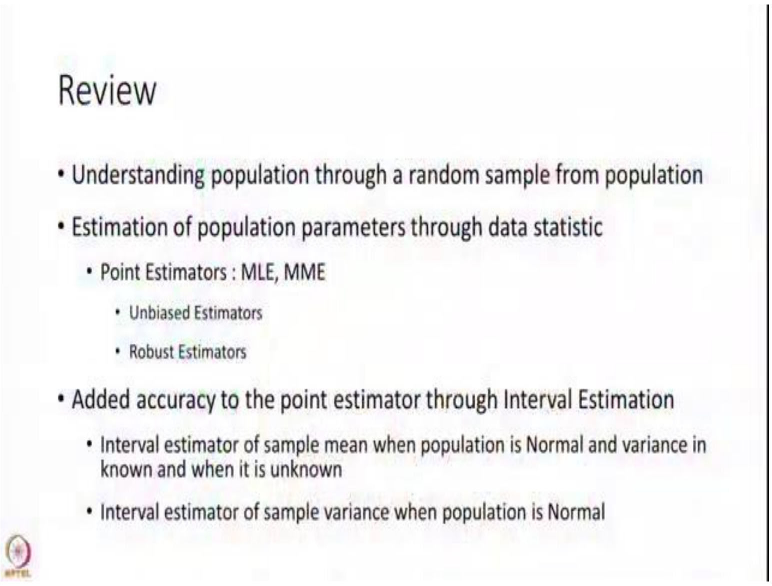
Dealing with Materials Data: Collection, Analysis and Interpretation
Professor M P Gururajan
Professor Hina A Gokhale
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay
Lecture 68 - Hypothesis Testing 1

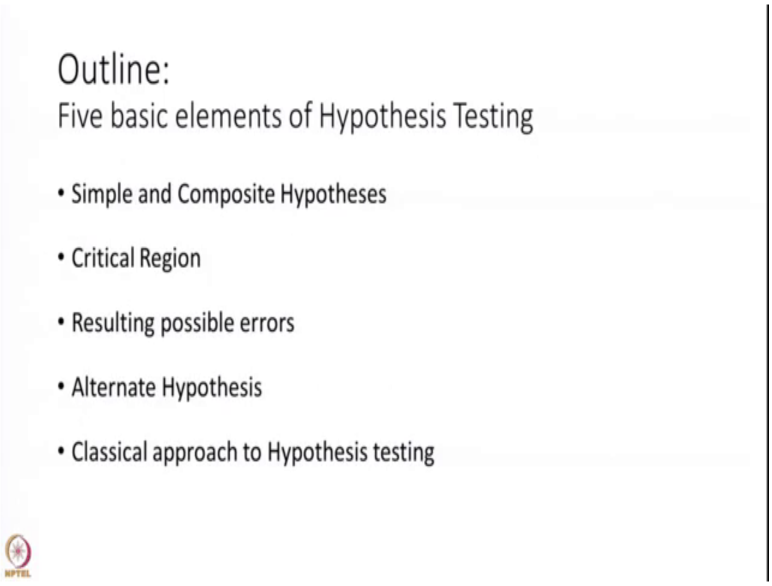
Hello and welcome to Dealing with Materials Data course. For past few sessions, we are trying to learn about the population through several statistical methods. Now, what we are going to work with is called hypothesis testing.

(Refer Slide Time: 0:40)



What we did in the past was our endeavor was to learn about the population from a small sample. So, what we did? We did the estimation of population parameter. We assumed that the data comes from a population with a certain distribution function and then that distribution function had certain parameters which were unknown. So, if we know the parameters, we know the population itself. So, we learned several techniques of estimation of distribution parameters, which is known as parametric estimations. We had point estimators and we added the accuracy to it by having an interval estimate.

