signed Mood score [1]

$$SM_{ij} = sign\left(R_{ij} - \frac{n+1}{2}\right)\left(R_{ij} - \frac{n+1}{2}\right)^2.$$

Note that, for the generic score S_{ij} are valid the following equalities:

1. $\sum_{j=1}^{n} S_{1j}^2 = \sum_{j=1}^{n} S_{2j}^2 = \cdots = \sum_{j=1}^{n} S_{mj}^2 = C_S$,
   where C_S is a non null constant that depends on the sample size n and the scores S_{ij}.
2. $\sum_{j=1}^{n} S_{1j} = \sum_{j=1}^{n} S_{2j} = \cdots = \sum_{j=1}^{n} S_{mj} = 0$, that is, $S_{i\bullet} = 0$, for $i = 1, 2, \ldots, m$.

   These happens because one has:
   - in the case of the van der Waerden and signed Klotz scores
     (a) if n odd,

$$\Phi^{-1}\left(\frac{k}{n+1}\right) = -\Phi^{-1}\left(\frac{n+1-k}{n+1}\right),$$

for $k = 1, 2, \ldots, \left[\frac{n}{2}\right]$
(where $[k]$ represents the integer part of k),
and

$$\Phi^{-1}\left(\frac{(n+1)/2}{n+1}\right) = \Phi^{-1}\left(\frac{1}{2}\right) = 0;$$

     (b) if n even,

$$\Phi^{-1}\left(\frac{k}{n+1}\right) = -\Phi^{-1}\left(\frac{n+1-k}{n+1}\right),$$

for $k = 1, 2, \ldots, \frac{n}{2}$;

   - in the case of signed Mood scores

$$\left(k - \frac{n+1}{2}\right)^2 = \left(n+1-k-\frac{n+1}{2}\right)^2,$$

for $k = 1, 2, \ldots, \left[\frac{n}{2}\right]$.

In a previous work [2], the authors proposed two new weighted concordance coefficients, built similarly to Kendall's coefficient of concordance W defined in Sect. 2.1. These new coefficients, as well as the van der Waerden coefficient, are based on the ratio between the variance of the n sums of m scores awarded to subjects, denoted by S_{SS}^2, and the maximum value that this variance can attain considering the values of m and n. So, the weighted coefficients of agreement to measure the concordance among m sets of ranks, giving more weight to the extremes ones, can be defined by

$$A_S = \frac{S_{SS}^2}{\max S_{SS}^2},$$

with

$$S_{ss}^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(S_{\bullet j} - \overline{S}_{\bullet\bullet}\right)^2,$$

where $S_{\bullet j}$ is the sum of the scores of m observers for the subject j and $\overline{S}_{\bullet\bullet}$ is the average of all $m \times n$ scores S_{ij}, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

Attending that $\overline{S}_{\bullet\bullet} = \sum_{j=1}^{n} S_{\bullet j}/n = 0$ then

$$S_{ss}^2 = \frac{1}{n-1}\sum_{j=1}^{n} S_{\bullet j}^2.$$

Therefore, making some calculus (see [2] for details), the weighted coefficients to measure the concordance among m sets of ranks, putting emphasis on the extremes ones (van der Waerden, Klotz and Mood coefficients), can be given by the general expression

$$A_S = \frac{\sum_{j=1}^{n}\left(\sum_{i=1}^{m} S_{ij}\right)^2}{m^2\, C_S}. \tag{1}$$

These coefficients take values in the interval $[0, 1]$.

Similarly to the Kendall's coefficient, these three weighted coefficients can be used as test statistics. Their asymptotic distributions were postulated in a theorem, proved in previous work by the authors [2], which states that: under the null hypotheses of no agreement or no association among the rankings, the statistic $m(n-1)A_S$ has an asymptotic chi-square distribution with $n-1$ degrees of freedom.

3 Simulation Study Design

Using R software, version 3.6.2 [8], a Monte Carlo simulation study was carried out to compare the performance of the three weighted coefficients to measure agreement among several sets of ranks emphasizing top and bottom ranks at the same time, presented in Sect. 2.2, with the Kendall's coefficient of concordance that awards equal weights to all rankings, shown in Sect. 2.1. Coefficients performance was estimated by the proportion of rejected null hypotheses, when testing whether the underlying population concordance coefficient is higher than zero. To simplify the writing, from now on, we will refer this proportion as power of coefficient.

The main objective of this simulation study is to contrast the virtue of each one of the three coefficients in identifying agreement among various sets of ranks given emphasis on both lower and upper ranks simultaneously, when compared with the Kendall's coefficient of concordance.

The data generation scheme adopted in this study was similar to the one followed by Legendre [7]. The first group of observations was produced through

the generation of a n-dimensional vector of standard random normal deviates. To obtain each one of the others groups of n observations, for some proportion $0 < p < 0.5$, and choosing an appropriate standard deviation σ, it was made the following:

1. random normal deviates with zero mean and standard deviation σ were added to the observations with the most extreme ranks in the first group of observations, that is for $i = 1, \ldots, [np]$ and $i = n - [np] + 1, \ldots, n$;
2. random normal deviates with zero mean and standard deviation σ were generated for the observations with the intermediate ranks in the first sample, that is for $i = [np] + 1, \ldots, n - [np]$.

