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

:

- Mean of all residual should be zero
- Residual should neither be correlated with independent variable or with each other
- Standard Deviation of the residuals should be constant (Homoscedasticity)
- Residual should be normally distributed

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

Correlation coefficient is technique of investigating the relationship between two quantitative, continuous variable for eg age and blood pressure. Statistical measure of the linear relationship (correlation) between a dependent-variable and an independent variable. Represented by the lowercase letter 'r', its value varies between -1 and 1 : 1 means perfect correlation, 0 means no correlation, positive values means the relationship is positive (when one goes up so does the other), negative values mean the relationship is negative (when one goes up the other goes down). Also called correlation coefficient

Coefficient of determination (r^2) is how much variance in an DV is explained by your IV.it measure the percentage of variability with in y-values that can be explained by regression model Therefore, values close to 100% means the model is useful and a value close to zero mean model is not useful.

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

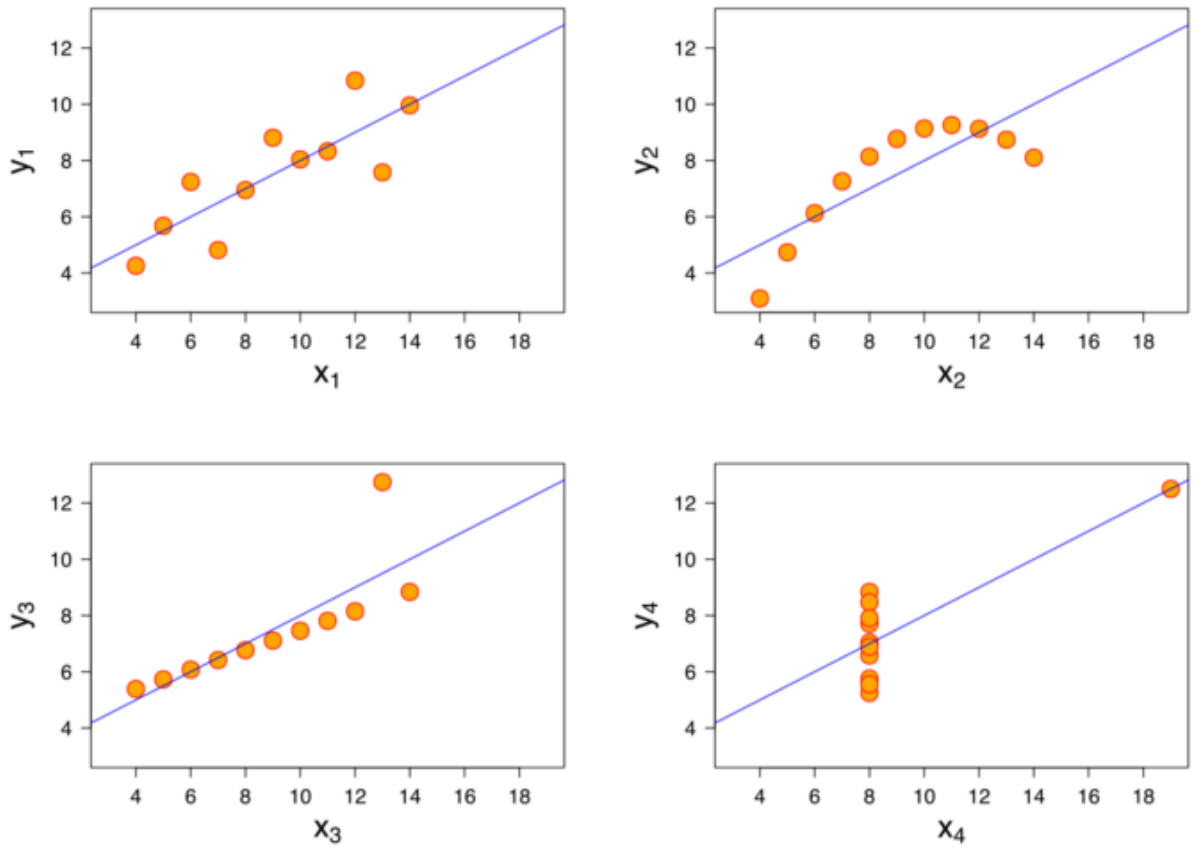
Anscombe's Quartet was developed by statistician Francis Anscombe. It comprises four datasets, each containing eleven (x,y) pairs. The essential thing to note about these datasets is that they share the same descriptive statistics. But things change completely, and I must emphasize **COMPLETELY,** when they are graphed. Each graph tells a different story irrespective of their similar summary statistics.

