Correlation (Pearson, Kendall, Spearman)

Correlation is a bivariate analysis that measures the strength of association between two variables and the direction of the relationship. In terms of the strength of relationship, the value of the correlation coefficient varies between +1 and -1. A value of ± 1 indicates 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. The direction of the relationship is indicated by the sign of the coefficient; a + sign indicates a positive relationship and a − sign indicates a negative relationship. Usually, in statistics, we measure four types of correlations: Pearson correlation, Kendall rank correlation, Spearman correlation, and the Point-Biserial correlation. The software below allows you to very easily conduct a correlation.

Discover How We Assist to Edit Your Dissertation Chapters

Aligning theoretical framework, gathering articles, synthesizing gaps, articulating a clear methodology and data plan, and writing about the theoretical and practical implications of your research are part of our comprehensive dissertation editing services.

Schedule Your FREE Consultation
with a Dissertation Expert Today

Bring dissertation editing expertise to chapters 1-5 in timely manner.

Track all changes, then work with you to bring about scholarly writing.

Ongoing support to address committee feedback, reducing revisions.

Quantitative Results in One-Hour
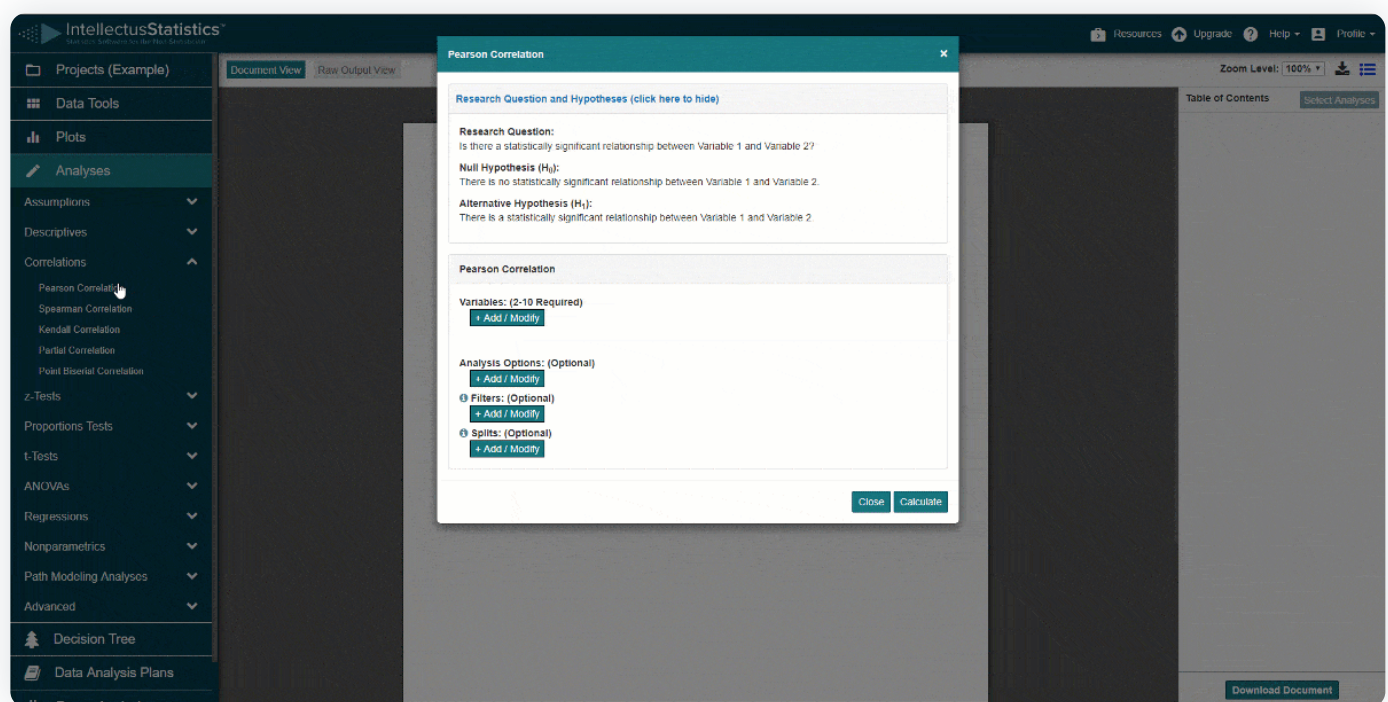
Screen share with a statistician as we walk you through conducting and understanding your interpreted analysis. Have your results draft complete in one hour with guaranteed accuracy.

Pearson r correlation: Pearson *r* correlation is the most widely used correlation statistic to measure the degree of the relationship between linearly related variables. For example, in the stock market, if we want to measure how two stocks are related to each other, Pearson *r* correlation is used to measure the degree of relationship between the two. The point-biserial correlation is conducted with the Pearson correlation formula except that one of the variables is dichotomous. The following formula is used to calculate the Pearson *r* correlation:

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

r_{xy} = Pearson r correlation coefficient between x and y
n = number of observations
x_i = value of x (for ith observation)
y_i = value of y (for ith observation)

Types of research questions a Pearson correlation can examine:

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

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

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

Assumptions

For the Pearson *r* correlation, both variables should be normally distributed (normally distributed variables have a bell-shaped curve). Other assumptions include linearity and homoscedasticity. Linearity assumes a straight line relationship between each of the two variables and homoscedasticity assumes that data is equally distributed about the regression line.

Conduct and Interpret a Pearson Correlation
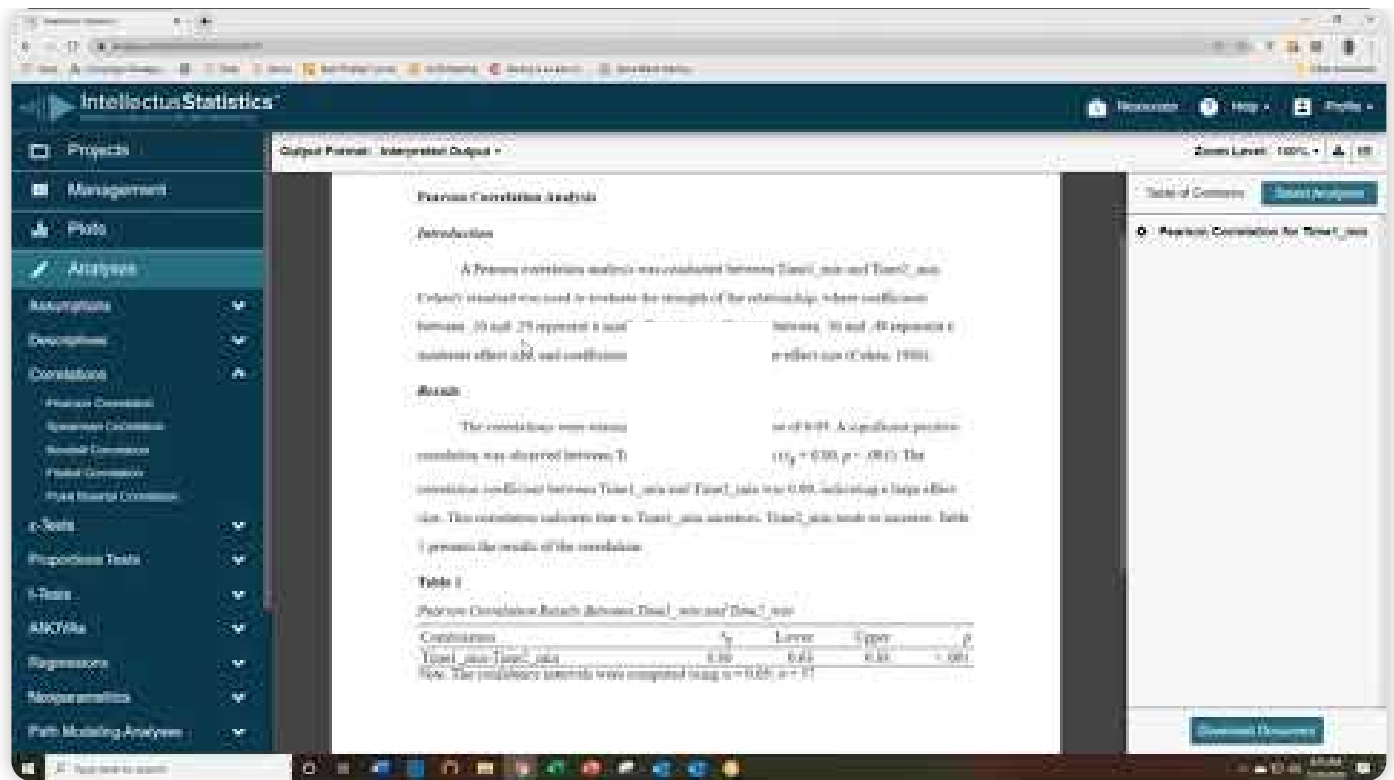
Key Terms

represent a medium association, and coefficients of .50 and above represent a large association or relationship.

Continuous data: Data that is interval or ratio level. This type of data possesses the properties of magnitude and equal intervals 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.

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 $n(n\text{-}1)/2$. The following formula is used to calculate the value of Kendall rank correlation:

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

Nc= number of concordant

Nd= Number of discordant

Conduct and Interpret a Kendall Correlation

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. The Spearman rank correlation test does not carry 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:

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

ρ= Spearman rank correlation

di= the difference between the ranks of corresponding variables

n= number of observations

**Types of research questions a Spearman Correlation can examine:**

Is there a statistically significant relationship between participants' level of education (high school, bachelor's, or graduate degree) and their starting salary?

