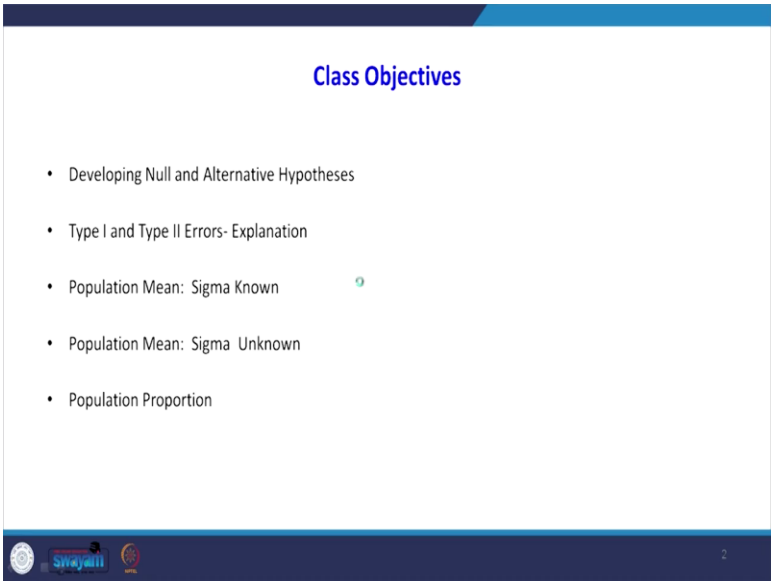
**Data Analytics with Python**
**Prof. Ramesh Anbanandam**
**Department of Management Studies**
**Indian Institute of Technology – Roorkee**

**Lecture – 16**
**Hypothesis Testing- I**

Welcome students today we are entering to a very, very interesting topic that is on hypothesis testing. Especially this topic is going to be fundamental for coming lectures. So, you are to carefully you should understand.

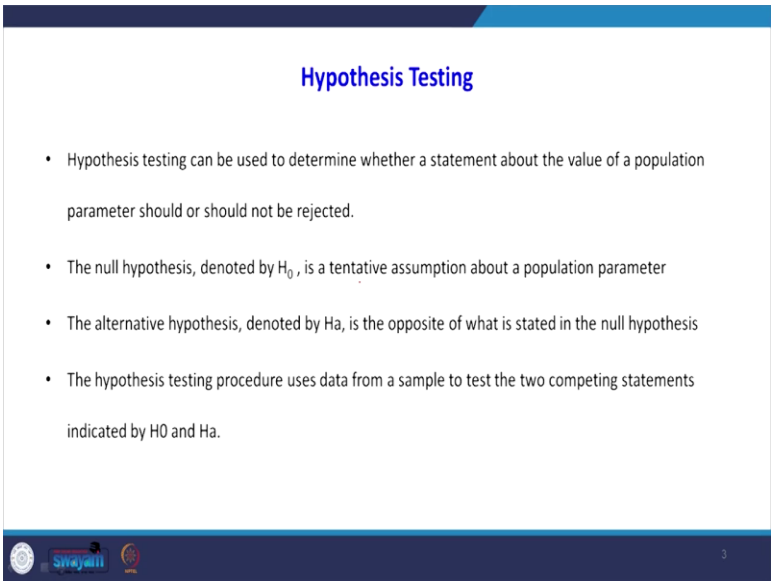
**(Refer Slide Time: 00:40)**




The class objectives are first I will to explain how to develop null and alternative hypotheses because solving a hypothesis problem is very easy the most important is how to formulate to the hypothesis. Once you are very good at the formulating the hypothesis solving the problem is very easy. Then I am going to explain what is a type 1 and type 2 error and how this context of type 1 type 2 error is connected with hypothesis.

Next we are going to do hypothesis testing when Sigma that is a population standard deviation is known. Next we will go to hypothesis testing when population standard deviation is not known. Then we will do hypothesis testing for the population proportion.

**(Refer Slide Time: 01:27)**

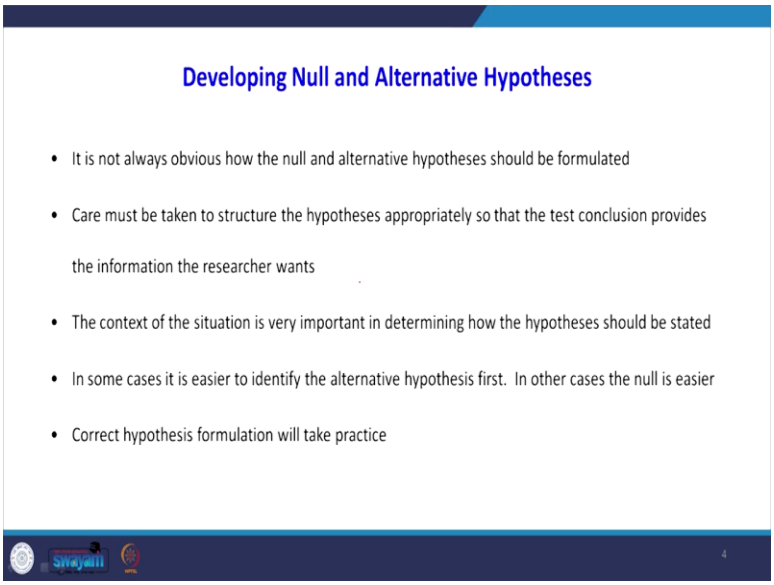



First we will go for what is hypotheses are testing hypothesis testing can be used to determine whether a statement about the value of the population parameter should or should not be rejected. So, hypothesis is nothing but some assumptions about the population parameter we know that most of the populations which we are going to do is going to follow a particular distribution for example a normal distribution.

So normal distribution having two parameter one is mean and variance. So, we can assume the population mean as a hypothesis and assumption otherwise you can have population variance also and hypothesis. The null hypothesis denoted by H0 is the tentative assumption about the population parameter. So, whatever assumptions which are having that will go to the population parameter the alternative hypothesis denoted by Ha is the opposite of what is stated in the null hypothesis.

The hypothesis testing procedure uses data from the sample to test to computing statement indicated by H0 and Ha. What are the two competing statement one is null hypothesis another one is alternative hypothesis.

