HYPOTHESIS TESTING

▸ **Hypothesis** : A premise or claim that we want to test or quantitative statement about population.

▸ **Null Hypothesis(H0)** : It is a claim or statement about population parameter that is assumed to be true until it is declared to be false.

▸ **Alternative Hypothesis(H1)** : Any hypothesis which is complementary to H0.

▸ **Hypothesis Test** : In hypothesis test we evaluate two exclusive statements on population using a sample of data.

STEPS OF HYPOTHESIS TESTING

1. **Making Initial Assumption.**
   ▶ Example:

   H0(Null hypothesis) – The mean data scientist salary is 2,50,000 dollars.

   H1(Alternate hypothesis) – The mean data scientist salary is not 2,50,000 dollars.

2. **Collecting The Data** :
   ▶ Also called evidences.

3. **Analyze or test the collected data whether to reject the H0 or to accept the H0.**
   ▶ Means whether the mean data scientist salary is 2,50,000 dollars or not.
   ▶ If we reject H0 means, The mean data scientist salary is not 2,50,000 dollars and vise versa.

HOW TO PERFORM HYPOTHESIS TESTING

▶ When we have huge amount of data, we try to analyze that data. We consider sample data of the population and we perform some analysis and testing on that sample to get important information about the data and population data.

▶ To test for hypothesis we will have both (Null and Alternate) hypothesis.

▶ **Steps To Perform Hypothesis Test.**

1. **P** value.
2. **Chi Square** Test.
3. **T** test.
4. **Anova** Test

- P value – Significant value
- P value <= 0.05 (reject the null hypothesis)

- The P value can be obtained from
1. Chi square test
2. T test
3. Anova test

DATA SET :

Hired student gender	stream	CGPA	package
M	ECE	8.5	850k
F	CSE	9.0	800k
F	ME	7.0	510k
M	ECE	8.0	900k
M	ECE	7.5	550k

Case 1: 2 - Categorical Feature

▸ Type of test : CHI square test.

▸ If P <=0.05, then we reject H0 and accept H1.

Case 2 : 1 Categorical Feature and 1 Continuous Feature

▸ Type of test – T test.

▸ P value<= 0.05(reject H0).

Case 3 : Comparing Continuous and Categorical Feature

➤ ANOVA test :This test should be done when there is more than two categories in a column.

➤ P <= 0.05 (reject H0).

Need For Hypothesis Testing

▶ Hypothesis testing is an important procedure in statistics. Hypothesis testing evaluates two mutually exclusive population statements to determine which statement is most supported by sample data. When we say that the findings are statistically significant.

Parameters Of Hypothesis Testing

▶ NULL hypothesis : In statistics, it is a basic assumption based on the given problem information.

▶ It refers to the P-value in which we accept or reject the null – hypothesis.

P-value

▸ In statistics, the p − value is the probability that a random chance generated the data that is equal or rarer (under H0).We calculate the p-value for the sample mean in our case. It is said to be the significant value.

Errors In Hypothesis Testing

▸ Type 1 error : When we reject the null hypothesis , although that hypothesis was true. Type 1 error is denoted by ALPHA.

▸ Type 2 errors : When we accept the null hypothesis but it is false. Type 2 errors are denoted by BETA.