The m sets of observations were converted into m sets of ranks considering an ascending order. These m sets of ranks are the related samples that were considered.

Simulating the data by this way, the agreement among the groups of ranks is higher in the lower and upper ranks than in the intermediate ones. The six values considered for σ ($\sigma = 0.25, 0.5, 1, 2, 3, 5$) allow to evaluate the performance of coefficients for several intensities of agreement, being that the lower values of σ correspond to higher degrees of agreement. For the proportion of ranks that are in agreement, two values were taken into account ($p = 0.1, 0.25$), which allow to compare the performance of the coefficients in a scenario where the concordance was concentrated on a higher proportion of extreme ranks (scenario 1: $p = 0.25$), with a scenario in which the concordance was focused on a lower proportion of extreme ranks (scenario 2: $p = 0.1$).

Synthesizing, in the simulation study were taken into account:

- three different number of independent observers or judges, $m = 3, 4, 5$;
- five different number of subjects, $n = 20, 30, 50, 100, 200$;
- two values considered for the proportion of ranks that are in agreement, $p = 0.1, 0.25$;
- and six intensities of agreement, $\sigma = 0.25, 0.5, 1, 2, 3, 5$.

Considering all possible combinations of these values, 180 simulated conditions were evaluated. For each simulated condition, 10000 replications were run in order to:

1. calculate the means and standard deviations of Kendall's, van der Waerden, signed Klotz and signed Mood coefficients for the simulated samples;
2. estimate the power of each coefficient by the percentage of rejected null hypotheses, assessed at 5% significance level, when testing whether the underlying population concordance coefficient is higher than zero.

4 Simulation Study Results

The simulation results related to the means and standard deviations of the concordance coefficients are given in Table 1, Table 2 and Table 3 for $m = 3, 4, 5$, respectively, while the corresponding powers are shown for $m = 3$ in Fig. 1 for

scenario 1 and in Fig. 2 for scenario 2, for $m = 4$ in Table 4, and for $m = 5$ in Table 5. The tables are in the Appendix. The most relevant results presented in the mentioned figures and tables were analyzed and are exposed below.

To achieve a better understanding of the results of the simulation study, means and standard deviations of the simulated concordance coefficients are presented in Table 1, Table 2 and Table 3. Analyzing these tables, in what concerns the mean estimates for the coefficients under study, we highlight the following results:

- Fixing n and σ, for all values of m considered, the means for all four coefficients estimates are higher in scenario 1 ($p = 0.25$) than the respective values in scenario 2 ($p = 0.1$); this occurs, as expected, because in scenario 1 there is a higher percentage of ranks that are in agreement;
- For both scenarios, for all values of m and n in study, and for all the coefficients, for higher degrees of agreement, corresponding to smaller values of σ, the means of coefficients estimates are higher than for smaller degrees of agreement; that is, as expected, the means of coefficients estimates decrease as σ increases;
- Setting the higher values of σ, for all values of m, for both scenarios, and for each coefficient, the mean estimates are very identical for any sample size n under study;
- Fixing the values of n and σ, for both scenarios and for all coefficients, the mean estimates decrease as m increases;
- For both scenarios, the means of coefficients estimates are higher for the three weighed coefficients A_W, A_K and A_M, when compared with the Kendall's coefficient W, being this difference more evident in scenario 2;
- In both scenarios, the means of van der Waerden coefficient estimates are slightly lower than the correspondent values of the other two weighted coefficients;
- Considering all values of m and n, for higher values of σ, the means of all the four coefficients estimates are similar, for both scenarios.
- Considering all values of m and n, for smaller values of σ, in scenario 1, the means of the Klotz and the Mood coefficients estimates remain practically the same, being slightly higher than the other two coefficients, with Kendall's having the lowest values as already mentioned; whereas in scenario 2, the means of the Klotz coefficient estimates are higher than the values of the other three coefficients, the means of van der Waerden and the Mood coefficients estimates are quite close, being however the values of van der Waerden slightly lower.

Concerning the empirical standard deviations, no pattern was found when comparing the four coefficients, the both scenarios and the six intensities of agreement. The standard deviations of the four coefficients decreased, as n and m increased.