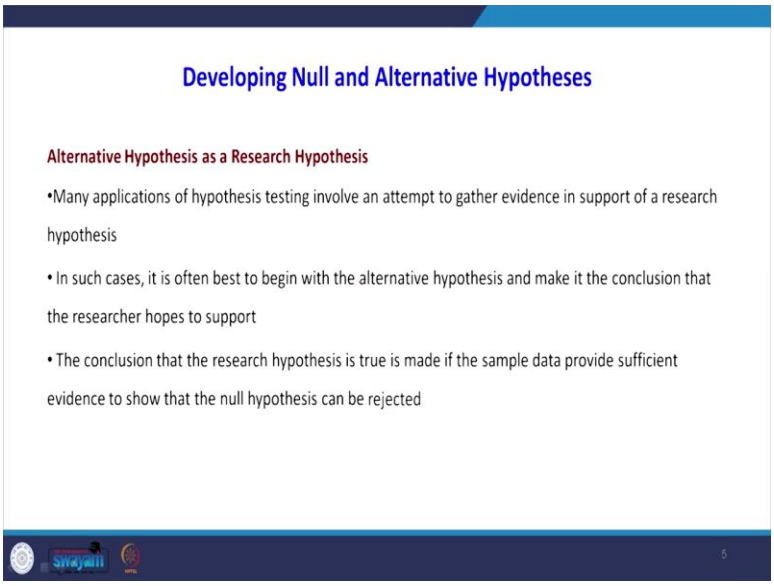
**(Refer Slide Time: 02:44)**




Next we will see how to develop null and alternative hypothesis, it is not always obvious how the null and alternative hypothesis should be formulated. We should be very careful to structure the hypothesis appropriately so that the test conclusion provides the information the researcher wants. The context of the situation is very important in determining how the hypothesis should be stated. In some cases it is easier to identify the alternative hypothesis first in other cases the null is easier.

So, correct hypothesis formulation, will take a practice in this lecture we are going to take some example and I am going to explain how to formulate the hypothesis whether it is null or altered hypothesis.

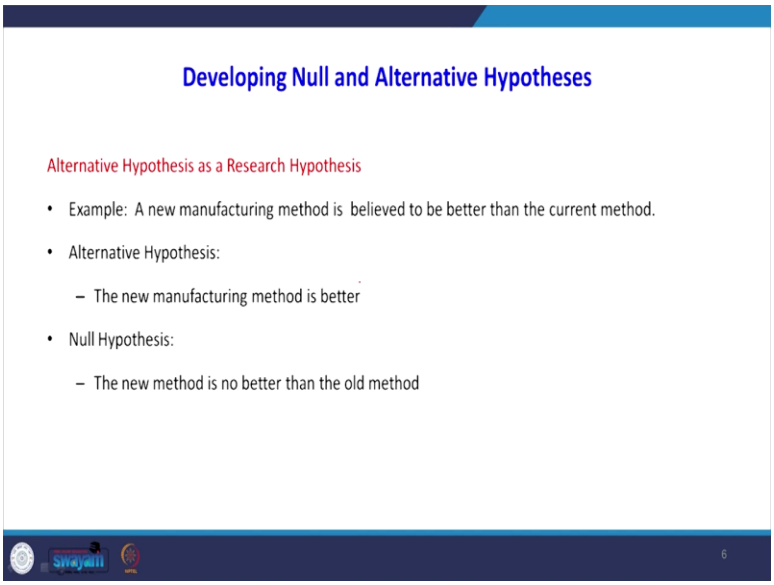
**(Refer Slide Time: 03:36)**




First you start with alternative hypothesis as a research hypothesis. Most of the time the researchers wanted to prove the alternate hypothesis. Many applications of hypothesis testing involve an attempt to gather evidence in support of research hypothesis. In such cases it is often best to begin with the alternative hypothesis and make it conclusion that the researcher hopes to support it because many of the time the researchers wanted to support his hypothesis.

So first you have to write the alternative hypothesis. The conclusion that the research hypothesis true is made if the sample data provide sufficient evidence to show that the null hypothesis can be rejected so, if we want to accept your alternative hypothesis the data which we are collected from the sample has to support the null hypothesis to reject it.

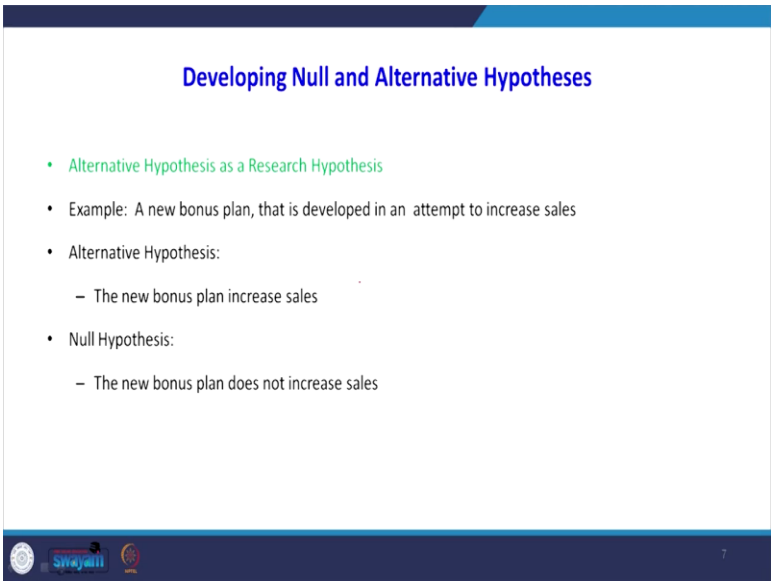
**(Refer Slide Time: 04:36)**




Next alternative hypothesis as a research hypothesis example we will see some examples here in example is a new manufacturing method is believed to be better than the current method. Assume that in your manufacturing context some is proposing a new way of doing work a new manufacturing method. So, we want to test this assumption so what is alternative hypothesis the new manufacturing method is better because that new method was given by the researcher always the researcher will believe that whatever he says is there is a support for that.

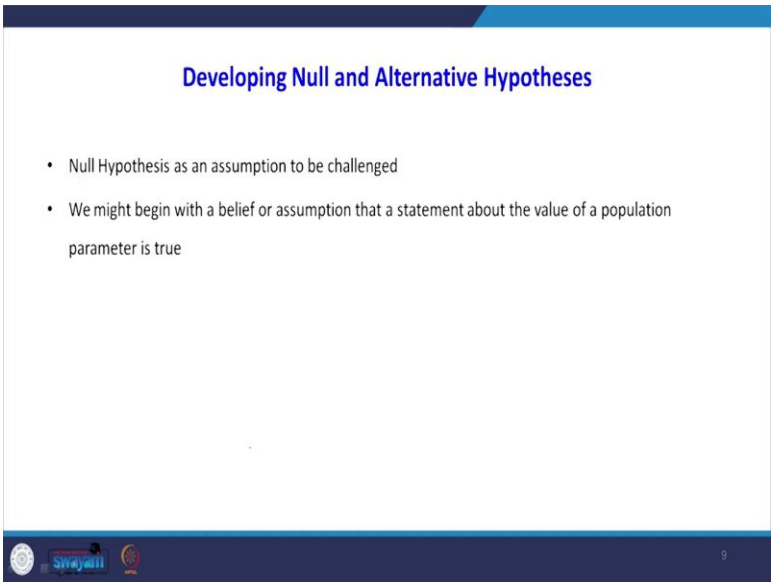
So, first we will formulate to the alternate hypothesis that is the new manufacturing method is better, the null hypothesis just to the complement of alternate hypothesis so the new method is no better than the old method.

