**Hypothesis Testing**

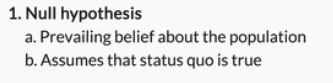
* Exploratory data analysis: Exploring data for insights and patterns
* Inferential statistics: Making inferences about the population using the sample data

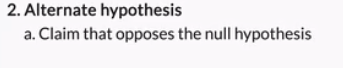
Now, these methods help you formulate a basic idea or conclusion about the population. Such assumptions are called “hypotheses”

Let’s understand the **basic difference between inferential statistics and hypothesis testing**.

**Inferential statistics** is used to find some population parameter (mostly population mean) when you have no initial number to start with. So, you start with the sampling activity and find out the sample mean. Then, you estimate the population mean from the sample mean using the confidence interval.

**Hypothesis testing** is used to confirm your conclusion (or hypothesis) about the population parameter (which you know from EDA or your intuition). Through hypothesis testing, you can determine whether there is enough evidence to conclude if the hypothesis about the population parameter is true or not.





The hypothesis you are trying to prove during hypothesis testing is the:

**Alternate Hypothesis**

**Feedback :***The null hypothesis is the status quo, i.e. the already existing situation. The alternate hypothesis is what you are trying to prove.*

**Null and Alternate Hypotheses**

Government regulatory bodies have specified that the maximum permissible amount of lead in any food product is **2.5 parts per million or 2.5 ppm**.

If you conduct tests on randomly chosen Maggi Noodles samples from the market, what would be the null hypothesis in this case?

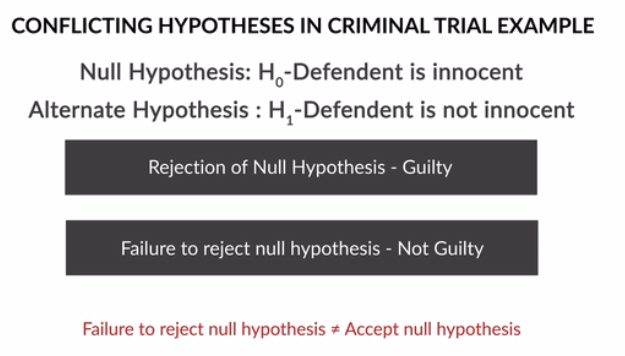
**The average lead content is less than or equal to 2.5 ppm**

**Feedback :***The null hypothesis is the status quo, i.e. the average lead content is within the allowed limit of 2.5 ppm.*

What would be the alternate hypothesis in the same case, when you conduct tests on randomly chosen Maggi Noodles samples from the market?

**The average lead content is more than 2.5 ppm**

**Feedback :***The alternate hypothesis is the opposite of the null hypothesis. Since the null hypothesis is that the average lead content is less than or equal to 2.5 ppm, the alternate hypothesis would be that the average lead content is more than 2.5 ppm.*



Hypothesis Testing starts with the formulation of these two hypotheses:

* **Null hypothesis** (H₀): The status quo
* **Alternate hypothesis** (H₁): The challenge to the status quo

In the Maggi Noodles example, if you **fail to reject the null hypothesis**, what can you conclude from this statement?

**Feedback :**The null hypothesis is that the average lead content is less than or equal to 2.5 ppm. Since you fail to reject the null hypothesis, you can conclude that Maggi Noodles do not contain excess lead. Please note than you can only fail to reject the null hypothesis, you can never accept the null hypothesis.

The null and alternative hypotheses divide all possibilities into:

**2 non-overlapping sets**

**Feedback :***Both the null and alternate hypotheses can’t be true at the same time. Only one of them will be true.*

The first step of hypothesis testing is the formulation of the null and alternate hypotheses for a given situation.

**Null and Alternate Hypotheses**

Flipkart claimed that its total valuation in December 2016 was $14 billion.

What would be the alternate hypothesis for the given situation?

**H₁: Total valuation ≠ $14 billion**

**Feedback :***The null hypothesis in this case would be that the total valuation is equal to $14 billion. Hence the alternate hypothesis is the opposite of that.*

You have seen examples where you can write the null hypothesis (or status quo) easily from the claim statement, like in the last question - Flipkart claimed that its total valuation in December 2016 was $14 billion.

But in some instances, if your claim statement has words like “at least”, “at most”, “less than”, or “greater than”, **you cannot formulate the null hypothesis just from the claim statement** (because it’s not necessary that the **claim is always about the status quo**).

You can use the following rule to formulate the null and alternate hypotheses:

**The null hypothesis** always has the following signs:  =  OR   ≤   OR    ≥

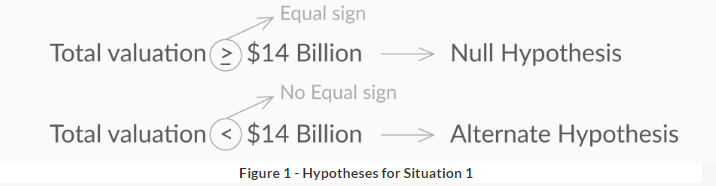
**The alternate hypothesis** always has the following signs:  ≠   OR  >   OR    <

For example:

**Situation 1:**  Flipkart claimed that its total valuation in December 2016 was at least $14 billion. Here, the claim contains ≥ sign (i.e. the at least sign), so **the null hypothesis is the original claim**.

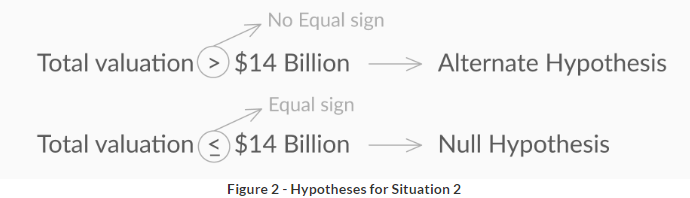
The hypothesis in this case can be formulated as:

**Figure 1 - Hypotheses for Situation 1**



**Situation 2:**  Flipkart claimed that its total valuation in December 2016 was greater than $14 billion. Here, the claim contains > sign (i.e. the ‘more than’ sign), so**the null hypothesis is the complement of the original claim**. The hypothesis in this case can be formulated as:

The hypothesis in this case can be formulated as:



**Null and Alternate Hypotheses**

The average commute time for an UpGrad employee to and from office is at least 35 minutes.

**H₀: μ ≥ 35 minutes and H₁: μ < 35 minutes**

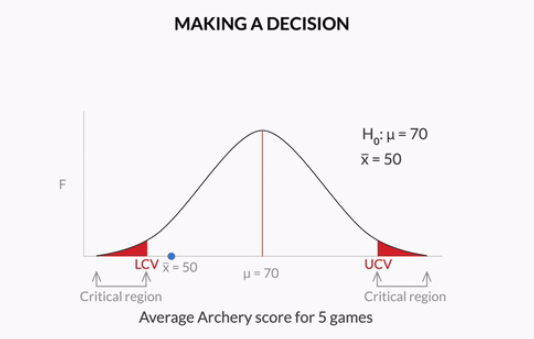
**Feedback :***The null hypothesis is always formulated by either = or ≤ or ≥ whereas the alternate hypothesis is formulated by ≠ or > or <. In this case, the average time taken was greater than or equal to 35 minutes. So, that becomes the null hypothesis. Less than 35 minutes becomes the alternate hypothesis.*

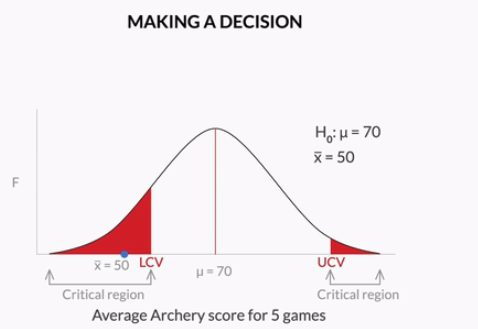
To summarize this, you cannot decide the status quo or formulate the null hypotheses from the claim statement, you need to take care of signs in writing the null hypothesis. Null Hypothesis never contains ≠ or > or < signs. It always has to be formulated using = or ≤ or ≥ signs.

