1. **What are the assumptions of linear regression regarding residuals?**

Few assumptions for liner regression residuals are

- There should be a linear relationship
- They are not multicollinear
- They are not Corelated

2. **What is the coefficient of correlation and the coefficient of determination?**

Correlation Coefficient: The value lies between -1 to 1, where, -1 indicates strongly inversely corelated. Which mean if one variable increases, the other variable decreases. Similarly, 1 indicates, strong directly corelated, which means, if one variable increases, other variable also increases.

It is denoted by the term R.

Determination coefficient: It is denoted by the term R2. In other words, Coefficient of Determination is the square of Coefficeint of Correlation.

3. **Explain the Anscombe's quartet in detail.**

Anscombe's quartet has 4 sets which has simple statistics in common, yet have different distributions. All 4 have different shapes when put in graphs. Each data set has 11 points which are intended to counter the impression among statisticians that "numerical calculations are exact, but graphs are rough

4. **What is Pearson's R?**

The term is used to define Pearson correlation coefficient. It is a measure of the linear correlation between two variables X and Y. Its values lies between -1 to 1, just like correlation as mentioned above. Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations

5. **What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?**

Scaaling is performed before doing any multi Liner regression model, to bring in all variables into same scale for a better model bulding.

Normalization rescales the values into a range of [0,1]. This might be useful in some cases where all parameters need to have the same positive scale. However, the outliers from the data set are lost.

Standardization rescales data to have a mean (µ) of 0 and standard deviation (σ) of 1 (unit variance).

6. **You might have observed that sometimes the value of VIF is infinite. Why does this happen?**

VIF show the degree to which a regression coefficient will be affected because of the variable's redundancy with other independent variables. As the squared multiple correlation of any predictor variable with the other predictors approaches unity, the corresponding VIF becomes infinite.