**(Refer Slide Time: 05:26)**




We will take another example a new bonus plan that is developed in an attempt to increase the sales. Now what is happening any other organization so we are introducing new bonus plan we are going to see that the bonus has any impact on the sales. So, what is alternative hypothesis is the new bonus plan increases sales. So, first to go for alternate hypothesis then what is the null hypothesis the new bonus plan does not increase the sales you see that whenever when you look at the null hypothesis it will say always does not increase the sales. That is why it is called the null nothing has happened there is a meaning of null.

**(Refer Slide Time: 06:13)**




We will go for another example of alternative hypothesis see a new drug is developed with a goal of lowering cholesterol level more than the existing drug. So, what is happening there are

already there are some drugs are available to lower the cholesterol that a researcher has found some drug so that is reducing the cholesterol better than the existing medicine drug.

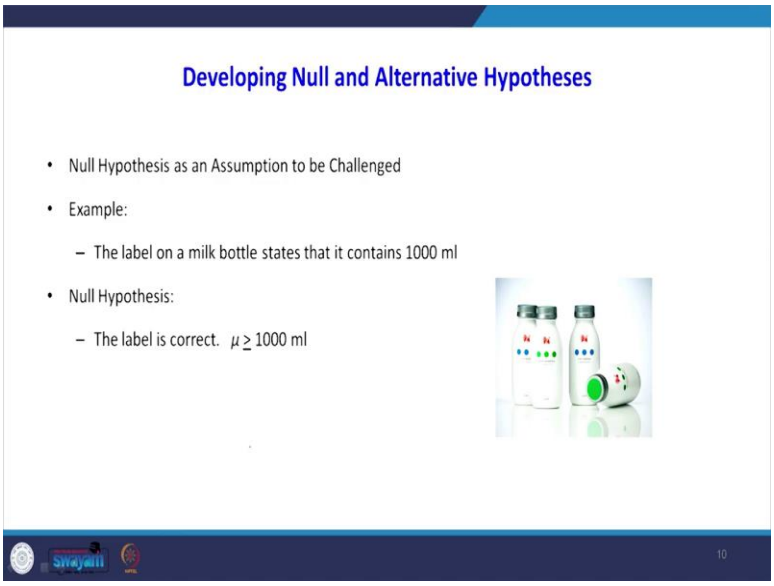
So, we will go for alternate hypothesis the new drug lowers the cholesterol level more than the existing drug. So, null hypothesis the new drug does not lower the cholesterol level more than the existing drug. You see that that does not so this does not represent the null so nothing significance has happened that is why we are calling it is null hypothesis. Then null hypothesis an assumption to be challenged.

We might begin with the belief or assumption that the statement about the value of the population parameter is true. So, in the hypothesis testing context always we will start the problem assuming that the null hypothesis is true. For example in India before starting a trail suppose somebody was accused so before starting a trial we the trial will be started assuming that the person is innocent person.

You see what is happening the trial will be started assuming that the person is innocent the police has to bring some evidence and they have to say that it is not innocent. When in other countries the person who is being suspected he has to prove his innocence. So, it is reverse so what is the meaning of this reverse that even though something has happened if there is no evidence that person is free.

We then using a hypothesis test to challenge the assumption and determine if there is a statistical evidence to conclude that the assumption is incorrect. In this situation it is helpful to develop the null hypothesis first. We will take an example of how to develop your null hypothesis. A null hypothesis is an assumption to be challenged.

**(Refer Slide Time: 08:23)**

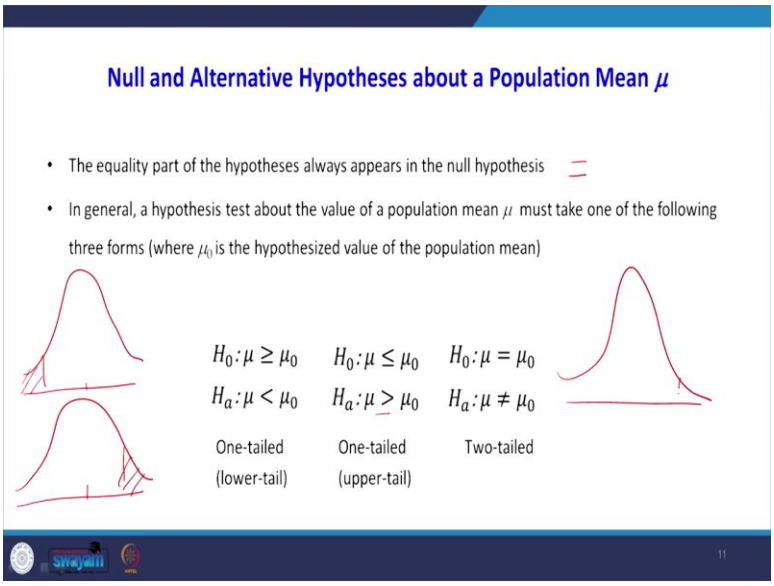



Example you see little the label on your milk bottle states that it contains 1000 ml null hypothesis the label is correct, so $\mu \geq 1000$ ml. Another hint is the null hypothesis is nothing but the status quo in null hypothesis always there will be '=' sign the null hypothesis is looked at a optimistic perspective. If somebody say the bottle contains the 1000 ml we are assuming that yes that assumption is correct so we formulating the null hypothesis is $\mu \geq 1000$ ml.

So, alternate hypothesis the label is incorrect $\mu < 1000$ ml you see that the signs are complementary since the null hypothesis it is greater than or equal to there is here less than. If the null hypothesis is less than or equal to so the alternate hypothesis it will be >. The null hypothesis is '=' the alternative hypothesis is '≠'. And the null hypothesis always will have equal to sign alternate hypothesis never contained equal to sign.

The status quo will go to null hypothesis we have to challenge the status quo that is nothing but your alternate hypothesis.

**(Refer Slide Time: 09:49)**




Okay you see how the nature of the null hypothesis. The Equality part of the hypothesis always appear in the null hypothesis that means in null hypothesis always there will be a equal to sign. So, when is equal to means it is equivalent to null nothing has happened that is the status quo is maintained as it is. In general the hypothesis test about the value of the population mean mu must we take one of the following 3 forms where $\mu_0$ is the hypothesis value of the population mean.

