

## Correlation (Pearson, Kendall, Spearman)

**Correlation** is a bivariate analysis that measures the strengths of association between two variables. In statistics, the value of the correlation coefficient varies between +1 and -1. When the value of the correlation coefficient lies around  $\pm 1$ , then it is said to be a perfect degree of association between the two variables. As the correlation coefficient value goes towards 0, the relationship between the two variables will be weaker. Usually, in statistics, we measure three types of correlations: [Pearson correlation](#), Kendall rank correlation and [Spearman correlation](#).

**Pearson r correlation:** Pearson r correlation is widely used in statistics to measure the degree of the relationship between linear related variables. For example, in the stock market, if we want to measure how two commodities are related to each other, Pearson r correlation is used to measure the degree of relationship between the two commodities. The following formula is used to calculate the Pearson r correlation:

$$r = \frac{N \sum xy - \sum (x)(y)}{\sqrt{N \sum x^2 - \sum (x^2)} [N \sum y^2 - \sum (y^2)]}$$

Where:

r = Pearson r correlation coefficient

N = number of value in each data set

?xy = sum of the products of paired scores

?x = sum of x scores

?y = sum of y scores

?x<sup>2</sup> = sum of squared x scores

?y<sup>2</sup> = sum of squared y scores

### Questions a Pearson correlation answers

Is there a statistically significant relationship between age, as measured in years, and height, measured in inches?

Is there a relationship between temperature, measure in degree Fahrenheit, and ice cream sales, measured by income?

Is there a relationship among job satisfaction, as measured by the JSS, and income, measured in dollars?

### Assumptions

For the Pearson  $r$  correlation, both variables should be normally distributed. Other assumptions include linearity and homoscedasticity. Linearity assumes a straight line relationship between each of the variables in the analysis and homoscedasticity assumes that data is normally distributed about the regression line.

### **Key Terms**

**Effect size:** Cohen's standard will be used to evaluate the correlation coefficient to determine the strength of the relationship, or the effect size, where coefficients between .10 and .29 represent a small association; coefficients between .30 and .49 represent a medium association; and coefficients above .50 represent a large association or relationship.

**Continuous data:** This type of data possess the properties of magnitude and equal interval between adjacent units. Equal intervals between adjacent units means that there are equal amounts of the variable being measured between adjacent units on the scale. An example would be age. An increase in age from 21 to 22 would be the same as an increase in age from 60 to 61; one year. In addition, we can perform mathematical functions on scale data to determine if  $X - Y = A - B$ ,  $X - Y > A - B$ , or if  $X - Y$

**Kendall rank correlation:** Kendall rank correlation is a non-parametric test that measures the strength of dependence between two variables. If we consider two samples, a and b, where each sample size is  $n$ , we know that the total number of pairings with a b is  $n(n-1)/2$ . The following formula is used to calculate the value of Kendall rank correlation:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

Where:

$N_c$ = number of concordant

$N_d$ = Number of discordant

### **Key Terms**

**Concordant:** Ordered in the same way

**Discordant:** Ordered differently.

**Spearman rank correlation:** Spearman rank correlation is a non-parametric test that is used to

measure the degree of association between two variables. It was developed by Spearman, thus it is called the Spearman rank correlation. Spearman rank correlation test does not assume any assumptions about the distribution of the data and is the appropriate correlation analysis when the variables are measured on a scale that is at least ordinal.

The following formula is used to calculate the Spearman rank correlation:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Where:

P= Spearman rank correlation

di= the difference between the ranks of corresponding values Xi and Yi

n= number of value in each data set

### Questions Spearman Correlation Answers

Is there a statistically significant relationship between participant responses to two Likert scales questions?

Is there a statistically significant relationship between how the horses place in the race and the horses' ages?

### Assumptions

Spearman rank correlation test does not make any assumptions about the distribution. The assumptions of Spearman rho correlation are that data must be at least ordinal and scores on one variable must be monotonically related to the other variable.

### Key Terms

**Effect size:** Cohen's standard will be used to evaluate the correlation coefficient to determine the strength of the relationship, or the effect size, where coefficients between .10 and .29 represent a small association; coefficients between .30 and .49 represent a medium association; and coefficients above .50 represent a large association or relationship.

**Ordinal data:** Ordinal scales rank order the items that are being measured to indicate if they possess more, less, or the same amount of the variable being measured. An ordinal scale allows us to determine if  $X > Y$ ,  $Y > X$ , or if  $X = Y$ . An example would be rank ordering the participants in a dance contest. The dancer who was ranked one was a better dancer than the dancer who was ranked two. The dancer ranked two was a better dancer than the dancer who was ranked three,

and so on. Although this scale allows us to determine greater than, less than, or equal to, it still does not define the magnitude of the relationship between units.

**For Assistance Conducting Analyses:**

Please [click here](#) for more information on our academic services.

**Correlation Resources:**

Algina, J., & Keselman, H. J. (1999). Comparing squared multiple correlation coefficients: Examination of a confidence interval and a test significance. *Psychological Methods*, 4(1), 76-83.

Bobko, P. (2001). *Correlation and regression: Applications for industrial organizational psychology and management* (2nd ed.). Thousand Oaks, CA: Sage Publications. [View](#)

Bonett, D. G. (2008). Meta-analytic interval estimation for bivariate correlations. *Psychological Methods*, 13(3), 173-181.

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage Publications. [View](#)

Cheung, M. W. -L., & Chan, W. (2004). Testing dependent correlation coefficients via structural equation modeling. *Organizational Research Methods*, 7(2), 206-223.

Coffman, D. L., Maydeu-Olivares, A., Arnau, J. (2008). Asymptotic distribution free interval estimation: For an intraclass correlation coefficient with applications to longitudinal data. *Methodology*, 4(1), 4-9.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates. [View](#)

Hatch, J. P., Hearne, E. M., & Clark, G. M. (1982). A method of testing for serial correlation in univariate repeated-measures analysis of variance. *Behavior Research Methods & Instrumentation*, 14(5), 497-498.

Kendall, M. G., & Gibbons, J. D. (1990). *Rank Correlation Methods* (5th ed.). London: Edward Arnold. [View](#)

Krijnen, W. P. (2004). Positive loadings and factor correlations from positive covariance matrices. *Psychometrika*, 69(4), 655-660.

Shieh, G. (2006). Exact interval estimation, power calculation, and sample size determination in normal correlation analysis. *Psychometrika*, 71(3), 529-540.

Stauffer, J. M., & Mendoza, J. L. (2001). The proper sequence for correcting correlation coefficients

for range restriction and unreliability. *Psychometrika*, 66(1), 63-68.

**Related Pages:**

- [Table of Critical Values: Pearson Correlation](#)
- [Conduct and Interpret a Spearman Rank Correlation](#)
- [Conduct and Interpret a Bivariate \(Pearson\) Correlation](#)