##1. What is Null hypothesis?

##ANSWER: Statistical hypothesis tests are based a statement called the null hypothesis that assumes nothing interesting is going on between whatever variables you are testing. The exact form of the null hypothesis varies from one type test to another: if you are testing whether groups differ, the null hypothesis states that the groups are the same. For instance, if you wanted to test whether the average age of voters in your home state differs from the national average, the null hypothesis would be that there is no difference between the average ages.

##2. What is alternate hypothesis?

##ANSWER: The alternative hypothesis is the hypothesis used in hypothesis testing that is contrary to the null hypothesis. It is usually taken to be that the observations are the result of a real effect (with some amount of chance variation superposed)

##3. Define Type 1 and Type 2 errors.

##ANSWER:

A type I error (false-positive) occurs if an investigator rejects a null hypothesis that is actually true in the population; a type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population

##4. When to go for Anova instead of t-test or chi square test?

##ANSWER: Chi-square test is used on contingency tables and more appropriate when the variable you want to test across different groups is categorical. It compares observed with expected counts. Both t test and ANOVA are used to compare continuous variables across groups. t test is used for only two groups and it compares the means of the two groups. Whereas ANOVA can be used for more than two groups and it compares the variance between group with the average variance within all groups, i.e. it only identifies if there is any group that is different from the average of all other groups. Afterwards you should use post-hoc tests to identify pair-wise differences among all groups. When the variables you want to compare are not divided by groups but instead are both numerical you should the use correlation tests.

##5. What role does Anova play in a ML Project pipeline?

##ANSWER:

ANOVA is used when we want to compare the means of a condition between more than two groups. ANOVA tests if there is a difference in the mean somewhere in the model (testing if there was an overall effect), but it does not tell us where the difference is (if there is one). Analysis of variance (ANOVA) is a statistical technique used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples.