Effective use of Spearman's and Kendall's correlation coefficients for association between continuously measured traits

Marie-Therese Puth[1], Markus Neuhäuser[1], Graeme D Ruxton[2]

1. Fachbereich Mathematik und Technik, RheinAhrCampus, Koblenz University of Applied Sciences, Joseph-Rovan-Allee 2, 53424 Remagen, Germany
2. School of Biology, University of St Andrews, St Andrews, Fife KY16 9TH, UK

Author for correspondence : GDR email: gr41@st-andrews.ac.uk; tel. +44 1334 464825 ; fax +44 1334 364825

Abstract

We examine the performance of the two rank order correlation coefficients (Spearman's rho and Kendall's tau) for describing the strength of association between two continuously measured traits. We begin by discussing when these measures should, and should not, be preferred over Pearson's product moment correlation coefficient on conceptual grounds. For testing the null hypothesis of no monotonic association, our simulation studies found both rank coefficients show similar performance to variants of the Pearson product-moment measure of association, and provide only slightly better performance than Pearson's measure even if the two measured traits are non-normally distributed. Where variants of the Pearson measure are not appropriate, there was no strong reason (based on our results) to select either of our rank-based alternatives over the other for testing the null hypothesis of no monotonic association. Further, our simulation studies indicated that for both rank coefficients there exists at least one method for calculating confidence intervals that supplies results close to the desired level if there are no tied values in the data. In this case, Kendall's coefficient produces consistently narrower confidence intervals, and might thus be preferred on that basis. However, as soon as there are any ties in the data, no matter whether this involves a small or larger percentage of ties, Spearman's measure returns values closer to the desired coverage rates; whereas Kendall's results differ more and more from the desired level as the number of ties increases, especially for large correlation values.

Keywords: confidence interval, null hypothesis testing, Pearson's product moment correlation coefficient, power, statistics, type 1 error

Highlights

- Kendall's and Spearman's coefficients measure monotonic (not linear) association.
- For testing the null hypothesis of no association both measures work well.
- Methods are highlighted for effective confidence interval construction for both.
- Ties in data do not affect hypothesis testing
- Ties in the data adversely affect construction of Kendall's confidence intervals.

Introduction

It is common in statistical analysis to want to explore and summarise the strength of association between two continuously measured traits on a number of experimental units. In a recent publication (Puth, Neuhäuser & Ruxton, 2014) we argued that Pearson's product-moment correlation coefficient (ρ) can often offer an effective description of linear association even when the traditional assumption that the underlying distribution being sampled is bivariate normal is violated. Specifically we demonstrated effective methods for calculating a confidence interval for ρ and for testing the null hypothesis that ρ is equal to any specified value. However, as classically defined, the Pearson's product-moment correlation coefficient is a parametric measure, and two nonparametric measures of association in common use are the Spearman rank order correlation coefficient r_S and Kendall's rank correlation coefficient τ. In 2013, 47 papers published in Animal Behaviour used Spearman's measure; 10 papers used Kendall's measure. Of these 57 papers only five discussed the motivation for selecting the measure used rather than Pearson's measure. Here we will discuss when such methods might be preferred over Pearson's product-moment correlation coefficient, and which of these alternatives performs best in different circumstances. We will do this both in the context of testing the null hypothesis of no association and of calculation of a confidence interval for the population value of these measures. First we briefly define the two measures.

Spearman rank order correlation coefficient (r_s)

The Spearman rank correlation coefficient is equivalent to Pearson's product-moment correlation coefficient performed on the ranks of the data rather than the raw data. Specifically, assume that we measure two traits *X* and *Y* on each of *n* subjects. Let x_i be the rank of the measurement of X taken on the *i*th individual; y_i being defined similarly. Identical values (ties) are assigned a rank equal to the average of their positions in the ascending order of the values. Then average ranks $\bar{x}$ and $\bar{y}$ are equal to $(n + 1)/2$ and

$$r_s = \frac{\sum_{i=1}^{n}\{(x_i - \bar{x})(y_i - \bar{y})\}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

Simpler formulations are possible for the case where there are no ties, but this method works in generality. This formulation will yield a value $-1 \leq r_s \leq 1$.The higher the absolute value of r_s the stronger the association between the two variables. Positive values suggest that higher values of one variable are associated with higher values of the other variable; whereas negative values suggest that higher values of one are associated with lower values of the other.

Kendall's rank correlation coefficient (τ)

If we compare two measurement units from our sample (indexed *i* and *j*), then any pair of observations (x_i, y_i) and (x_j, y_j) are said to be concordant if the ranks for both elements agree: i.e. if both ($x_i > x_j$ and $y_i > y_j$) or if both ($x_i < x_j$ and $y_i < y_j$). They are said to be discordant, if ($x_i > x_j$ and $y_i < y_j$) or if ($x_i < x_j$ and $y_i > y_j$). If ($x_i = x_j$ and/or $y_i = y_j$), the pair is neither concordant nor discordant.

For a sample of size *n* there are n_0 unique unordered pairs of observations where $n_0 = 0.5n(n-1)$. Let n_c be the number of these pairs that are concordant and n_d the number of discordant pairs. In the simple case where there are no tied ranks then τ is simply given by

$$\tau = \frac{n_c - n_d}{n_0}.$$

Where there are ties, a number of different formulations have been suggested, by far the most commonly used is termed τ_b. For the quantity *X*, there will be a number (*p*) of groups of unique ranks less than or equal to *n*. Let t_i be the number of tied values in the *i*th group, we then define n_1 as follows:

$$n_1 = 0.5 \sum_{i=1}^{p} t_i(t_i - 1),$$

Note that “tied groups” with $t_i = 1$ are possible.