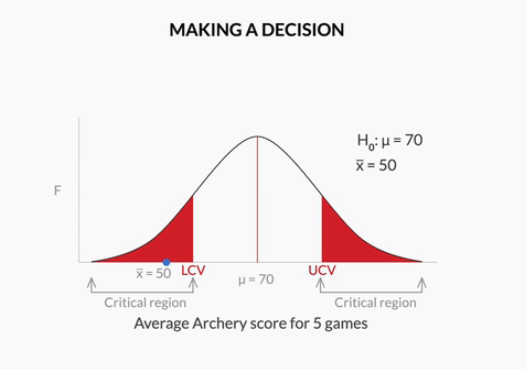
# Making a Decision

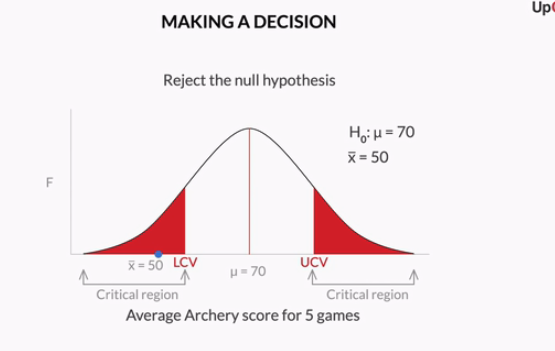
Once you have formulated the null and alternate hypotheses, let’s turn our attention to the most important step of hypothesis testing — **making the decision to either reject or fail to reject the null hypothesis**

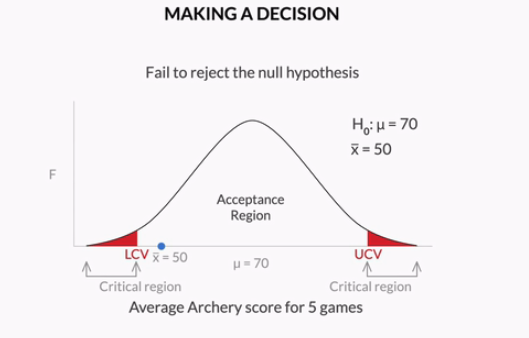




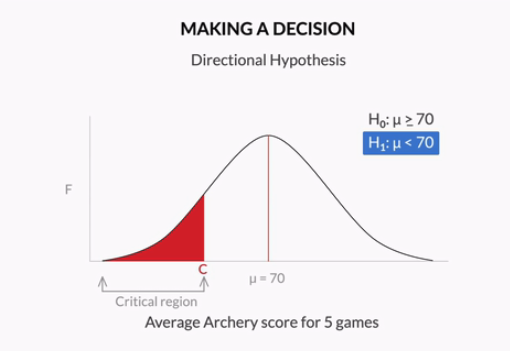


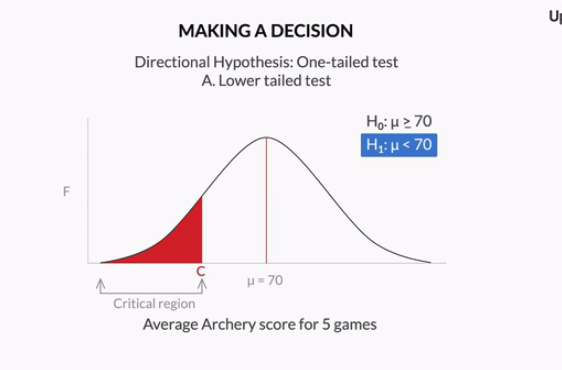


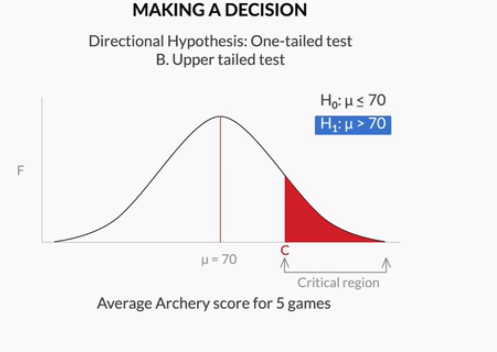




If your sample mean lies in the acceptance region, you fail to reject the null hypothesis because it is not beyond the critical point and you can consider that sample mean is equal to the population mean statistically.







**Types of tests**

Government regulatory bodies have specified that the maximum permissible amount of lead in any food product is 2.5 parts per million or 2.5 ppm.

If you conduct tests on randomly chosen Maggi Noodles samples from the market to see if its lead content is above the permissible amount of 2.5 ppm, what type of test this would be?



**Upper-tailed test**

**Feedback :***The alternate hypothesis in this case would be that the average lead content is more than 2.5 ppm, so the critical region would lie on right side of distribution. So this would be an upper-tailed test. Here, you can notice that alternate hypothesis is formulated with “more than” argument (equivalently > sign), which justifies it being a right-tailed test.*

The formulation of the null and alternate hypotheses determines the type of the test and the position of the critical regions in the normal distribution.

You can tell the type of the test and the position of the critical region on the basis of the ‘**sign’ in the alternate hypothesis.**

       ≠ in H₁    →   Two-tailed test        →     Rejection region on **both sides** of distribution

       < in H₁    →   Lower-tailed test     →     Rejection region on **left side** of distribution

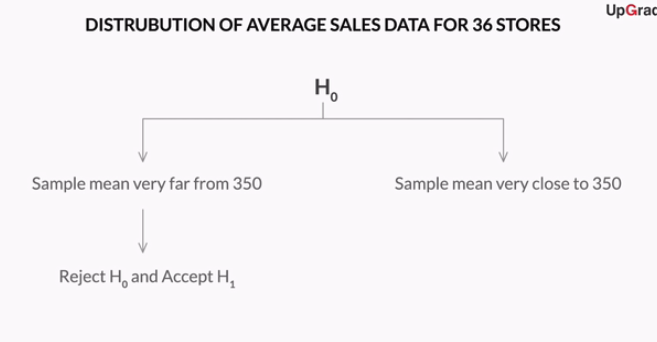
       > in H₁    →   Upper-tailed test     →     Rejection region on **right side** of distribution

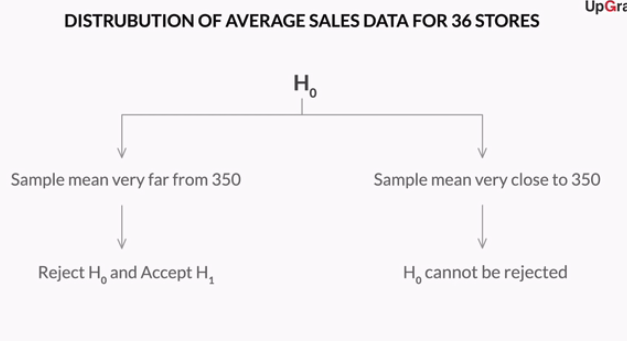
The average commute time for an UpGrad employee to and from office is at least 35 minutes.

If this hypothesis has to be tested, select the type of the test and the location of the critical region.

**Lower-tailed test, with the rejection region on the left side**

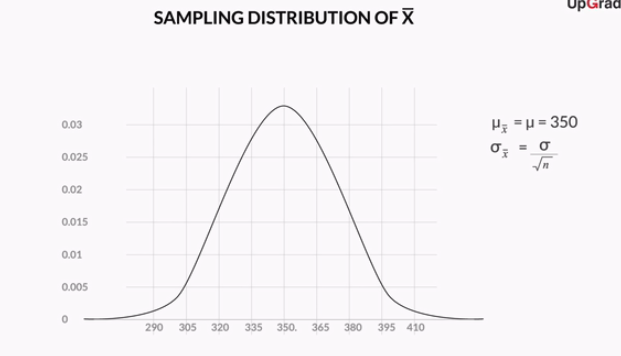
**Feedback :***For this situation, the hypotheses would be formulated as H₀: μ ≥ 35 minutes and H₁: μ < 35 minutes. As < sign is used in alternate hypothesis, it would be a lower-tailed test and the rejection region would be on the left side of the distribution.*

