Correlation vs regression

Both of these terms of statistics that are used to measure and analyze the connections between 2 different variables and used to make prediction.

# What is correlation?

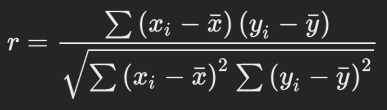
In statistics, correlation or dependency is any statistical relationship, whether causal or not, between two random variables or bivariate data.

Any change in one variable directly relates to a change in the other.

## The formula of correlation

A correlation coefficient is a numerical measure of some type of correlation, meaning a statistical relation ship between two variables. The variables may be two columns of a given data set of observations, often called a sample, or two components of a multivariate random variable with a known distribution.

A correlation coefficient is a number between -1 and +1 calculated so as to represent the linear dependence of two variables or sets of data.



* R = correlation coefficient
* Xi = values of the x-variable in a sample
* X bar = the mean(average) of the values of the x-variable
* Yi = values of the y-variable in a sample
* Y bar = the mean(average) of the values of the y-variable

There are a few common types of tests to measure correlation, these are:

* Pearson
* Spearman rank
* Kendal Tau.

Each have their own assumptions about the data set that needs to be met in order for the test to be able to accurately measure the level of correlation.

When you sample contains sufficient evidence, you can reject the null and conclude that the effect is statistically significant. Statisticians often denote the null hypothesis as H0 or Ha

H0 is similar to a default theory about the data. It does not exist until it is proven to exist. Innocent until proven guilty as they say in law trials

* H0 or null hypothesis
  + There is no difference between groups or no relationship between variable. It is one of two mutually exclusive hypotheses about a population in a hypothesis test
* Alternative hypothesis or Ha
  + The effect exists in the population