
upGrad
Campus

upGrad

Course : Statistics for Business

Lecture On : Hypothesis Testing

Instructor : Santosh Kumar

upGrad

Session Agendas

In this session we will learn and revise the concepts of Hypothesis Testing which are among the most prominent notions in statistics for data science.

What is a Hypothesis?

Hypothesis: a claim or an assumption that you make about one or more population parameters that you need to validate.

There are 5 main steps in hypothesis testing:

- State your research hypothesis as a null (Ho) and alternate (Ha) hypothesis.
- Collect data in a way designed to test the hypothesis.
- Perform an appropriate statistical test.
- Decide whether to reject or fail to reject your null hypothesis.
- Present the findings in your results and discussion section.

Types of Hypothesis:

Null Hypothesis:

- Makes an assumption about the status quo
- Always contains the symbols ‘=’, ‘≤’ or ‘≥’

Alternate Hypothesis:

- Challenges and complements the null hypothesis
- Always contains the symbols ‘≠’, ‘<’ or ‘>’

p-value method for a right-tailed hypothesis test:

- Calculate the Z-score using the formula $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, where $\bar{X}$ is the sample mean, μ is the population mean, σ is the population standard deviation (which can be approximated to S, the sample standard deviation like you learnt during inferential statistics), and n is the sample size.

- From the Z-score, calculate the Z-value using the Z-table. Now, you won't be required to do this step as this is done automatically in tools like Excel but it's good to know the method.

- To get the final p-value, subtract the Z-value that you get from 1.

Types of Errors:

The two types of errors are:

- Type-I Error: α is the acceptable probability of making a Type I error (also called the significance level). Alternatively, (1 – α) is called the confidence level. This occurs when your null hypothesis is actually true but you reject it.

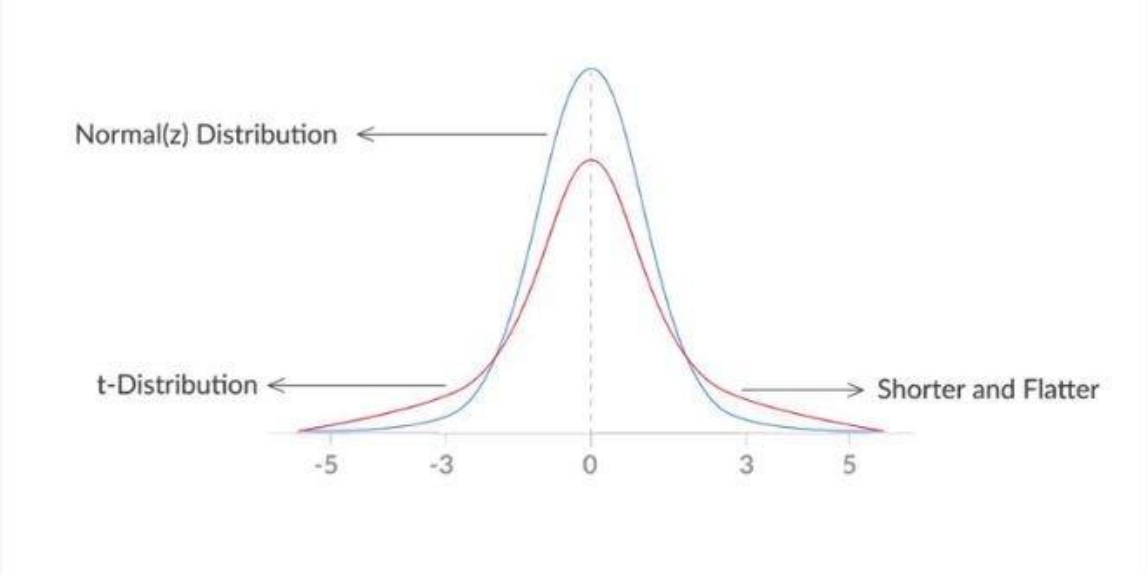
- Type-II Error: β is the probability of making a Type II error. Alternatively, (1 – β) is called the power of the test. This occurs when your alternate hypothesis is true but you still fail to reject your null hypothesis.

Basic steps you need to perform to evaluate any hypothesis:

- Calculate the value of Z-score for the sample mean point on the distribution
- Calculate the p-value from the cumulative probability for the given Z-score using the Z-table
- Make a decision on the basis of the p-value (multiply it by 2 for a two-tailed test) with respect to the given value of α (significance value).