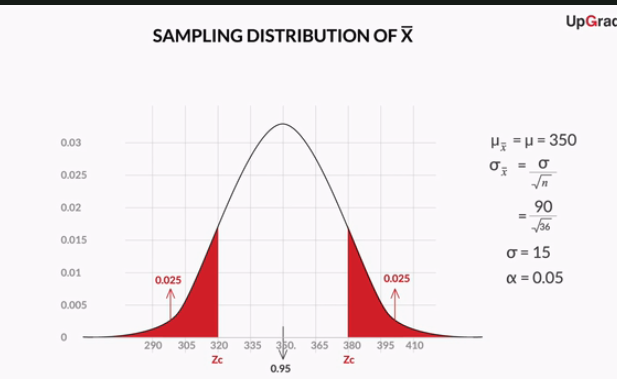


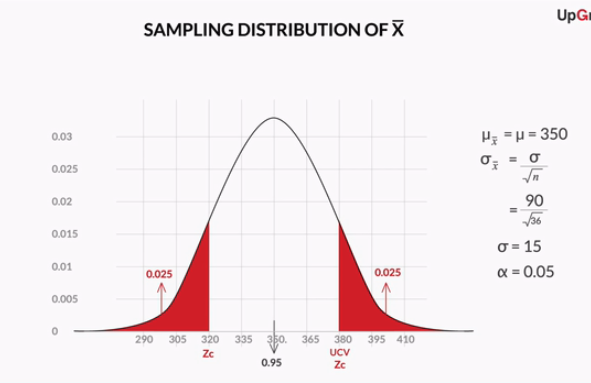


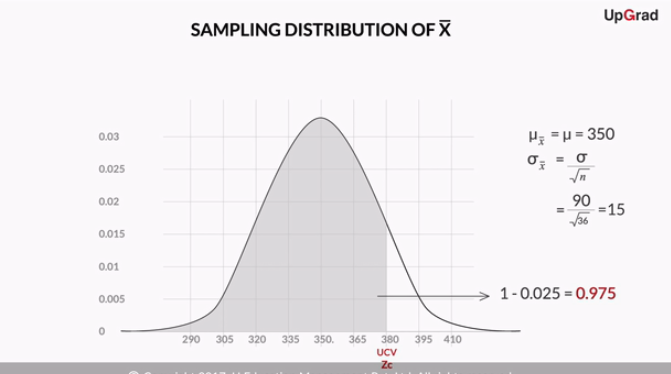
What will be the standard deviation of the **distribution of sample means** if the population has a mean of 350 and a standard deviation of 90, a sample mean of 370.16, and a sample size of 36?

he standard deviation of a distribution of sample means is obtained by dividing the population standard deviation by the square root of the sample size. So, 90/ ​√3636​ = 90/6 = 15.









Before you proceed with finding the Zc and finally the critical values, let’s revise the steps performed in this method till now.

1. First, you define a new quantity called α, which is also known as the significance level for the test. It refers to the proportion of the sample mean lying in the critical region. For this test, α is taken as 0.05 (or 5%).
2. Then, you calculate the cumulative probability of UCV from the value of α, which is further used to find the z-critical value (Zc) for UCV.

Attempt the following questions before you go ahead and learn the remaining steps in this method.

What will be the area of the critical region on the right-hand side of the distribution if the significance level (α) for a two-tailed test is 3%?

**0.015**

**Feedback :***Here, value of α is 0.03 (of 3%), so the area of the rejection region would be 0.03 and the area of the acceptance region would be 0.97. In addition, since this is a two-tailed test, the area of the critical region on the right-hand side would be half of 0.03, i.e. 0.015.*

What would be the area of the critical region on the right-hand side of the distribution if the significance level (α) for an upper-tailed test is 3%?

Top of Form



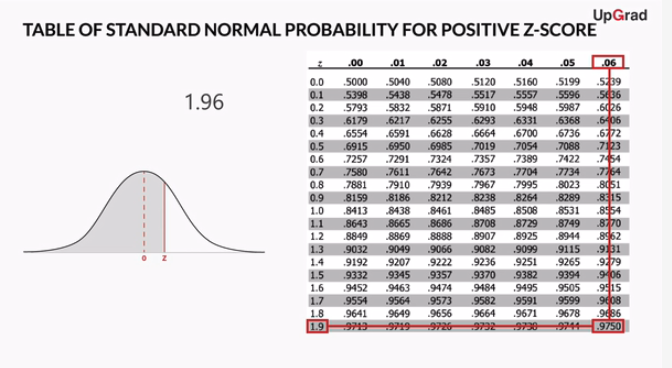
**0.03**

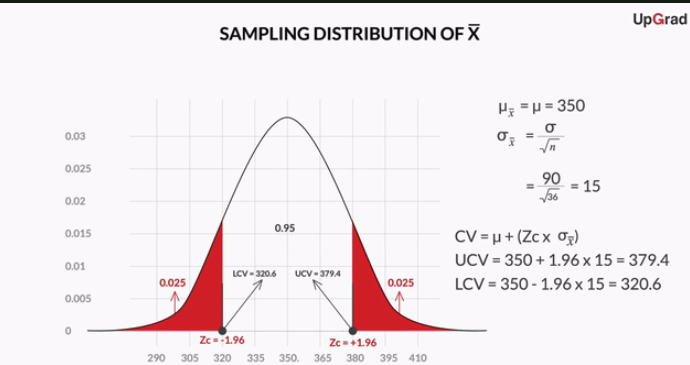
**Feedback :***Here, the value of α is 0.03 (of 3%), so the area of the critical region would be 0.03 and the area of the acceptance region would be 0.97. Since this is an upper-tailed test, the critical region is only on the right-hand side of the distribution, and the area of the critical region would be 0.03.*

What would be the value of the cumulative probability of UCV if the significance level (α) for an upper-tailed test is 3%?

**0.97**

**Feedback :***The area of the critical region in this case would be 0.03 (as calculated in the last question), which would be the area beyond the UCV point in the distribution. So, the area till the UCV point would be 1 - 0.03, i.e. 0.97. This would be the cumulative probability of that point, going by the definition of cumulative probability.*





After formulating the hypothesis, the steps you have to follow to **make a decision** using **the critical value method** are as follows:

1. Calculate the value of Zċ from the given value of α (significance level). Take it a 5% if not specified in the problem.
2. Calculate the critical values (UCV and LCV) from the value of Zċ.
3. Make the decision on the basis of the value of the sample mean x with respect to the critical values (UCV AND LCV).

You can download the z-table from the attachment below. It will be useful in the subsequent questions.

**Critical Value Method**

**1st step:**Calculate the value of Zc from the given value of α (significance level).

Calculate the z-critical score for the two-tailed test at 3% significance level.

**2.17**

**Feedback :***For 3% significance level, you would have two critical regions on both sides with a total area of 0.03. So, the area of the critical region on the right side would be 0.015, which means that the area till UCV (cumulative probability of that point) would be 1 - 0.015 = 0.985. So, you need to find the z-value of 0.985. The z-score for 0.9850 in the z-table is 2.17 (2.1 on the horizontal axis and 0.07 on the vertical axis).*

**2nd step:**Calculate the critical values (UCV and LCV) from the value of Zc.

Find out the UCV and LCV values for Zc = 2.17.

μ = 36 months        σ = 4 months       N (Sample size) = 49

Top of Form



**UCV = 37.24 and LCV = 34.76**

**Feedback :***The critical values can be calculated from μ ± Zc x (σ/​*√NN*​) as 36 ± 2.17(4/​*√4949*​) = 36 ± 1.24 which comes out to be 37.24 and 34.76.*

Bottom of Form

**Critical Value Method**

**3rd step:**Make the decision on the basis of the value of the sample mean ​¯xx¯ with respect to the critical values (UCV AND LCV).

What would be the result of this hypothesis test?

UCV = 37.24 months                 LCV = 34.76 months              Sample mean (​¯xx¯) = 34.5 months

**Reject the null hypothesis**

**Feedback :***The UCV and LCV values for this test are 37.24 and 34.76. The sample mean in this case is 34.5 months, which is less than LCV. So, this implies that the sample mean lies in the critical region and you can reject the null hypothesis.*

Government regulatory bodies have specified that the maximum permissible amount of lead in any food product is 2.5 parts per million or 2.5 ppm. Let’s say you are an analyst working at the food regulatory body of India FSSAI. Suppose you take 100 random samples of Sunshine from the market and have them tested for the amount of lead. The mean lead content turns out to be 2.6 ppm with a standard deviation of 0.6.

One thing you can notice here is that the standard deviation of the sample is given as 0.6, instead of the population’s standard deviation. In such a case, you can approximate the population’s standard deviation to the sample’s standard deviation, which is 0.6 in this case.

Answer the following questions in order to find out if a regulatory alarm should be raised against Sunshine or not, at 3% significance level.Select the correct null and alternate hypotheses in this case.

