Assumptions of Spearman's Correlation / Spearman's Rho

→ It is also known as Rank Correlation.
→ It was propounded by Charles Edward Spearman (1904)

→ In this method we give ranks to particular variables.

→ This method is used when there is no normal distribution.

→ Another assumption is that it is used where the data is qualitative; eg like the ~~beauty~~ beauty of a woman beauty can not be measured with actual figures and so we need to give preferences or orders/ranks for it (like it is less beautiful or more beautiful).

(Formula:)

$$R = 1 - \frac{6\Sigma D^2}{N^3 - N} \quad \text{or} \quad \frac{1 - 6\Sigma D^2}{N(N^2 - 1)}$$

* Spearman's rho is non parametric.

* There are no distributional assumptions

* Remember it only has to do with the bivariate pair (X, Y) that you are calculating correlations for.

* It is also called Rank Correlation which measures the degree of association between two variables.

* Spearman rank correlation test doesn't 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.

M T W T F S S
Date __/__/__

☆ Spearman correlation is often used to evaluate relationships involving ordinal variables. For example, you might use a Spearman correlation to evaluate whether the order in which employees complete a test exercise is related to the number of months they have been employed.

☆ The Spearman correlation evaluates the monotonic relationship between two continuous or ordinal variables.

In a monotonic relationship, the variables tend to change together, but not necessarily at a constant rate.

☆ This type of correlation coefficient is based on the ranked values for each variable rather than the raw data.

S M T W T F S S
Date __/__/__

☆ The Spearman Correlation is less sensitive than the Pearson Correlation to strong outliers that are in the tails of both samples. That is because Spearman's ρ limits the outlier to the value of its rank.

☆ Intuitively, the Spearman Correlation between two variables will be high when observations have a similar {or identical for a Correlation of 1} rank (i.e relative position label of the observations within the variable: 1st, 2nd, 3rd etc) between the two variables, and low when observations have a dissimilar {or fully opposed for a correlation of -1} rank between the two variables.

M T W T F S S
Date __/__/__

Interpretation of Correlation coefficient.

The commonly interpretable range of coefficient of correlation is :-

Value of Correlation Coefficient	Interpretation
0 {Zero}	Zero Correlation
0.01 to ± 0.20	Negligible relationship
0.21 to ± 0.40	Low correlation

0.41 to ± 0.70	Moderate Correlation
0.71 to ± 0.90	High Correlation
0.91 to ± 0.99	Very high Correlation
± 1	Perfect Correlation