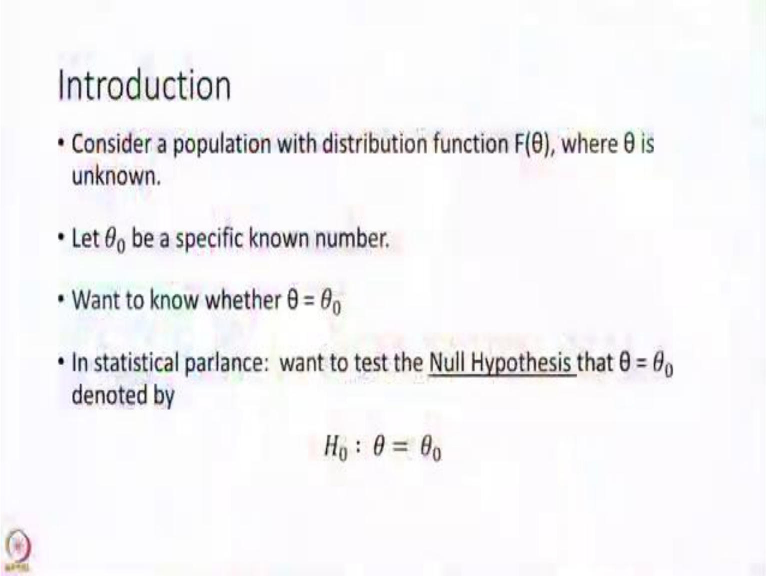
(Refer Slide Time: 1:25)



What we want to do now is what is called hypothesis testing. What does it mean by hypothesis testing? Well, if you have the data and you know that it comes from a certain population and you also know that, that population has a certain mean. Let us take an example. Suppose, a person wants to buy or a company wants to buy a certain kind of an alloy and it is interested in having that alloy should have some good property of yield strength. The company tells him the, that the supplier tells them that I guess our alloy has huge strength value Y, how does the buyer know the customer know that truly what they are getting as a consignment has the yield strength Y?

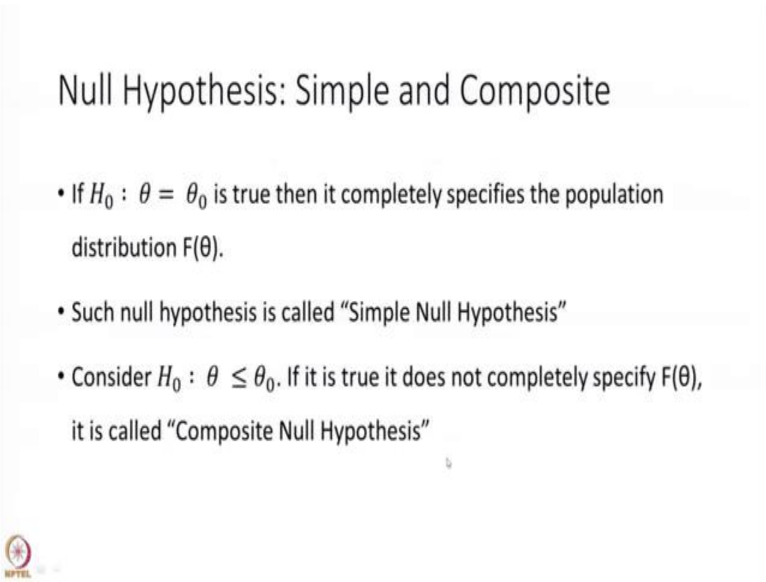
So, what they would like to do is take a sample of from the consignment, test their yield strengths and then check or test if this yield strength is same as y, this is called hypothesis testing. So, in this today's session, we are going to look into 5 basic elements of hypothesis testing. The first one is simple and composite hypothesis. The second one is a critical region, resulting possible errors while you do hypothesis testing. What is called an alternate hypothesis and classical approach to hypothesis testing?

