Sample Size Charts for Spearman and Kendall Coefficients

Justine O. May1 and Stephen W. Looney2*

1Augusta University Health IT, 1120 15th Street, AE-1014 Augusta, GA 30912-4900, USA
2Department of Population Health Sciences, Medical College of Georgia at Augusta University, 1120 15th Street, AE-1014 Augusta, GA 30912-4900, USA

Abstract

Bivariate correlation analysis is one of the most commonly used statistical methods. Unfortunately, it is generally the case that little or no attention is given to sample size determination when planning a study in which correlation analysis will be used. For example, our review of clinical research journals indicated that none of the 111 articles published in 2014 that presented correlation results provided a justification for the sample size used in the correlation analysis. There are a number of easily accessible tools that can be used to determine the required sample size for inference based on a Pearson correlation coefficient; however, we were unable to locate any widely available tools that can be used for sample size calculations for a Spearman correlation coefficient or a Kendall coefficient of concordance. In this article, we provide formulas and charts that can be used to determine the required sample size for inference based on either of these coefficients. Additional sample size charts are provided in the Supplementary Materials.

Keywords: Confidence interval • Fisher z-transform • Hypothesis testing • Power • Significance level

Introduction

Bivariate correlation analysis is one of the most commonly used statistical methods. There are a number of widely available tools that can be used for sample size determination when the planned analysis will be based on the Pearson product moment correlation coefficient (PCC). These include, for example, tables [1], formulas [2], software packages (R, PASS, nQuery, GPower), and internet-based tools (http://www.sample-size.net/correlation-sample-size/). For example, in R, the pwr package can be used to perform power calculations in many inferential situations. In particular, pwr.t.test can be used to perform power calculations for the PCC. However, the pwr package does not have the capability to perform power calculations for either the Spearman rank correlation coefficient (SCC) or the Kendall coefficient of concordance (KCC). In fact, as best we can determine, there are no widely available tools for sample size calculation when the planned analysis will be based on either the SCC or the KCC.

Even though the PCC is the most commonly used measure of association, the SCC and KCC are also quite popular. For example, Brough et al. [3] used Spearman correlation to examine the associations between peanut protein levels found in various household environments, including dust, surfaces, bedding, furnishings and air. Heist et al. [4] used Kendall’s coefficient in their investigation of the use of bevacizumab as a chemotherapeutic agent for the treatment of advanced non-small cell lung cancer. The KCC is also commonly used in the analysis of environmental data. For example, Helsel [5] used the KCC to determine if the concentrations of dissolved iron collected from the Brazos River, Texas, during summers from 1977-1985 followed a trend.

In this article, we provide charts that can be used for sample size determination for hypothesis tests and confidence intervals for Spearman or Kendall coefficients. In Section 2, we describe our methods and present results for the Spearman and Kendall coefficients; in Section 3, we provide some notes on the use of the sample size charts; and, in Section 4, we discuss our results.

Methods for Sample Size Determination

We first describe methods that are commonly used for sample size determination for the Pearson coefficient. Even though we do not present any tools for finding the sample size for the PCC in this article (primarily because they are already so widely available), the methods we describe can be used to perform power calculations for the Spearman and Kendall coefficients are based on modifications of methods commonly used for sample size determination for the PCC.

Traditional approach for Pearson correlation

Let \( \rho \) denote the population value of the PCC, and let \( H_0 \) denote the null hypothesis and \( H_1 \) the alternative hypothesis. Assuming that the two variables being correlated have a bivariate normal distribution, an exact test of the hypotheses.

\[
H_0: \, \rho = 0 \quad \text{vs.} \quad H_1: \, \rho \neq 0 \tag{1}
\]

Can be based on the test statistic,

\[
t = \frac{r \sqrt{n - 2}}{\sqrt{1 - r^2}} \tag{2}
\]

Where \( r \) denotes the sample value of the PCC and \( n \) denotes the sample size. Under the null hypothesis, the test statistic \( t \) in eqn. (2) follows a \( t \) distribution with \( n-2 \) degrees of freedom.

