**1. What is a Linear Regression?**

In simple terms, linear regression is adopting a linear approach to modeling the relationship between a dependent variable (scalar response) and one or more independent variables (explanatory variables). In case you have one explanatory variable, you call it a simple linear regression. In case you have more than one independent variable, you refer to the process as multiple linear regressions.

**2. Can you list out the critical**[**assumptions of linear regression**](https://www.digitalvidya.com/blog/assumptions-of-linear-regression/)**?**

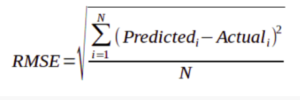
There are three crucial assumptions one has to make in linear regression. They are,

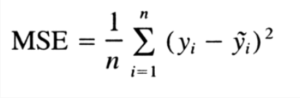
1. It is imperative to have a linear relationship between the dependent and independent A scatter plot can prove handy to check out this fact.
2. The independent variables in the dataset should not exhibit any multi-collinearity. In case they do, it should be at the barest minimum. There should be a restriction on their value depending on the domain requirement.
3. Homoscedasticity is one of the most critical: It states that there should be an equal distribution of errors.

### 4.    What is the primary difference between R square and adjusted R square?

In [linear regression](https://www.digitalvidya.com/blog/linear-regression/), you use both these values for model validation. However, there is a clear distinction between the two. R square accounts for the variation of all independent variables on the dependent variable. In other words, it considers each independent variable for explaining the variation. In the case of Adjusted R square, it accounts for the significant variables alone for indicating the percentage of variation in the model. By significant, we refer to the P values less than 0.05.

### 5.    Can you list out the formulas to find RMSE and MSE?





The most common measures of accuracy for any linear regression are RMSE and MSE. MSE stands for Mean Square Error whereas RMSE stands for Root Mean Square Error. The formulas of RMSE and MSE are as hereunder.

### 6. What is the difference between Mean Squared Error and Mean Absolute Error?

### The ****Mean Squared Error**** measures the variance of the residuals and is used when we want to punish the outliers in the dataset.

### The ****Mean Absolute Error**** measures the average of the residuals in the dataset. Is used when we don’t want outliers to play a big role. It can also be useful if we know that our distribution is multimodal, and it’s desirable to have predictions at one of the modes, rather than at the mean of them.

### 7.    Can you name a possible method of improving the accuracy of a linear regression model?

You can do so in many ways. One of the most common ways is ‘The Outlier Treatment.’

Outliers have great significance in [linear regression](https://www.digitalvidya.com/blog/what-is-regression/) because regression is very sensitive to outliers. Therefore, it becomes critical to treat outliers with appropriate values. It can also prove useful if you replace the values with mean, median, mode or percentile depending on the distribution.

### 8.    What are outliers? How do you detect and treat them?

An outlier is an observation point distant from other observations. It might be due to a variance in the measurement. It can also indicate an experimental error. Under such circumstances, you need to exclude the same from the data set. If you do not detect and treat them, they can cause problems in statistical analysis.

There is no strict mathematical calculation of how to determine an outlier. Deciding whether an observation is an outlier or not, is itself a subjective exercise. However, you can detect outliers through various methods. Some of them are graphical and are known as normal probability plots whereas some are model-based. You have some hybrid techniques such as Boxplots.

Once you have detected the outlier, you should either remove them or correct them to ensure accurate analysis. Some of the methods of eliminating outliers are the Z-Score and the IQR Score methods.

### \*9.    How do you interpret a Q-Q plot in a linear regression model?

As the name suggests, the Q-Q plot is a graphical plotting of the quantiles of two distributions with respect to each other. In other words, you plot quantiles against quantiles.

Whenever you interpret a Q-Q plot, you should concentrate on the ‘y = x’ line. You also call it the 45-degree line in statistics. It entails that each of your distributions has the same quantiles. In case you witness a deviation from this line, one of the distributions could be skewed when compared to the other.

### 10.    What is the importance of the F-test in a linear model?

The F-test is a crucial one in the sense that it tests the goodness of the model. When you reiterate the model to improve the accuracy with the changes, the F-test proves its utility in understanding the effect of the overall regression.

### 11.  What are the disadvantages of the linear regression model?

One of the most significant demerits of the linear model is that it is sensitive and dependent on the outliers. It can affect the overall result. Another notable demerit of the linear model is overfitting. Similarly, underfitting is also a significant disadvantage of the linear model.