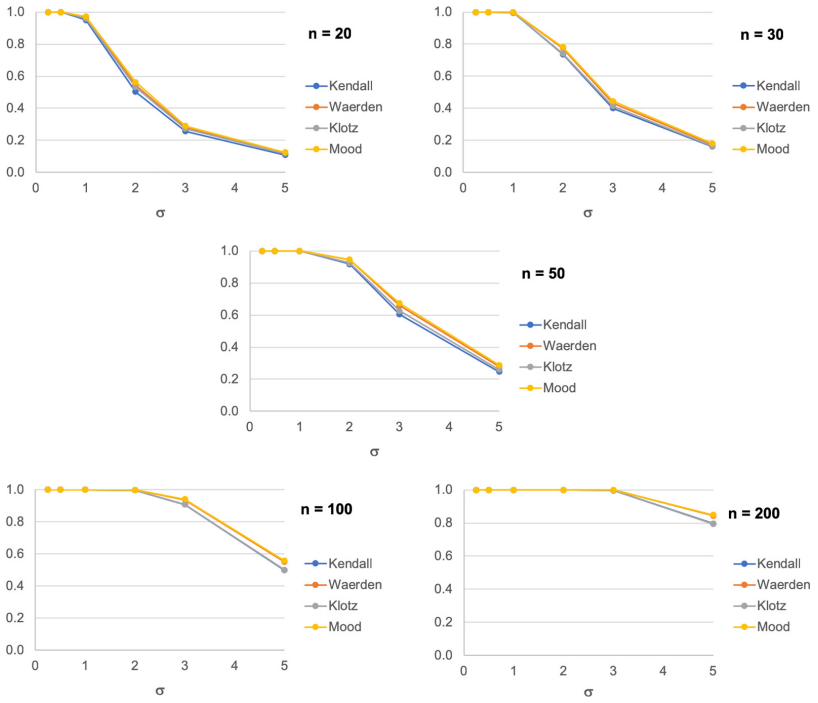

Fig. 1. Powers (%) of Kendall's coefficient of concordance, van der Waerden, signed Klotz, and signed Mood weighted coefficients for $\boldsymbol{m = 3}$ in the scenario 1 ($\boldsymbol{p = 0.25}$), in which the concordance was targeted for a higher proportion of extreme ranks.

Regarding the powers of the four coefficients, analyzing Fig. 1 for scenario 1 and in Fig. 2 for scenario 2, for $m = 3$, and Table 4 and Table 5 for $m = 4$ and $m = 5$ respectively, it can be stated that:

- Fixing n and σ, for all values of m considered, the powers of all indexes are higher in scenario 1 ($p = 0.25$) than the respective values in scenario 2 ($p = 0.1$); this occurs, as expected, because in scenario 1 there is a higher percentage of ranks that are in agreement;
- For both scenarios and for all the number of judges considered ($m = 3, 4, 5$), the powers of all coefficients are higher for smaller values of σ when compared to the higher ones;
- For both scenarios, the three weighed coefficients A_W, A_K and A_M have higher powers than the Kendall's coefficient of concordance W; nevertheless, this difference is more evident in scenario 2 in which the concordance was concentrated on a lower proportion of extreme ranks;
- In scenario 1 (respectively, in scenario 2), for higher values of σ, the Klotz powers are slightly lower (respectively, higher) than the correspondent values of the other two weighted coefficients, what is more evident for lower values of n;

- For all values of m and n, for both scenarios, towards smaller values of σ, the powers attained the maximum value (100%), but for higher values of σ, the powers decrease as σ increases, being this wane more pronounced for small values of n, where the achieved powers are far from been reasonable; for higher values of σ, the powers are not reasonable (because they are much below 70%), and this is worse in scenario 2;
- For all values of m and n, in scenario 1, the powers of all the four coefficients have a similar behavior for each one of the six intensities of agreement in study, whereas, in scenario 2, for higher values of σ, the powers of all coefficients show a distinct behavior, being the powers of the Klotz coefficient the ones that reach always the highest values for the same n and σ, while the van der Waerden and the Mood weighted coefficients have lower and quite similar values;
- For a certain value of n and a fixed σ, the powers of the four coefficients increase as m increases.

