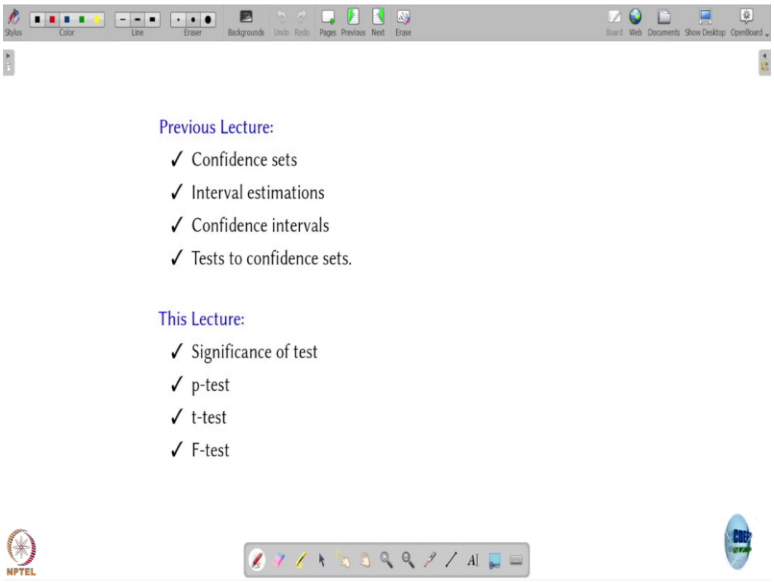
Engineering Statistics
Professor Manjesh Hanawala
Department of Industrial Engineering and Operations Research
Indian Institute of Technology Bombay
Week 11
Lecture 53
p-value, p-test of significance of a statistical test

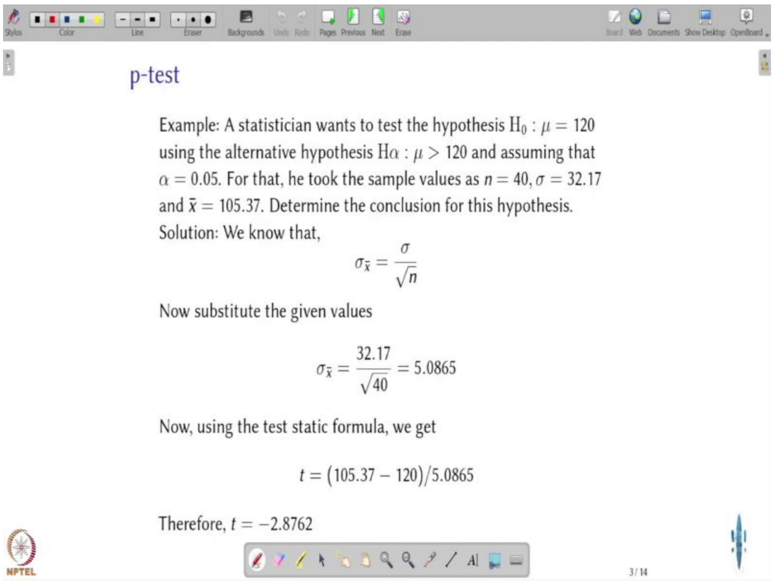
(Refer Slide Time: 00:27)

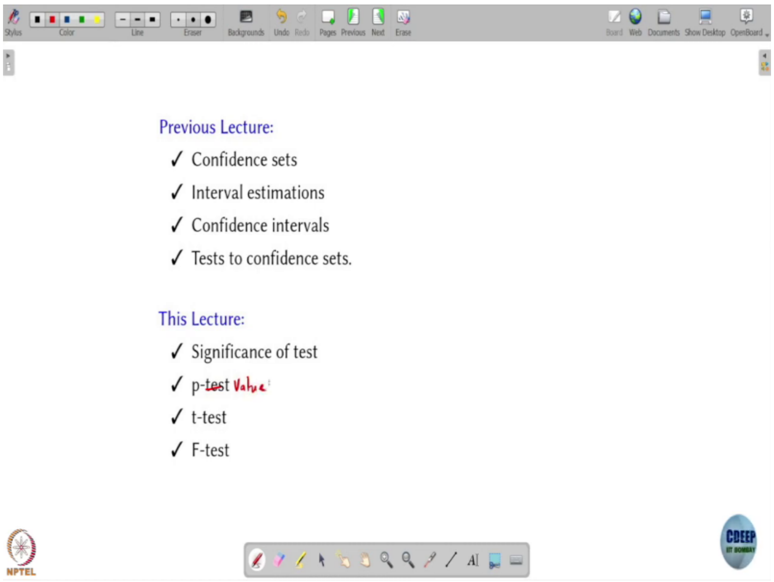


Let us get started with today's lecture. So, in the previous lecture, after talking hypothesis testing, we discussed about how to construct confidence sets, how to get interval estimations and from that how to get confidence intervals. And we also discussed about how hypothesis test can be used to get confidence test.