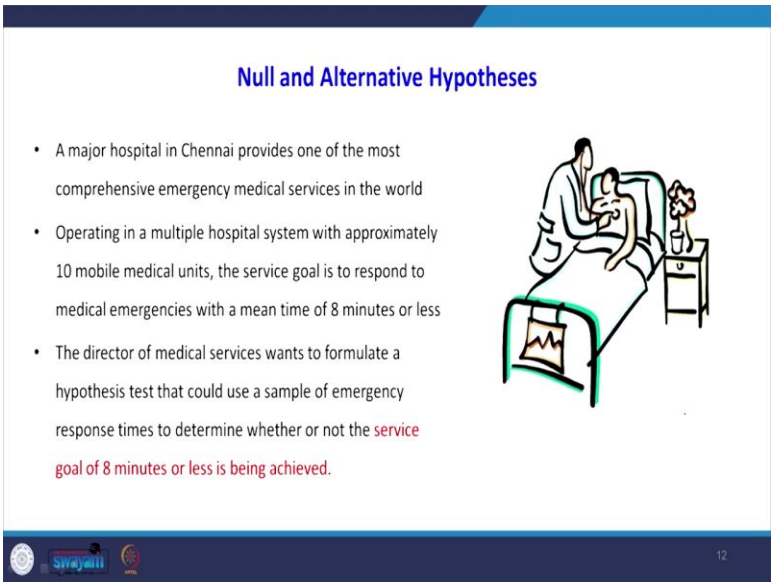
You see that the hypothesis may take different forms for example μ greater than or equal to $\mu_0$ the $\mu_0$ is what do you have a similar population mean. You see that the null hypothesis there is a greater than or equal to so we are writing in the alternative hypothesis less than because the signs are complementary. So, this test is one tailed test that is called lower tailed test. So, how we are calling it to lower tailed test is for example if I am drawing here.

We have to look at the sign of your alternate hypothesis. The sign of for alternate hypothesis is less than $\mu_0$ so it is left tailed test. If anything goes beyond the left hand side we will reject it. See there is another context μ, H0: μ less than or equal to $\mu_0$ so alternative hypothesis is μ greater than $\mu_0$ here also you look at this this is a less than or equal to so complement sign is greater than.

So it is one tailed it is called per tail test look at the sign of our alternate hypothesis it is greater than, so if it is greater than so it is called right tailed test. If anything beyond this point suppose

this mean beyond this point will be rejected the last one is equal to sign μ equal to $\mu_0$, so Ha: μ not equal to $\mu_0$, this is called a two tailed test. So, two tailed test is you see that the rejection area will be on both side so if the value goes below this will reject it the value goes about this rejected what is the meaning of value.

**(Refer Slide Time: 12:11)**

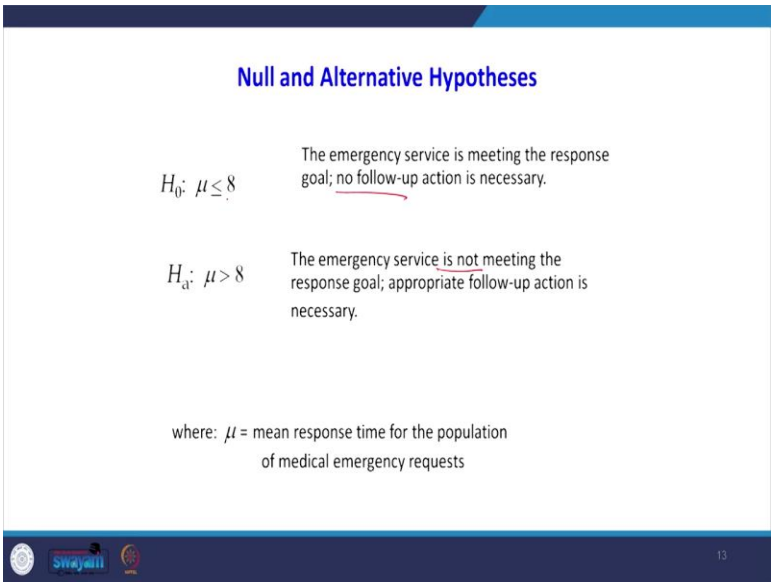



I will explain we will take an example to do a hypothesis testing a major hospital in Chennai provides one of the most comprehensive emergency medical services in the world. Operating in here multiple hospital system with approximately 100 mobile medical units that hospital is having 100 it is not 100, it is 10 mobile medical units the service goal is to respond to medical emergencies with a mean time of 8 minutes or less.

So, the problem is that they have 10 mobile medical units they are too whenever there is a emergency they have to respond 8 minutes or less. The director of medical services want to formulate a hypothesis test that could you see a sample of emergency response times to determine whether or not the service goal of the goal of 8 minutes or less is being achieved. Look at this problem the director wanted to test the service goal of 8 minutes or less is being achieved.

See now it is like here alternative hypothesis the researchers wanted to test whether the service goal of 8 minutes or less is achieved. So, what will happen now the status quo the status quo is 8 minutes or less.

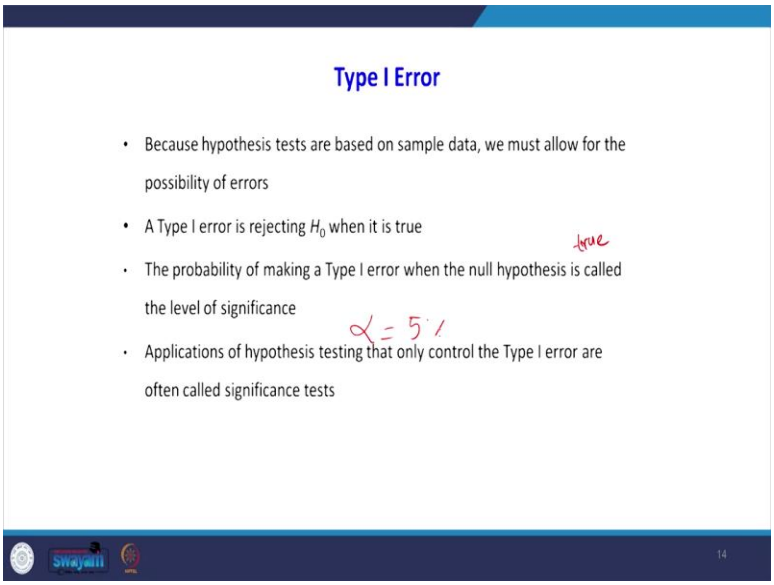
**(Refer Slide Time: 13:40)**




So what happened the status quo will go to null hypothesis so what is the null hypothesis the emergency service meeting the response goal. So, no follow-up action is required the another name why it is called null hypothesis is when you accept a null hypothesis no follow-up action is required, no course of action is required. So, why we are saying mu less than or equal to 8 because that is the status quo, so, always null hypothesis null hypothesis look at at their optimistic perspective.

So when I say $\mu \leq 8$ you see that the opposite of this what is that the emergency service is not meeting the response goal that is appropriate follow-up action is necessary that is why it is called alternate hypothesis, so mu greater than equal to 8 you see that here it is a less than or equal to 8. So, the sign is complimentary it is greater than equal to 8 while we are writing mu less than or equal to the status go should go to null hypothesis okay where the mu is the mean response time for the population of medical emergency request.

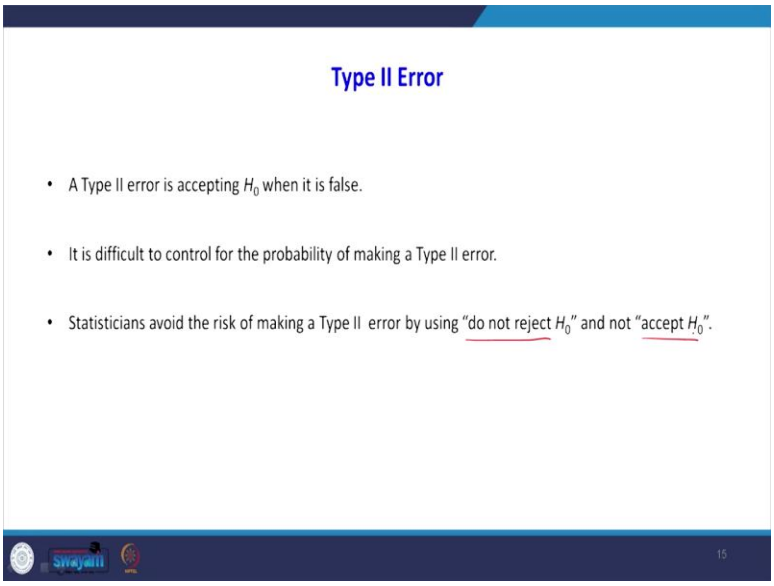
**(Refer Slide Time: 14:49)**




So, we will go to what is a type 1 error because hypothesis tests are based on the sample data we must allow for possibility of errors because the conclusion of hypothesis that is to accept a reject is based on sample data. So, always there is a possibility of error. Here type 1 error is rejecting H0 when it is true, as he told you in the code context somebody is pleading that is innocent but the judge is not accepting his innocent but really is innocent but he was his innocence was rejected that is incorrect rejection, that is a type 1 error.

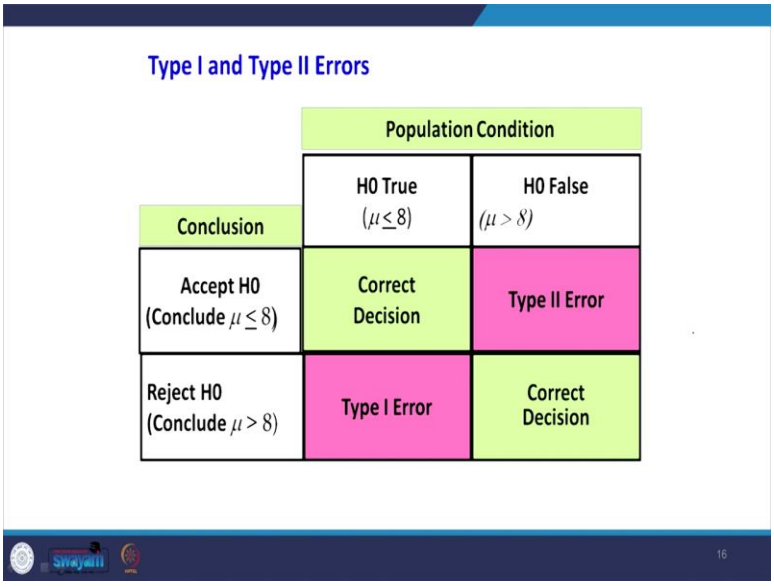
The probability of making a type 1 error is when the null hypothesis is true when the null hypothesis is called the level of significance. So, level of significance we call it is alpha most of the time it is 5% what is the meaning of this 5% is the probability of incorrectly rejection is only 5%, application of hypothesis testing that one that only control the type one error are often called significance test.

**(Refer Slide Time: 16:10)**




Type 2 error a type two error is accepting H0 when it is false, it is difficult to control the probability of making a type 2 error status easy and avoid there is a risk of making type 2 error by using do not reject H0 instead of accept null hypothesis because in the hypothesis context when we concluded we will not say accept null hypothesis we will say do not reject null hypothesis. Because there is no proof for that null hypothesis is true.

**(Refer Slide Time: 16:53)**




See the context see the population condition is H0 is true you see that in the conclusion H0 is true we will see this when you reject H0 that is called your type 1 error so that is called incorrect rejection. You see the other case the H0 is false but you have accepted so that is called your type 2 error. So, another name for a type 1 error is incorrect rejection, for type 2 error it is false

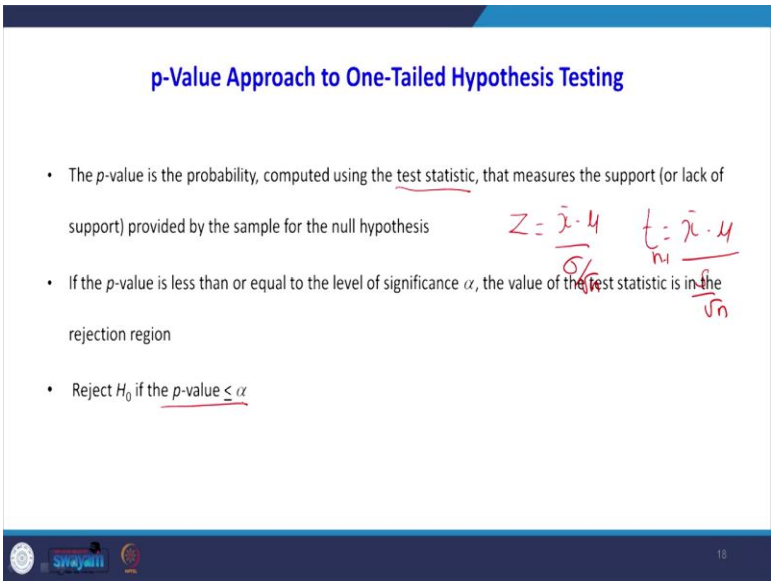
acceptance. We can say another example the producer risk we call this alpha, alpha is called type 1 error, beta consumer risk is called type 2 error.