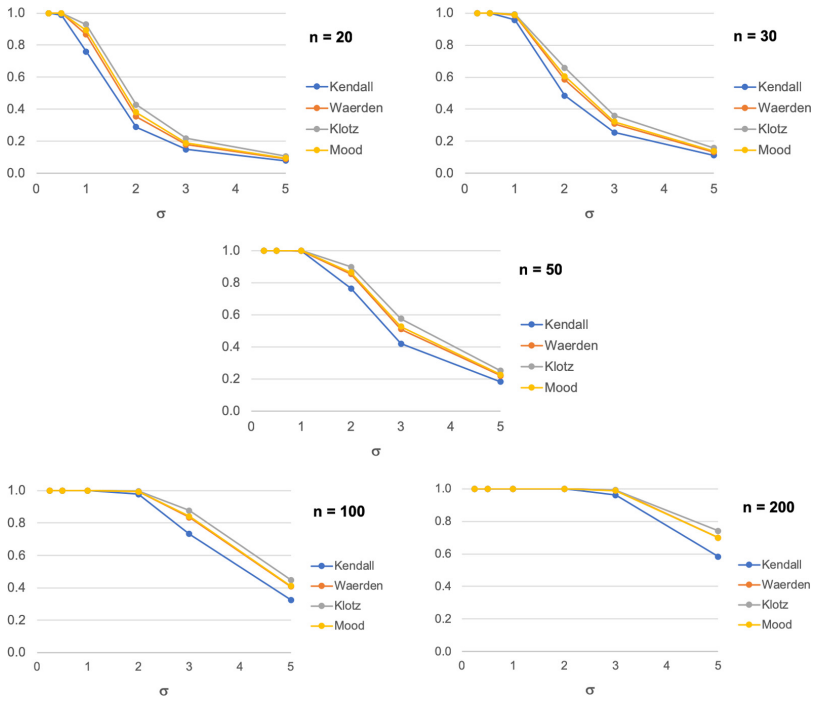

Fig. 2. Powers (%) of Kendall's coefficient of concordance, van der Waerden, signed Klotz, and signed Mood weighted coefficients for $\boldsymbol{m = 3}$ in the scenario 2 ($\boldsymbol{p = 0.1}$) in which the concordance was targeted for a lower proportion of extreme ranks.

5 Conclusions

In this paper, the performance of the three generalized weighted agreement coefficients A_S to measure the concordance among several sets of ranks putting emphasis on the extreme ranks, the two new ones proposed by the authors [2] — the generalized signed Klotz and the signed Mood coefficients— and the van der Waerden coefficient, was compared with the behavior of Kendall's coefficient of concordance W that assigns equal weights to all rankings.

In order to assess whether the generalized weighted agreement coefficients, especially the signed Klotz and the signed Mood coefficients, can be a benefit in assessing the concordance among several sets of ranks, when one intends to place more emphasis on the lower and upper ranks simultaneously, a Monte Carlo simulation study was carried out.

The simulation study allowed to show some interesting results. It can be noticed that all coefficients in study, the three generalized weighted ones and the non-weighted, showed a better performance when the agreement is focused on a higher proportion of extreme ranks than when it is located on a smaller proportion. It can also be stated that the behavior of all coefficients is better for higher degrees of agreement when compared to the smaller ones, for both proportions of extreme ranks under study and for all the number of judges considered. It can still be seen that for a fixed number of subjects and intensity of agreement, the powers of the four coefficients increase as the number of judges increases.

In situations where it is important to give relevance to the agreement on the lower and upper ranks at the same time, we realized that one of the three generalized weighted agreement coefficients should be used instead of Kendall's coefficient of concordance. Besides that, we emphasize that the choice of one of the three types of generalized weighted agreement coefficients depends on the amount of extreme ranks that are the focus of the assessment. It is very interesting to verify that, in cases where the concordance was focused on a lower proportion of extreme ranks, the Klotz coefficient show the best performance when compared with the other two generalized weighted coefficients. But on the contrary, this coefficient is the worst in occurrences where the agreement is located in a higher proportion of extreme ranks, being the Signed Mood and van der Waerden coefficients the ones that reveal the best performance in this case. In both scenarios, the Signed Mood and van der Waerden coefficients present quite similar powers to the several intensities of agreement considered in this study, for a fixed number of subjects and a fixed number of judges.

We believe that this simulation study helps to understand the profits that can result from the use of the generalized weighted coefficients of agreement A_S, instead of the non weighted Kendall's coefficient of concordance, in situations where it is intended to weight the extreme ranks simultaneously.

Acknowledgments. Sandra M. Aleixo was partly supported by the Fundação para a Ciência e Tecnologia, under Grant UIDB/00006/2020 to CEAUL - Centre of Statistics and its Applications. Júlia Teles was partly supported by the Fundação para a Ciência e Tecnologia, under Grant UIDB/00447/2020 to CIPER - Centro Interdisciplinar para o Estudo da Performance Humana (unit 447). *Conflict of Interest*: None declared.

Appendix

Table 1. Mean (standard deviation) of Kendall's W, van der Waerden A_W, signed Klotz A_K, and signed Mood A_M coefficients estimates, with $m = 3$, for two scenarios: on the left, the concordance was targeted for a higher proportion ($p = 0.25$) of extreme ranks, and on the right, the concordance was focused on a lower proportion ($p = 0.1$) of extreme ranks.