(Refer Slide Time: 3:24)



Consider we, as I said let us have a population with a distribution function f theta where theta is the parameter and it is unknown. If we know theta, we know the distribution, we know the population then there is no question of doing any statistics but we do not know theta. Let theta not be a specific known number or known value. So, as I said in the previous example, if what I said is that the company claimed that our material has yield strength Y then this Y is theta naught then we would like to know whether the unknown parameter theta is truly theta naught. In statistical parlance, we said that we want to test the null hypothesis that theta is equal to theta naught and it is denoted by a null hypothesis so it is called h 0 is theta is equal to theta naught.

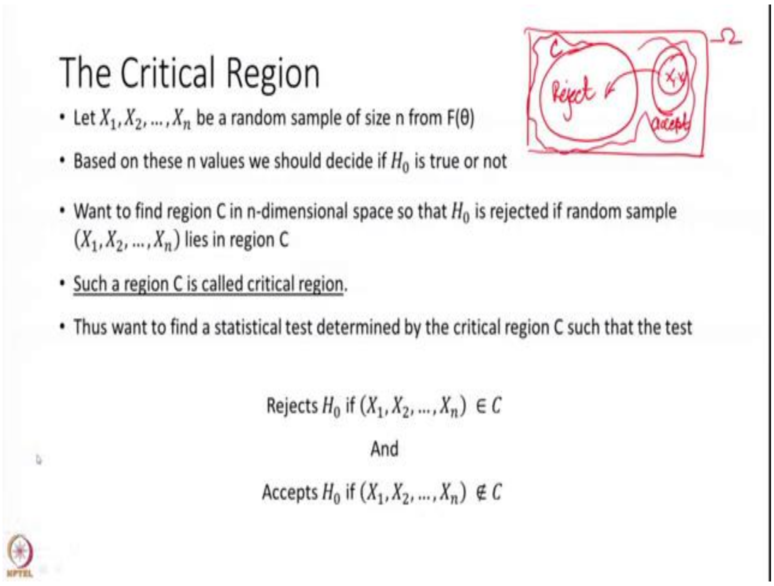
(Refer Slide Time: 4:33)



This null hypothesis there are 2 kinds, if the null hypothesis of the kind that h 0 says that theta is equal to theta naught then if you actually put theta equal to theta naught then it completely it would specify the population distribution. Again, you know the whole population what the population behaves like. So, if we say that if null hypothesis is true and if it completely specifies the population, it is called a simple null hypothesis.

While if you take another hypothesis which says that, for example company may say that my yield strength is at the most so much. It means that your null hypothesis is that theta is less than or equal to theta naught. And if this is true, it does not completely specify the distribution and therefore the population and therefore, it is called a composite null hypothesis. Next, what we would like to do is we want to find a critical region.

(Refer Slide Time: 5:51)

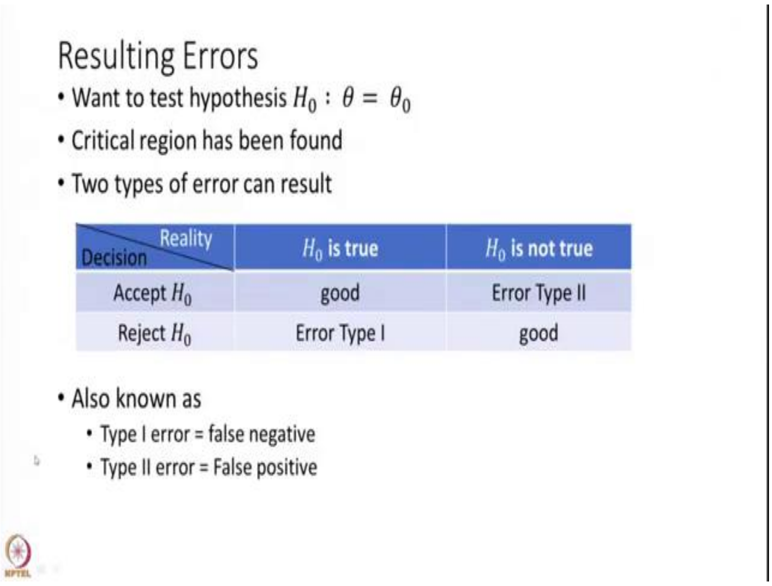


Now, here what really we would like to do is something like this, we have a sample space, please do not forget this old terms, we have a sample space which is a result of all the experiments that you can conduct, all possible results of the experiment that you conduct. So, here for example, if we are continuing with the yield strength and all possible yield strength that it can have will sit in this. Now, X 1, X 2, X 3, X n is a random sample. So, it will give you some value of this is your sample X 1, X 2, X 3, X n, it is some set within the sample set. We would like to define a region within the sample space, which I called C in such a way that if this sample belongs to this critical region, I can confidently say that reject the null hypothesis.

I would say that reject the null hypothesis. I am defining a critical region to reject the null hypothesis. If it is not in this and if it is sitting outside, I said it okay. I do not have enough evidence; I cannot reject the null hypothesis or accept the null hypothesis. Why are we going on the rejection? In day-to-day life you see to accept anything requires a lot of justification, lot of groups and lot of convincing, rejecting anything is much easier. So, is true even in terms of statistics and mathematics. Therefore, we try to find a region which will reject the null hypothesis.

So, here in this region, we are going to reject the null hypothesis, everywhere else we are going to accept. Generally, in statistics, we do not say we accept it, we only say that we do not have sufficient reasons to reject it, because there is always a doubt had my sample been larger, had it been another sample, possibly it would have been rejected. And therefore, we always say that we did not have sufficient evidence to say that the null hypothesis can be rejected. So, this is how we define a critical region.

(Refer Slide Time: 8:40)



Now, when you do such a hypothesis testing, there are 2 kinds of errors that we can commit. Let us look at it in a systematic way. The reality says that null hypothesis is true or it says that null hypothesis is not true. Based on the hypothesis testing finding a critical region, we can take a decision, accept null hypothesis or reject a null hypothesis. If you accept the null hypothesis which is true, there is nothing to worry. Similarly, if you reject the null hypothesis when it is not true, there is nothing to worry, you have done the right thing. But, if we accept the null hypothesis when it is not true, it is called type 2 error.

And the another error is when you reject the null hypothesis when it is true, it is called a type 1 error. These days with respect to medical testing, it is very common to come across 2 terms.

One is called false negative and the other is called false positive. Type 1 error refers to the false negative, it is a negative answer wrongly given, it is a negative answer which is wrongly given. False positive is that it is a positive answer, you accept the null hypothesis but it is wrongly given. Therefore, it is called false positive, the type 2 error and type 1 error is called false negative.

At this point, I would also like to bring another issue, which is very common with the machine learning technique which is getting more and more prevalent in all subject areas of science and technology. We use something called neural networks. It is in a way, a very sophisticated curve fitting technique where we have a very highly nonlinear curve which we have a neural network as a tool to estimate it and what we generally do, we generally have a set, which we call the test set, the fitting set and then we have a test set. So, we take one set of observations, we do the iterations and we find a neural network model and then we have a test set in which we put that values in it and see that it truly fits within the boundaries or not.

We have something like a least squares error there and then we try to see that even the test set falls within the boundaries of the specified minimum error limits. In all this process, it is something which is not done is finding the type 1 error and type 2 error that effort should also, it is not very difficult to do. Not a part of this particular course, so it is not being covered. But it is just to bring it to your notice that in the euphoria of applying machine learning and therefore, using techniques such as neural network, please remember before using this heuristic method also devise a method to identify the 2 types of errors which can occur in any decision making procedure.

Any decision this is a decision making procedure and this is kind of a error table for a decision table. So, in such situations, even when you use any heuristic method, heuristic method is like neural networks or any other method. These two kinds of errors should also be tried to be evaluated. Anyway, let us come back to our present context of hypothesis testing.