Top of Form



H₀: Average lead content ≤ 2.6 ppm and H₁: Average lead content > 2.6 ppm



**H₀: Average lead content ≤ 2.5 ppm and H₁: Average lead content > 2.5 ppm**

**Feedback :***The null hypothesis is your assumption about the population — it is based on the status quo. It always makes an argument about the population using the equality sign. The null hypothesis in this case would be that the average lead content in the food material is less than or equal to 2.5 ppm. And the alternate hypothesis is that the average lead content is greater than 2.5 ppm.*

Bottom of Form

Calculate the z-critical score for this test at 3% significance level.

Top of Form



**1.88**

**Feedback :***This is a one-tailed test. So, for 3% significance level, you would have only one critical region on the right side with a total area of 0.03. This means that the area till the critical point (the cumulative probability of that point) would be 1 - 0.030 = 0.970. So, you need to find the z-value of 0.970. The z-score for 0.9699 (~0.970) in the z-table is 1.88.*

Now, you need to find out the critical values and make a decision on whether to raise a regulatory alarm against Sunshine or not. Select the correct option.

**Critical value = 2.61 ppm and Decision: Don’t raise a regulatory alarm**

**Feedback :***The critical value can be calculated from μ + Zc x (σ/​*√NN*​), as 2.5 + 1.88(0.6 /​*√100100*​) = 2.61 ppm . You need to use the + sign since the critical value is on the right-hand side (upper-tailed test). Since the sample mean 2.6 ppm is less than the critical value (2.61 ppm), you fail to reject the null hypothesis and don’t raise a regulatory alarm against Sunshine.*

The critical value for this test at 3% significance level comes out to be 2.61 ppm. If you take more than 100 samples (with the same sample mean and standard deviation), how would the z-score and critical value change?



**The z-score would remain the same but the critical value would decrease**

**Feedback :***Since Zc is calculated from the given value of α (3%), it remains the same. Critical value is calculated using the formula: μ + Zc x (σ/​*√NN*​), since it is an upper-tailed test. If you increase the value of N, the critical value would decrease according to the formula*

Consider this problem — H₀: μ ≤ 350 and H₁: μ > 350

In case of a two-tailed test, you find the z-score of 0.975 in the z-table, since 0.975 was cumulative probability of UCV in that case. In this problem, what would be the cumulative probability of critical point in this example for the same significance level of 5%?



**0.950**

**Feedback :***In this problem, the area of the critical region beyond the only critical point, which is on the right side, is 0.05 (in the last problem, it was 0.025). So, the cumulative probability of the critical point (the total area till that point) would be 0.950.*



**1.645**

**Feedback :***0.950 is not there in the z-table. So, look for the numbers nearest to 0.950. You can see that the z-score for 0.9495 is 1.64 (1.6 on the horizontal bar and 0.04 on the vertical bar), and the z-score for 0.9505 is 1.65. So, taking the average of these two, the z-score for 0.9500 is 1.645.*

Consider this problem, H₀: μ ≤ 350 and H₁: μ > 350

So, the Zc comes out to be 1.645. Now, find the critical value for the given Zc and make the decision to accept or reject the null hypothesis.

μ = 350     σ = 90       N (Sample size) = 36    ¯xx¯= 370.16

Consider this problem, H₀: μ ≤ 350 and H₁: μ > 350

So, the Zc comes out to be 1.645. Now, find the critical value for the given Zc and make the decision to accept or reject the null hypothesis.

μ = 350     σ = 90       N (Sample size) = 36    ¯xx¯= 370.16



**Critical value = 374.67 and Decision = Fail to reject the null hypothesis**

**Feedback :***The critical value can be calculated from μ + Zc x (σ/*√NN*). 350 + 1.645(90/*√3636*) = 374.67. Since 370.16 (*¯xx¯*) is less than 374.67,*¯xx¯*lies in the acceptance region and you fail to reject the null hypothesis.*

**Summary**

So what did you learn in this session?

1. Hypothesis — a claim or an assumption that you make about one or more population parameters
2. Types of hypothesis:
   * **Null hypothesis** (H₀) - Makes an assumption about the status quo  
                                          - Always contains the symbols ‘=’, ‘≤’ or ‘≥’
   * **Alternate hypothesis** (H₁) - Challenges and complements the null hypothesis

                                                                  - Always contains the symbols ‘≠’, ‘<’ or ‘>’

1. Types of tests:
   * **Two-tailed test**- The critical region lies on both sides of the distribution  
                                  - The alternate hypothesis contains the ≠ sign
   * **Lower-tailed test**- The critical region lies on the left side of the distribution  
                                     - The alternate hypothesis contains the < sign
   * **Upper-tailed test**- The critical region lies on the right side of the distribution

                                                 - The alternate hypothesis contains the > sign

1. Making a decision - Critical value method:
   * Calculate the value of Zc from the given value of α (significance level)
   * Calculate the critical values (UCV and LCV) from the value of Zc
   * Make the decision on the basis of the value of the sample mean ¯xx¯ with respect to the critical values (UCV AND LCV)

A house owner claims that the current market value of his house is at least Rs.40,00,000.  60 real estate agents are asked independently to estimate the house's value. The hypothesis test that is conducted ends with the decision of "reject H₀".  Which of the following statements accurately states the conclusion?



**The house owner is wrong, the house is worth less than Rs. 40,00,000**

**Feedback :***Rejection of the null hypothesis means rejection of the status quo or the earlier assumption of the house owner that his house is worth at least Rs. 40,00,000. As the null hypothesis is H₀: House market value ≥ 40,00,000, the alternate hypothesis would be opposite of that.*

The null and alternative hypotheses are statements about:

Top of Form



**Population parameters**

**Feedback :***The hypothesis is always made about the population parameters. The sample parameters are only used as evidence to test the hypothesis.*

Which of the following options hold true for null hypothesis?



**The claim with the “less than or equal to” sign**

**Feedback :**The null hypothesis is always written with the “equal to” or “less than or equal to” or “more than or equal to” sign.

**Correct**



**The claim with the “equal to” sign**

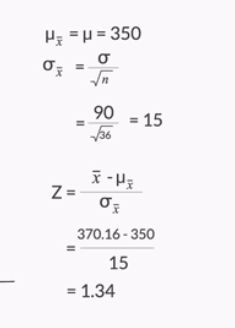
**Feedback :**The null hypothesis is always written with the “equal to” or “less than or equal to” or “more than or equal to” sign.



**UCV = 62.49 g, LCV = 57.51 g and Result = Don’t pass the test**

**Feedback :***The critical values can be calculated from μ ± Zc x (σ/​*√NN*​) as 60g ± 2.33(10.7 /*√100100*) = 60g ± 2.49g which comes out to be 62.49 g and 57.51 g. Since 62.6 g is greater than the UVC of 62.49, which means the sample mean lies outside the range of the critical values, you reject the null hypothesis that the average weight of the chocolate is 60 g. So, the QA would not pass this test.*

# p-value Method





**In the critical value method, the z-score is calculated for the critical points, which is called Zc**

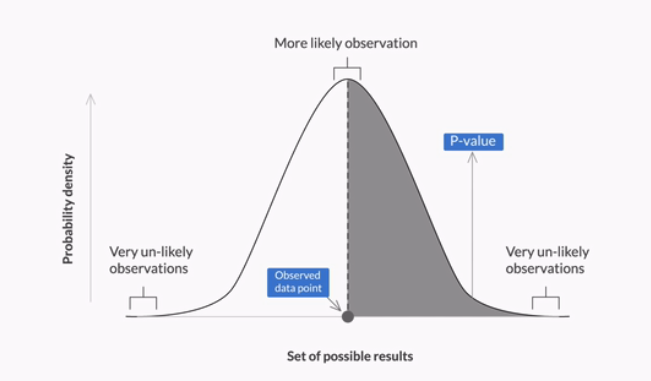
**Feedback :**α is used to calculate the Zc value for the critical points in this method.