The usual approach for the PCC is to determine the sample size needed to guarantee that a level \( \alpha \) test of the hypotheses in eqn. (1) based on the test statistic in eqn. (2) will achieve power of at least \( \varphi \) to detect a pre-specified alternative value \( \rho_1 \). The most commonly used value for \( \alpha \) is 0.05 and commonly used values for \( \varphi \) are 0.80 and 0.90. One may repeat the sample size calculation using several different values of \( \rho_1 \) and then construct a power curve by plotting the required sample size \( n \) vs. \( \rho_1 \).

For example, suppose that one wishes to determine the sample size needed to detect an alternative value of 0.40 in a test of eqn. (1) with 80% power using a 0.05 level test. One can use table with number 3.4.1 in Cohen [1, p. 102] or one of the other tools mentioned previously to determine the required sample size. Cohen’s table indicates that \( n = 46 \) will yield 80% power to detect an alternative value of 0.40 when performing a test of the hypotheses in eqn. (1) with \( \alpha = 0.05 \). A sample size of \( n=18 \) will be required to detect an alternative value of 0.60 under the same testing conditions.

Fisher z-transform

Sample size determination based on the test of \( H_0: \, \rho = \rho_0 \quad \text{vs.} \quad H_1: \, \rho \neq \rho_0 \)
where \( \rho_o \) using the test statistic in eqn. (2) can be performed only if the null value is \( \rho_o=0 \). If one wishes to test a non-zero null value, the most common approach is to use a test statistic based on the Fisher z-transform of \( r \) [6].

For any \(-1 < r < 1\), the Fisher z-transform of \( r \) is given by:

\[
z(r) = \text{tanh}^{-1} r = \frac{\ln(1+r/1-r)}{2},
\]

which is asymptotically distributed as \( N\left( \text{tanh}^{-1} \rho, \sigma_z^2 \right) \), where \( \rho \) denotes the population value of the PCC and \( \sigma_z^2 \) denotes the asymptotic variance of \( z(r) \). The same transformation can be applied to the sample value of a Spearman or Kendall coefficient, yielding an approximately normally distributed transformed coefficient. For the PCC, \( \sigma_z^2 = 1/(n-3) \) [6] and, for the KCC, \( \sigma_z^2 = 0.4377/(n-4) \) [7]. For Spearman’s coefficient, the value of \( \sigma_z^2 \) depends on \( \rho \), the population value of the SCC: for \( |\rho| < 0.95 \), \( \sigma_z^2 = (1 + \rho^2 + \rho^4)/2/(n-3) \) [2]; for \( |\rho| \geq 0.95 \), \( \sigma_z^2 = 1.06/(n-3) \) [7].

### Sample size based on hypothesis testing

The asymptotic distributional result for \( z(r) \) can be used to derive a test statistic for testing \( H_0: \rho = \rho_o \):

\[
z_{\rho_0} = \frac{z(r) - z(\rho_o)}{\sqrt{1/(n-3)}},
\]

where \( n \) is the sample size, \( z(r) \) is the Fisher z-transform applied to the sample value of the PCC, and \( z(\rho_o) \) is the Fisher z-transform applied to the hypothesized value of the PCC. An approximate \( p \)-value is then obtained by calculating the appropriate tail probability for \( z_{\rho_0} \) using the standard normal distribution. For example, suppose one wishes to test \( H_0: \rho_o=0.2 \) vs. \( H_1: \rho_o \neq 0.2 \), and that a sample of size \( n=30 \) yields \( r=0.52 \). Then:

\[
z_{\rho_0} = \frac{z(0.52) - z(0.20)}{\sqrt{1/(30-3)}} = \frac{0.576 - 0.203}{\sqrt{1/27}} = 1.94.
\]

With an appropriate two-tailed \( p \)-value of \( 2(1 - \Phi(1.94)) = 0.0522 \), where \( \Phi(*) \) denotes the distribution function of the standard normal.

More generally, let \( \xi \) denote the population value of the measure of association to be tested (either the PCC, SCC, or KCC). Then a test statistic based on the Fisher z-transform is given by:

\[
z_{\xi} = \frac{z(r) - z(\xi)}{\sqrt{1/(n-b)}},
\]

where \( \xi \) denotes the sample estimate, \( \xi \) denotes the hypothesized value, \( n \) denotes the sample size, \( z(*) \) denotes the Fisher z-transform, and \( b \) and \( c^2 \) are obtained from Table 1.

An approximate \( p \)-value is then obtained by calculating the appropriate tail probability for \( z_{\xi} \) using the standard normal distribution.

Suppose we wish to use the test statistic in eqn. (5) to test \( H_0: \xi = \xi_o \):

\[
\xi_o = \xi_o
\]

Where \(-1 < \xi_o < 1\) is the pre-specified null value of the desired measure of association (either the PCC, SCC, or KCC). Let \( \xi_o \) denote the alternative value of the measure of association that we wish to detect with our planned hypothesis test. For a two-tailed test of the null hypothesis in eqn. (6), the required sample size for detecting the value \( \xi_o \) with power \( \alpha \) using a test with significance level \( \alpha \) is given by:

\[
n = b + c^2 \left( \frac{(z_{\xi_o} + z_{\xi_o})}{2} \right)^2,
\]

where \( z_{\xi_o} = \text{upper } \gamma \)-percent point of the standard normal distribution, \( z(*) \) denotes the Fisher z-transform, and \( b \) and \( c^2 \) are obtained from Table 1.

For a one-tailed test of eqn. (6), replace \( z_{\xi_o} \) in eqn. (7) by \( z_{\xi_o} \). If the formula in eqn. (7) does not yield an integer value, the next larger integer is used (Table 1).

A chart based on eqn. (7) for finding the sample size required to achieve 80% power for a two-tailed test of Spearman’s coefficient at significance level 0.05 is given in Figure 1; charts for upper-tailed and lower-tailed tests of Spearman’s coefficient are provided in Figures S1 and S2 in the Supplementary Materials. A chart for finding the sample size required to achieve 80% power for a two-tailed test of Kendall’s coefficient at significance level 0.05 is given in Figure 2; a chart for a one-tailed test of Kendall’s coefficient is provided in Figure S3 in the Supplementary Materials.

Note that the sample size chart for the two-tailed test for the SCC (Figure 1) is formatted differently than that for the KCC (Figure 2). Because the value of \( c^2 \) given in Table 1 for the SCC depends on the null value \( \rho_o \) if \( \rho_o < 0.95 \), the values of \( \xi_o \) and \( \xi_o \) do not enter the formula in eqn. (7) symmetrically, as they do for the KCC. Therefore, to simplify the sample size curves for the two-tailed test of the SCC, we split the chart into two pieces: one for alternative values of \( \rho_o \) greater than the null value \( \rho_o \) (Figure 1a) and one for alternative values of \( \rho_o \) less than the null value \( \rho_o \) (Figure 1b).

To use the sample size chart for the SCC, first locate the null value of the SCC along the \( x \)-axis in either Figure 1a (\( \rho_o < \rho_o \)) or Figure 1b (\( \rho_o > \rho_o \)). Then, draw a vertical line that intersects with the sample size curve corresponding to \( \rho_o \) and the desired alternative value in the hypothesis test. Finally, draw a horizontal line from the curve to the \( y \)-axis. The point of intersection is the required sample size. To use the sample size chart for the KCC, first locate the smaller of the null value (\( r_o \)) and the alternative value (\( r_o \)) of the KCC along the \( x \)-axis. Then, draw a vertical line that intersects with the sample size curve corresponding to the larger of \( r_o \) and \( r_o \). Finally, draw a horizontal line from the curve to the \( y \)-axis. The point of intersection is the required sample size.