(Refer Slide Time: 13:12)

Decision \ Reality	H_0 is true	H_0 is not true
Accept H_0	good	Error Type II
Reject H_0	Error Type I	good

- Would like to minimise both types of errors
- Mathematically not possible
- Classical approach is to
  - fix α at the minimum possible level and,
  - Set up test so that probability of Type I error of the test is ≤ α

These two errors frankly, we would like to minimize both the errors so that we do not commit any great error. Mathematically, it is not possible. The classical approach to hypothesis testing says that fix some value alpha at the minimum possible level and then set up a test so that your probability of type 1 error is smaller than alpha.

And then you study how much is the type 2 error. This is the classical approach to hypothesis testing. And with this, we have 2 values now, as you can see we are going to fix some value below which the type 1 error should be. And we will study the type 2 error accordingly later.

(Refer Slide Time: 14:13)

Significance Level and Power of Test

$$P[type\ I\ error] \leq \alpha \ \Rightarrow P[Reject\ H_0\ |H_0\ is\ true] \leq \alpha$$

$$\Rightarrow P[(X_1, X_2, \ldots, X_n\ \in C)|H_0] \leq \alpha$$

- α is called significance level of the test
- 1- α is called confidence level of the test

$$P[type\ II\ error] = \beta \ \Rightarrow \ P[Accept H_0|H_0\ is\ not\ true] = \beta$$

$$\Rightarrow \ P[Rejet\ H_0\ |H_0 is\ not\ true] = 1 - \beta$$

$$\Rightarrow P[(X_1, X_2, \ldots, X_n\ \in C)|H_0\ is\ not\ true] = 1 - \beta$$

- 1- β is called Power of the test

So, in this case there are two important aspect come into picture, what we are trying to say is that you make probability of type 1 error smaller than alpha. It means that you are going to reject the null hypothesis when it is true, that probability you would like to make it less than alpha or it is same as saying that your sample will lie in the critical region when the null hypothesis true that probability is also less than alpha. Such an alpha is called a significance level of the test. The alpha level that you have fixed, it is called a significance level of the test. And 1 minus alpha is called the confidence level of the test.

$$P[type\ I\ error] \leq \alpha \ \ \Rightarrow P[Reject\ H_0\ |H_0\ is\ true] \leq \ \alpha$$

$$\Rightarrow P[(X_1,\ X_2,\ \ldots,\ X_n\ \in C)|H_0] \leq \ \alpha$$

α is called significance level of the test

1- α is called confidence level of the test

$$P[type\ II\ error] = \ \beta \ \ \Rightarrow \ \ P[AcceptH_0|H_0\ is\ not\ true] = \beta$$

$$\Rightarrow \ \ P[Rejet\ H_0\ |H_0 is\ not\ true] = 1 - \beta$$

$$\Rightarrow P[(X_1,\ X_2,\ \ldots,\ X_n\ \in C)|H_0\ is\ not\ true] = 1 - \beta$$