n	σ	Scenario 1 ($p = 0.25$)				Scenario 2 ($p = 0.1$)			
		W	A_W	A_K	A_M	W	A_W	A_K	A_M
20	0.25	.87(.03)	.90(.02)	.94(.03)	.94(.02)	.59(.06)	.66(.05)	.80(.03)	.70(.04)
	0.5	.82(.05)	.84(.05)	.87(.06)	.87(.05)	.58(.06)	.64(.06)	.76(.06)	.67(.06)
	1	.68(.08)	.70(.08)	.70(.09)	.71(.08)	.52(.08)	.55(.08)	.62(.10)	.57(.09)
	2	.53(.10)	.54(.10)	.54(.10)	.54(.10)	.43(.09)	.45(.09)	.47(.10)	.46(.10)
	3	.46(.10)	.47(.10)	.47(.10)	.47(.10)	.40(.09)	.41(.09)	.42(.10)	.41(.09)
	5	.41(.09)	.41(.10)	.41(.10)	.41(.10)	.37(.09)	.38(.09)	.39(.10)	.38(.09)
30	0.25	.90(.02)	.92(.02)	.95(.02)	.95(.01)	.61(.05)	.69(.04)	.84(.02)	.73(.03)
	0.5	.84(.04)	.86(.04)	.87(.05)	.88(.04)	.60(.05)	.67(.04)	.80(.04)	.70(.04)
	1	.70(.07)	.72(.06)	.71(.07)	.72(.07)	.54(.07)	.58(.07)	.65(.08)	.59(.07)
	2	.54(.08)	.55(.08)	.54(.08)	.55(.08)	.45(.07)	.47(.08)	.49(.08)	.47(.08)
	3	.47(.08)	.48(.08)	.47(.08)	.48(.08)	.41(.08)	.42(.08)	.43(.08)	.42(.08)
	5	.41(.08)	.42(.08)	.41(.08)	.42(.08)	.38(.07)	.38(.08)	.39(.08)	.39(.08)
50	0.25	.89(.01)	.92(.01)	.95(.02)	.95(.01)	.63(.03)	.72(.03)	.87(.02)	.75(.02)
	0.5	.83(.03)	.86(.03)	.88(.03)	.88(.03)	.62(.04)	.70(.03)	.83(.03)	.73(.03)
	1	.69(.05)	.72(.05)	.72(.06)	.72(.05)	.55(.05)	.60(.05)	.67(.06)	.61(.05)
	2	.54(.06)	.55(.06)	.54(.06)	.55(.06)	.46(.06)	.48(.06)	.50(.07)	.48(.06)
	3	.47(.06)	.48(.06)	.47(.06)	.48(.06)	.41(.06)	.43(.06)	.44(.06)	.43(.06)
	5	.41(.06)	.42(.06)	.41(.06)	.42(.06)	.38(.06)	.39(.06)	.39(.06)	.39(.06)
100	0.25	.90(.01)	.93(.01)	.96(.01)	.96(.01)	.64(.02)	.74(.02)	.89(.01)	.77(.01)
	0.5	.85(.02)	.87(.02)	.88(.02)	.89(.02)	.64(.02)	.72(.02)	.85(.02)	.74(.02)
	1	.70(.03)	.73(.03)	.72(.04)	.73(.03)	.56(.03)	.62(.04)	.69(.05)	.62(.04)
	2	.54(.04)	.55(.04)	.54(.05)	.56(.04)	.46(.04)	.49(.04)	.51(.05)	.49(.04)
	3	.47(.04)	.48(.04)	.47(.04)	.48(.04)	.42(.04)	.43(.04)	.45(.05)	.44(.04)
	5	.41(.04)	.42(.04)	.42(.04)	.42(.04)	.38(.04)	.39(.04)	.40(.04)	.39(.04)
200	0.25	.91(.01)	.94(.00)	.96(.01)	.96(.00)	.65(.02)	.75(.01)	.90(.01)	.77(.01)
	0.5	.85(.01)	.88(.01)	.89(.02)	.90(.01)	.64(.02)	.73(.01)	.86(.01)	.75(.01)
	1	.71(.02)	.73(.02)	.72(.03)	.74(.02)	.57(.02)	.63(.02)	.70(.03)	.63(.03)
	2	.54(.03)	.56(.03)	.55(.03)	.56(.03)	.47(.03)	.49(.03)	.51(.03)	.49(.03)
	3	.47(.03)	.48(.03)	.47(.03)	.48(.03)	.42(.03)	.44(.03)	.45(.03)	.44(.03)
	5	.42(.03)	.42(.03)	.42(.03)	.42(.03)	.38(.03)	.39(.03)	.40(.03)	.39(.03)

Table 2. Mean (standard deviation) of Kendall's W, van der Waerden A_W, signed Klotz A_K, and signed Mood A_M coefficients estimates, with $m = 4$, for two scenarios: on the left, the concordance was targeted for a higher proportion ($p = 0.25$) of extreme ranks, and on the right, the concordance was focused on a lower proportion ($p = 0.1$) of extreme ranks.

