

Spearman's Rank-Order Correlation

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									
Englis	5	7	4	7	6	6	5	8	7	6

	Marks									
h	6	5	5	1	1	4	8	0	6	1
Maths	6	7	4	6	6	5	5	7	6	6
	6	0	0	0	5	6	9	7	7	3

The procedure for ranking these scores is as follows:

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

English (mark)	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.