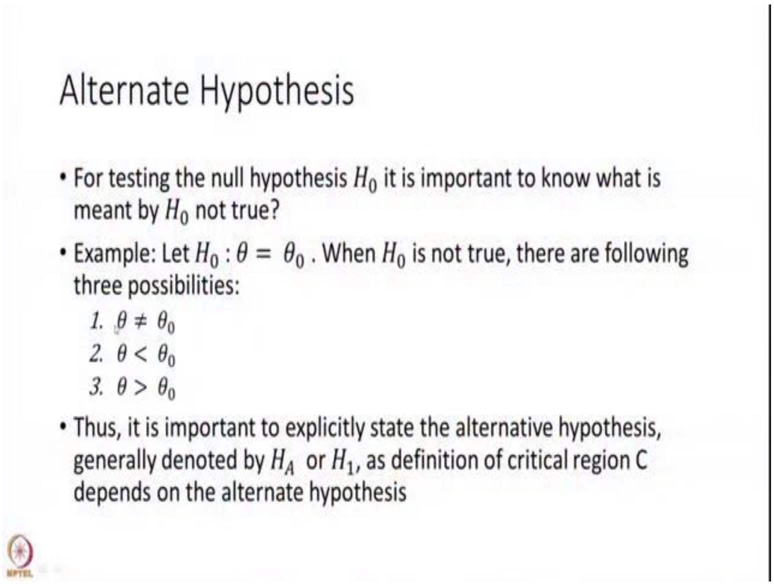
1- β is called Power of the test

Generally, in practice, you would have heard it and you might have used it in the past, this alpha is kept at 1 percent, 5 percent or 10 percent level. Please remember, this is kind of arbitrary, no specific meaning is attached to y 10 percent and y 1 percent and y 5 percent. But, important thing is before you do anything in the testing of hypothesis, the level of significance alpha is to be fixed. It should not get affected by looking at your further analysis, data analysis results, it should be fixed up right in the beginning. 1 minus alpha is called the confidence level.

So, if your alpha is 5 percent then you have a 95 percent confidence level of the test. Look at the type 2 error, probability of type 2 error is generally denoted by beta. It means that you are going to accept the null hypothesis when it is not true, it is same as reject the null hypothesis when hypothesis is not true is 1 minus beta. And therefore, we say that 1 minus beta is the power of the test that you are able to reject the hypothesis when it is not true that is called the power of the test.

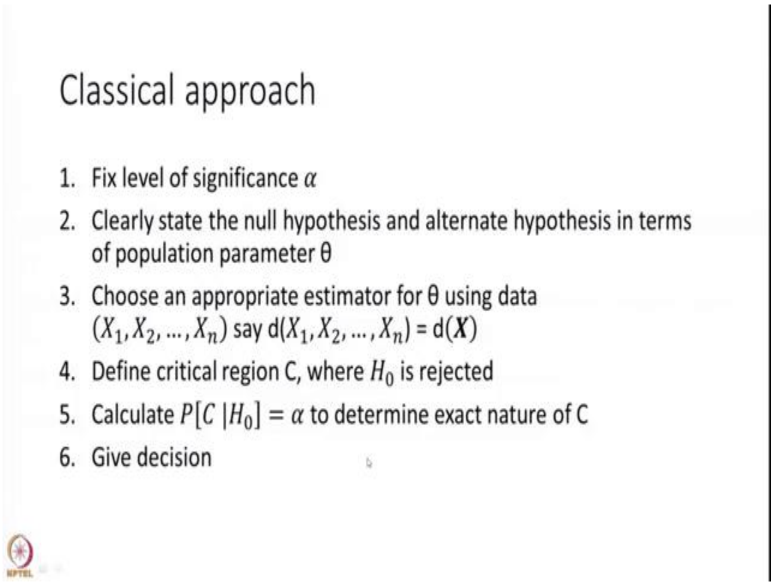
Now, when I say that okay, my hypothesis is that theta is equal to theta naught. Now, I define a critical region that where hypothesis is not true. So, when theta is not equal to theta naught, there are many alternatives.

(Refer Slide Time: 16:59)



It could be that theta is just not equal to theta naught. It is neither greater nor less or we can say that theta is not equal to theta naught but actually theta is smaller than theta naught or I can say that theta is not equal to theta naught but actually theta is greater than theta naught. These are the called alternate hypothesis. The notation says that just as we write null hypothesis as H naught or H_0, alternate is written as H_A or H_1, H_2, H_3, etc.