n	σ	Scenario 1 ($p = 0.25$)				Scenario 2 ($p = 0.1$)			
		W	A_W	A_K	A_M	W	A_W	A_K	A_M
20	0.25	.86(.03)	.89(.02)	.93(.03)	.93(.02)	.54(.05)	.61(.04)	.78(.03)	.66(.04)
	0.5	.79(.05)	.81(.05)	.84(.06)	.84(.05)	.52(.05)	.59(.05)	.73(.06)	.63(.05)
	1	.62(.08)	.64(.08)	.64(.09)	.65(.09)	.45(.07)	.48(.08)	.55(.10)	.50(.08)
	2	.44(.09)	.45(.09)	.45(.09)	.46(.09)	.35(.08)	.37(.08)	.38(.09)	.37(.08)
	3	.37(.09)	.38(.09)	.38(.09)	.38(.09)	.31(.08)	.32(.08)	.33(.08)	.33(.08)
	5	.32(.08)	.32(.08)	.32(.08)	.32(.08)	.29(.08)	.29(.08)	.30(.08)	.29(.08)
30	0.25	.89(.02)	.91(.02)	.94(.03)	.94(.02)	.57(.04)	.65(.03)	.82(.02)	.69(.03)
	0.5	.81(.04)	.84(.04)	.85(.05)	.86(.04)	.55(.04)	.63(.04)	.77(.04)	.66(.04)
	1	.64(.07)	.66(.07)	.65(.07)	.66(.07)	.47(.06)	.51(.06)	.58(.08)	.53(.07)
	2	.45(.07)	.46(.07)	.46(.08)	.47(.08)	.37(.06)	.38(.07)	.40(.08)	.39(.07)
	3	.38(.07)	.39(.07)	.38(.07)	.39(.07)	.32(.06)	.33(.07)	.34(.07)	.34(.07)
	5	.32(.07)	.33(.07)	.32(.07)	.33(.07)	.29(.06)	.30(.06)	.30(.06)	.30(.06)
50	0.25	.87(.01)	.91(.01)	.94(.02)	.94(.01)	.59(.03)	.68(.02)	.85(.01)	.72(.02)
	0.5	.80(.03)	.84(.03)	.85(.04)	.86(.03)	.57(.03)	.66(.03)	.80(.03)	.69(.03)
	1	.63(.05)	.66(.05)	.66(.06)	.66(.05)	.49(.05)	.54(.05)	.61(.07)	.55(.05)
	2	.45(.06)	.46(.06)	.46(.06)	.47(.06)	.37(.05)	.39(.05)	.41(.06)	.40(.05)
	3	.38(.05)	.39(.05)	.38(.06)	.39(.05)	.33(.05)	.34(.05)	.35(.05)	.34(.05)
	5	.32(.05)	.33(.05)	.32(.05)	.33(.05)	.29(.05)	.30(.05)	.31(.05)	.30(.05)
100	0.25	.89(.01)	.92(.01)	.95(.01)	.95(.01)	.60(.02)	.71(.01)	.87(.01)	.74(.01)
	0.5	.82(.02)	.85(.02)	.86(.03)	.87(.02)	.59(.02)	.68(.02)	.82(.02)	.71(.02)
	1	.65(.04)	.67(.03)	.66(.04)	.67(.04)	.50(.03)	.56(.03)	.63(.05)	.56(.04)
	2	.46(.04)	.47(.04)	.46(.04)	.47(.04)	.38(.04)	.40(.04)	.42(.04)	.40(.04)
	3	.38(.04)	.39(.04)	.38(.04)	.39(.04)	.33(.03)	.35(.04)	.36(.04)	.35(.04)
	5	.33(.04)	.33(.04)	.33(.04)	.33(.04)	.30(.03)	.30(.03)	.31(.04)	.30(.03)
200	0.25	.89(.01)	.93(.00)	.95(.01)	.95(.00)	.61(.01)	.72(.01)	.88(.01)	.74(.01)
	0.5	.82(.01)	.86(.01)	.86(.02)	.87(.01)	.60(.01)	.70(.01)	.83(.01)	.72(.01)
	1	.65(.03)	.68(.02)	.66(.03)	.68(.03)	.51(.02)	.57(.02)	.64(.03)	.57(.03)
	2	.46(.03)	.47(.03)	.46(.03)	.47(.03)	.38(.03)	.41(.03)	.43(.03)	.41(.03)
	3	.38(.03)	.39(.03)	.38(.03)	.39(.03)	.34(.02)	.35(.03)	.36(.03)	.35(.03)
	5	.33(.03)	.33(.03)	.32(.03)	.33(.03)	.30(.02)	.31(.02)	.31(.03)	.31(.02)

Table 3. Mean (standard deviation) of Kendall's W, van der Waerden A_W, signed Klotz A_K, and signed Mood A_M coefficients estimates, with $m = 5$, for two scenarios: on the left, the concordance was targeted for a higher proportion ($p = 0.25$) of extreme ranks, and on the right, the concordance was focused on a lower proportion ($p = 0.1$) of extreme ranks.

