Lecture Notes

Hypothesis Testing

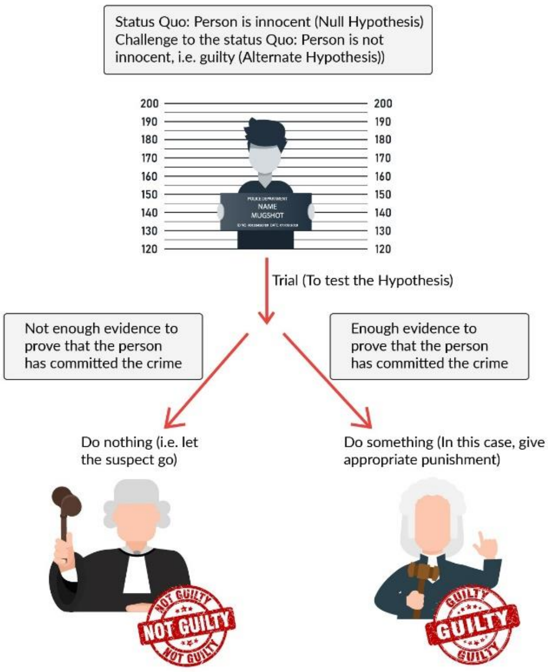
Framing a Hypothesis

The statistical analysis learnt in Inferential Statistics enable you to try making inferences about the population mean from the sample data when you have no idea about the population mean. However, sometimes you have some starting assumption about the population mean and you want to confirm those assumptions using the sample data. It is here that hypothesis testing comes into the picture.

A hypothesis is just a starting point that is open to a test and can be accepted or rejected in light of strong evidence. There are two terms that you came across:

1. Null Hypothesis (H0): The status quo.
2. Alternate Hypothesis (HA): The challenge to the status quo.

You also saw the example of the justice system to understand the null and the alternate hypothesis better.



Apart from that, you also learnt that the general convention for framing the hypothesis is -

- Null Hypothesis (H_0): Always contains an equality ($=, \leq, \geq$)
- Alternate Hypothesis (H_A): Always contains an inequality ($\neq, <, >$)

Hypothesis Testing Process

Next, you learnt about the five steps involved in any hypothesis test. These were:

1. Begin by assuming that H0, i.e. the status quo is true. In our case, the status quo is that the average spending is less than 120 since this does not require any action.
2. Put the onus on the data to contradict H0 beyond a **reasonable doubt**. In our case, the sample mean that we got was ₹130. So we need to prove that this mean is significantly different from ₹120. This means that using this sample mean of ₹130, we need to prove that the chances of the population mean being less than ₹120 is **very low**.
3. Thirdly, we also need to quantify what 'reasonable doubt' or 'very low' highlighted above are. For that, we use something known as a p-value.
4. Calculate the actual probability of observing the sample, i.e., calculate the p-value.
5. Conclude and take appropriate action, i.e., either reject or fail to reject the null hypothesis.

Left, Right, and Two-Tailed Tests

After formulating the null and alternate hypotheses, the steps to follow in order to **make a decision** using the **p-value method** are as follows:

1. Calculate the value of Z-score for the sample mean point on the distribution
2. Calculate the p-value from the cumulative probability for the given Z-score using the Z-table
3. 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).

To find the correct p-value from the z-score, first find the **cumulative probability** by simply looking at the z-table, which gives you the area under the curve till that point.

Situation 1: The sample mean is on the right side of the distribution mean, i.e., the Z-score is positive (Right-tailed test).

Example: z-score for sample point = + 3.02

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Hence, the cumulative probability of sample point = 0.9987

For one-tailed test → p-value = 1 - 0.9987 = 0.0013

For two-tailed test → p-value = 2 (1 - 0.9987) = 2 * 0.0013 = 0.0026

Situation 2: The sample mean is on the left side of the distribution mean, i.e., the Z-score is negative (Left-tailed test).

Example: Z-score for sample point = -3.02

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026

Hence, the cumulative probability of sample point = 0.0013

For one-tailed test → p = 0.0013

For two-tailed test → p = 2 * 0.0013 = 0.0026

Types of Errors in Hypothesis Testing

There are essentially two types of errors that you can commit while performing a hypothesis test.

Reality \ Decision	Do not reject H_0	Reject H_0
H_0 is True	Correct Decision	Type I Error (α)
H_a is True	Type II Error (β)	Correct Decision

These two errors are:

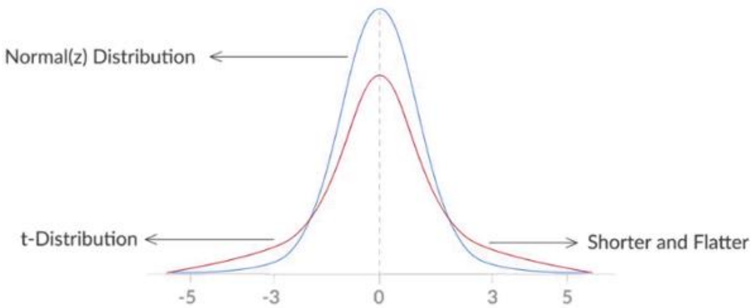
1. **Type-I Error:** α is the acceptable probability of making a Type I error (also called the significance level). Alternatively, (1 – α) is called the confidence level (Recall that you had learnt about confidence level in your inferential statistics module.). This occurs when your null hypothesis is actually true but you reject it.
2. **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.