	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

The summary statistics show that the means and the variances were identical for x and y across the groups :

- Mean of x is 9 and mean of y is 7.50 for each dataset.
- Similarly, the variance of x is 11 and variance of y is 4.13 for each dataset
- The correlation coefficient (how strong a relationship is between two variables) between x and y is 0.816 for each dataset

When we plot these four datasets on an x/y coordinate plane, we can observe that they show the same regression lines as well but each dataset is telling a different story :



- Dataset I appears to have clean and well-fitting linear models.
- Dataset II is not distributed normally.
- In Dataset III the distribution is linear, but the calculated regression is thrown off by an outlier.

- Dataset IV shows that one outlier is enough to produce a high correlation coefficient.

This quartet emphasizes the importance of visualization in Data Analysis. Looking at the data reveals a lot of the structure and a clear picture of the dataset.

A computer should make both calculations and graph. Both sorts of output should be studied; each will contribute to understanding.

4. What is Pearson's R?

The Pearson correlation coefficient, also called Pearson's R, is a statistical calculation of the strength of two variables' relationships. In other words, it's a measurement of how dependent two variables are on one another.

The Pearson product-moment correlation coefficient depicts the extent that a change in one variable affects another variable. This relationship is measured by calculating the slope of the variables' linear regression.

The value of Person r can only take values ranging from +1 to -1 (both values inclusive). If the value of r is zero, there is no correlation between the variables.

If the value of r is greater than zero, there is a positive or direct correlation between the variables. Thus, a decrease in first variable will result in a decrease in the second variable.

If the value of r is less than zero, there is a negative or inverse correlation. Thus, a decrease in the first variable will result in an increase in the second variable.

When plotted on a diagram, a positive correlation will see a line which slopes downwards from left to right and a negative correlation will see a line which slopes downwards from right to left.

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

Feature scaling is a method used to standardize the range of independent variables or features of data
Since the range of values of raw data varies widely, in some machine learning algorithms, objective functions will not work properly without normalization. For example, the majority of classifiers calculate the distance between two points by the distance. If one of the features has a broad range of values, the distance will be governed by this particular feature. Therefore, the range of all features should be normalized so that each feature contributes approximately proportionately to the final distance.

Another reason why feature scaling is applied is that gradient descent converges much faster with feature scaling than without it.

Both Normalization and Standardization are the feature scaling techniques which help in dealing with variables of different units and scales. This is a very important step in the data preprocessing and data wrangling.

For example, consider an Employee dataset. It contains features like Employee Age and Employee Salary. Now Age feature contains values on the scale 22-60 and Salary contains values on the scale 10000-100000. As these two features are different in scale, these need to be normalized and standardized to have common scale while building any Machine Learning model. Some algorithms have this feature built-in, but for some algorithms you must do it.

Normalization

Normalization scales the values of a feature into a range of [0,1].

Xnew = (X – Xmin) / (Xmax – Xmin)

It will be useful when we are sure enough that there are no anomalies (i.e. outliers) with extremely large or small values. For example, in a recommendation system, the ratings made by users are limited to a small finite set like {1, 2, 3, 4, 5}

Standardization

Standardization refer to the subtraction of the mean (μ) and then dividing by its standard deviation (σ). Standardization transforms your data such that the resulting distribution has a mean of 0 and a standard deviation of 1.

Xnew = (X - μ) / σ

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

The infinity value of VIFs mean that there is perfect collinearity so if variable return infinite VIF value then it that are exactly collinear or linear transformations of each other.