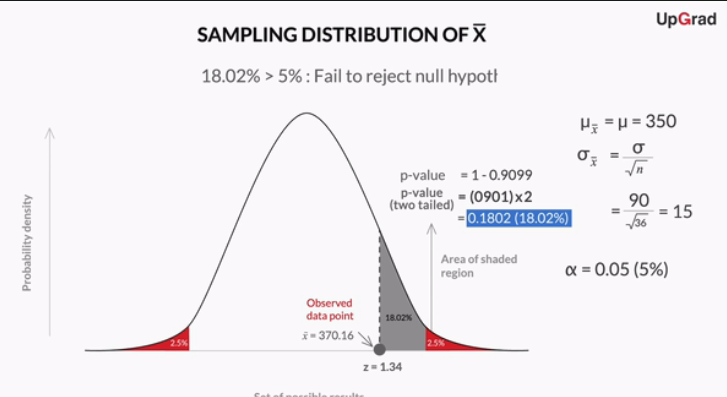


**In the p-value method, the z-score is calculated for the sample mean**

**Feedback :**In the p value method, the z-score is calculated for the sample mean.

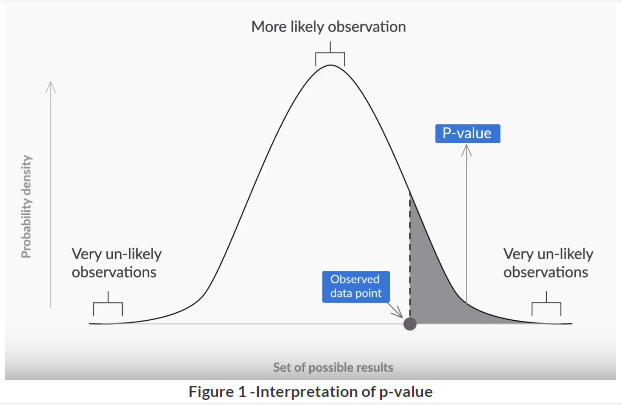
**:**p-value is equivalent to the probability of the null hypothesis being accepted (or more aptly, not being rejected). So, the smaller the p-value, the farther will be the sample mean from the hypothesised population mean, which indicates more evidence that the sample mean lies in the critical region, and the alternate hypothesis is accepted.





Prof Tricha has defined the **p-value** as the **probability of the null hypothesis** being accepted (or more aptly, not being rejected). This statement is not technically correct (or formal) definition of p-value, but it is used for better understanding of the p-value.

Higher the p-value, higher is the probability of failing to reject a null hypothesis. On the other hand, lower the p-value, higher is the probability of the null hypothesis being rejected.



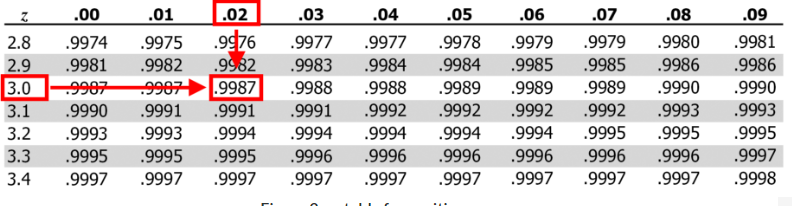
After formulating the null and alternate hypotheses, the steps to follow in order to**make a decision**using the **p-value method** are as follows:

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 (the z-score is positive)

**Example:**z-score for sample point = + 3.02



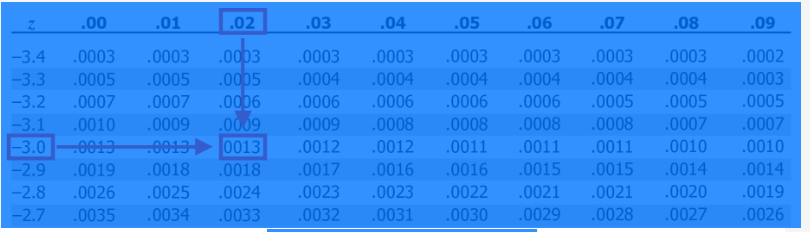
Cumulative probability of sample point = 0.9987

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

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

**Situation 2:**The sample mean is on the left side of the distribution mean (the z-score is negative)

**Example:** z-score for sample point = -3.02          



**Figure 3 - z-table for positive z-scores**

Cumulative probability of sample point = 0.0013

For one-tailed test  →    p = 0.0013

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

You can download the z-table from the attachment below. It will be useful in the subsequent questions.

**Let’s solve the following problem stepwise** to consolidate your learning on how to make a decision about any hypothesis using the p-value method.

You are working as a data analyst at an auditing firm. A manufacturer claims that the average life of its product is 36 months. An auditor selects a sample of 49 units of the product, and calculates the average life to be 34.5 months. The population standard deviation is 4 months. Test the manufacturer’s claim at 3% significance level using the p-value method.

First, **formulate the hypotheses** for this two-tailed test, which would be:

                                   H₀: μ = 36 months and H₁: μ ≠ 36 months

Now, you need to follow the three steps to **find the p-value and make a decision**.

Try out the three-step process by answering the following questions.

**p-value Method**

**Step 1:**Calculate the value of z-score for the sample mean point on the distribution. Calculate z-score for sample mean (¯xx¯) = 34.5 months.



**-2.62**

**Feedback :***You can calculate the z-score for sample mean 34.5 months using the formula: (​​*¯xx¯*​ - μ) / (σ/​*√xx*​). This gives you (34.5 - 36) / (4/*√4949*) = (-1.5) \* 7/4 = -2.62. Notice that, since the sample mean lies on the left side of the hypothesised mean of 36 months, the z-score comes out to be negative.*

**p-value Method**

**Step 2:**Calculate the p-value from the cumulative probability for the given z-score using the z-table.

Find out the p-value for the z-score of -2.62 (corresponding to the sample mean of 34.5 months).

Hint: The sample mean is on the left side of the distribution and it is a two-tailed test.



**0.0088**

**Feedback :***The value in the z-table corresponding to -2.6 on the vertical axis and 0.02 on the horizontal axis is 0.0044. Since the sample mean is on the left side of the distribution and this is a two-tailed test, the p-value would be 2 \* 0.0044 = 0.0088.*

**p-value Method**

**Step 3:** Make the decision on the basis of the p-value with respect to the given value of α (significance value).

What would be the result of this hypothesis test?



**Reject the null hypothesis**

**Feedback :***Here, the p-value comes out to be 2 \* 0.0044 = 0.0088. Since the p-value is less than the significance level (0.0088 < 0.03), you reject the null hypothesis that the average lifespan of the manufacturer's product is 36 months.*

# p-value Method - Examples

**Comprehension**

Let’s revisit an example we looked at earlier.

Let’s say you work at a pharmaceutical company that manufactures an antipyretic drug in tablet form, with paracetamol as the active ingredient. An antipyretic drug reduces fever. The amount of paracetamol deemed safe by the drug regulatory authorities is 500 mg. If the value of paracetamol is too low, it will make the drug ineffective and become a quality issue for your company. On the other hand, a value that is too high would become a serious regulatory issue.

There are 10 identical manufacturing lines in the pharma plant, each of which produces approximately 10,000 tablets per hour.

Your task is to take a few samples, measure the amount of paracetamol in them, and test the hypothesis that the manufacturing process is running successfully, i.e. the paracetamol content is within regulation. You have the time and resources to take about 900 sample tablets and measure the paracetamol content in each.

Upon sampling 900 tablets, you get an average content of 510 mg with a standard deviation of 110. What does the test suggest, if you set the significance level at 5%? Should you be happy with the manufacturing process or should you ask the production team to alter the process? Is it a regulatory alarm or a quality issue?

**Solve the following questions** in order to find out the answers to the questions stated above.

One thing you can notice here is that the standard deviation of the sample of 900 is given as 110, instead of the population’s standard deviation. In such a case, you can **approximate the population’s standard deviation to the sample’s standard deviation, which is 110** **in this case**.

**p-value Method**

Calculate the z-score for sample mean (¯xx¯) = 510 mg.



**2.73**

**Feedback :***You can calculate the z-score for the sample mean 510 mg using the formula: (​*¯xx¯*​ - μ) / (σ /​*√NN*​). This gives you (510 - 500) / (110 /*√900900*) = (10) / (110 / 30) = 2.73. Notice that, since the sample mean lies on the right side of the hypothesised mean of 500 mg, the z-score comes out to be positive.*

Find out the p-value for the z-score of 2.73 (corresponding to the sample mean of 510 mg).