What is the meaning of this producer risk and consumer is case assume that I am the manufacturer I am producer I am producing shaft, so whose diameter is for example the shaft diameter is say 50 mm. Suppose there is a supplier is coming the supplier has taken some sample from my production lot then he is rejected my lot, he says that you were your production level is not meeting our specification that is 50mm.

There is a 2 possibilities there the supplier who has the way he measured is wrong otherwise I made the sample which have kept is not correct. So, that is incorrect rejection even though I have quality good products they have rejected that is an incorrect rejection that is called to produce a risk. So, there is another possibility assume that I am making only 49 mm of shaft again the supplier came he measured is 50, it is 49 but he is measure it is 50 then he has accepted my lot so that is false acceptance that is called a type 2 error.

There are two possibility one is the sample which I have kept that meet all his requirements but my whole lot does not meet meeting his requirement. So, that means my sample is not the representative of the population that is one possibility otherwise the way they have measured it that is wrong, so that is called false acceptance that is a type 2 error. In the next lecture we will see the application of type 2 error in detail.

**(Refer Slide Time: 19:29)**

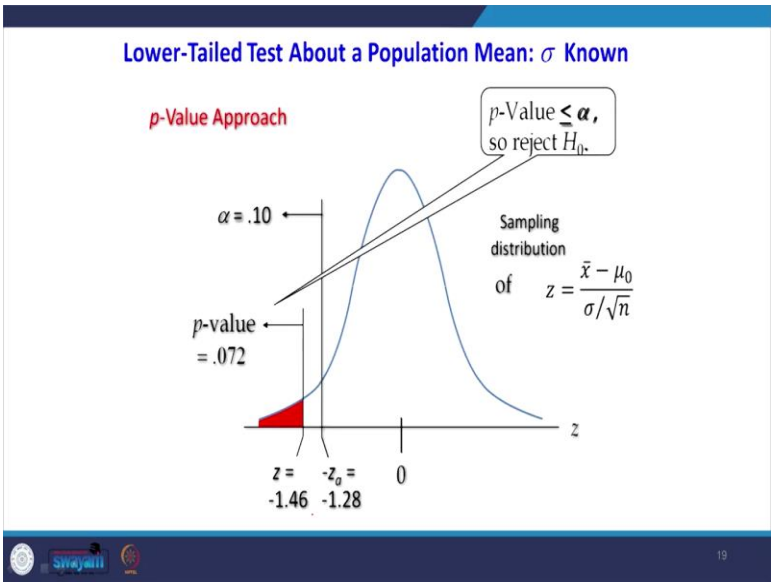



There are 3 approaches for hypothesis testing first approach is p-value approach most of the statistical package follow this method. Second method is critical value method, the third one is confidence interval value method. The confidence interval value method mostly used for 2-tailed test. First we will go for p-value approach that is a one-tailed hypothesis testing. What is the p-value the p-value is the probability computed using test statistic. You should be careful test the statistic that measures the support our lack of support provided by the sample for the null hypothesis.

So the p-value says whether it is supporting the null hypothesis or it is not supporting null hypothesis if the p-value is very high it will support null hypothesis you will accept null hypothesis. If the p-values be less it will not support null hypothesis we will reject the null hypothesis. Say we say that what is the test statistic the test statistic? For example in the Z context the test statistic nothing but this one (X bar – mu) /(σ/√n) that is the test statistic for Z test if it is a t-test this is (X bar – mu)/( s /√n).

So, n - 1 degrees of freedom, so, whatever value which have calculated with the help of sample that is called a test statistic, if the p-value is less than or equal to the value of the alpha then the value of the test statistics in the rejection region. I will show you in the next slide. Reject H0 if the p-value is less than alpha. The p-value is very less it is not supporting null hypothesis here to reject it where alpha is significant level.

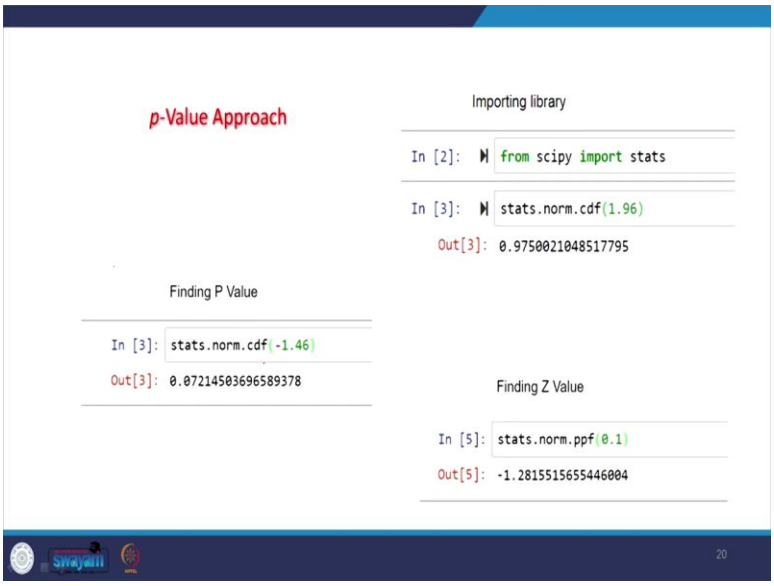
**(Refer Slide Time: 21:36)**




We will see how to use hypothesis, how to do hypothesis testing using p-value approach. The p-value approach the first one is, see assume that the problem alpha equal to 10% it is given this was alpha so we have to calculate this test statistic that that is your Z value. So, X bar might be given X bar is the sample mean, minus μ is the population mean what we have assumed Sigma value must be given, root of n.

For example this value assumed that it is -1.46 okay. So, this -1.46 corresponding what is the left side area. So, this value is our p-value, okay how to get this one so when Z value is -1.46 we can get corresponding area of your normal distribution on the left hand side.

