Correlation Analysis

By Nilosree Sengupta,
Coursera Project Network

Correlation refers to the statistical concept that studies the relationship between 2 quantitative variables.

Here we are going to do the **Pearson's product moment correlation**.This is the most popular correlation that is used to find the degree/strength of association of relationship between 2 variables.

Here are some important parameters/concepts of Correlation which **you need to know, when you start with Correlation Analysis, i.e. before you start with task 9**.

1. **r** : It is the Correlation Coefficient.It refers to the level of association between 2 variables from a given data.
    - It shows the association of points of a scatter plot to linear regression from those points.
    - The range of r is from -1 to +1.

Value of r	Meaning
+1	Perfectly linear positive correlation.i.e. Perfectly positive association
0	Very weak and No association
-1	Perfectly linear negative correlation.i.e. Perfectly negative association

    - The closer the value of r towards +1, stronger is the positive association
    - The closer the value of r towards -1, stronger is the negative association
    - When the value of r is close to 0, it means very weak association or no association at all.

2. **p-value**: It is a statistical measure that determines whether a null hypothesis is to be accepted or rejected.

p-value	Significance
When the p-value <= Critical Value(alpha)	It signifies to reject the hypothesis.
When the p-value > Critical Value(alpha)	It signifies to support the hypothesis.

    - Critical value is generally considered to be **0.05**.
    - Lesser the p-value, greater is the chances for an alternative hypothesis test.i.e.greater tendencies to reject the null hypothesis and vice versa.

3. **t** : This is the t-test statistical measure.It is used to understand the statistical difference between 2 different groups from sample data.

4. **df** :This is the degrees of freedom that refers to the maximum number of values which have the freedom to vary.

5. **Samples estimates** : It refers to the correlation coefficient(r).

6. **Correlation Matrix** : The matrix table that shows the correlation.