n	σ	Scenario 1 ($p = 0.25$)				Scenario 2 ($p = 0.1$)			
		W	A_W	A_K	A_M	W	A_W	A_K	A_M
20	0.25	.85(.03)	.88(.02)	.92(.03)	.92(.02)	.51(.04)	.59(.03)	.76(.02)	.64(.03)
	0.5	.77(.05)	.80(.05)	.82(.06)	.82(.06)	.49(.05)	.56(.05)	.71(.06)	.60(.05)
	1	.58(.08)	.60(.08)	.60(.09)	.61(.09)	.40(.07)	.44(.07)	.51(.10)	.46(.08)
	2	.39(.08)	.40(.08)	.40(.09)	.40(.08)	.30(.07)	.32(.07)	.33(.08)	.32(.08)
	3	.32(.08)	.32(.08)	.32(.08)	.33(.08)	.26(.07)	.27(.07)	.28(.07)	.27(.07)
	5	.26(.07)	.27(.07)	.26(.07)	.27(.07)	.23(.06)	.24(.06)	.24(.07)	.24(.07)
30	0.25	.88(.02)	.91(.02)	.93(.03)	.94(.02)	.54(.03)	.63(.02)	.81(.02)	.67(.02)
	0.5	.79(.04)	.82(.04)	.83(.05)	.84(.04)	.52(.04)	.60(.03)	.75(.05)	.64(.04)
	1	.60(.07)	.62(.07)	.61(.07)	.63(.07)	.43(.05)	.47(.06)	.54(.08)	.49(.07)
	2	.40(.07)	.41(.07)	.40(.07)	.41(.07)	.32(.06)	.33(.06)	.35(.07)	.34(.06)
	3	.33(.06)	.33(.06)	.33(.06)	.33(.06)	.27(.06)	.28(.06)	.29(.06)	.28(.06)
	5	.27(.06)	.27(.06)	.27(.06)	.27(.06)	.24(.05)	.24(.05)	.25(.05)	.24(.05)
50	0.25	.86(.01)	.90(.01)	.94(.02)	.94(.01)	.56(.02)	.66(.02)	.84(.01)	.70(.02)
	0.5	.79(.03)	.82(.03)	.84(.04)	.84(.03)	.54(.03)	.63(.02)	.79(.03)	.67(.03)
	1	.60(.05)	.62(.05)	.62(.06)	.63(.05)	.45(.04)	.50(.05)	.57(.07)	.51(.05)
	2	.40(.05)	.41(.05)	.40(.05)	.41(.05)	.32(.04)	.34(.05)	.36(.06)	.35(.05)
	3	.32(.05)	.33(.05)	.33(.05)	.33(.05)	.28(.04)	.29(.04)	.30(.05)	.29(.05)
	5	.27(.04)	.27(.04)	.27(.04)	.27(.04)	.24(.04)	.25(.04)	.25(.04)	.25(.04)
100	0.25	.88(.01)	.92(.01)	.94(.01)	.95(.01)	.57(.02)	.69(.01)	.86(.01)	.72(.01)
	0.5	.80(.02)	.83(.02)	.84(.03)	.85(.02)	.56(.02)	.66(.02)	.81(.02)	.69(.02)
	1	.61(.04)	.63(.04)	.62(.04)	.64(.04)	.46(.03)	.52(.03)	.59(.05)	.52(.04)
	2	.41(.04)	.42(.04)	.41(.04)	.42(.04)	.33(.03)	.35(.03)	.37(.04)	.35(.03)
	3	.33(.03)	.34(.03)	.33(.03)	.34(.03)	.28(.03)	.29(.03)	.30(.03)	.29(.03)
	5	.27(.03)	.27(.03)	.27(.03)	.27(.03)	.24(.03)	.25(.03)	.25(.03)	.25(.03)
200	0.25	.89(.00)	.92(.00)	.95(.01)	.95(.00)	.58(.01)	.70(.01)	.87(.01)	.73(.01)
	0.5	.81(.02)	.84(.01)	.84(.02)	.86(.01)	.57(.01)	.67(.01)	.82(.02)	.70(.01)
	1	61(.02)	.64(.02)	.62(.03)	.64(.03)	.47(.02)	.53(.02)	.60(.03)	.53(.03)
	2	.41(.03)	.42(.03)	.41(.03)	.42(.03)	.33(.02)	.36(.02)	.37(.03)	.36(.02)
	3	.33(.02)	34(.02)	.33(.02)	.34(.02)	.28(.02)	.30(.02)	.30(.02)	.30(.02)
	5	.27(.02)	.28(.02)	.27(.02)	.28(.02)	.24(.02)	.25(.02)	.25(.02)	.25(.02)

Table 4. Estimated powers (%) of Kendall's W, van der Waerden A_W, signed Klotz A_K, and signed Mood A_M coefficients, with $m = 4$, for two scenarios: on the left, the concordance was targeted for a higher proportion ($p = 0.25$) of extreme ranks, and on the right, the concordance was focused on a lower proportion ($p = 0.1$) of extreme ranks.