Similarly, for the quantity *Y*, there will be a number (*q*) of groups of unique ranks less than or equal to *n*. Let u_j be the number of tied values in the *j*th group, we then define n_2 as follows:

$$n_2 = 0.5 \sum_{j=1}^{q} u_j(u_j - 1),$$

again, $u_j = 1$ is possible. Then

$$\tau_b = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}}.$$

This formulation will yield a value $-1 \leq \tau_b \leq 1$, and this measure as well as τ is interpreted in an analogous manner to Spearman's r_s. Specifically, the higher the absolute value of τ_b the stronger the association between the two variables. Positive values suggest that higher values of one variable are associated with higher values of the other variable; negative values suggest that higher values of one are associated with lower values of the other.

When might these measures be preferred over Pearson's product-moment correlation coefficient (*r*)

Pearson's and the two rank correlation coefficients defined above measure different types of association. Pearson's coefficent measures linear association only, whereas the other two measure a broader class of association: a high absolute value of Spearman or Kendall correlation coefficient indicates that there is a monotonic (but not necessarily linear) relationship between the two variables. Sometimes scientists may have good theoretical reason for testing this broader hypothesis. Further, the Pearson correlation coefficient was designed to work with variables measured on a continuous scale, if the variables are measured on an ordinal scale it cannot be applied; then Spearman's or Kendall's measure could be used instead. Finally, it may also sometimes be appropriate to use Spearman's or Kendall's measure over Pearson's if this gives easier comparison with a previous study that used that method.

It seems common practice in the literature to select between measures of the data on the basis of examination of the sampled data. Specifically, if the distributions of the samples of either or both of the variables deviates from normality then one of the rank measures is used, with Pearson being adopted otherwise (this approach was taken in all five 2013 *Animal Behaviour* papers mentioned above). However we have argued previously (Puth et al. 2014) that the robustness of approaches based on Pearson's measure makes this approach unnecessary. Investigators should be able to decide on whether to use Pearson's or a rank measure on the basis of the nature of the hypotheses they are interested in and how they intend to collect the data. Once the data is collected, there should be no need to switch from one measure to another on the basis of visual inspection or preliminary testing of the data. Given our discussion immediately above, researchers who have switched intended analysis on this basis should bear in mind that the Pearson and rank coefficients measure different types of association.

Testing the null hypothesis of no association

We explored the performance of these two alternative rank measures in a simulation study. Specifically we explored the performance of the two in terms of estimated type 1 error rate and power from samples of size n drawn from underlying distributions of specified marginal distributions of the two variables and association ρ between then. We used sample sizes of n

= 10, 20, 40 and 80, ρ values of 0.0, 0.1 and 0.5, and distributions of the two variables that were either normal, symmetric and heavy tailed, or asymmetric and heavy tailed. The details of the method used are provided in Appendix 1 and the results (based on 10,000 samples in each case) presented in tables 1-3. We utilized a nominal type I error rate (α) of 0.01; however recent work by Bishara & Hittner (2012) suggest that our conclusions should hold in essentially unchanged form for α = 0.05. For comparison purposes we also include the performance of two methods of implementing the Pearson product moment correlation coefficient that we have previously found to perform well. Specifically we recommended (Puth et al. 2014) the permutation test based on this measure when sample sizes are small (less than twenty) and using the RIN transformation prior to implementing the standard t-test procedure on this measure otherwise. *P*-values associated with Spearman's and Kendall's coefficients were calculated using the *cor.test* function of the *stats* package of *R*. For Spearman this is exact if n < 10 and an approximation to the exact *P*-value using the algorithm of Best & Roberts (1975) otherwise. For Kendall's coefficient, the *P*-value is exact providing n < 50 and if there were no ties, otherwise it is evaluated under the assumption that under the null hypothesis

$$\frac{3\tau\sqrt{n(n-1)}}{\sqrt{2(2n+5)}}$$

is normally distributed with mean zero and unit variance.

From evaluation of tables 1-3, both Spearman's r_s and Kendall's τ preserve type 1 error rates close to the nominal 1% values throughout all combinations of distributions. There is no consistent pattern as to which measure is superior in this regard. The RIN and permutation methods associated with Pearson's measure also provide good control of the type 1 error rate.

Concerning power, our two non-parametric measures are generally inferior to either the RIN and permutation methods associated with Pearson's measure, but generally not by a large margin. For all methods, power to detect low levels of association (ρ = 0.1) are not high for any method. Comparing Spearman's *r* and Kendall's τ, there is never a substantial difference in the power of the two measures. When sample sizes are as low as ten, all four methods offer relatively low power even for detecting relatively large levels of association (ρ = 0.5).

For brevity, we omit the details, but found that our qualitative conclusions above were unaffected if we rounded values to one or two decimal places prior to including them in our sample so as to create between 11 and 56% ties within samples (see Appendix 2).

Calculation of confidence intervals

As an alternative or complement to null-hypothesis statistical testing, we often want to present a confidence internal for the population value of the statistic under investigation. In this section, we will explore how this might be achieved for both both of our rank measures of association.

Spearman's r_s

The key to obtaining an effective confidence intervals for the Spearman correlation coefficient is a good estimation for its sample variance. Once this has been obtained, we can exploit the fact that Fisher's z-transformation for a sample correlation coefficient r is defined by

$$z = 0.5 \ln \left(\frac{1+r}{1-r} \right) = \tanh^{-1}(r)$$

and converts r into an approximately standard normally distributed value z. Applied to Spearman's r_s, the lower and upper limits of the (1-α) confidence interval for the transformed value are given by

$$\tilde{L} = 0.5 \ln \left(\frac{1+\hat{r}_{s,}}{1-\hat{r}_{s,}} \right) - z_{1-\alpha/2}\hat{\sigma} \quad \text{and} \quad \tilde{U} = 0.5 \ln \left(\frac{1+\hat{r}_{s,}}{1-\hat{r}_{s,}} \right) + z_{1-\alpha/2}\hat{\sigma}$$

where $\widehat{r_s}$ denotes the estimated Spearman correlation, $z_{1-\alpha/2}$ represents the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution, and $\hat{\sigma}$ describes the standard deviation. We can then obtain the lower and upper limits (*L* and *U*) of the confidence interval for the population value of r_s from the conversions below.

$$L = \frac{\exp(2\tilde{L})-1}{\exp(2\tilde{L})+1} \text{ And } U = \frac{\exp(2\tilde{U})-1}{\exp(2\tilde{U})+1}.$$

There is no universally-agreed method for obtaining the appropriate variance to use in these calculations. One estimate of the variance (denoted Method A) by Fieller, Hartley and Pearson (1957) is defined by

$$\hat{\sigma}_A{}^2 = \frac{1.06}{n-3},$$

where *n* denotes the sample size.

The next one (Method B) was proposed by Bonett & Wright (2000) and is defined by

$$\hat{\sigma}_B{}^2 = \frac{1+\frac{\widehat{r_S}^2}{2}}{n-3}.$$

Another commonly used method (Method C) is given by

$$\hat{\sigma}_C{}^2 = \frac{1}{n-2} + \frac{|\hat{\xi}|}{6n+4n^{1/2}},$$

where $\hat{\xi} = \tanh^{-1}(\widehat{r_s})$. This method was introduced by Caruso & Cliff (1997).

Here we will examine the relative performance of three alternatives. Additionally we examined two different bootstrap methods for producing a confidence interval: the BCa method (see Efron & Tibshirani 1993 and Manly 2007 for details of this methodology), and the bootstrap variance estimation method. The latter is based on an asymptotic normal $(1-\alpha)$ confidence interval of form $\widehat{r_s} \pm z_{(1-\alpha/2)} * \hat{\sigma}_{Boot}$, where $\widehat{r_s}$ is the Spearman correlation of the original data set and $\hat{\sigma}_{Boot}$ denotes the standard deviation of the bootstrap estimates of r_s.

A Monte Carlo simulation with 20,000 samples for Methods A, B & C and 1,000 samples using 1,000 resamples for the two bootstrap methods was performed for several values of *ρ* and *n*. The coverage probabilities for a 95% confidence interval using these five methods are summarized in table 4. That is, we calculated how often the 95% confidence interval calculated on the basis of a sample enclosed the specified underlying population value. Results are based on bivariate normal random variables, but will hold for any other monotonic transformation since the rank order correlation coefficient is invariant under monotonic transformations.

Examination of Table 4 suggests that all five methods generally offer reasonable estimation of the confidence interval. The bootstrap methods are never sufficiently superior to justify their much higher computational costs. Method B is the best performing method for very high levels of association ($\rho \geq 0.9$); but otherwise Method C is generally (but not always) the best performing method. Method C can perhaps be recommended, since it offers the most consistently good performance over all the scenarios we explored.

Kendall's tau

For Kendall's tau we examined four different methods to construct confidence intervals including the same two bootstrap methods as described above and two other variance estimation methods that could be used in the same Fisher-transformation approach as described previously for Spearman's measure.

The first variance estimation (Method A) by Fieller et al. (1957) is given by

$$\hat{\sigma}_A{}^2 = \frac{0.437}{n-4}\,.$$

This estimation is only accurate for values of |ρ|< 0.8, therefore we considered another variance estimation (method C) given in Xu, Hou, Hung & Zou (2013), which is defined by

$$\sigma_C{}^2 = \frac{2}{n(n-1)}\left[1 - \frac{4S_1{}^2}{\pi^2} + 2(n-2)(\frac{1}{9} - \frac{4S_2{}^2}{\pi^2})\right],$$

where $S_1 = \sin^{-1}\rho$, $S_2 = \sin^{-1}\frac{\rho}{2}$ and ρ denotes the correlation coefficient of the bivariate sample data and can be estimated using the relationship $\hat{\rho} = \sin(\frac{\pi}{2}\tau)$.

Again, a Monte Carlo simulation with 20,000 samples for the two variance estimation methods and 1,000 samples with 1.000 resamples were used for the bootstrap methods. The coverage probabilities for a 95%-confidence interval for Kendall’s tau are summarized in table 5.

All four methods approach the desired 95%-coverage rate well for values of $|\rho|$ < 0.8. As soon as ρ gets larger, only the variance estimation (C) introduced by Xu et al. (2013) provides nearly accurate values. Both bootstrap methods return values even higher than the desired 0.95, whereas the other variance estimation (A) tends to values less than 0.95, especially for small sample sizes. From this perspective, we can recommend the Fisher transformation approach combined with Method C for variance estimation as an effective way to calculate confidence intervals for Kendall’s tau.

Comparing the two measures

Comparing the performance of Spearman's rho and Kendall's tau in confidence interval construction for bivariate normal data without ties, our results indicate that both methods seem to have at least one variance-estimation method which provides nearly accurate results for all values of ρ. To make a clearer recommendation, we explored the average width of the 95%-confidence intervals to see if there are any consistent differences between the two methods. A small width is desirable as this indicates less variation and a more precise interval estimation. We calculated the difference between the upper and the lower limits and determined the mean of these differences in order to generate the average width values. We only considered the two variance estimations which performed best: meaning that we used variance estimation (B) of Bonett & Wright (2000) for Spearman's rho and variance estimation (C) of Xu et al (2013) for Kendall's tau. The average width values for these two methods are summarized in Table 6. Since with increasing sample sizes the estimation values become more precise, the widths for the large sample size of n=200 are smaller than for the small sample size of n=20 no matter whether we look at Spearman's or Kendall's measure. But comparing the two methods, it is obvious that Kendall's measure supplies smaller intervals for all different values of ρ and n. Based on these results for data without ties, Kendall's measure seems preferable.

