Pearson Correlation Coefficient

The Pearson Correlation Coefficient is a very helpful statistical formula that measures the strength between variables and relationships. In the field of statistics, this formula is often referred to as the Pearson R test. When conducting a statistical test between two variables, it is a good idea to conduct a Pearson Correlation Coefficient value to determine just how strong that relationship is between those two variables.

Formula

In order to determine how strong the relationship is between two variables, a formula must be followed to produce what is referred to as the coefficient value. The coefficient value can range between -1.00 and 1.00. If the coefficient value is in the negative range, then that means the relationship between the variables is negatively correlated, or as one value increases, the other decreases. If the value is in the positive range, then that means the relationship between the variables is positively correlated, or both values increase or decrease together. Let's look at the formula for conducting the Pearson Correlation Coefficient value.

Step 1: First, you make a chart with your data for two variables, labeling the variables (x) and (y). Then you must add 3 more columns labeled (xy), (x^2), and (y^2).

Source:

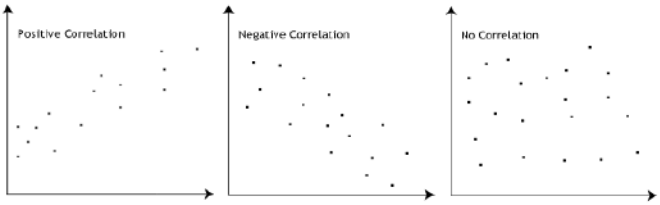
Pearson Product-Moment Correlation

What does this test do?

The Pearson product-moment correlation coefficient (or Pearson correlation coefficient, for short) is a measure of the strength of a linear association between two variables and is denoted by *r*. Basically, a Pearson product-moment correlation attempts to draw a line of best fit through the data of two variables, and the Pearson correlation coefficient, *r*, indicates how far away all these data points are to this line of best fit (how well the data points fit this new model/line of best fit).

What values can the Pearson correlation coefficient take?

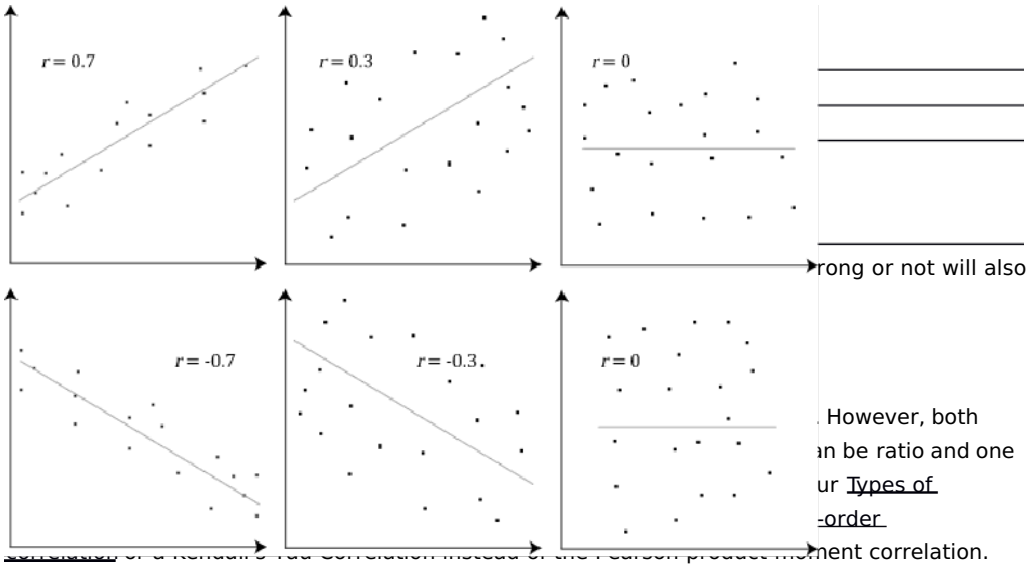
Th om +1 to -1. A value of 0
in value greater than 0 indicates
a s, so does the value of the
ot ; that is, as the value of one
va



How can we determine the strength of association based on the Pearson correlation coefficient?

The stronger the association of the two variables, the closer the Pearson correlation coefficient, *r*, will be to either +1 or -1 depending on whether the relationship is positive or negative, respectively. Achieving a value of +1 or -1 means that all your data points are included on the line of best fit - there are no data points that show any variation away from this line. Values for *r* between +1 and -1 (for example, *r* = 0.8 or -0.4) indicate that there is variation around the line of best fit. The closer the value of *r* to 0 the greater the variation around the line of best fit.