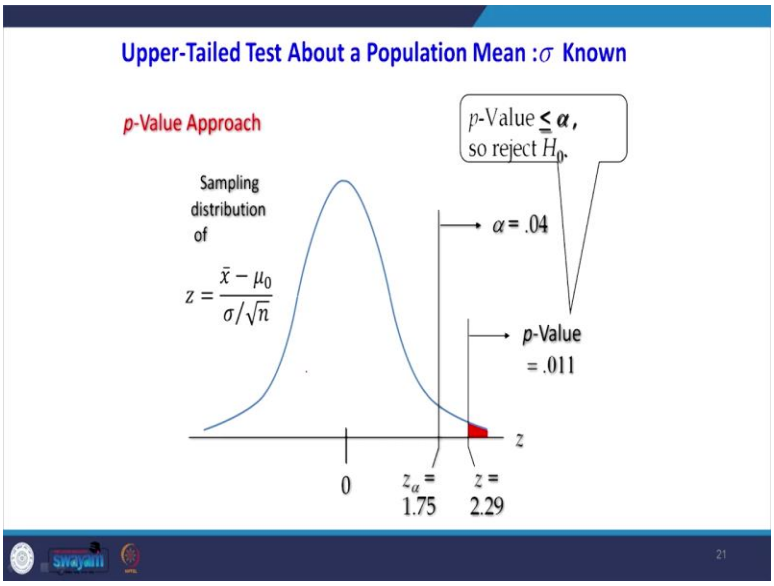
**(Refer Slide Time: 22:39)**




So, that one we can do with the help of Python for that first you have to import scipy so importing library from scipy import stats. Then the left side area is say, -1.46 the left side is Z statics -1.46 so when you put a minus this one stats.norm.cdf cumulative distribution function -1.46 we are getting the probability is 0.07 so that is nothing but 0.07 you see that this alpha is 10% so the p-value is less than the Alpha so way out to this region is a rejection region.

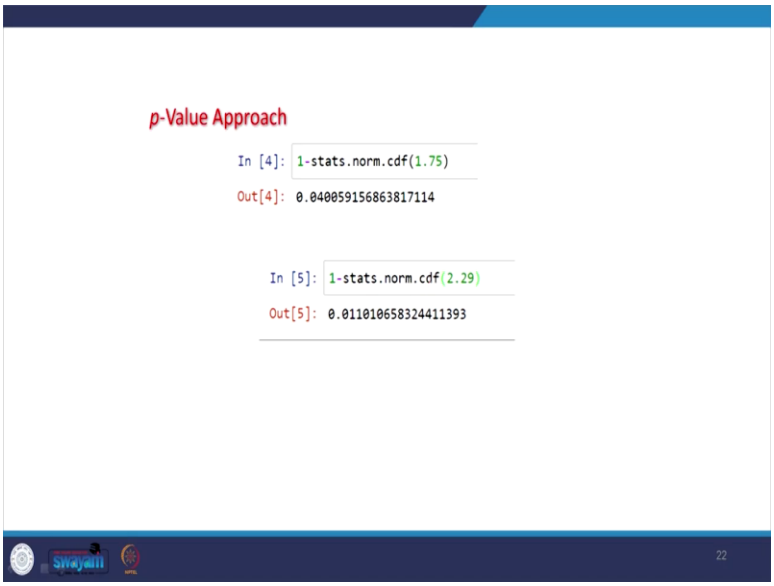
So, this region is acceptance region beyond this point it is the rejection region. So, the value of the P that is when Z value equal to -1.46 since we are standing on the left hand side that is we are standing on the rejection side we have to reject the null hypothesis. This is left side test lower tailed test this way explain.

**(Refer Slide Time: 24:00)**




Now so suppose if it is a right tailed test say the calculated Z value is 2.29 we got some X bar value, μ value, Sigma by root n value, so suppose this is giving 2.29 so for alpha assume that alpha equal to 4% when alpha equal to 4, we go to mark it, alpha equal to 4% from the right to left. So, when alpha equal to 0.4% when Z values 2.29 we would what is the corresponding area towards the right side so what you have to do we can find out 1.75 also see that when Z values 2.29.

**(Refer Slide Time: 24:46)**

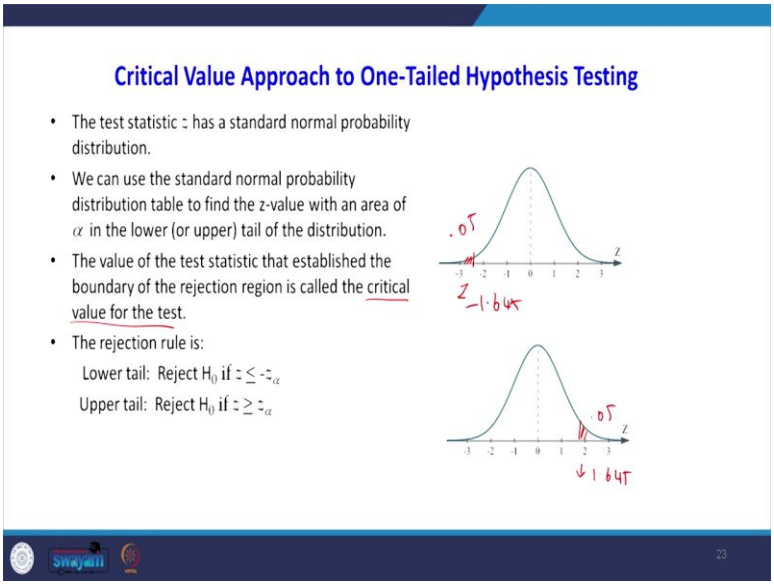



So stat.norm.cdf will give you the right side area when you put 1 – stats.norm.cdf (2.29) will give you the left side area. So, this first one actually it is not required here because alpha will is directly given here this is only for proof this is for testing, how to use from Z, how this is Z value

from that we have find out the probability value. So, now the Z value is 2.29 so we want to know the right side area so 1 minus corresponding area that will give the right side area that 0.011.

So, this area I am saying this area is 0.011 now look at this alpha so the p-value is less than the alpha otherwise you see the p-value so this side is the rejection region this side the acceptance region. So, when the p-value is 0.01 still you are standing in the rejection region so how to reject a null hypothesis. In case if the p-value is 0.05 you might crossed the boundary after crossing the boundary you will be landing on the acceptance region so we have to accept a null hypothesis.

**(Refer Slide Time: 26:07)**



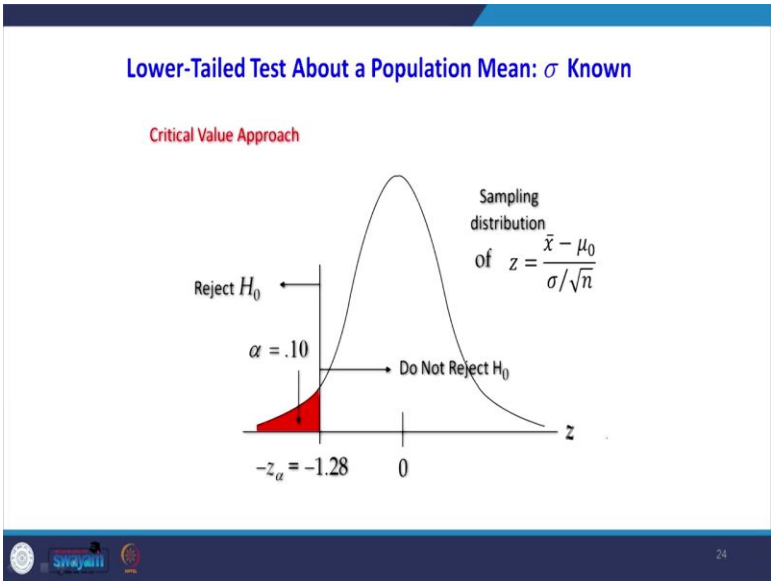

I will go to another method critical value approach for one tailed hypothesis are testing. The test statistic Z has the standard normal probability distribution we can use the standard normal probability distribution table to find out the Z value with an area of alpha in the lower tail or upper tail of the distribution. For example we know the Alpha value, say alpha value is 0.05, so this side area is 0.05 with the help of Python when alpha is 0.05 you can get the Z value this is lower tailed test.