Presence of ties

Finally, we explore how the performance of our confidence interval estimation methods change if the data contains ties. We generated bivariate normal random variables using the method described in Appendix 1 and then rounded these random samples to one decimal place for small sample sizes (n=20, 50) and to two decimal places for large sample sizes (n=100,200). This led to different percentages of ties depending on the sample size, on average we have about 22% ties for a sample size of 20, 42% ties for a sample size of 50, 12% ties for a sample size of 100 and 22% ties for a sample size of 200.

Our results for the coverage probability with respect to ties are summarized in tables 7 & 8.

Table 7 presents the performance of Spearman's rho. As with our simulations without ties, variance estimation B generally provides values closest to the nominal 0.95 for different combinations of ρ and n. This observation fits well to the fact that the variance estimation B is dependent on the estimated correlation and the observation that the average correlation of the original bivariate data set and the average correlation of the rounded bivariate data set are very similar. They often just show differences in the fourth or fifth decimal place.

By contrast, Kendall's τ_b, generally provides values less than 0.95 especially for large correlation values and a high percentage of ties (n=50). This can be due to higher differences between the original correlation and the correlation of the rounded bivariate data set. There are often differences in the second decimal place between the average original correlation and the average rounded correlation estimate, especially for large values of ρ. As soon as there are ties in the data, our analysis showed that Spearman's rho provides better coverage rates, especially for large correlation values and a high percentage of ties than Kendall's tau.

Appendix 3 shows that we draw essentially equivalent conclusions to those described above for two continuous variables if we restrict one of the variables to being ordinal with only five levels. However, previous work suggests that our conclusions do not hold if both variables are restricted to four or five levels. In this case Woods (2007) found that confidence intervals were more reliable for Kendall's than Spearman's measure. However, if a confidence interval for Spearman's measure is required for data involving such restricted variables, then Ruscio (2008) suggests that bootstrapping can produce reasonably accurate confidence intervals provided $n > 25$.

Conclusion

As an alternative to Pearson's product moment correlation coefficient, we examined the performance of the two rank order correlation coefficients: Spearman's rho and Kendall's tau.

Concerning hypothesis testing, both rank measures show similar results to variants of the Pearson product-moment measure of association and provide only slightly better values than Pearson if the two random samples are both non-normally distributed. Where variants of the Pearson measure are not appropriate, there is no strong reason (based on our results) to select either of our rank-based alternatives over the other for testing the null hypothesis of no monotonic association. Concerning confidence interval estimation, our analysis indicates that both of them provide at least one method concerning confidence intervals construction which supplies results close to the desired level if ties do not exist. Additionally we looked at the average width of the confidence intervals and found out that Kendall's intervals are narrower and therefore should be preferred. But as soon as there are any ties in the data, no matter whether this involves a small or larger percentage of ties, Spearman's method should be considered superior. Spearman’s measure returns values closer to the desired coverage rates whereas Kendall’s results differ more and more from the desired level as the number of ties increases, especially for large correlation values.

Acknowledgement

We thank two referees and the editor for helpful comments.

References

Best, D. J. & Roberts, D. E. (1975) Algorithm AS89: the upper tail probabilities of Spearman’s rho. *Applied Statistics*, 24, 377-379

Bishara, A. & Hittner, J. (2012) Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, Transformation and Resampling approaches. *Psychological Methods*, 17, 399-417.

Bonett, D. & Wright, T. (2000) Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika*, 65, 23-28.

Caruso, J. & Cliff, N. (1997) Empirical size, coverage, and power of confidence intervals for Spearman's rho. *Educational and Psychological Measurements*, 57, 637-654.

Fieller, E. C., Hartley, H. O. & Pearson E. S. (1957) Tests for rank correlation coefficients I. *Biometrika*, 44, 470-481.

Headrick, T. & Sawilowsky, S. (1999) Simulating correlated multivariate nonnormal distributions: extending the Fleishman power method. *Psychometrika*, 64, 25-35.

Puth, M. T., Neuhäuser, M. & Ruxton, G. D. (2014) Effective use of Pearson's product-moment correlation coefficient. *Animal Behaviour*, 93, 183-189.

Ruscio J. (2008) Constructing confidence intervals for Spearman's rank correlation with ordinal data: a simulation study comparing analytic and bootstrap methods. *Journal of Modern Applied Statistical Methods*, 7, 416-434.

Woods, C. M. (2007) Confidence intervals for gamma-family measures of ordinal association. *Psychological Methods*, 7, 416-434

Xu, W., Hou, Y., Hung, Y. S. & Zou Y. (2013) A comparative analysis of Spearman's rho and Kendall's tau in normal and contaminated normal models. *Signal Processing*, 93, 261-276.

Zopluoglu, C. (2011) Applications in R: Generating Multivariate Non-normal Variables. [pdf] http://www.tc.umn.edu/~zoplu001/resim/gennonnormal.pdf.

Appendix 1: Generation of bivariate random deviates

We used the method of Headrick & Sawilowsky (1999). First we obtain the Fleishman constants *a*, *b*, *c* and *d* for both variables (say *X* and *Y*) by solving the Fleishman's equations:

a = -*c*

$b^2 + 6bd + 2c^2 + 15d^2 - 1 = 0$

$2c(b^2 + 24bd + 105d^2 + 2) - \gamma_1 = 0$

$24[bd + c^2(1 + b^2 + 28bd) + d^2(12 + 48bd + 141c^2 + 225d^2)] - \gamma_2 = 0$

where γ_1 denotes the desired skewness and γ_2 is the desired excess kurtosis.