n	σ	Scenario 1 ($p = 0.25$)				Scenario 2 ($p = 0.1$)			
			A_W	A_K	A_M	W	A_W	A_K	A_M
20	0.25	100.00	100.00	100.00	100.00	99.97	100.00	100.00	100.00
	0.5	100.00	100.00	100.00	100.00	98.82	99.85	100.00	99.92
	1	99.38	99.61	99.46	99.62	76.02	86.59	92.92	89.32
	2	70.68	73.53	72.11	74.45	28.84	35.44	42.81	38.02
	3	39.23	42.13	40.78	43.04	14.88	17.80	21.85	18.83
	5	17.19	18.43	18.11	18.92	7.89	8.96	10.60	9.41
30	0.25	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	0.5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	1	99.99	100.00	99.99	100.00	95.64	98.63	99.37	98.95
	2	87.98	90.41	88.13	90.78	48.46	58.49	65.87	60.64
	3	58.38	61.89	58.28	62.74	25.37	30.87	36.00	32.20
	5	24.57	25.69	24.37	26.17	11.14	13.16	15.75	13.76
50	0.25	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	0.5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	1	100.00	100.00	100.00	100.00	99.86	99.98	99.99	99.99
	2	97.90	98.60	97.88	98.69	76.43	85.40	89.92	86.50
	3	77.04	80.87	78.19	81.63	41.96	50.91	57.51	52.66
	5	35.67	38.90	37.17	39.94	18.07	21.86	25.11	22.62
100	0.25	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	0.5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	1	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	2	100.00	100.00	100.00	100.00	97.71	99.39	99.65	99.36
	3	97.48	98.34	97.18	98.39	73.21	83.37	87.64	84.05
	5	64.82	69.66	64.88	69.84	32.52	40.78	44.83	41.14
200	0.25	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	0.5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	1	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	2	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	3	99.99	99.99	99.98	100.00	96.07	98.98	99.25	98.91
	5	90.69	93.29	89.34	93.50	58.23	69.97	74.09	69.89

Table 5. Estimated powers (%) of Kendall's W, van der Waerden A_W, signed Klotz A_K, and signed Mood A_M coefficients, with $m = 5$, for two scenarios: on the left, the concordance was targeted for a higher proportion ($p = 0.25$) of extreme ranks, and on the right, the concordance was focused on a lower proportion ($p = 0.1$) of extreme ranks.

n	σ	Scenario 1 ($p = 0.25$)				Scenario 2 ($p = 0.1$)			
		W	A_W	A_K	A_M	W	A_W	A_K	A_M
20	0.25	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	0.5	100.00	100.00	100.00	100.00	99.85	99.97	99.98	99.97
	1	99.75	99.85	99.84	99.87	89.92	94.90	97.01	95.75
	2	81.48	83.75	82.46	84.69	41.44	48.16	55.11	50.95
	3	49.25	52.06	50.31	53.00	20.28	23.97	28.27	25.49
	5	21.36	22.46	21.79	23.01	9.94	11.19	13.20	11.82
30	0.25	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	0.5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	1	99.99	99.99	99.98	99.99	99.13	99.75	99.89	99.76
	2	94.60	95.62	94.53	95.70	64.25	72.51	78.17	74.40
	3	68.94	71.98	68.54	72.35	32.68	39.11	44.73	40.49
	5	31.42	33.07	31.26	33.66	13.74	16.44	19.78	17.18
50	0.25	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	0.5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	1	100.00	100.00	100.00	100.00	99.98	100.00	100.00	100.00
	2	99.51	99.73	99.49	99.72	88.61	93.48	95.14	93.74
	3	86.89	89.39	86.91	89.77	54.84	63.74	68.83	65.00
	5	44.97	48.75	45.99	49.75	23.71	28.19	31.38	28.94
100	0.25	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	0.5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	1	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	2	100.00	100.00	100.00	100.00	99.53	99.88	99.96	99.87
	3	99.23	99.61	99.07	99.61	84.57	91.73	93.38	91.95
	5	74.44	78.68	73.60	79.10	41.73	50.47	54.31	51.60
200	0.25	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	0.5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	1	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	2	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	3	100.00	100.00	100.00	100.00	98.88	99.69	99.82	99.71
	5	95.35	97.04	94.83	97.02	68.76	79.32	82.43	79.24

References

1. Aleixo, S.M., Teles, J.: Weighting lower and upper ranks simultaneously through rank-order correlation coefficients. In: Gervasi, O., et al. (eds.) ICCSA 2018. LNCS, vol. 10961, pp. 318–334. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-95165-2_23
2. Aleixo, S.M., Teles, J.: Weighted coefficients to measure agreement among several sets of ranks emphasizing top and bottom ranks at the same time. In: Misra, S., et al. (eds.) ICCSA 2019. LNCS, vol. 11620, pp. 23–33. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24296-1_3
3. Coolen-Maturi, T.: New weighted rank correlation coefficients sensitive to agreement on top and bottom rankings. J. Appl. Stat. **43**(12), 2261–2279 (2016)
4. Friedman, M.: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc. **32**, 675–701 (1937)
5. Hájek, J., Šidák, Z.: Theory of Rank Tests. Academic Press, New York (1972)
6. Kendall, M.G., Babington-Smith, B.: The problem of m rankings. Ann. Math. Stat. **10**, 275–287 (1939)
7. Legendre, P.: Species associations: the Kendall coefficient of concordance revisited. J. Agric. Biol. Environ. Stat. **10**(2), 226–245 (2005)
8. R Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2019). https://www.r-project.org
9. Teles, J.: Concordance coefficients to measure the agreement among several sets of ranks. J. Appl. Stat. **39**, 1749–1764 (2012)
10. van der Waerden, B.L.: Order tests for the two-sample problem and their power. Proc. Koninklijke Nederlandse Akademie van Wetenschappen **55**, 453–458 (1952)
11. Wallis, W.A.: The correlation ratio for ranked data. J. Am. Statist. Assoc. **34**, 533–538 (1939)