Is there a statistically significant relationship between horse's finishing position a race and horse's age?

Assumptions

The assumptions of the Spearman correlation are that data must be at least ordinal and the scores on one variable must be monotonically related to the other variable.

Conduct and Interpret a Spearman Correlation
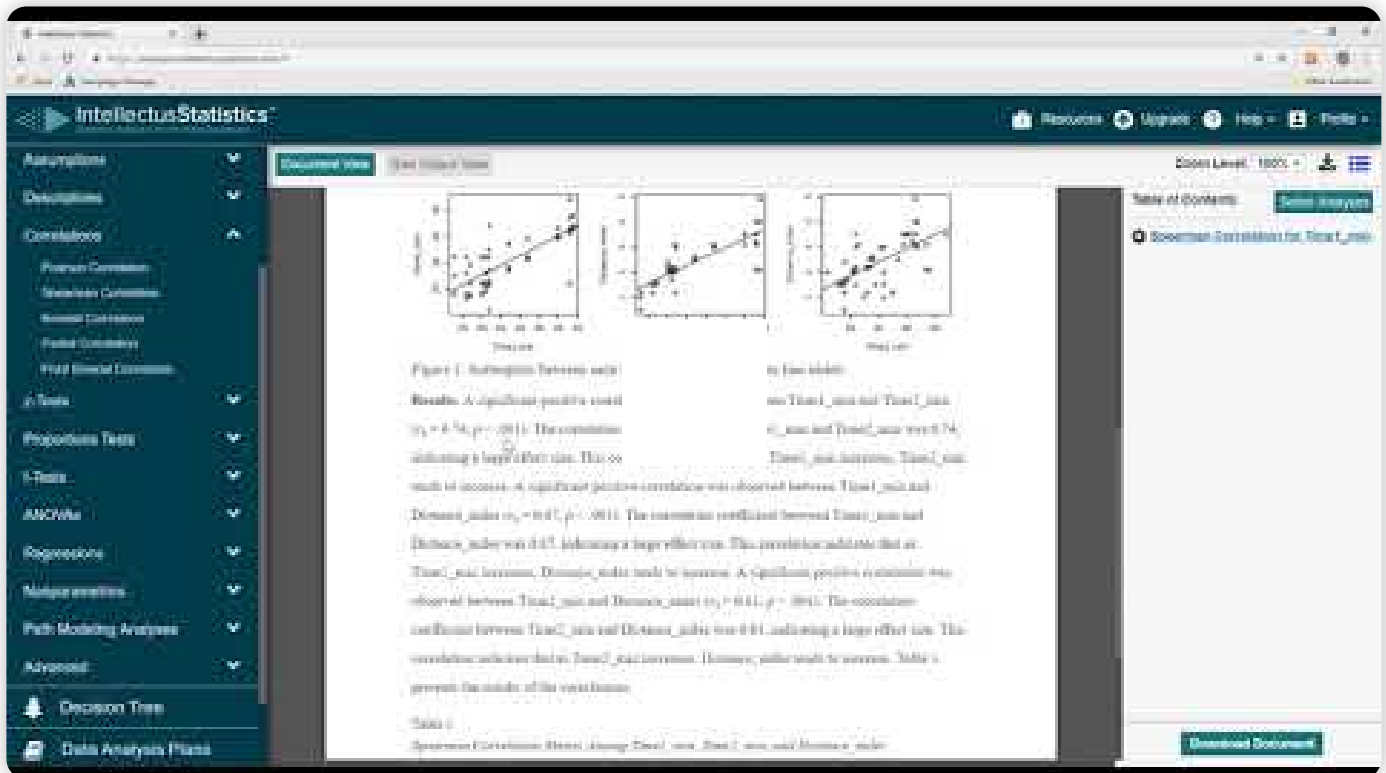
Key Terms

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

Ordinal data: In an ordinal scale, the levels of a variable are ordered such that one level can be considered higher/lower than another. However, the magnitude of the difference between levels is not necessarily known. An example would be rank ordering levels of education. A graduate degree is higher than a bachelor's degree, and a bachelor's degree is higher than a high school diploma. However, we cannot quantify how much higher a graduate degree is compared to a bachelor's degree. We also cannot say that the difference in education between a graduate degree and a bachelor's degree is the same as the difference between a bachelor's degree and a high school diploma.

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

Step Boldly to Completing your Research

If you're like others, you've invested a lot of time and money developing your dissertation or project research. Finish strong by learning how our dissertation specialists support your efforts to cross the finish line.

Learn More

Dissertation Consulting

Webinars

Free Dissertation Resources

Blog

About Us

Sitemap

Our Privacy Policy

Call

727-442-4290

Email

info@statisticssolutions.com

Search dissertation topics

Contact Us