Now let's take a look at the business implications of the type-I error. In this case, you're rejecting your null hypothesis even if it's true, i.e., you're accepting the alternate hypothesis and taking an action. But since this action requires time, money, and other resources, you cannot afford to make a mistake here very often which is why managers are more concerned with the type-I error.

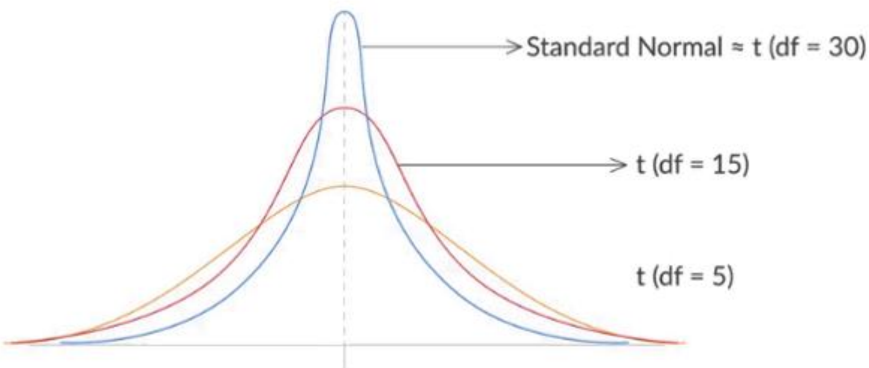
Insight Overview

So far you performed hypothesis testing using a Z-test, i.e., the normal distribution. But many a time it happens that either your sample size is small (<30), or the population parameters are not known. In such cases, you need to evaluate a hypothesis test using a T-distribution instead of a Z-distribution.

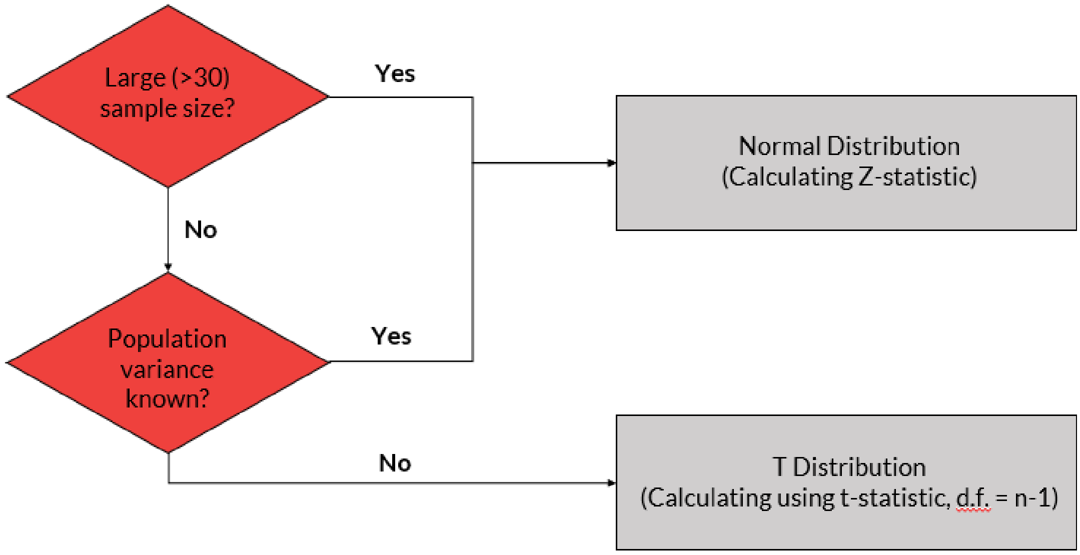
A T-distribution is not very different from a Z-distribution. It's just that the T-curve is shorter and flatter than the Z-curve.



You also learnt about **degrees of freedom** associated with the T-distribution. Degrees of freedom is nothing but the sample size minus 1. For example, **if your sample size is n, the degrees of freedom will be n-1**. You also saw that as you keep increasing the sample size, the T-distribution becomes narrower and steeper. In fact, for a large sample size (>30), it can be closely approximated to a normal distribution.



Finally, you learnt when you should use a Z-test and when you should use a T-test. Basically, 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. You can use the following flowchart to decide whether you should employ a T-test or a Z-test.



You should be able to:

- Frame null and alternate hypotheses
- Identify the type of test, i.e., left-tailed, right-tailed, or two-tailed, basis the hypotheses framed
- Evaluate the hypothesis using the 5-step process
- Identify whether to use a Z-test or a T-test based on the sample size and known parameters

Industry Demonstrations

- **Two-sample mean test - paired** is used when your sample observations are from the same individual or object. During this test, you are testing the same subject twice. For example, if you are testing a new drug, you would need to compare the sample before and after the drug is taken to see if the results are different.
- **Two-sample mean test - unpaired** is used when your sample observations are independent. During this test, you are not testing the same subject twice. For example, if you are testing a new drug, you would compare its effectiveness to that of the standard available drug. So, you would take a sample of patients who consumed the new drug and compare it with another sample who consumed the standard drug.
- **Two-sample proportion test** 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. For example, if you are comparing the effectiveness of two drugs, you would define the desired outcome of the drug as the success. So, you would take a sample of patients who consumed the new drug and record the number of successes and compare it with successes in another sample who consumed the standard drug. You also learnt how this test is the basis of A/B testing.

UpGrad