Suppose we are planning a study in which the SCC will be used as the measure of association and that we will test \( H_0: \rho_o=0.4 \) vs. \( H_1: \rho_o \neq 0.4 \) once we have obtained the data. With a sample of \( n=129 \), we could detect an alternative value of \( \rho_o \) as small as 0.6 with 80% power using a two-tailed test with significance level 0.05 (Figure 1a). With a sample of \( n=180 \), we could detect an alternative value of \( \rho_o \) as large as 0.2 with 80% power using a two-tailed test with significance level 0.05 (Figures 1b and 2).

### Table 1. Constants Needed to Apply Fisher z-Transform to Measures of Association.

<table>
<thead>
<tr>
<th>Measure of Association</th>
<th>( b )</th>
<th>( c^2 )</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>3</td>
<td>1</td>
<td>[6]</td>
</tr>
<tr>
<td>Spearman</td>
<td>3</td>
<td>1 + (( \rho_o^2 )/2)</td>
<td>[2]</td>
</tr>
<tr>
<td>Spearman</td>
<td>3</td>
<td>1.06</td>
<td>[7]</td>
</tr>
<tr>
<td>Kendall</td>
<td>4</td>
<td>0.437</td>
<td>[7]</td>
</tr>
</tbody>
</table>

*For sample size determination based on a confidence interval, \( \rho_o \) is a "planning value," where \( \rho_o \) denotes the population value of the Spearman rank correlation coefficient. For sample size determination based on a hypothesis test, \( \rho_o \) is the null value \( \rho_o \). The value \( c^2 = 1 + (\rho_o^2)/2 \) is used if \( \rho_o \) < 0.95; otherwise 1.06 is used.*
Figure 1a. Curves for finding the required sample size that will yield 80% power for a two-tailed test of a single Spearman coefficient at significance level 0.05, to be used when the alternative value (\(\rho_{s1}\)) is greater than the null value (\(\rho_{s0}\)). To use the chart, first locate the null value of the SCC along the x-axis. Then, draw a vertical line that intersects with the sample size curve corresponding to the desired alternative value. Finally, draw a horizontal line from the curve to the y-axis. The point of intersection with the y-axis is the required sample size.

Figure 1b. Curves for finding the required sample size that will yield 80% power for a two-tailed test of a single Spearman coefficient at significance level 0.05, to be used when the alternative value (\(\rho_{s1}\)) is less than the null value (\(\rho_{s0}\)). To use the chart, first locate the null value of the SCC along the x-axis. Then, draw a vertical line that intersects with the sample size curve corresponding to the desired alternative value. Finally, draw a horizontal line from the curve to the y-axis. The point of intersection with the y-axis is the required sample size.
If we intend to use the KCC as the measure of association in the study being planned, the required sample sizes would be \( n = 52 \) and \( n = 75 \), respectively, for a two-tailed test of
\[
H_0^*: \tau = 0.4 \text{ vs. } H_a^*: \tau \neq 0.4,
\]
assuming alternative values of \( \tau = 0.6 \) and 0.2 (Figure 2). A comparison of the sample size results for the SCC and KCC in this example illustrates the phenomenon described by Helsel [5] that, with all other things (including sample size) being equal, smaller values of \( \tau \), the sample value of the KCC, relative to \( r \) and \( r_s \) are required to yield a given degree of statistical significance. With regard to sample size estimation in this article, this means that smaller sample sizes are required for the KCC than for the SCC to achieve the same level of power, all other things (including null and alternative values) being equal.

Sample size based on confidence interval estimation

The Fisher z-transform can also be used to derive approximate 100(1-\( \alpha \))% confidence limits for \( \rho \). To obtain these limits, one must first derive a Wald-type approximate confidence interval (C.I.) for the Fisher-z transformed value of the population value of the PCC, \( z(\rho) = \ln \left( \frac{1+\rho}{1-\rho} \right) / 2 \).

