

1. Where you have used Hypothesis Testing in your Machine learning Solution.

Ans. When you have n set of features and you want to know if your assumptions of taking all features as independent. To test that all given features are independent or not we use Hypothesis testing. We can use OLS (Ordinary Least Square) library to see p-value, t-values, etc. and see if features that we are considering is independent features or not.

2. What kind of statistical tests you have performed in your ML Application.

Ans. We have performed following statistical tests:

- P-value
- T-test
- Chi Square
- Anova test
- Z-test
- F- test

3. What do you understand by P Value? And what is use of it in ML?

Ans.

- P value gives us probability of occurrence of data in given sample or population. Say p value if .95, means that the probability of occurring of a dataset is 95%.
- P-value plays a crucial role in selecting an independent feature. We can use OLS library to obtain values of all features in our dataset.

4. Which type of error is severe Error, Type 1 or Type 2? And why with example.

Ans. Best answer to this question is, it depends on problem statement. Below are two examples to assist my statement.

Null	Type 1 Error: H0 true, but rejected	Type 2 Error: H0 false, but not rejected
Medicine A does not relieve Condition B.	Medicine A does not relieve Condition B, but is not eliminated as a treatment option.	Medicine A relieves Condition B, but is eliminated as a treatment option.
Consequences	Patients with Condition B who receive Medicine A get no relief. They may experience worsening condition and/or side effects, up to and including death. Litigation possible.	A viable treatment remains unavailable to patients with Condition B. Development costs are lost. Profit potential is eliminated.

In above case Type 1 becomes a serious situation.

We commit a Type 1 error if we reject the null hypothesis when it is true. This is a false positive, like a fire alarm that rings when there's no fire.

A Type 2 error happens if we fail to reject the null when it is not true. This is a false negative—like an alarm that fails to sound when there is a fire.

In above case Type 2 becomes a serious situation.

5. Where we can use chi square and have used this test anywhere in your application?
Ans. A Chi-Square test is a test of statistical significance for categorical variables.
A research scholar is interested in the relationship between the placement of students in the statistics department of a reputed University and their C.G.P.A (their final assessment score). Above test is used in admission prediction dataset problem.
6. Can we use Chi square with Numerical dataset? If yes, give example. If no, give Reason?
Ans. No, Chi square cannot be used for numerical dataset. It measures the degree of association between two categorical variables. If both are numeric, we can use Pearson's product-moment correlation, and if the attribute is numerical and there are two classes we can use a t-test if more than two classes we can use ANOVA.
7. What do you understand by ANOVA Testing?
Ans. Analysis of variance, or ANOVA, is a statistical method that separates observed variance data into different components to use for additional tests
Analysis of variance (ANOVA) is a statistical technique that is 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.
8. Give me a scenario where you can use Z test and T test.
Ans. Below are two scenarios for z and t test:
 - z-test is used for it when sample size is large, generally $n > 30$.
 - t-test is used for hypothesis testing when sample size is small, usually $n < 30$ where n is used to quantify the sample size.
9. What do you understand by inferential Statistics?
Ans. In this type of stats, from a population or sample, a sample of data is taken and different experiment are performed on this data to present solutions of same. It includes doing hypothesis testing, p-value, t-test, f-test, etc.
10. When you are trying to calculate Std Deviation or Variance, why you used N-1 in Denominator? (Hint: Basel Connection)
Ans. SD: tells how far element is away from mean.
Variance: is used to measure spread of curve.

In Basel connection we use $(n-1)$ in denominator because it provided better result to scientist when they performed experiment using $(n-1)$ in denominator using Basel Connection.