**Pearson’s correlation**

**Correlation coefficient**

Pearson’s correlation coefficient is a statistical measure of the strength of a linear relationship between paired data. In a sample it is denoted by r and is by design constrained as follows

**-1 <= r => 1**

Furthermore:

* Positive values denote positive linear correlation;
* Negative values denote negative linear correlation;
* A value of 0 denotes no linear correlation;
* The closer the value is to 1 or –1, the stronger the linear correlation.

Note:

* The correlation coefficient does not relate to the gradient beyond sharing its **+ve** or **–ve** sign!
* The correlation coefficient is a measure of **r = 0** linear relationship and thus a value of does not imply there is no relationship between the variables. For example in the following scatterplot which implies no (linear) correlation however there is a perfect quadratic relationship:

Correlation is an effect size and so we can verbally describe the strength of the correlation using the guide that Evans (1996) suggests for the absolute value of r:

* .00-.19 “very weak”
* .20-.39 “weak”
* .40-.59 “moderate”
* .60-.79 “strong”
* .80-1.0 “very strong”

**Assumptions**

The calculation of Pearson’s correlation coefficient and subsequent significance testing of it requires the following data assumptions to hold:

* Interval or ratio level;
* Linearly related;
* Bivariate normally distributed.

In practice the last assumption is checked by requiring both variables to be individually normally distributed (which is a by-product consequence of bivariate normality). Pragmatically Pearson’s correlation coefficient is sensitive to skewed distributions and outliers, thus if we do not have these conditions we are content.

If your data does not meet the above assumptions then use Spearman’s rank correlation!

**Caution –** the existence of a strong correlation does not imply a causal link between the variables.

* **Reference -** <http://www.statstutor.ac.uk/resources/uploaded/pearsons.pdf>

**Spearman’s correlation**

**Assumption**

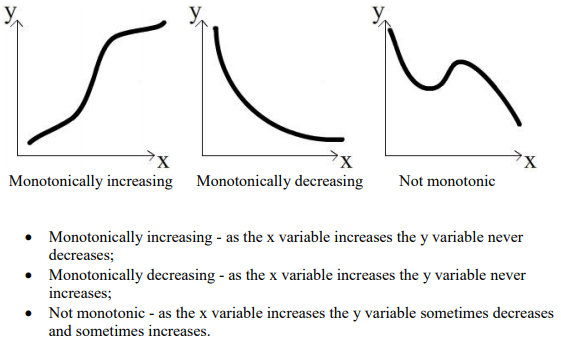
Its calculation and subsequent significance testing of it requires the following data assumptions to hold:

* Interval or ratio level;
* Linearly related;
* Bivariate normally distributed.

If your data does not meet the above assumptions then use Spearman’s rank correlation!

**Monotonic function**

A monotonic function is one that either never increases or never decreases as its independent variable increases.



**Spearman’s correlation coefficient**

Spearman’s correlation coefficient is a statistical measure of the strength of a monotonic relationship between paired data. In a sample it is denoted by and is by design constrained as follows:

**-1 <= rs => 1**

And its interpretation is similar to that of Pearson’s, e.g. the closer is to the stronger the monotonic relationship. 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:

* .00-.19 “very weak”
* .20-.39 “weak”
* .40-.59 “moderate”
* .60-.79 “strong”
* .80-1.0 “very strong”

The calculation of Spearman’s correlation coefficient and subsequent significance testing of it requires the following data assumptions to hold:

* interval or ratio level or ordinal;
* Monotonically related. Note, unlike Pearson’s correlation, there is no requirement of normality and hence it is a nonparametric statistic.
* **Reference -** <http://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf>