Spearman's correlation

The Spearman rank-order correlation coefficient (**Spearman's Rho**) is a statistical test for measuring the dependence between the rankings of two variables; it is used to understand the strength of the relationship between these variables that can be continuous or ordinal and should have a monotonic relationship.
This test is a nonparametric measure, denoted by the symbol r_s (or the Greek letter ρ, pronounced rho) of the strength and direction of association that exists between two variables measured on at least an ordinal scale. The test is used for either ordinal variables or for continuous data that has failed the assumptions necessary for conducting the Pearson's correlation and its interpretation is similar to that of Pearson's, e.g. the closer ρ is to ±1 the stronger the monotonic relationship.

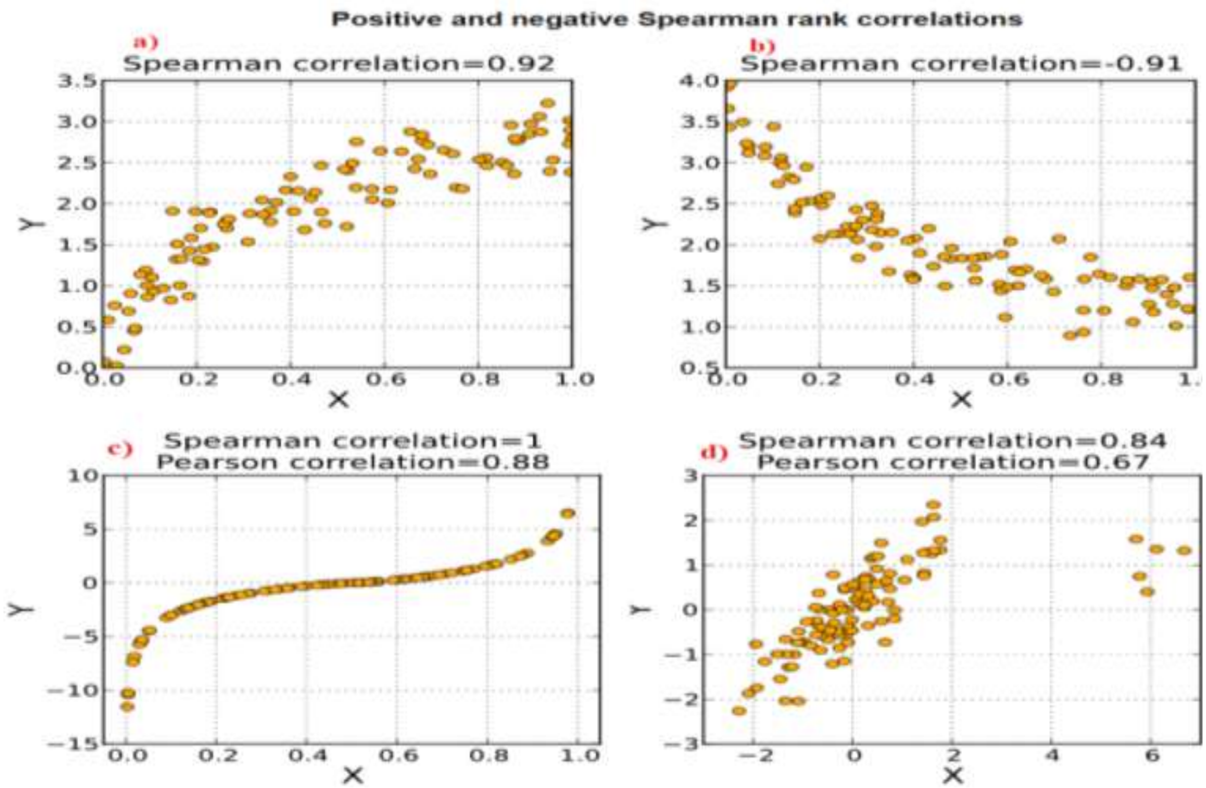
The Spearman Rank Correlation Coefficient (SRCC) for **ordinal**, **interval** or **ratio scale** data is the nonparametric version of Pearson's Correlation Coefficient (PCC). Here nonparametric means a statistical test where it's not required for your data to follow a normal distribution. They're also known as distribution-free tests and can provide benefits in certain situations.

Spearman's rank correlation requires a **monotonic relationship** between the two variables. 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:

Correlation is an effect size and so we can verbally describe the strength of the correlation using the following guide for the absolute value of

0.01-0.19	"very weak"
0.20- 0.39	"weak"
0.40-0.59	"moderate"
0.60-0.79	"strong"
0 .80-1.0	very strong

Spearman's rank correlation requires ordinal data. Examples of ordinal data are:
1st, 2nd, 3rd,
Small, Medium, Large, XL,
Strongly agree, Agree, Neutral, Disagree, Strongly Disagree
Very often, Often, Not Often, Not at all



A Spearman's rank correlation between variables X and Y is calculated by

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

Where d is the difference between ranks.

Example: The scores for nine students in physics and math are as follows:
Physics: 35, 23, 47, 17, 10, 43, 9, 6, 28
Mathematics: 30, 33, 45, 23, 8, 49, 12, 4, 31
Compute the student's ranks in the two subjects and compute the Spearman rank correlation.

Solution:

Step 1: Find the ranks for each individual subject: order the scores from greatest to smallest; assign the rank 1 to the highest score, 2 to the next highest and so on:

Step 2: Add a fourth column, d, to your data. The d is the difference between ranks. For example, the first student's physics rank is 3 and math rank is 5, so the difference is -2 points. In a sixth column, square your d values.

Physics(X)	Rank (X)	Math(Y)	Rank (Y)	d= (rank x – rank y)	d^2
35	3	30	5	-2	4
23	5	33	3	2	4
47	1	45	2	-1	1
17	6	23	6	0	0
10	7	8	8	-1	1
43	2	49	1	1	1
9	8	12	7	1	1
6	9	4	9	0	0
28	4	31	4	0	0
					$\sum$=12

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

$$\rho = 1 - \frac{6 * 12}{9(81 - 1)}$$

$\rho = 0.9$, very strong correlation.

Note: Compared to the Pearson correlation coefficient, the Spearman correlation does not require continuous-level data (interval or ratio), because it uses ranks instead of assumptions about the distributions of the two variables. This allows us to analyze the association between variables of ordinal measurement levels. Moreover, the Spearman correlation does not assume that the variables are normally distributed. A Spearman correlation analysis can therefore be used in many cases in which the assumptions of the Pearson correlation (continuous-level variables, linearity, and normality) are not met.