An approximate 100(1-\( \alpha \))% C.I. for \( z(\rho) \) is given by:
\[
z(r) \pm z_{\alpha/2} \sqrt{1/(n-3)},
\]
where \( z_{\alpha/2} \) is the upper \( \alpha/2 \)-percentage point of the standard normal distribution.

Next, the endpoints of the C.I. in eqn. (8) are "back transformed" using the inverse of the Fisher-z transform. This yields the following approximate 100(1-\( \alpha \))% C.I. for \( \rho \):
\[
\left( \exp\left(2(LCL) - 1\right) / \exp\left(2(UCL) + 1\right) \right) \left( \exp\left(2(UCL) - 1\right) / \exp\left(2(LCL) + 1\right) \right),
\]
where
\[
LCL = z(r) - z_{\alpha/2} \sqrt{1/(n-3)}.
\]
And
\[
UCL = z(r) + z_{\alpha/2} \sqrt{1/(n-3)}.
\]

More generally, let \( \xi \) denote the population value of the desired measure of association (either the PCC, SCC, or KCC). Then approximate 100(1-\( \alpha \))% confidence limits for \( z(\xi) \) are given by:
\[
LCL = z(\hat{\xi}) - z_{\alpha/2} \sqrt{c^2/(n-b)} \text{ and } UCL = z(\hat{\xi}) + z_{\alpha/2} \sqrt{c^2/(n-b)}, \tag{10}
\]
where \( b \) and \( c^2 \) are obtained from Table 1. The limits in eqn. (10) are then back transformed using eqn. eqn. (9) to obtain approximate 100(1-\( \alpha \))% confidence limits for \( \xi \).

Bonett and Wright [2] proposed a two-stage process for determining \( n \) based on the desired width of a two-sided 100(1-\( \alpha \))% C.I. for the true value of a measure of association between two variables \( X \) and \( Y \). This method can be applied to Pearson, Spearman, and Kendall coefficients.

Stage 1: Compute \( n_o = 4c^2 \left( 1 - \phi^2 \right) (z_{\alpha/2} / w)^2 + b \).

Stage 2: Compute
\[
n = (n_o - b)(w_o / w)^2 + b
\]
if the desired confidence interval width has not been attained using \( n_o \). In these formulas,
\[
\hat{\xi} \text{ is a planning value for the coefficient of interest,}
\]
\( w = \) desired width of the C.I. for the population value \( \xi \),
\( z_{\alpha/2} \) upper \( \alpha/2 \)-percentage point of the standard normal distribution,
\( b \) and \( c^2 \) are obtained from Table 1,
\( n_o \) is the sample size estimate from Stage 1,
\( w_o \) is the width of the Stage 1 interval based on \( n_o \).

Figure 2. Curves for finding the required sample size that will yield 80% power for a two-tailed test of a single Kendall coefficient at significance level 0.05. To use the chart, first locate the smaller of the null value (10) and the alternative value (11) of the KCC along the x-axis. Then, draw a vertical line that intersects with the sample size curve corresponding to the larger of 10 and 11. Finally, draw a horizontal line from the curve to the y-axis. The point of intersection with the y-axis is the required sample size.
If the value of $n$ resulting from Stage 2 is not an integer, the next larger integer is used. Charts for finding the required sample size for estimating a SCC or KCC using a 95% two-sided C.I. that is based on this two-stage procedure are provided in Figures 3 and 4, respectively.

To use the charts in Figures 3 and 4, first locate the planning value for the measure of association along the $x$-axis. Then, draw a vertical line that intersects with the curve corresponding to the desired width of the confidence interval. Finally, draw a horizontal line from the curve to the $y$-axis. The point of intersection is the required sample size. For example, using the chart in Figure 3, we find that a sample size of $n=177$ would yield a 95% C.I. for $\rho_s$ of width 0.20 based on a planning value of 0.70. For the KCC, a sample size of $n=62$ would yield a 95% C.I. for $r$ of width 0.20 based on a planning value of 0.70 (Figure 4).

