

When should I use Spearman's Rank Correlation?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *When should I use Spearman's Rank Correlation?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=157890>

Spearman's Rank Correlation is a statistical method used to measure the strength and direction of the relationship between two variables. It is typically used when the data is non-parametric, meaning it is not normally distributed, and when the relationship between the variables is not linear. This method is particularly useful when analyzing ordinal or ranked data, as it allows for the comparison of two sets of rankings. It is commonly used in social sciences, psychology, and market research to determine the degree of association between variables. Overall, Spearman's Rank Correlation is a valuable tool for analyzing and understanding relationships between variables in non-parametric data sets.

When to Use Spearman's Rank Correlation (2 Scenarios)

The most common way to quantify the linear association between two variables is to use the r , which always takes on a value between -1 and 1 where:

-1 indicates a perfectly negative linear correlation
0 indicates no linear correlation
1 indicates a perfectly positive linear correlation

However, this type of correlation coefficient works best when the true underlying relationship between the two variables is *linear*.

There is another type of correlation coefficient known as Spearman's rank correlation that is better to use in two specific scenarios:

Scenario 1: When working with ranked data.

An example could be a dataset that contains the rank of a student's math exam score along with the rank of their science exam score in a class.

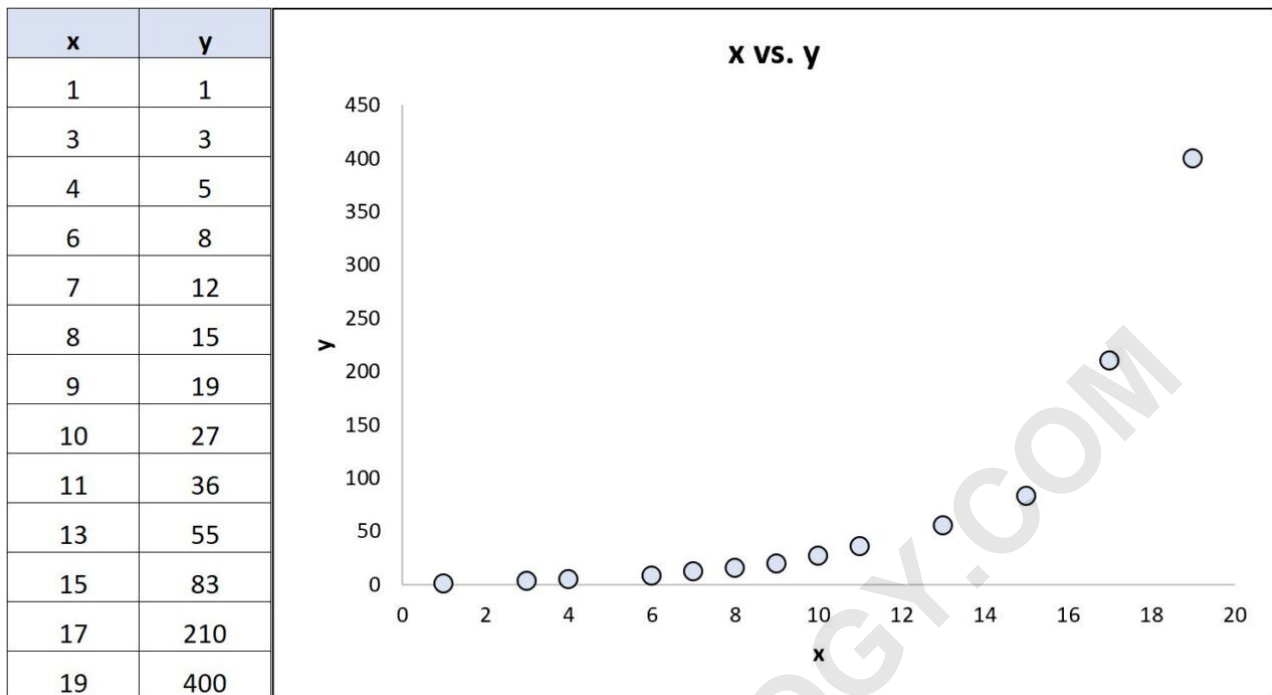
Scenario 2: When one or more extreme outliers are present.