**0.0064**

**Feedback :***The value in the z-table corresponding to 2.7 on the vertical axis and 0.03 on the horizontal axis is 0.9968. Since the sample mean is on the right side of the distribution and this is a two-tailed test (because we want to test if the value of the paracetamol is too low or too high), the p-value would be 2 \* (1 - 0.9968) = 2 \* 0.0032 = 0.0064.*

What decision would you make about the manufacturing process from this hypothesis test?



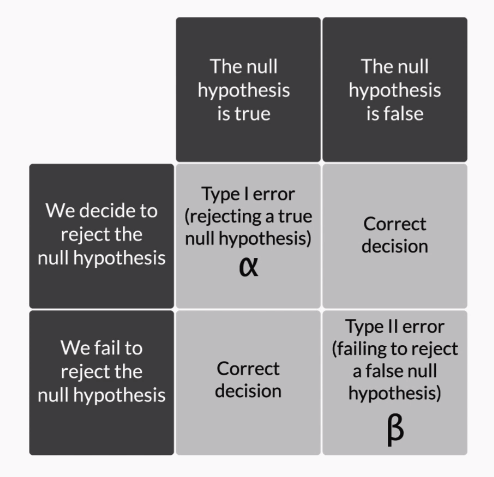
**The manufacturing process is not fine and changes need to be made**

**Feedback :***Here, the p-value comes out to be 0.0064. Since the p-value is less than the significance level (0.0064 < 0.05) and smaller p-value gives you greater evidence against the null hypothesis. So you reject the null hypothesis that the average amount of paracetamol in medicines is 500 mg. So, this is a regulatory alarm for the company and the manufacturing process needs to change.*

# Types of Errors

While doing hypothesis testing, there is always the possibility of making the wrong decision about your hypothesis. These instances of a wrong decision being made are referred to as errors. Let’s learn about the different types of errors during hypothesis testing.

There are two types of errors that can result during the hypothesis testing process — type-I error and type-II error.



**Figure 1 - Types of errors**

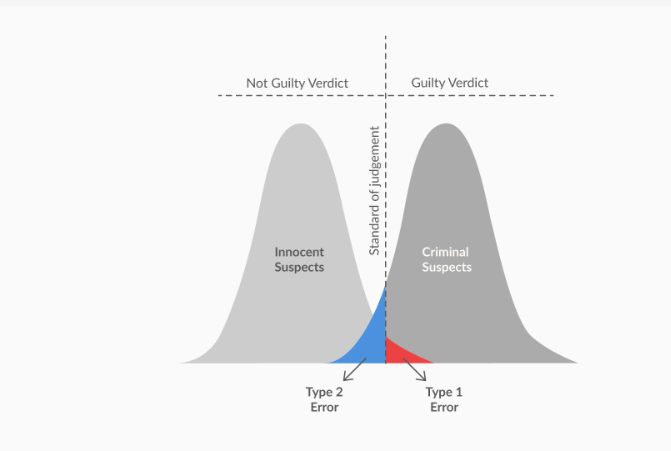
A **type I-error** represented by α occurs when you reject a true null hypothesis.

A **type-II error** represented by β occurs when you fail to reject a false null hypothesis.

The power of any hypothesis test is defined by 1 - β. Power of the test or calculation of β is beyond the scope of this course. You can study more about power of a test from [this](http://www2.fiu.edu/~howellip/Power.pdf)link.

If go back to the analogy of the **criminal trial example**, you would find that the probability of making a type-I error would be more if the jury convicts the accused even on less substantial evidence. The probability of a type-I error can be reduced if the jury adopts more stringent criteria to convict an accused party.

However, reducing the probability of a type-I error may increase the probability of making a type-II error. If the jury becomes very liberal in acquitting the people on trial, there would be a higher probability that an actual criminal is able to walk free.





**Type-I error occurs when the null hypothesis is rejected when it is in fact correct**

**Feedback :**Type-I error occurs when the null hypothesis is true (i.e. the sample mean lies in the acceptance region) but you incorrectly reject it.



**Type II error occurs when the null hypothesis is not rejected when it is in fact incorrect**

**Feedback :**Type-II error occurs when the null hypothesis is not true (i.e. the sample mean lies in the critical region) but you incorrectly fail to reject the null hypothesis.

Suppose the null hypothesis is that a particular new process is as good as or better than the old one. A type-I error is to conclude that:

**The old process is better than the new one, when it is not**

**Feedback :***Type-I error means incorrectly rejecting a true null hypothesis. So, type-1 error means that the null hypothesis is true, i.e. the new process is as good as or better than the old one, but you reject it, i.e. you conclude that the old process is better.*

**Summary**

So what did you learn in this session?

1. Making a decision - p-value method:

* 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 the decision on the basis of the p-value with respect to the given value of α (signIficance level)

1. Types of errors:

* **Type-I error**    - Occurs when you reject a null hypothesis even when it is true  
                            - Its probability is represented by α
* **Type-II error** - Occurs when you fail to reject the null hypotheses even though it is false

                                   - Its probability is represented by β

**p-value Method**

Suppose you conduct a hypothesis test and observe that the values of the sample mean and sample standard deviation when n = 25 do not lead to the rejection of the null hypothesis. You calculate the p-value as 0.0667. What would happen to the p-value if you observe the same sample mean and sample standard deviation for a larger sample size, say greater than 50?



**Decrease**

**Feedback :***With an increase in the sample size, the denominator of the z-score decreases, and thus the absolute value of Z-score increases, which means that the sample mean would move away from the central tendency towards the tails. This means that the p-value would actually decrease. Conceptually, Increasing the sample size will make the distribution of sample means narrower, and chance of sample mean falling in the critical region decreases. So p-value will decrease.*

Consider the null hypothesis that a process produces no more than the maximum permissible rate of defective items. In this situation, a type-II error would be:



**To conclude that the process does not produce more than the maximum permissible rate of defective items, when it actually does**

**Feedback :***Type-II error means not rejecting the incorrect null hypothesis. So, a type-II error would signify that the null hypothesis is actually incorrect, i.e. the process actually produces more than the maximum permissible rate of defective items, but you fail to reject it, i.e. you think it does not produce more than the maximum permissible rate of defective items.*

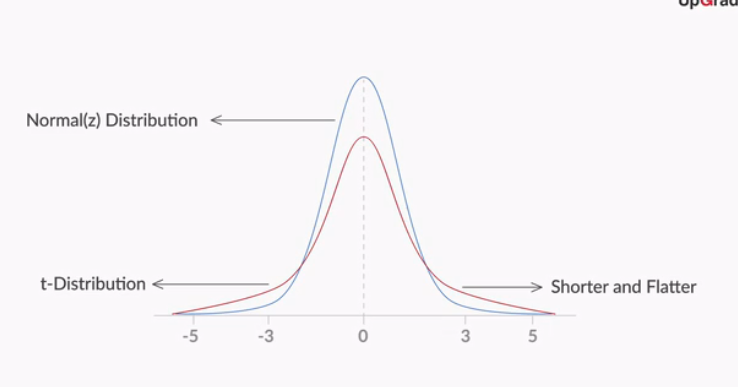
A test to screen for a serious but curable disease is similar to hypothesis testing. In this instance, the null hypothesis would be that the person does not have the disease, and the alternate hypothesis would be that the person has the disease. If the null hypothesis is rejected, it means that the disease is detected and treatment will be provided to the particular patient. Otherwise, it will not. Assuming the treatment does not have serious side effects, in this scenario, it is better to increase the probability of:

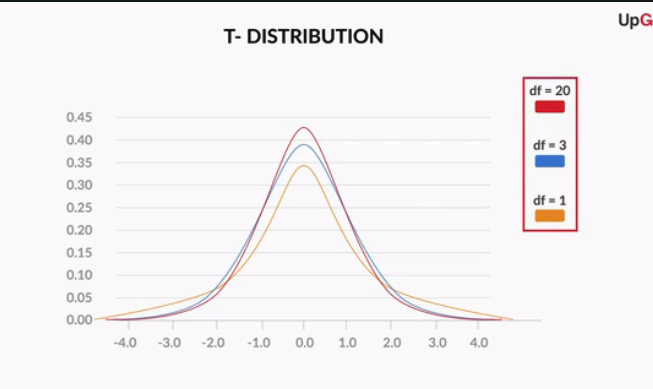


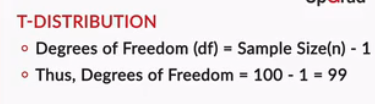
**Making a type-I error, i.e. providing treatment when it is not needed**