Notes on Using the Charts

Zero null value

By far the most commonly used null value in a test of a single PCC, SCC, or KCC, is $\xi_0=0$. We have designed our charts so that it is relatively easy to determine the sample sizes required to detect several alternative values if the null hypothesis is $H_0: \xi_s=0$. For example, suppose that we are planning a study in which the primary analysis will be to perform a two-tailed test of $H_0: \rho_s=0$. By simply reading the values from the $y$-axis in Figure 1a, we can identify the sample sizes needed for each of the alternative values given in Table 2.

Negative coefficients

The sample size charts in this article can be used only when $\xi_1$ and $\xi_0$ have the same sign. For negative values of $\xi_1$ and $\xi_0$, one simply enters the appropriate chart with $|\xi_1|$ and $|\xi_0|$. If $\xi_1$ and $\xi_0$ are of opposite signs, the formula in eqn. (7) is still valid; however, the charts provided in this article do not apply.

Minimum detectable alternative value

In addition to finding the sample size needed for planning statistical inference for Spearman or Kendall coefficients, the charts can also be used to find the minimum detectable alternative value for a given sample size. For example, suppose that the primary analysis in the study being planned will involve a two-tailed test of $H_0: \rho_s=0.4$ vs. $H_a: \rho_s \neq 0.4$ at significance level 0.05, and that the largest possible sample size available to the investigators is $n=50$. One could determine the minimum detectable alternative value ($\rho_{s1}$) greater than the null value ($\rho_{s0}=0.4$) from Figure 1a by first drawing a horizontal line from $n=50$ on the $y$-axis, and then drawing a vertical line from 0.4 on the $x$-axis. One then reads off the alternative values larger than 0.4.
from this vertical line until it intersects with the horizontal line drawn from the y-axis. Note that graphical interpolation will be required if the horizontal line does not intersect with one of the sample size curves. Using Figure 1a, we see that with n=50, we could detect any alternative value $\rho_{s1}$ greater than or equal to 0.7 with 80% power using a two-tailed test with significance level 0.05. A similar method could be used with Figure 1b to determine the maximum detectable alternative value ($\rho_{s1}$) less than the null value in a test of $H_0: \rho_s = \rho_{s0}$ vs. $H_a: \rho_s \neq \rho_{s0}$.

Interpolation errors

As with any graphical method, these charts are subject to error. For example, one must interpolate graphically if the horizontal line drawn from the appropriate curve in the charts intersects the y-axis at a value between tick marks. This is more of a problem for larger sample sizes since the distances between tick marks are much larger. However, the error for the interpolated value will be no larger than the difference between the values corresponding to the tick marks, and an adjustment can be made by slightly inflating the value read from the charts. We have found that an 6-inch ruler marked off in millimeters is particularly useful for carrying out any necessary graphical interpolation in the charts.

Discussion

In this article, we have presented charts that can be used for sample size determination when planning a study in which the primary analysis involves hypothesis testing or confidence interval estimation for a Spearman or Kendall coefficient. In addition to the charts, we have provided simple sample size formulas that can be used for more accurate calculations. We have found Microsoft Excel® to be particularly useful for performing these calculations, and a spreadsheet that accomplishes this is available from the second author. Freely available SAS code for performing these sample size calculations can be found in the on-line Software Appendix for Looney and Hagan [8], which can be downloaded from http://bcs.wiley.com/he-bcs/Books?action=index&itemId=1118027558&bcsId=9351.

Despite the widespread use of correlation analysis, it is generally the case that little or no attention is given to sample size determination when planning a study in which correlation will be the primary analysis.

Conclusion

In fact, our review of clinical research journals indicated that none of the 111 articles published in 2014 that used correlation as the primary analysis provided a sample size justification or power calculation. It is hoped that availability of the tools presented in this article will encourage analysts to perform sample size calculations for correlation coefficient inference. It is vitally important that one do this, given the adverse consequences when studies are either under- or over-powered. This can easily happen if one does not perform a sample size calculation prior to beginning the study.

References