When extreme outliers are present in a dataset, Pearson's correlation coefficient is highly affected.

The following examples show how to calculate the Spearman Rank Correlation in each of these scenarios.

Scenario 1: Spearman's Rank Correlation with Ranked Data

Consider the following dataset (and corresponding scatter plot) that shows the relationship between two variables:



Using statistical software, we can calculate the following correlation coefficients for these two variables:

Pearson's correlation: 0.79 Spearman's rank correlation: 1

In this scenario, if we only care about the ranks of the data values (when the rank of x increases, does the rank of y also increase?) then Spearman's rank correlation would provide us with a better idea of the correlation between the two variables.

In this particular dataset, as the rank of x increases the

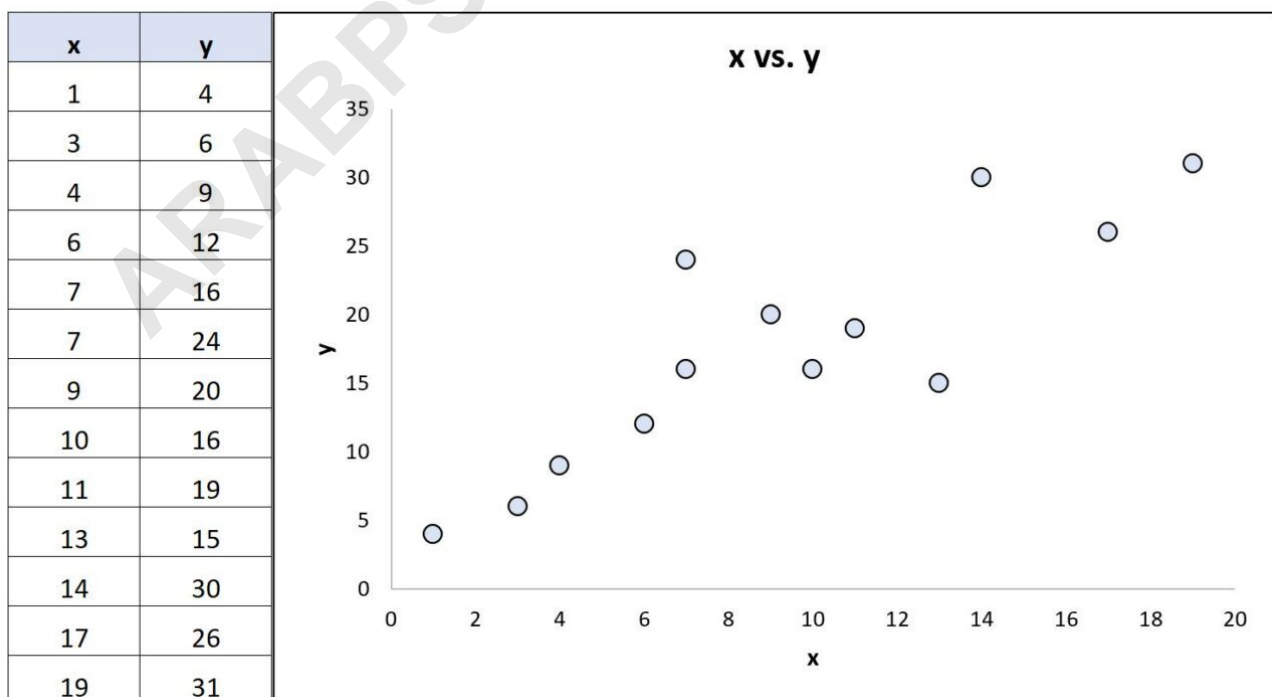
rank of y *always* increases.

Spearman's rank correlation captures this behavior perfectly by telling us that there is a perfect positive relationship ($\rho = 1$) between the ranks of x and the ranks of y .

This is true, but it's not useful if we only care about the relationship between the ranks of x and the ranks of y .

