**Dhir Thacker | 17070122019 | CSE-1**

**SD lab assignment 7 (R-assignment 3)**

**Q1) What is the need of correlation analysis?**

A1) Correlation analysis is used to determine and evaluate the relationship between two or more variables. In terms of the strength of relationship, the value of the correlation coefficient varies between +1 and -1 where +1 indicates perfect positive relation and -1 for negative or inverse relation. Pearson correlation, Kendall rank correlation, Spearman correlation, and the PointBiserial correlation are a few examples of correlation analysis techniques

**Q2) Discuss one real world scenario where correlation helps to take decision.**

A2) Take an example of a celestial body like a Star. We can correlate its size and the gravitational force it exhibits on nearby objects by calculating the correlation coefficient. In this case, it is obvious to expect a coefficient that tends to +1. A negative coefficient can be seen by considering the distance from the star, the greater the distance, lesser the gravitational force.

**Q3) Write about function in R to compute correlation.**

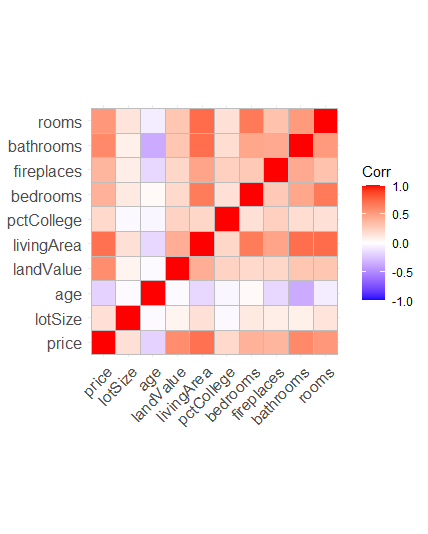
A3) A simple cor(x) can be used which returns the correlation coefficient. This function also allows us to choose the method of correlation as well as an option to handle missing data.

To test the relation between pairs, cor.test() is used. It returns both the correlation coefficient and the significance level (p-value) of the correlation.

**Q4) Which kind of visualization (chart/graph) is suitable to represent correlation analysis graphically.**

A4) A Scatter Chart is commonly used to visualize the correlation and distribution between two variables. These charts can show clusters that can help in detecting outliers.

Another widely used map is the Heat Map. This is more suitable when we want to visualize the correlation between multiple variables within a single plot



Q.5 Write a note on regression analysis.

A5) Regression analysis is a statistical method for estimating the relationships between variables in a data set. A dependent variable and one or more independent variables are required to perform the analysis. Generally, a dependent variable is what we wish to predict based on the values/parameters provided by the independent variables. Using this, a linear regression model can be generated which can help in predicting the outcome based on independent variable parameters.

There are 3 main types of regression analysis –

1. **Linear**

Dependent and a single independent variable show a linear relationship

1. **Multilinear**

Similar to linear, but the difference here is that multiple independent variables are involved

1. **Non-linear**

Commonly used for complicated datasets where the variables show a nonlinear. Unlike linear models which show a simple curve line, non-linear models can have a variety of curve lines indicating the complexity of the data.

Correlation analysis in R

