https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php

Spearman's Rank-Order Correlation

This guide will tell you when you should use Spearman's rank-order correlation to analyse your data, what assumptions you have to satisfy, how to calculate it, and how to report it. If you want to know how to run a Spearman correlation in SPSS Statistics, go to our guide here.

When should you use the Spearman's rank-order correlation?

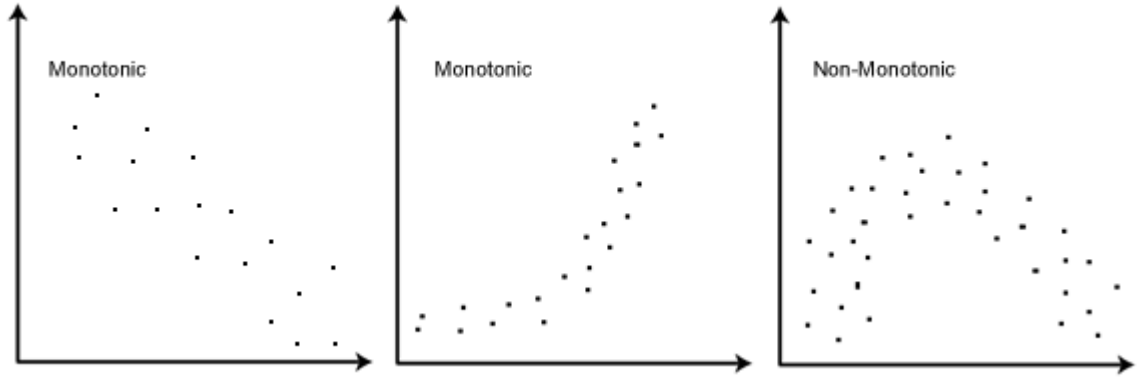
The Spearman's rank-order correlation is the nonparametric version of the Pearson product-moment correlation. Spearman's correlation coefficient, (ρ, also signified by r_s) measures the strength and direction of association between two ranked variables.

What are the assumptions of the test?

You need two variables that are either ordinal, interval or ratio (see our Types of Variable guide if you need clarification). Although you would normally hope to use a Pearson product-moment correlation on interval or ratio data, the Spearman correlation can be used when the assumptions of the Pearson correlation are markedly violated. However, Spearman's correlation determines the strength and direction of the **monotonic relationship** between your two variables rather than the strength and direction of the linear relationship between your two variables, which is what Pearson's correlation determines.

What is a monotonic relationship?

A monotonic relationship is a relationship that does one of the following: (1) as the value of one variable increases, so does the value of the other variable; or (2) as the value of one variable increases, the other variable value decreases. Examples of monotonic and non-monotonic relationships are presented in the diagram below:



Why is a monotonic relationship important to Spearman's correlation?

Spearman's correlation measures the strength and direction of monotonic association between two variables. Monotonicity is "less restrictive" than that of a linear relationship. For example, the middle image above shows a relationship that is monotonic, but not linear.

A monotonic relationship is not strictly an assumption of Spearman's correlation. That is, you can run a Spearman's correlation on a non-monotonic relationship to determine if there is a **monotonic component** to the association. However, you would normally pick a measure of association, such as Spearman's correlation, that fits the pattern of the observed data. That is, if a scatterplot shows that the relationship between your two variables looks monotonic you would run a Spearman's correlation because this will then measure the strength and direction of this monotonic relationship. On the other hand if, for example, the relationship appears linear (assessed via scatterplot) you would run a Pearson's correlation because this will measure the strength and direction of any linear relationship. You will not always be able to visually check whether you have a monotonic relationship, so in this case, you might run a Spearman's correlation anyway.

How to rank data?

In some cases your data might already be ranked, but often you will find that you need to rank the data yourself (or use SPSS Statistics to do it for you). Thankfully, ranking data is not a difficult task and is easily achieved by working through your

data in a table. Let us consider the following example data regarding the marks achieved in a maths and English exam:

	Marks									
English	56	75	45	71	61	64	58	80	76	61
Maths	66	70	40	60	65	56	59	77	67	63

The procedure for ranking these scores is as follows:

First, create a table with four columns and label them as below:

	Maths (mark)	Rank (English)	Rank (maths)
56	66	9	4
75	70	3	2
45	40	10	10
71	60	4	7
61	65	6.5	5
64	56	5	9
58	59	8	8
80	77	1	1
76	67	2	3
61	63	6.5	6

You need to rank the scores for maths and English separately. The score with the highest value should be labelled "1" and the lowest score should be labelled "10" (if your data set has more than 10 cases then the lowest score will be how many cases you have). Look carefully at the two individuals that scored 61 in the English exam (highlighted in bold). Notice their joint rank of 6.5. This is because when you have two identical values in the data (called a "tie"), you need to take the average of the ranks that they would have otherwise occupied. We do this because, in this example,

we have no way of knowing which score should be put in rank 6 and which score should be ranked 7. Therefore, you will notice that the ranks of 6 and 7 do not exist for English. These two ranks have been averaged ((6 + 7)/2 = 6.5) and assigned to each of these "tied" scores.

