Things to Remember - Hypothesis Testing

Why Hypothesis Testing:

- Having done the summary statistics alone does not help in decision making. It is the interpretation of data which churn out meaning results from the data.
- Using **Hypothesis Testing**, we try to interpret or draw conclusions about the population using sample data.
- A **Hypothesis Test** evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data.

Terminologies to Understand

Population - Population: Population is a collection of objects that we want to study/test. The collection of objects could be Cities, Students, Factories, etc. It depends on the study at hand.

Parameter - is a summary description of a fixed characteristic or measure of the target population. A Parameter denotes the true value that would be obtained if a census rather than a sample were undertaken.

Sample - Sample is a subset of population.

Statistic - A Statistic is a summary description of a characteristic or measure of the sample. The Sample Statistic is used as an estimate of the population parameter. E.g. - Sample Mean, Sample variance, sample standard deviation.

Types of Hypothesis

Null Hypothesis - A statement in which no difference or effect is expected. If the null hypothesis is not rejected, no changes will be made.

Alternative Hypothesis - A statement that some difference or effect is expected. Accepting the alternative hypothesis will lead to changes in opinions or actions. It is the opposite of the null hypothesis.

All statistical conclusions are made in reference to the null hypothesis.

We either reject the null hypothesis or fail to reject the null hypothesis; we do not accept the null hypothesis. From the start, we assume the null hypothesis to be true, later the assumption is rejected, or we fail to reject it.

• When we reject the null hypothesis, we can conclude that the alternative hypothesis is supported.

• If we fail to reject the null hypothesis, it does not mean that we have proven the null hypothesis is true.

– Failure to reject the null hypothesis does not equate to proving that it is true.

– It just holds up our assumption or the status quo.

One tailed and two tailed tests

A one-tailed test is a statistical hypothesis test in which the critical area of a distribution is one-sided so that it is either greater than or less than a certain value, but not both. If the sample being tested falls into the one-sided critical area, the alternative hypothesis will be accepted instead of the null hypothesis A one-tailed test is also known as a directional hypothesis or directional test

Critical Region: The critical region is the region of values that corresponds to the rejection of the null hypothesis at some chosen probability level.

A two-tailed test is a method in which the critical area of a distribution is two-sided and tests whether a sample is greater than or less than a certain range of values. If the sample being tested falls into either of the critical areas, the alternative hypothesis is accepted instead of the null h hypothesis convention, two-tailed tests are used to determine significance at the 5% level, meaning each side of the distribution is cut at 2.5%.

Two-Tailed Versus One-Tailed Hyphothesis Tests

Figure A:
Two-Tailed Test

Figure B:
One-Tailed Test
(Left-Tailed Test)

2.5%   2.5%

5.0%

0   0
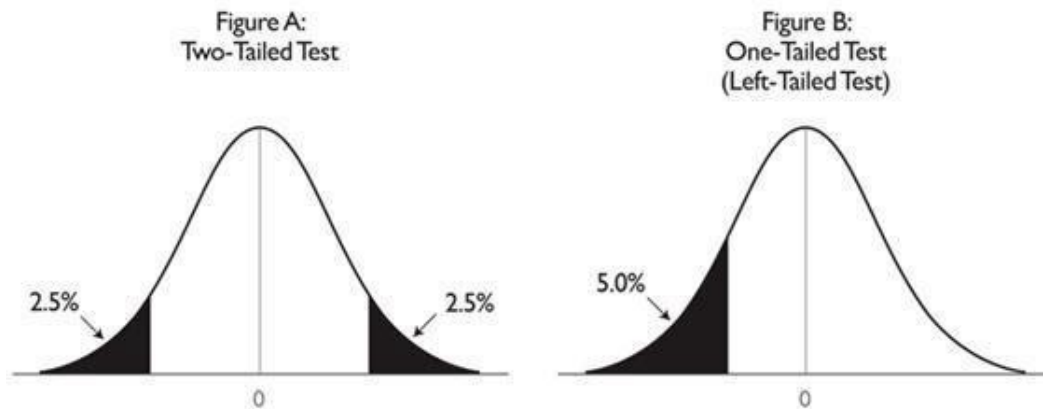
Types of error
Type 1 Error - If a true null hypothesis is rejected it is called a Type 1 - error while if a false null hypothesis is accepted by the test, this is called **Type 2 errors**

The probability of type 1 error is specified by α while the probability of type 2 error is represented by β. The complement of α is referred to as **level of significance** while the complement of β is called as **power of test.**

P Value approach to hypothesis testing
P-values evaluate how well the sample data support the argument that the null hypothesis is true. It measures how compatible your data are with the null hypothesis.