**Feedback :***Here, type-I error would be providing treatment upon detecting the disease, when the person does not actually have the disease. And type-II error would be not providing treatment upon failing to detect the disease, when the person actually has the disease. Since the treatment has no serious side effects, type-I error poses a lower health risk than type-II error, as not providing treatment to a person who actually has the disease would increase his/her health risk.*

# T Distribution





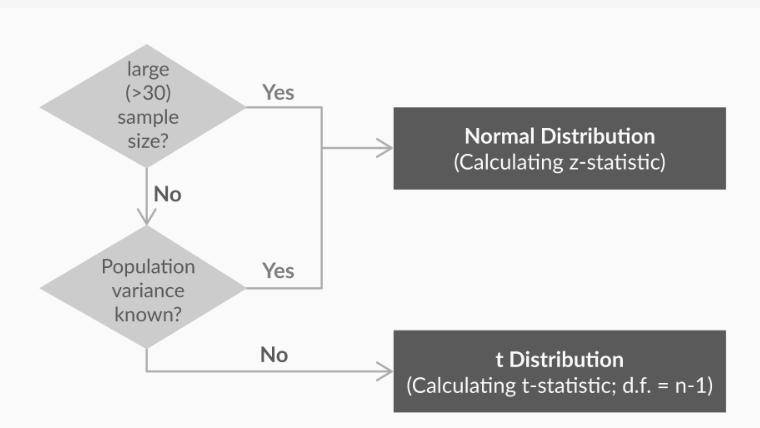




A t-distribution is also referred to as **Student’s T distribution**. A t-distribution is similar to the normal distribution in many cases; for example, it is symmetrical about its central tendency. However, it is shorter than the normal distribution and has a flatter tail, which would eventually mean that it has a larger standard deviation.

At a sample size beyond 30, the t-distribution becomes approximately equal to the normal distribution.

The most important use of the t-distribution is that you can approximate the value of the **standard deviation of the population (σ)**from the**sample standard deviation (s)**. However, as the sample size increases more than 30, the t-value tends to be equal to the z-value. Thus, if you want to summarise the decision-making in a flowchart, this is what you would get.



If the sample size is 10 and the standard deviation of the population is known, which distribution should be used to calculate the critical values and make the decision during hypothesis testing?

Top of Form



**Standard normal distribution (z-distribution)**

**Feedback :***Whenever the standard deviation of the population is known, you have to use z-distribution, irrespective of the value of the sample size (N).*

If the sample size is 10 and the standard deviation of the population is unknown, which distribution should be used to calculate the critical values and make the decision during hypothesis testing?



**T distribution**

**Feedback :***T distribution is used whenever the standard deviation of the population is unknown and the sample size is less than 30.*

Let’s look at how the method of **making a decision** changes if you are using the sample’s standard deviation instead of the population’s. If you recall the critical value method, the first step is as follows:

1. Calculate the value of Zc from the given value of α (significance level). Take it as 5% if not specified in the problem.

So, to find Zc, you would use the **t-table** instead of the z-table. The **t-table** contains values of Zc for a given degree of freedom and value of α (significance level). Zc, in this case, can also be called as t-statistic (critical).

Download the t-table given below and attempt the following questions to understand how to use the t-table to find Zc.

**t-test**

You are given the standard deviation of a sample of size 25 for a two-tailed hypothesis test of a significance level of 5%.

**Use the t-table** given above to find the value of Zc.



**2.064**

**Feedback :***For sample size = 25, your degrees of freedom would become 25 - 1 = 24. So, if you look for the value in the t-table corresponding to d.f. = 24 and α = 0.05 for a two-tailed test, you would get the t-value as 2.064.*

**t-test**

You are given the standard deviation of a sample of size 32 for a two-tailed hypothesis test of a significance level of 5%.

**Use the t-table** given above to find the value of Zc



**1.96**

**Feedback :***For sample size = 32, your degrees of freedom would become 32 - 1 = 31. So, if you look for the value in the t-table corresponding to d.f. > 29 and α = 0.05 for a two-tailed test, you would get a value of 1.96.*

In the second question, you used the t-table to find the value of Zc for sample size = 32 and a significance level of 5%. If you use the z-table for the same, you would get the same value of Zc, since, **for sample size ≥ 30, the t-distribution is the same as the z-distribution.**

Practically you would not need to refer to the z-table or t-table when doing hypothesis testing in the industry. Going forward when you need to do hypothesis testing in demonstrations of Excel or R, you would use the term **t-test**since that is mostly performed in the industry. All calculations and results of a t-test are same as the z-test whenever the sample size ≥ 30.

# Two-Sample Mean Test

**Hypotheses in a Two-Sample Mean Test**

In this two-sample mean test, suppose the null hypothesis is stated as follows: H₀: μ₁ - μ₂ = 0  
  
What would be the alternate hypothesis in this case?

**Hypotheses in a Two-Sample Mean Test**

In this two-sample mean test, suppose the null hypothesis is stated as follows: H₀: μ₁ - μ₂ = 0  
  
What would be the alternate hypothesis in this case?



**H₁: μ₁ - μ₂ ≠ 0**

**Feedback :***The alternate hypothesis is the claim which opposes the null hypothesis. If the null hypothesis states that the difference between both the means is zero, then the alternate hypothesis would be that the difference between both the means is not zero, i.e. μ₁ - μ₂ ≠ 0.*

**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 - Paired**

There is a hypothesis that Virat Kohli performs better or as good in the second innings of a test match as the first innings. This would be a**two-sample mean test**, where sample 1 would contain his score from the first innings and sample 2 would contain his score from the second innings. This would be a **paired test**since each row in the data would correspond to the same match.

What would be the null hypothesis in this case?



**H₀: μ₂ - μ₁ ≥ 0**

**Feedback :***Here, the assumption is that Virat Kohli performs better or as good as in the second innings, which means his average in the second innings is assumed to be greater than or equal to his average in the first innings. So, the null hypothesis would be: μ₂ ≥ μ₁ or μ₂ - μ₁ ≥ 0*

**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.

In the Excel file, go to the third tab ‘2-sample mean test - Unpaired’ and perform the required test. Answer the questions below after performing this test, taking a significance level of 5%.

**Two-Sample Mean Test - Unpaired**

What would be the p-value for this test if you consider it as a two-tailed test?



**0.022**

**Feedback :***If you perform the steps shown in the demonstration above and look at the second-last row of the output table, you would see that the p-value of a two-tailed test is 0.022.*

**Two-Sample Mean Test - Unpaired**

What can you conclude from this test about the two unpaired samples if you take a significance level of 5%



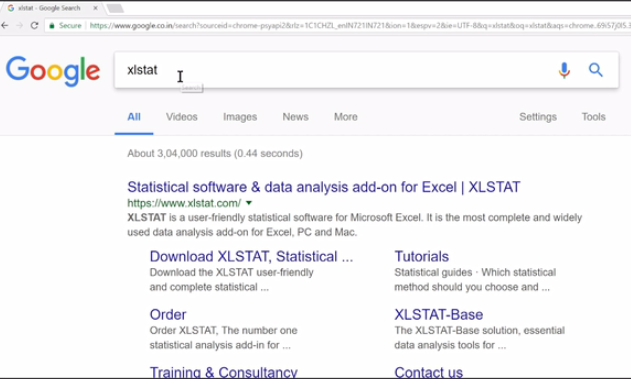
**There is some significant difference between the means of the two samples**

**Feedback :***The null hypothesis in this test is that the means of the two samples are the same, and there is no significant difference between them. As the p-value (0.022) is less than 0.05 (value of α), you can reject the null hypothesis and conclude that there is some significant difference between the means of the two samples.*

# Two-Sample Proportion Test

So, you now know how to compare the mean of a sample to a particular value using the one-sample mean test, and the means of two different samples using the two-sample mean test.

One thing you should observe in these tests is that the data from the sample is always numeric in nature. But what would you do if the data is categorical in nature, i.e. 1 or 0; Yes or No, etc.? Let’s take a look.