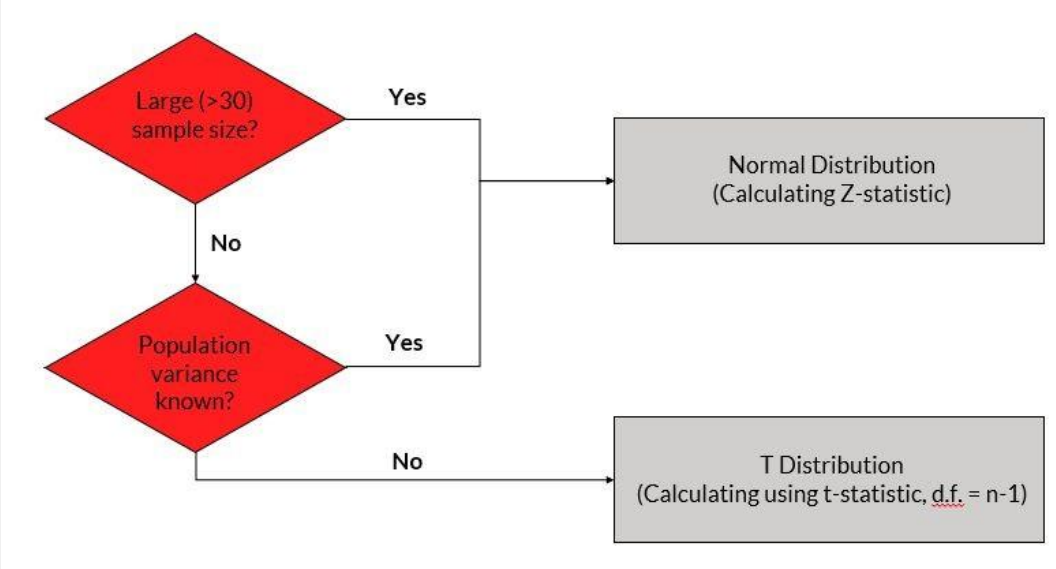
T-distribution:

- nothing but a shorter and flatter normal distribution

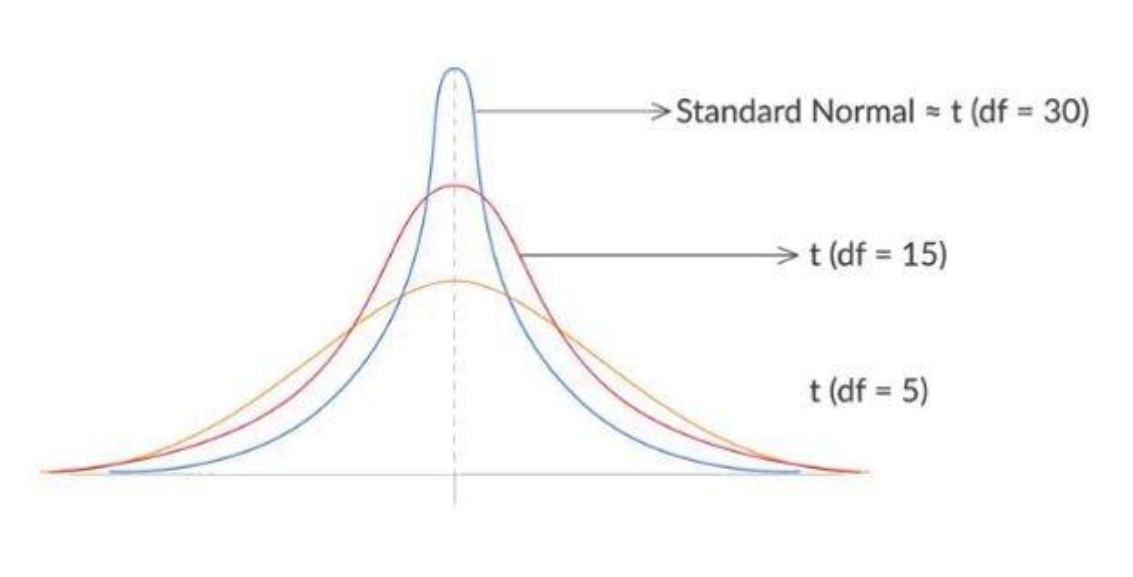


Important points:

- If your sample size is small (<30) and the population standard deviation is not known, you employ a T-test. In all other cases, you can simply go ahead and use a Z-test.



T distribution for different sample sizes:



Two-sample mean test - paired:

- It is used when your sample observations are from the same individual or object
- During this test, you are testing the same subject twice

Two-sample mean test - unpaired:

- During this test, you are not testing the same subject twice
- It is used when your sample observations are independent

Two-sample proportion test:

- It is used when your sample observations are categorical, with two categories
- It could be True/False, 1/0, Yes/No, Male/Female, Success/Failure, etc.

A/B Testing:

- A/B testing is a direct industry application of the two-sample proportion test
- It is a widely used process in digital companies in the e-commerce, manufacturing and advertising domains
- It provides a way to test two different versions of the same element and see which one performs better

Any Queries?

Thank You!