For upper tail test when alpha equal to 0.05 you can find out corresponding Z value. In Python what you have to do if you want to know this right side, upper tail test you want to know the Z value you have to is 1- 0.05 for that probability you were to find out corresponding Z value. The

value of the test statistic that established the boundary of the rejection region is called critical value of the test. If it is for the 5% age you will get 1.645 here also -1.645.

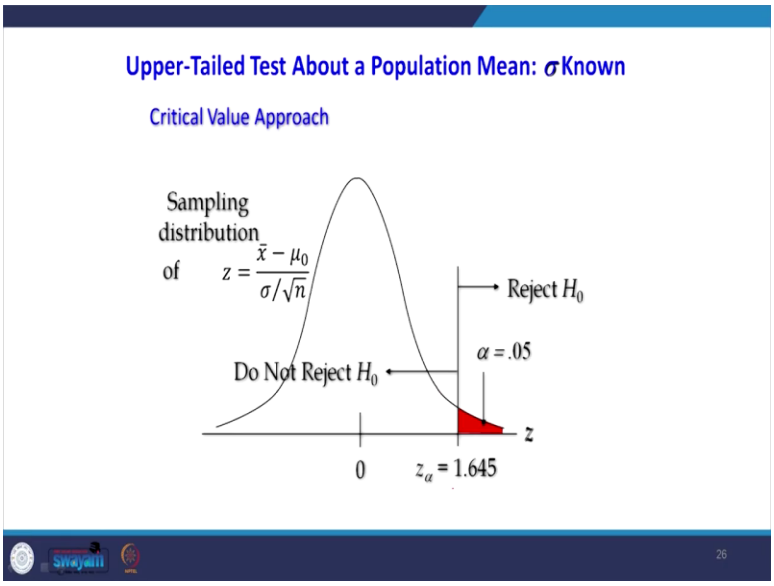
So, this -1.645 is called a critical region. Rejection rule if it is a left tailed test reject if the Z value that means your calculated Z value is less than this -1.645 because you will be standing on the rejection side. If it is a right tailed test the calculated Z values greater than your table value then you have to reject it.

**(Refer Slide Time: 27:57)**



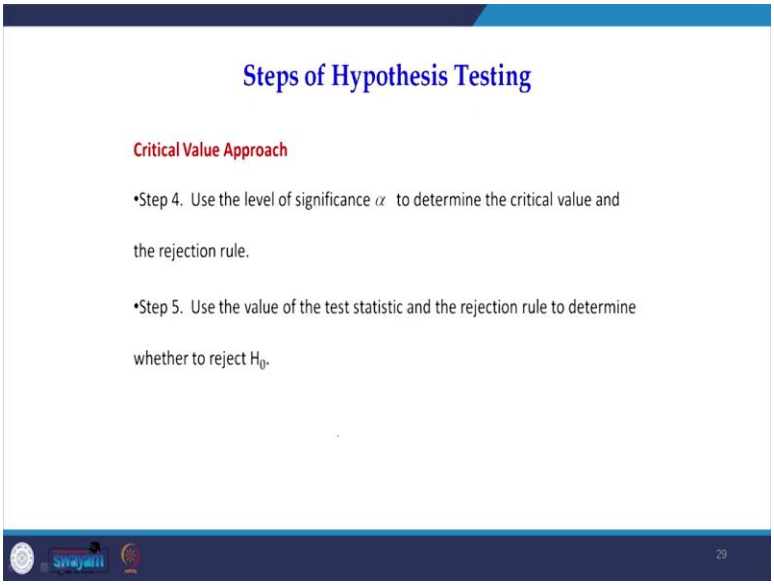

See for example Sigma is known the Alpha equal to 10% when alpha equal to 10% corresponding Z value is -1.28 this is our critical region this is our critical region. So, with the help of sample data you have to find out the Z value if the Z value is lying on this side you have to reject it. If the Z value is lying on that side you have to accept it.

**(Refer Slide Time: 28:25)**




For example when area equal to 0.1 the corresponding Z value is -1.28 and going back to previous slides -1.28. Now we will go for upper tail test when alpha equal to 0.05 so the right side area is 0.05 so this side is 0.95, so when the left side here is 0.95 corresponding Z values 1.645. If any calculated Z if this Z value is lying on this side for example 1.7 you have to reject the null hypothesis but is lying this side you have to accept the null hypothesis.

**(Refer Slide Time: 29:01)**




For example 0.95 the value is 1.65 what is the see how, now I now we have done hypothesis testing with help of p-value approach and critical value approach what is important point you to note that is the decision whether to reject or accept a null hypothesis in p-value method is decided by comparing the probability. Probability of what is the probability P value versus alpha

value. But in your critical value approach the decision is done by comparing the critical value and calculated Z value.

Decision will be same, only for comparison purpose sometimes we compare probability sometimes we compare critical value but the end result will be same. So, what is the first step will develop null and alternative hypothesis step 2 specify the level of significance alpha this is very important. Before starting of the test you have to decide the significance level. Step 3 collect the sample data and compute the test statistic.

This test statistic maybe your t value or it may be z value. The p-value approach what will you do use the value of test statistics to compute the p-value, if the p-value is less than or equal to alpha you rejected the same step for critical value method use the level of significance alpha, to determine the critical value and rejection role. Use the value of test statistic and the rejection rule to determine whether to reject H0.

Dear students in this lecture so far what we have seen we have seen what is hypothesis what does the null and alternative hypothesis. We have learnt how to formulate hypotheses then we have seen hypothesis testing. In the hypothesis testing we have seen what is left tail test, what is the right tail test. What is the two tail test then we have seen the theory of how to test the hypothesis by using p-value approach and by using the critical value approach.

The next class will take one problem with the help of that problem will formulate the hypothesis then we will test the hypothesis with help of p-value approach and critical value approach then we will compare the result. And one more thing one more method that I did not cover in this lecture is that is testing the hypothesis with the help of confidence interval that we will do in the next class, thank you very much.