High P values - data is likely with a true null
Low p values - data is unlikely with true null

Summary steps for Hypothesis testing
  i.   Formulate the null and alternative hypothesis
  ii. Select the appropriate test
  iii. Choose the level of significance
  iv. Collect the data and calculate the test statistic
  v.   Determine the probability associated with test statistic
  vi. Determine the critical value associated with test statistic
  vii. Compare with level of significance
  viii. Determine if critical value of the test statistic falls into non - rejection region
  ix. Reject or do not reject the null hypothesis
  x.   Interpretation and conclusion of the business problem

Summary of different types of test

Tree Diagram (CI) Taxonomy of Inferential Techniques
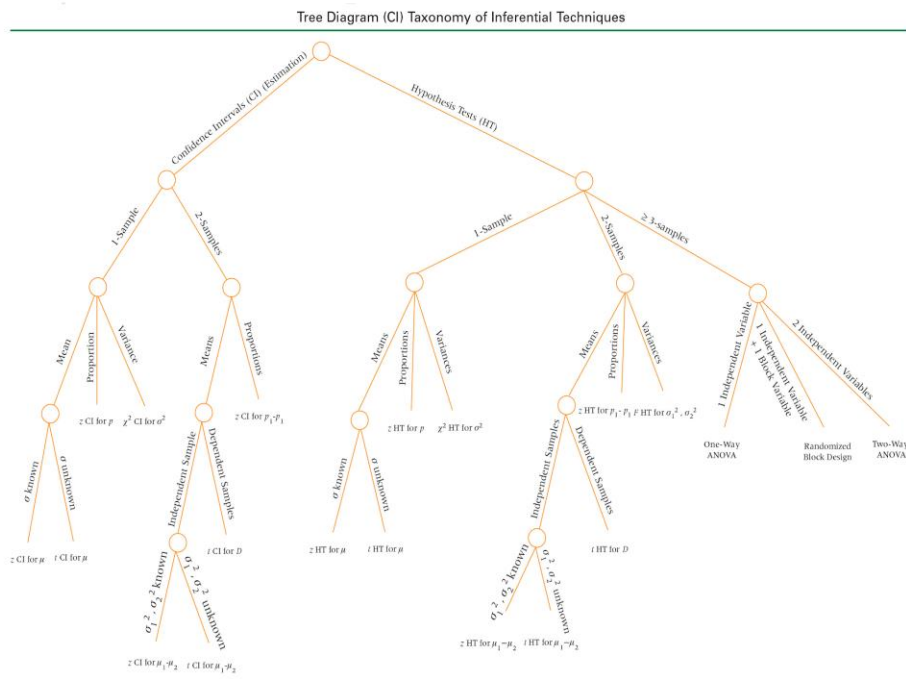
Test your understanding for formulating hypothesis

i. A study claims that mean income of the senior executives in the amazon is 25,00,000 per annum. To test this claim it is decided to take a sample of 300 executives and obtain their mean income. Set up the appropriate hypothesis.

ii. ICICI bank is experiencing an era of expansion. In considering whether or not to open a branch in an area, it considers the proportion of business coming from the area. For this purpose, it takes a sample of the business conducted on different days and uses the cutoff point as 5 percent, so that 5 percent or more business is coming from a particular area, it shall open a branch here. Formulate the appropriate null and alternative hypothesis for the decision.

iii. A random sample of 80 bank employees is taken to test the claim that the mean salary of the bank executives in a certain state is Rs 48,400 per month. Further, from a related study undertaken recently, it is known that the SD of the distribution of the salaries of the bank executives in the state is Rs 5870 and it is believed to be true. The sample has yielded an average monthly salary of Rs 47,456. Is this claim correct? Test at 1 percent level of significance.