What is the definition of Spearman's rank-order correlation?

There are two methods to calculate Spearman's correlation depending on whether: (1) your data does not have tied ranks or (2) your data has tied ranks. The formula for when there are no tied ranks is:

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

where d_i = difference in paired ranks and n = number of cases. The formula to use when there are tied ranks is:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

where i = paired score.

Spearman's Rank-Order Correlation (cont...)

What values can the Spearman correlation coefficient, r_s, take?

The Spearman correlation coefficient, r_s, can take values from +1 to -1. A r_s of +1 indicates a perfect association of ranks, a r_s of zero indicates no association between ranks and a r_s of -1 indicates a perfect negative association of ranks. The closer r_s is to zero, the weaker the association between the ranks.

An example of calculating Spearman's correlation

To calculate a Spearman rank-order correlation on data without any ties we will use the following data:

	Marks									
English	56	75	45	71	62	64	58	80	76	61
Maths	66	70	40	60	65	56	59	77	67	63

We then complete the following table:

	Maths (mark)	Rank (English)	Rank (maths)	d	d^2
56	66	9	4	5	25
75	70	3	2	1	1
45	40	10	10	0	0
71	60	4	7	3	9
62	65	6	5	1	1
64	56	5	9	4	16
58	59	8	8	0	0
80	77	1	1	0	0

	Maths (mark)	Rank (English)	Rank (maths)	d	d^2
76	67	2	3	1	1
61	63	7	6	1	1

Where d = difference between ranks and d^2 = difference squared.

We then calculate the following:

$$\sum d_i^2 = 25 + 1 + 9 + 1 + 16 + 1 + 1 = 54$$

We then substitute this into the main equation with the other information as follows:

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

$$\rho = 1 - \frac{6 \times 54}{10(10^2 - 1)}$$

$$\rho = 1 - \frac{324}{990}$$

$$\rho = 1 - 0.33$$

$$\rho = 0.67$$

as n = 10. Hence, we have a ρ (or r_s) of 0.67. This indicates a strong positive relationship between the ranks individuals obtained in the maths and English exam. That is, the higher you ranked in maths, the higher you ranked in English also, and vice versa.

How do you report a Spearman's correlation?

How you report a Spearman's correlation coefficient depends on whether or not you have determined the statistical significance of the coefficient. If you have simply run the Spearman correlation without any statistical significance tests, you are able to simple state the value of the coefficient as shown below:

$$\rho = coefficient \quad OR \quad r_s = coefficient$$

e.g.

$$\rho = 0.67 \quad OR \quad r_s = 0.67$$

However, if you have also run statistical significance tests, you need to include some more information as shown below:

$$\rho(df) = \rho\ coefficient, P = P\ value$$

$$\rho(8) = 0.67, P = 0.033$$

where *df* = *N* – 2, where *N* = number of pairwise cases.

How do you express the null hypothesis for this test?

The general form of a null hypothesis for a Spearman correlation is:

H_0: There is no [monotonic] association between the two variables [in the population].

Remember, you are making an inference from your sample to the population that the sample is supposed to represent. However, as this a general understanding of an inferential statistical test, it is often not included. A null hypothesis statement for the example used earlier in this guide would be:

H_0: There is no [monotonic] association between maths and English marks.

How do I interpret a statistically significant Spearman correlation?

It is important to realize that statistical significance does not indicate the strength of Spearman's correlation. In fact, the statistical significance testing of the Spearman correlation does not provide you with *any* information about the strength of the relationship. Thus, achieving a value of $p = 0.001$, for example, does not mean that the relationship is stronger than if you achieved a value of $p = 0.04$. This is because the significance test is investigating whether you can reject or fail to reject the null hypothesis. If you set $\alpha = 0.05$, achieving a statistically significant Spearman rank-order correlation means that you can be sure that there is less than a 5% chance that the strength of the relationship you found (your ρ coefficient) happened by chance if the null hypothesis were true.