We them determine the intermediate correlation r^2 using

$r^2(b_1b_2 + 3b_2d_1 + 3b_1d_2 + 9d_1d_2 + 2a_1a_2r^2 + 6d_1d_2r^4) = \rho$

where ρ is the desired post-correlation and a_1, a_2, b_1, b_2, d_1 and d_2 are the calculated Fleishman constants of the two variables *X* and *Y*. With this intermediate correlation we were able to generate standard random normal deviates of the form $\tilde{X} = rZ_1 + \sqrt{1-r^2}\,E_1$ and $\tilde{Y} = rZ_1 + \sqrt{1-r^2}\,E_2$, where Z_1, E_1 and E_2 are normally distributed independent random variables with zero mean and unit variance. Finally we generate the desired nonnormal variables X^* and Y^* using the Fleishman transformation equation: $X^* = a_1 + b_1\tilde{X} + c_1\tilde{X}^2 + d_1\tilde{X}^3$.

Our code for generating bivariate nonnormal random samples is based on Zopluoglu's R-Script (2011). We adopted his method to obtain the Fleishman constants and his idea for a method to solve the equation to find the intermediate correlation. His function to obtain the Fleishman constants is based on a Newton-Iteration with a Jacobian matrix. The only thing we corrected was the first partial derivative of the third Fleishman-equation in his Jacobian matrix.

We obtained a normal distribution by using the parameters (γ_1 = 0, γ_2 = 0); heavy tailed but symmetric distribution (γ_1 = 0, γ_2 = 6), and heavy tailed and asymmetric distribution (γ_1 = 2, γ_2 = 6).

Appendix 2: Hypothesis testing for Spearman/Kendall with ties:

We created correlated data sets with ties by generated bivariate normal random variables using the method described in Appendix 1 and then rounded these random samples to one decimal place. This gave on average the following fractions of ties (11% for n = 10, 22% for n = 20, 38% for n = 40 and 56% for n = 80). In addition, for n = 80 we repeated our analysis this time rounding to the second decimal place, producing 10% ties on average. We then performed similar analyses to those used for the untied data described in the main text. For Kendall's measure we used the *cor.test* function in *R* used previously but also the *Kendall* function in the package *Kendall* which was designed to produce more accurate *P*-values than *cor.test* in the event of ties.

Comparing Table 1 with Table A1 below, we see no strong evidence of introduction of ties leading to loss of control of type 1 error rates for all the measures considered. Comparing Table 3 with table A2, we see a similar lack of strong effect on power.

Tables A1 then A2 here

Appendix 3: Results for 95%- and 99%- CI for Spearman's correlation coefficient when one variable is restricted to only five possible values

We generated samples of one variable *x* with length n by randomly sampling from the values 1,..,5 with replacement using the sample-function in *R*. To recreate a correlated variable *y*, we used the *corgen*-function of the R package *ecodist*. This function generates a correlated variable *y* within the range of a given epsilon to a given (Pearson) correlation to *x*. In our code, we do not specify an epsilon to ensure some variation in the correlation between samples. We then explored the coverage of 95% and 99% confidence intervals calculated in exactly the same way as in the main paper: see tables A3 & A4.

For 95% confidence intervals, for small correlation values (0.1 and 0.3) all methods perform well. For larger correlation values the coverage probability is higher than desired, especially for small samples sizes and medium correlation values, and for large correlation values for all sample sizes. Bootstrap methods provide values less than the desired 0.95 for high correlation values and large sample sizes (n=200, correlation: 0.8, 0.9, 0.95) For 99% confidence intervals, for small and medium correlation values (0.1 up to 0.7) all methods perform well with results. For large correlations the results are still generally satisfactory but they are higher than desired.

Tables A3 and A4 here

TABLES

Table 1: Type-I Error Rate (α=0.01) for different sample sizes (n) and combinations of distribution shapes, evaluated for the Spearman and Kendall measures as well as Permutation and RIN (rank-based inverse normal) transform implementation of Pearson's measure.

distribution	n	Spearman	RIN	Permutation	Kendall
normal & normal	10	0.0098	0.0115	0.0105	0.0090
	20	0.0122	0.0113	0.0106	0.0101
	40	0.0093	0.0106	0.0101	0.0078
	80	0.0106	0.0113	0.0104	0.0096
normal & heavy-tailed	10	0.0098	0.0115	0.0105	0.0090
	20	0.0122	0.0113	0.0109	0.0101
	40	0.0093	0.0106	0.0097	0.0078
	80	0.0106	0.0113	0.0106	0.0096
normal & asymmetric-heavy-tailed	10	0.0100	0.0105	0.0091	0.0089
	20	0.0097	0.0109	0.0109	0.0097
	40	0.0099	0.0104	0.0089	0.0083
	80	0.0111	0.0091	0.0085	0.0096
heavy-tailed & heavy-tailed	10	0.0096	0.0117	0.0090	0.0090
	20	0.0108	0.0119	0.0119	0.0101
	40	0.0092	0.0098	0.0105	0.0079
	80	0.0094	0.0097	0.0101	0.0096
asymmetric-heavy-tailed & asymmetric-heavy-tailed	10	0.0096	0.0111	0.0096	0.0091
	20	0.0095	0.0104	0.0095	0.0093
	40	0.0093	0.0098	0.0110	0.0081
	80	0.0098	0.0099	0.0103	0.0095

Table 2: Power with small effect size (ρ=0.1) for different sample sizes (n) and combinations of distribution shapes, evaluated for the Spearman and Kendall measures as well as Permutation and RIN (rank-based inverse normal) transform implementation of Pearson's measure.