(Refer Slide Time: 17:33)

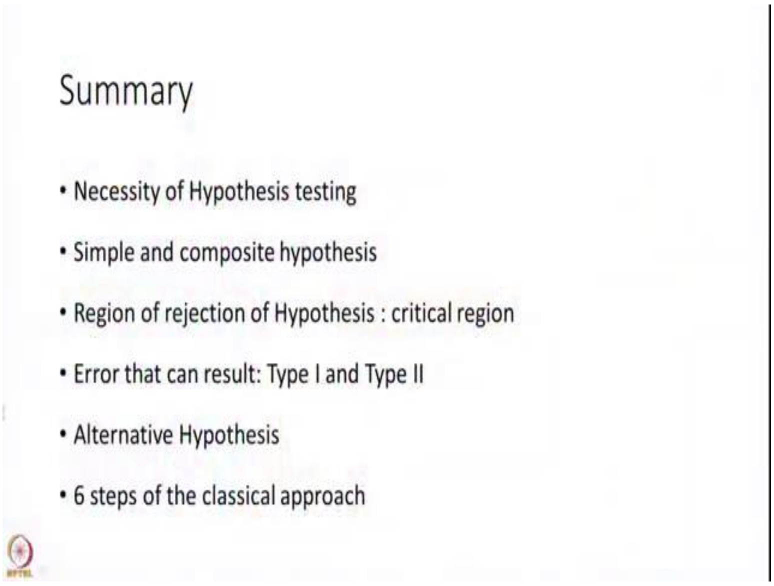


Now, what is I would like to give you a very basic classical approach to this hypothesis testing. It covers 6 steps. And I would like to suggest that whenever you do or solve any problem, please go one by one through each step because as you will see in future or when you do it yourself, you will realize that it kind of tends to get confusing in the middle. So, it is a good practice to follow this classical approach of 6 steps step by step. So, first step is fixed the significance level of alpha, very important. Before you do anything, significance level of alpha should be fixed.

Then clearly state what is your null hypothesis and what is your alternate hypothesis. Third, choose an appropriate estimator of theta using the data X 1, X 2, X 3, X n let us call it d, a function d of X 1, X 2, X 3, X n or we call it d X vector. Then we define a critical region using this d where H naught is rejected. Then we calculate the probability of critical region when H naught is accepted. We compare it with alpha and we determine the exact nature of C and then we know how to give the decision.

Once you know this probability, you know the nature of C, you know the probability and then it is simply says that if it is less than alpha or equal to alpha, you are going to reject it, otherwise you do not have sufficient evidence to reject it. So, you accept it. So, let us review this quickly.

(Refer Slide Time: 19:34)



We saw the necessity of hypothesis testing, that is there are situations where the estimation of a population parameter is not just sufficient, you have to check or you have to test whether it equals to the value which is pre-decided or already given to you, whether equates to that, it is smaller than that, it is greater than that whatever. If it defines a completely the your population for your population distribution by the null hypothesis, we call it a simple null hypothesis.

If it does not for example, null hypothesis could be that theta is less than theta naught. In that case, it does not define the population completely at the end of testing of null hypothesis and therefore, it is called a composite hypothesis. Then we said that it is much easier if we can find out a region in the sample space, where if our observed sample falls, then we can say that reject the null hypothesis. That is called a critical region. Then we said that, well what can happen, 2 types of errors can occur. The first type of error is when you reject the null hypothesis when in reality it is true, it is called false negative.

The other type of error is you accept the null hypothesis when it is not true, then it is called false negative. This, did I say false, it will be false positive. And then we talked about what are the alternate, I am finding I am proceeding the way that first I have a region where I reject the null hypothesis. So in that case, I must know if I reject the null hypothesis, what is the alternative? That alternative I call it alternative hypothesis and then we have defined 6 steps of classical approach.

In this summary, I would like to emphasize once again two things, the type 1 and type 2 errors are of extreme importance, they drive the whole classical approach and they should be driving even the new fresh approach that we take in variety of problems of decision making. These two errors should be taken into account that gives the strength to your test. Second thing, I would like to say is that the type 1 and type 2 error are a part of decision making procedure and therefore, they should be applied as and when whenever a decision procedure is taking place. However, once again I repeat, in this particular course, we are going to restrict the meaning and interpretation of type 1 and type 2 error only up to the hypothesis testing in the classical statistical manner. Thank you.