Are there guidelines to interpreting Pearson's correlation coefficient?



rong or not will also

. However, both
an be ratio and one
ur Types of
-order
correlation.

Do the two variables have to be measured in the same units?

No, the two variables can be measured in entirely different units. For example, you could correlate a person's age with their blood sugar levels. Here, the units are completely different; age is measured in years and blood sugar level measured in mmol/L (a measure of concentration). Indeed, the calculations for Pearson's correlation coefficient were designed such that the units of measurement do not affect the calculation. This allows the correlation coefficient to be comparable and not influenced by the units of the variables used.

What about dependent and independent variables?

The Pearson product-moment correlation does not take into consideration whether a variable has been classified as a dependent or independent variable. It treats all variables equally. For

example, you might want to find out whether basketball performance is correlated to a person's height. You might, therefore, plot a graph of performance against height and calculate the Pearson correlation coefficient. Lets say, for example, that $r = .67$. That is, as height increases so does basketball performance. This makes sense. However, if we plotted the variables the other way around and wanted to determine whether a person's height was determined by their basketball performance (which makes no sense), we would still get $r = .67$. This is because the Pearson correlation coefficient makes no account of any theory behind why you chose the two variables to compare.

Does the Pearson correlation coefficient indicate the slope of the line?

It is important to realize that the Pearson correlation coefficient, r, does not represent the slope of the line of best fit. Therefore, if you get a Pearson correlation coefficient of +1 this does not mean that for every unit increase in one variable there is a unit increase in another. It simply means that there is no variation between the data points and the line of best fit.

What assumptions does Pearson's correlation make?

There are five assumptions that are made with respect to Pearson's correlation:

1. The variables must be either interval or ratio measurements (see our Types of Variable guide for further details).

2. The variables must be approximately normally distributed (see our Testing for Normality guide for further details).

3. There is a linear relationship between the two variables. We discuss this later in this guide (jump to this section here).

4. Outliers are either kept to a minimum or are removed entirely. We also discuss this later in this guide (jump to this section here).

5. There is homoscedasticity of the data. This is discussed later in this guide (jump to this section here).

How can you detect a linear relationship?

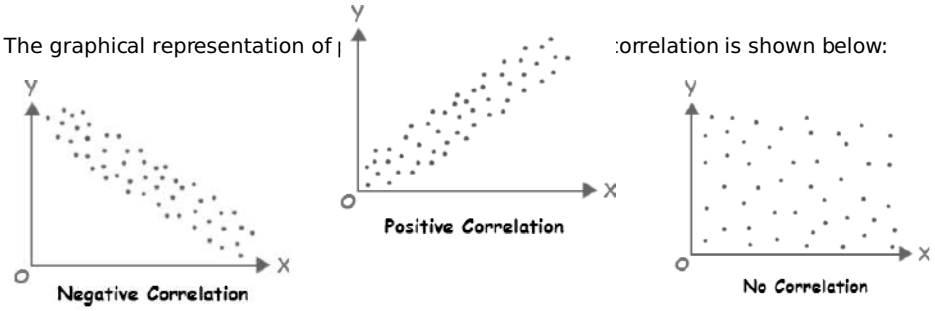
To test to see whether your two variables form a linear relationship you simply need to plot them on a graph (a scatterplot, for example) and visually inspect the graph's shape. In the diagram below (click image to enlarge), you will find a few different examples of a linear relationship and some non-linear relationships. It is not appropriate to analyse a non-linear relationship using a Pearson product-moment correlation.

Source: https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide-2.php

Pearson Correlation Formula

Correlation is the relationship between two variables. Correlation coefficient is the measurement of correlation. It indicates how well the two set of data are interconnected. Pearson correlation coefficient measures the linear dependence of two variables upon each other. It is also referred as Pearson product-moment correlation coefficient. The value of Pearson correlation coefficient lies between -1 to +1. If the coefficient of correlation is zero, then there is no correlation between given two variables. On the other hand, the perfectly positive correlation has a value of +1, while a perfectly negative correlation has a value of -1.

The graphical representation of p :orrelation is shown below:



Pearson correlation coefficient for sample data is denoted by "r". The formula for Pearson correlation coefficient r is given by:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{\left[n\sum x^2 - (\sum x)^2\right]\left[n\sum y^2 - (\sum y)^2\right]}}$$

Where,
r = Pearson correlation coefficient
x = Values in first set of data
y = Values in second set of data
n = Total number of values.

Source: http://formulas.tutorvista.com/math/pearson-correlation-formula.html