distribution	n	Spearman	RIN	Permutation	Kendall
normal & normal	10	0.0127	0.0153	0.0113	0.0131
	20	0.0161	0.0164	0.0187	0.0136
	40	0.0244	0.0247	0.0251	0.0224
	80	0.0415	0.0459	0.0476	0.0376
normal & heavy-tailed	10	0.0126	0.0152	0.0120	0.0132
	20	0.0162	0.0157	0.0190	0.0139
	40	0.0247	0.0254	0.0271	0.0227
	80	0.0450	0.0481	0.0475	0.0391
normal & asymmetric-heavy-tailed	10	0.0123	0.0151	0.0120	0.0141
	20	0.0173	0.0175	0.0192	0.0148
	40	0.0267	0.0272	0.0257	0.0244
	80	0.0495	0.0529	0.0465	0.0455
heavy-tailed & heavy-tailed	10	0.0129	0.0157	0.0123	0.0133
	20	0.0164	0.0184	0.0179	0.0136
	40	0.0261	0.0260	0.0257	0.0232
	80	0.0475	0.0517	0.0437	0.0411
asymmetric-heavy-tailed & asymmetric-heavy-tailed	10	0.0125	0.0146	0.0115	0.0135
	20	0.0179	0.0192	0.0177	0.0167
	40	0.0287	0.0292	0.0252	0.0274
	80	0.0565	0.0573	0.0395	0.0522

Table 3: Power with large effect size (ρ=0.5) for different sample sizes (n) and combinations of distribution shapes, evaluated for the Spearman and Kendall measures as well as Permutation and RIN (rank-based inverse normal) transform implementation of Pearson's measure.

distribution	n	Spearman	RIN	Permutation	Kendall
normal & normal	10	0.0844	0.1027	0.1204	0.0888
	20	0.3079	0.3326	0.3780	0.2985
	40	0.7159	0.7565	0.7798	0.7089
	80	0.9741	0.9854	0.9854	0.9744
normal & heavy-tailed	10	0.0931	0.1099	0.1314	0.0976
	20	0.3385	0.3654	0.3983	0.3281
	40	0.7576	0.7936	0.7960	0.7512
	80	0.9835	0.9910	0.9878	0.9844
normal & asymmetric-heavy-tailed	10	0.1097	0.1219	0.1418	0.1128
	20	0.4058	0.4270	0.4138	0.3932
	40	0.8335	0.8551	0.8101	0.8259
	80	0.9944	0.9963	0.9914	0.9939
heavy-tailed & heavy-tailed	10	0.0970	0.1133	0.1302	0.1016
	20	0.3502	0.3790	0.3768	0.3435
	40	0.7732	0.8111	0.7573	0.7681
	80	0.9874	0.9910	0.9817	0.9877
asymmetric-heavy-tailed & asymmetric-heavy-tailed	10	0.1005	0.1106	0.1101	0.1009
	20	0.3717	0.3896	0.2880	0.3581
	40	0.8023	0.8148	0.6233	0.7959
	80	0.9916	0.9928	0.9398	0.9912

Table 4: containment probability values for a 95% confidence interval for Spearman's correlation coefficient using three different variance estimation methods (A, B & C) defined in the text in combination with Fisher's z-transformation as well bootstrap variance estimation and the BCa bootstrapping method

ρ	n	A	B	C	Boot	Bca
0.1	20	0.95480	0.95845	0.94850	0.929	0.955
	50	0.95640	0.95515	0.95250	0.943	0.962
	100	0.95660	0.95265	0.95230	0.941	0.956
	200	0.95590	0.95195	0.95220	0.952	0.950
0.3	20	0.95420	0.95965	0.94820	0.947	0.957
	50	0.95030	0.95135	0.94660	0.955	0.946
	100	0.95195	0.95200	0.94920	0.952	0.941
	200	0.95460	0.95395	0.95305	0.947	0.954
0.5	20	0.94780	0.95630	0.94225	0.928	0.948
	50	0.94565	0.95375	0.94630	0.941	0.943
	100	0.95345	0.95905	0.95420	0.944	0.941
	200	0.95115	0.95805	0.95360	0.940	0.950
0.7	20	0.94135	0.95395	0.94005	0.951	0.946
	50	0.94140	0.95855	0.94630	0.950	0.944
	100	0.93995	0.95745	0.94745	0.955	0.944
	200	0.94090	0.96000	0.94885	0.947	0.943
0.8	20	0.94040	0.95810	0.94120	0.966	0.955
	50	0.93260	0.95550	0.94055	0.949	0.947
	100	0.93570	0.95895	0.94585	0.944	0.953
	200	0.93025	0.95645	0.94265	0.954	0.956
0.9	20	0.92830	0.95475	0.93170	0.986	0.962
	50	0.92505	0.95660	0.93925	0.967	0.941
	100	0.92125	0.95640	0.94020	0.961	0.938
	200	0.92040	0.95590	0.94150	0.945	0.958
0.95	20	0.89800	0.94100	0.90965	0.993	0.967
	50	0.90260	0.94525	0.92625	0.989	0.944
	100	0.90990	0.95030	0.93470	0.978	0.938
	200	0.90905	0.95155	0.93630	0.966	0.941

Table 5: containment probability values for a 95% confidence interval for Kendall's correlation coefficient using different variance estimation methods using two different variance estimation methods (A & C) defined in the text in combination with Fisher's z-transformation as well bootstrap variance estimation and the BCa bootstrapping method