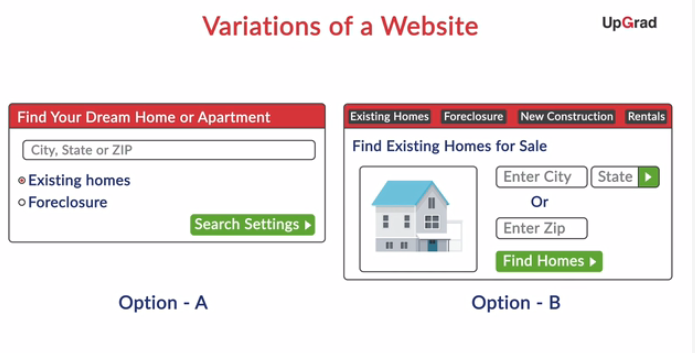
**Two-sample proportion test**is used when your sample observations are categorical, with two categories. It could be True/False, 1/0, Yes/No, Male/Female, Success/Failure etc.

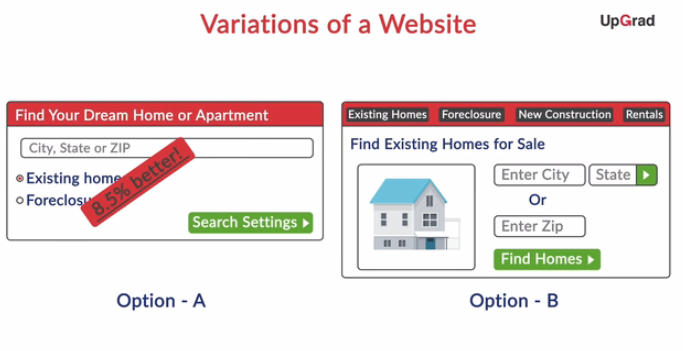
For example, if you are comparing the effectiveness of two drugs, you would define the desired outcome of the drug as the success. So, you would take a sample of patients who consumed the new drug and record the number of successes and compare it with successes in another sample who consumed the standard drug.

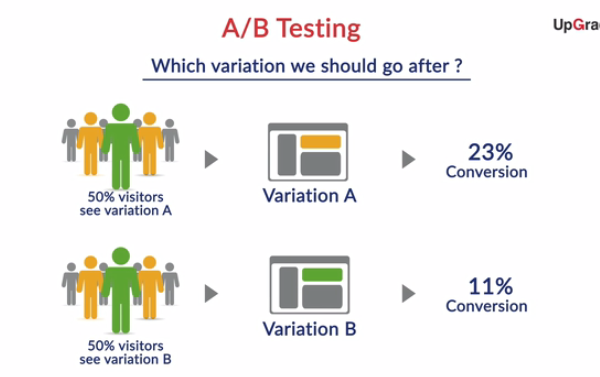
You can download the Excel file given below and play around with the two-sample proportion-test after installing the trial version of the XLSTAT add-in from [this](https://www.xlstat.com/en/) link. Later, you will learn to conduct the two-sample proportion test on R as well.

# A/B Testing Demonstration

 a process that is widely used in digital and online companies —**A/B testing.**A/B testing is a direct industry application of the two-sample proportion test sample you have just studied.







While developing an e-commerce website, there could be different opinions about the choices of various elements, such as the shape of buttons, the text on the call-to-action buttons, the colour of various UI elements, the copy on the website, or numerous other such things.

Often, the choice of these elements is very subjective and is difficult to predict which option would perform better. To resolve such conflicts, you can use A/B testing. **A/B testing** provides a way for you to test two different versions of the same element and see which one performs better. You can read more about A/B testing from this [link](https://www.optimizely.com/ab-testing/).

You can see a few more case studies and applications of A/B testing in the real world [here](https://blog.optimizely.com/2015/06/04/ecommerce-conversion-optimization-case-studies/).

The two-sample proportion test is used when you want to compare the proportions of two different samples. Let’s now see how A/B testing is entirely based on the two-sample proportion test, and go through an Excel demonstration of A/B testing with Deepak.

You can see a few more case studies and applications of A/B testing in the real world [here](https://blog.optimizely.com/2015/06/04/ecommerce-conversion-optimization-case-studies/).

**A/B Testing Demonstration**

618 out of 2,620 people filled the application using the old feature, which makes it 23.59% conversions in the control group. 618 out of 2,380 people filled the application using the new feature, which makes it 25.25% conversions in the variation group.



**Nothing can be concluded by just looking at this data**

**Feedback :***The number of conversions is more in case of the new feature. But there needs to be some statistical difference between the two values to arrive at a conclusion about this test, which would depend upon the significance level you choose. So, you cannot conclude anything just from this data.*

**A/B Testing Demonstration**

The p-value you get for this one-tailed A/B testing data is 0.091. What can you conclude about this test at a 5% significance level?



**Fail to reject the null hypothesis (the old feature is better or as good as the new feature)**

**Feedback :***You can make a decision on a test based on the p-value and the significance level of the test. As the p-value (0.091) is greater than the significance level (0.05), you fail to reject the null hypothesis. In this example of A/B testing, the null hypothesis is that the old feature is better or as good as the new feature.*

**pinion**

Skip

**Interesting Fact:**A few years ago, a team at Google couldn’t decide between two shades of blue, so the team conducted tests on 41 different shades of blue to gather data and see which shades of blue performed better (i.e. to see which shades would get more clicks).

This is called multivariate testing (A/B testing for multiple numbers of variations/features).

You can use various other tools also to conduct A/B testing, like R, [Optimizely](https://www.optimizely.com/" \t "_blank) etc.

# Industry Relevance

Different industry experts have different views on the applications and usability of hypothesis testing because of different reasons.

Let us first hear from Madhukar about the applications of hypothesis testing in the real world and get some insights into how this will be relevant in the next courses of Predictive Analysis, when you reach the model building stage of the Crisp-DM framework.

**Industry Relevance**

For this example of Domino’s campaign for Peppy Paneer Pizza, test population is considered as it is purchasing that particular pizza because of the campaign and control population is the one which buys that pizza normally.  
  
What would be the null hypothesis in this case?



**The difference in response rate of control and test population is statistically insignificant**

**Feedback :***This is a 2-sample proportion test. The null hypothesis would contain the equality parameter. So you would assume that difference between response rates of both groups is equal to zero, or you can say that the difference is statistically insignificant.*

Anand has a very different take on hypothesis testing, as he believes that, nowadays, all the data is available digitally for most industries. This makes the usability of hypothesis testing a little limited in lot of scenarios.

To summarise, hypothesis testing still holds importance in the following two types of industries, even if all the data is available digitally:

1. Manufacturing processes in the food, pharmaceuticals, chemicals industries, where it is not practically possible to gather information on the entire population.
2. E-commerce, advertising and digital marketing companies, where the amount of data collected from various samples is so huge that analysing all data becomes very difficult without having big data systems in place.

**Hypothesis Testing in R**

You can easily do hypothesis testing in R also through following two built-in functions:

1. t.test() - A built-in function to do **one-sample or two-sample mean test**.
2. prop.test() - A built-in function to do **one-sample or two-sample proportion test**.

These functions just take the sample data as arguments and return you the output, which contains p-value and a lot of other information. You can then make your decision based on the p-value obtained.

Attempt the following **SWIRL exercise** to learn how to use the above mentioned functions in R.

The files given in the attachment below is **hypothesis testing.swc** and **ttest.csv**. Download both the files and place them in your working directory. The following set of commands will help you get started with this lesson.

library(swirl)

install\_course() #Select the hypothesistesting.swc file after this command

swirl()

**Note:** If you conduct two-sample proportion test on R, you would observe slight difference in output results as compared to results of XLSTAT. This is because of some statistical assumptions made in XLSTAT add-in. But the final result of rejecting or failing to reject the null hypothesis would remain the same in both the tools. Going forward, **you would be using R to conduct hypothesis testing** in next courses.

**Summary**

So what did you learn in this session?

1. **T-distribution**:
   * A T-distribution is used whenever the standard deviation of the population is unknown
   * The degrees of freedom of a T-distribution is equal to sample size n - 1
   * For sample size ≥ 30, the T-distribution becomes the same as the normal distribution
   * The output values and results of both t-test and z-test are same for sample size ≥ 30
2. **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
3. **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

1. **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.
2. **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 ecommerce, manufacturing and advertising domains
   * It provides a way to test two different versions of the same element and see which one performs better

You can download the lecture notes for the module from below. The lecture notes include a summary of the entire module.

Bottom of Form

Top of Form

Bottom of Form

Bottom of Form

Bottom of Form

Bottom of Form

Top of Form

Bottom of Form