Scenario 2: Spearman's Rank Correlation with Extreme Outliers

Consider the following dataset (and corresponding scatter plot) that shows the relationship between two variables:

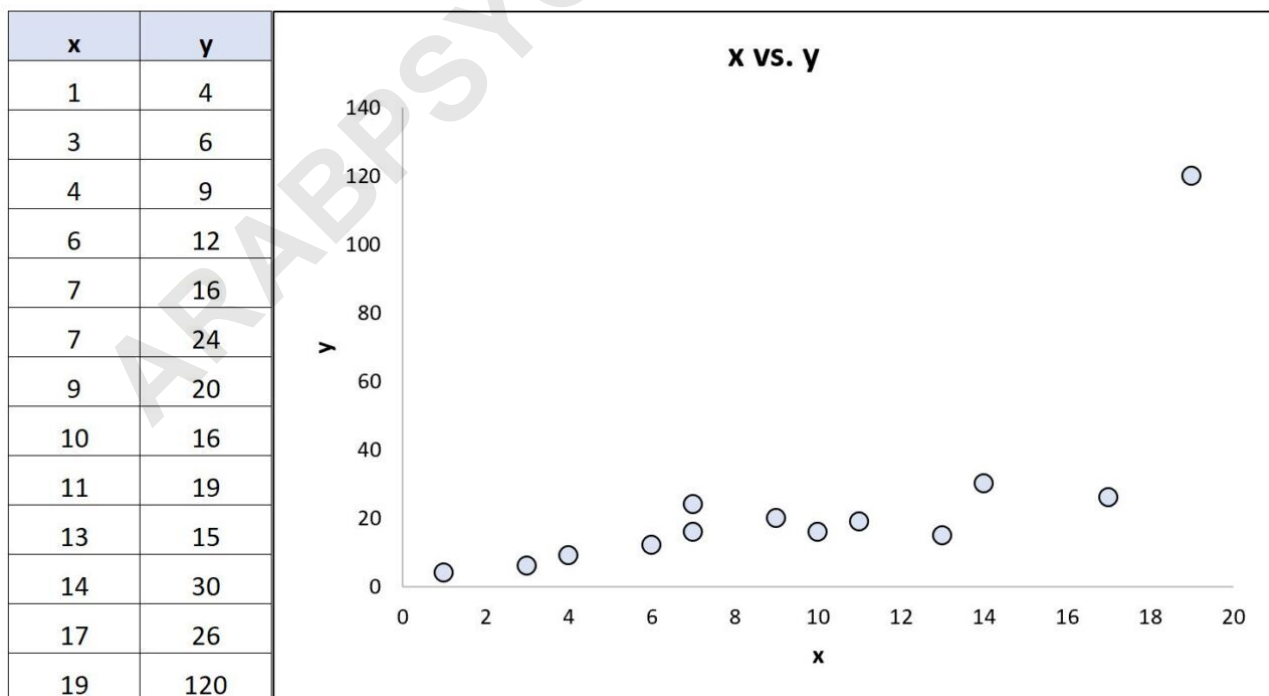


Using statistical software, we can calculate the following correlation coefficients for these two variables:

Pearson's correlation: 0.86
Spearman's rank correlation: 0.85

The correlation coefficients are nearly identical because the underlying relationship between the variables is roughly linear and there are no extreme outliers.

Now suppose we change the last y value in the dataset to be an extreme outlier:



Using statistical software, we can calculate the correlation coefficients once again:

Pearson's correlation: 0.69 Spearman's rank correlation: 0.85

Pearson's correlation coefficient changed dramatically while Spearman's rank correlation coefficient remained the same.

Using statistical jargon, we would say that the relationship between x and y is (as x increases, y generally increases) but not linear since the outlier influences the data so much.

In this scenario, Spearman's rank correlation does a good job of quantifying this monotonic relationship, while Pearson's correlation does a poor job because it's attempting to calculate the linear relationship between the two variables.

Related:

Additional Resources

The following tutorials explain how to calculate the

Spearman Rank Correlation using different software:

ARABPSYCHOLOGY.COM