ρ	n	A	Boot	Bca	C
0.1	20	0.94795	0.948	0.968	0.95225
	50	0.94925	0.955	0.951	0.94965
	100	0.94845	0.946	0.951	0.95040
	200	0.94970	0.945	0.950	0.94795
0.3	20	0.95195	0.951	0.967	0.95535
	50	0.94970	0.954	0.957	0.94765
	100	0.95090	0.950	0.955	0.94970
	200	0.95020	0.946	0.944	0.95085
0.5	20	0.94205	0.935	0.957	0.95330
	50	0.94960	0.948	0.962	0.95075
	100	0.95295	0.961	0.960	0.95090
	200	0.95330	0.951	0.952	0.95120
0.7	20	0.93325	0.963	0.974	0.95295
	50	0.95465	0.960	0.966	0.95025
	100	0.95850	0.958	0.958	0.95250
	200	0.96055	0.954	0.961	0.94980
0.8	20	0.92335	0.971	0.988	0.94985
	50	0.95005	0.966	0.972	0.95325
	100	0.96015	0.959	0.961	0.94795
	200	0.96170	0.966	0.954	0.95000
0.9	20	0.84055	0.963	0.995	0.94630
	50	0.92870	0.981	0.984	0.95100
	100	0.95295	0.983	0.982	0.95395
	200	0.96475	0.966	0.965	0.95070
0.95	20	0.75995	0.949	0.999	0.95920
	50	0.87450	0.977	0.992	0.94960
	100	0.93350	0.987	0.988	0.94795
	200	0.95630	0.974	0.980	0.95025

Table 6: average width of 95%-confidence intervals using Spearman's measure and variance estimate B and Kendall's measure and variance estimate C

ρ	n	Spearman	Kendall
0.1	20	0.8509	0.6295
	50	0.5449	0.3784
	100	0.3870	0.2629
	200	0.2744	0.1843
0.3	20	0.8089	0.5789
	50	0.5143	0.3459
	100	0.3639	0.2398
	200	0.2577	0.1679
0.5	20	0.7216	0.4796
	50	0.4474	0.2819
	100	0.3146	0.1942
	200	0.2215	0.1355
0.7	20	0.5646	0.3364
	50	0.3336	0.1901
	100	0.2296	0.1289
	200	0.1599	0.0891
0.8	20	0.4455	0.2507
	50	0.2515	0.1359
	100	0.1700	0.0903
	200	0.1174	0.0618
0.9	20	0.2849	0.1551
	50	0.1462	0.0773
	100	0.0956	0.0491
	200	0.0651	0.0326
0.95	20	0.1769	0.1004
	50	0.0826	0.0462
	100	0.0520	0.0276
	200	0.0346	0.0175

Table 7: containment probability values for a 95% confidence interval for Spearman's correlation coefficient using data with ties generated as previously then rounded to one decimal place (for n = 20, 50) or two decimal places (for n = 100, 200).

ρ	n	A	B	C	Boot	Bca
0.1	20	0.95265	0.95670	0.94615	0.917	0.971
	50	0.95340	0.95180	0.94920	0.936	0.950
	100	0.95645	0.95260	0.95205	0.959	0.950
	200	0.95570	0.95115	0.95160	0.950	0.958
0.3	20	0.95340	0.95995	0.94800	0.917	0.947
	50	0.95160	0.95330	0.94790	0.936	0.966
	100	0.95295	0.95190	0.94950	0.947	0.950
	200	0.95375	0.95290	0.95200	0.946	0.942
0.5	20	0.95015	0.95810	0.94560	0.920	0.949
	50	0.94650	0.95330	0.94670	0.936	0.952
	100	0.94905	0.95465	0.95035	0.954	0.940
	200	0.94705	0.95420	0.94960	0.947	0.953
0.7	20	0.94255	0.95550	0.94115	0.939	0.946
	50	0.94195	0.95775	0.94640	0.938	0.952
	100	0.94305	0.96000	0.94920	0.949	0.939
	200	0.94025	0.95795	0.94855	0.940	0.953
0.8	20	0.94110	0.96005	0.94260	0.954	0.968
	50	0.93230	0.95450	0.94060	0.934	0.953
	100	0.93205	0.95565	0.94195	0.940	0.941
	200	0.92980	0.95665	0.94340	0.950	0.955
0.9	20	0.92555	0.95440	0.93255	0.978	0.975
	50	0.91940	0.95325	0.93480	0.969	0.943
	100	0.92445	0.95875	0.94215	0.965	0.942
	200	0.92015	0.95705	0.94120	0.961	0.946
0.95	20	0.90145	0.94035	0.91350	0.991	0.964
	50	0.90350	0.94530	0.92525	0.978	0.947
	100	0.90740	0.94705	0.93145	0.966	0.950
	200	0.91140	0.95340	0.93895	0.958	0.936

Table 8: containment probability values for a 95% confidence interval for Kendall's correlation coefficient using data with ties generated as previously then rounded to one decimal place (for n = 20, 50) or two decimal places (for n = 100, 200).

ρ	n	A	Boot	Bca	C
0.1	20	0.94115	0.931	0.963	0.94665
	50	0.94120	0.951	0.962	0.94205
	100	0.94635	0.952	0.960	0.94870
	200	0.9468	0.959	0.957	0.94775
0.3	20	0.94075	0.933	0.965	0.94575
	50	0.94150	0.942	0.956	0.94520
	100	0.94845	0.954	0.956	0.95220
	200	0.94780	0.954	0.951	0.95000
0.5	20	0.93440	0.946	0.976	0.95110
	50	0.94130	0.943	0.958	0.94725
	100	0.95475	0.955	0.961	0.95140
	200	0.95205	0.950	0.952	0.95145
0.7	20	0.91800	0.935	0.982	0.95545
	50	0.92315	0.914	0.958	0.94045
	100	0.95865	0.941	0.946	0.95040
	200	0.95930	0.951	0.961	0.95130
0.8	20	0.88330	0.934	0.977	0.96325
	50	0.87965	0.900	0.952	0.92925
	100	0.95820	0.956	0.968	0.95220
	200	0.96250	0.948	0.956	0.95085
0.9	20	0.78065	0.887	0.984	0.97555
	50	0.72375	0.812	0.931	0.87060
	100	0.94530	0.971	0.982	0.95145
	200	0.9498	0.944	0.964	0.94430
0.95	20	0.64555	0.782	0.871	0.99545
	50	0.49360	0.679	0.895	0.81675
	100	0.90145	0.963	0.981	0.94670
	200	0.90855	0.934	0.963	0.92320

Table A1: The same approach as table 1 except that we rounded these random samples to one decimal place; In addition, for n = 80 we repeated our analysis this time rounding to the second decimal place (shown in the last line).

distribution	N	Spearman	RIN	Permutation	Kendall(cor.test)	Kendall (Kendall)
normal & normal	10	0.0104	0.0098	0.0092	0.0065	0.0048
	20	0.0108	0.0108	0.0113	0.0089	0.0077
	40	0.0094	0.0102	0.0089	0.0082	0.0080
	80	0.0111	0.0104	0.0101	0.0107	0.0106
	80	0.0112	0.0100	0.0100	0.0110	0.0108

Table A2: The same approach as table 3 except that we rounded these random samples to one decimal place; In addition, for n = 80 we repeated our analysis this time rounding to the second decimal place (shown in the last line).

distribution	N	Spearman	RIN	Permutation	Kendall(cor.test)	Kendall (Kendall)
normal & normal	10	0.1084	0.1033	0.1197	0.0768	0.0601
	20	0.3179	0.3336	0.3764	0.2906	0.2776
	40	0.7168	0.7544	0.7786	0.7038	0.6990
	80	0.9743	0.9821	0.9845	0.9738	0.9735
	80	0.9745	0.9825	0.9852	0.9735	0.9732

Table A3: containment probability for 95% confidence interval for Spearman's coefficient using the same approach as table 4 except that one variable is restricted to taking only five values.

ρ	n	A	B	C	Boot	Bca
0.1	20	0.9539	0.9585	0.9491	0.9210	0.9610
	50	0.9570	0.9549	0.9527	0.9460	0.9650
	100	0.9562	0.9525	0.9520	0.9470	0.9450
	200	0.9552	0.9505	0.9507	0.9450	0.9520
0.3	20	0.9581	0.9639	0.9528	0.9380	0.9620
	50	0.9608	0.9613	0.9570	0.9550	0.9650
	100	0.9595	0.9595	0.9573	0.9580	0.9500
	200	0.9597	0.9591	0.9582	0.9480	0.9600
0.5	20	0.9701	0.9776	0.9678	0.9390	0.9630
	50	0.9699	0.9765	0.9706	0.9540	0.9720
	100	0.9670	0.9741	0.9689	0.9420	0.9720
	200	0.9579	0.9668	0.9611	0.9490	0.9540
0.7	20	0.9838	0.9921	0.9852	0.9560	0.9730
	50	0.9803	0.9904	0.9842	0.9370	0.9750
	100	0.9703	0.9854	0.9769	0.9330	0.9590
	200	0.9471	0.9695	0.9577	0.9130	0.9490
0.8	20	0.9883	0.9950	0.9906	0.9760	0.9800
	50	0.9852	0.9939	0.9894	0.9550	0.9710
	100	0.9758	0.9901	0.9833	0.9310	0.9590
	200	0.9356	0.9725	0.9547	0.8690	0.9030
0.9	20	0.9933	0.9979	0.9949	0.9780	0.9830
	50	0.9941	0.9986	0.9968	0.9770	0.9840
	100	0.9848	0.9957	0.9925	0.9370	0.9490
	200	0.9447	0.9837	0.9715	0.8380	0.8640
0.95	20	0.9961	0.9988	0.9968	0.9920	0.9660
	50	0.9979	0.9997	0.9993	0.9840	0.9760
	100	0.9970	0.9996	0.9993	0.9620	0.9630
	200	0.9832	0.9981	0.9952	0.8830	0.8890

Table A4: As table A3 but for a 99% confidence interval

ρ	n	A	B	C	Boot	Bca
0.1	20	0.9887	0.9919	0.9874	0.9650	0.9900
	50	0.9906	0.9912	0.9898	0.9830	0.9910
	100	0.9913	0.9901	0.9903	0.9840	0.9930
	200	0.9926	0.9918	0.9919	0.9860	0.9910
0.3	20	0.9909	0.9939	0.9896	0.9820	0.9900
	50	0.9933	0.9941	0.9929	0.9820	0.9940
	100	0.9924	0.9928	0.9919	0.9850	0.9940
	200	0.9933	0.9929	0.9928	0.9940	0.9900
0.5	20	0.9942	0.9968	0.9940	0.9760	0.9960
	50	0.9952	0.9976	0.9959	0.9820	0.9960
	100	0.9946	0.9966	0.9953	0.9880	0.9960
	200	0.9938	0.9965	0.9949	0.9860	0.9920
0.7	20	0.9974	0.9992	0.9977	0.9840	0.9920
	50	0.9968	0.9993	0.9982	0.9860	0.9950
	100	0.9953	0.9988	0.9969	0.9820	0.9930
	200	0.9912	0.9969	0.9949	0.9710	0.9870
0.8	20	0.9981	0.9994	0.9988	0.9900	0.9990
	50	0.9986	0.9995	0.9991	0.9830	0.9990
	100	0.9976	0.9998	0.9987	0.9810	0.9950
	200	0.9921	0.9983	0.9965	0.9690	0.9830
0.9	20	0.9989	0.9998	0.9993	0.9980	0.9970
	50	0.9995	0.9999	0.9999	0.9960	0.9960
	100	0.9992	0.9999	0.9998	0.9820	0.9960
	200	0.9966	0.9996	0.9989	0.9560	0.9820
0.95	20	0.9994	0.9998	0.9994	0.9990	0.9880
	50	0.9999	0.9999	0.9999	0.9970	0.9980
	100	0.9999	0.9999	0.9999	0.9960	0.9950
	200	0.9994	0.9999	0.9998	0.9790	0.9840