So, in this test, we are going to talk about given a test to estimate some parameter, how good is it like how confident we are in that about our null hypothesis, for that we are going to define study various tests like p-test, t-test and F-test and how to quantify them using something called significance of a test.

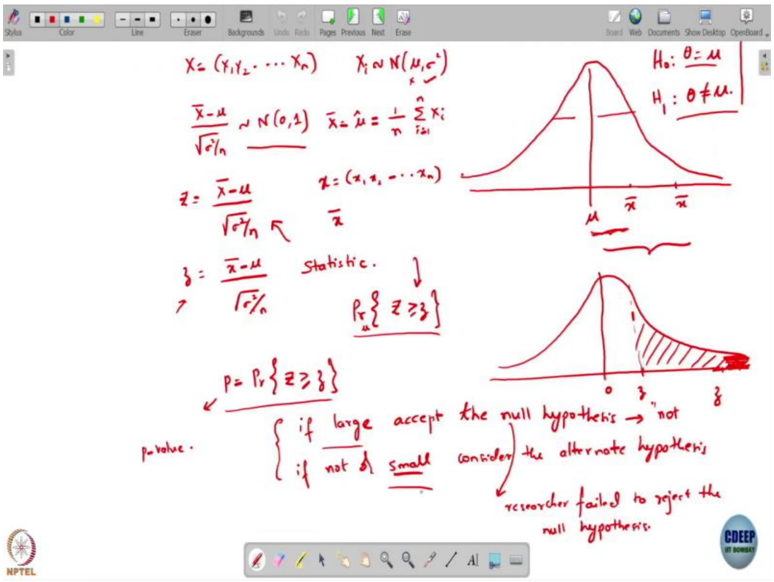
(Refer Slide Time: 01:13)





So, let us get started with p-test or rather p-value. Maybe it should be more appropriate to call it a p-value rather than p-test.

(Refer Slide Time: 01:37)



As a motivation suppose let us say you have samples and let us assume that each of these Xi is coming from gaussian with parameter mu and sigma square. We have already seen that if I take the empirical mean, do the centralization. And this normalization, assume that right now I do not know mu, but I know sigma square, I know that this is going to be gaussian distributed with mean 0, 1.

And now, I am interested in estimating mu hat, I know one estimate is 1 by n summation xi i 1 to n. Now, the question is now our claim is mu hat is the population mean. But now the gaussian is why one should take it as population value, what you are giving is simply an estimate. Somebody has to believe in whatever you are giving is going to be a good representation or in fact the true value of the population mean how one can believe it or what should be the criteria.

And that is where p-values comes into picture. Now, we are also denoting it as x bar. What is one possibility? We know that x bar minus mu divided by square root of sigma square n is normally distributed with parameter mu and 1. And we know well about its distribution, especially about its CDF and its tail behavior. Maybe we can use this property to see how good is my claim that mu hat is a good representation of the underlying population mean.

So, for that one possible way is let us say I am going to define z to be x bar mu and sigma square by n. And now, what I am interested in is what is the probability that my x bar is going to be close to the population mean. Now, think about this. Let us draw normal distribution and let the mean is here mu. Now, for a given sample.

So, this is like I said, let us say, for a given x, this is my realized sample, I got my x bar. And I get a particular realization of this, which I am going to denote it has mu by sigma square by mu, sorry, n, where n is the number of samples we have. Now, if suppose I get for a given realization z here, maybe this is a good representation of my population mean. But suppose this z happens to be here, instead of here that means it is a little far away from mu, if z happens to be here, maybe this is actually not a good representation of my underlying population mean.

Now, depending on how far this z value, the estimated value I got is going to be away from the true population mean, maybe I can say something about this. Suppose, let us say, if this quantity is very small, maybe I have a good confidence that yes, whatever the z value estimate I got it is going to be a good representative of my underlying population mean. And if this is large, maybe that is not the case.

So, to do that, instead of directly looking into z, what we, instead of directly looking into the x bar, so maybe this should have been x bar here not z like so suppose, let us say is x bar is here. Now, if they are close, I have a maybe more confidence, or I may be more happy that this is the representation of my true population mean. And if this happens to be a little far away, that is not the case.

So, naturally, what we should be looking into is that difference, like if the difference is small maybe it is good, if the difference is large maybe that is not a true representation of my population mean. And also, in this case, the variance plays the role, like how does is the dispersion around your population mean? So, because of that instead of just taking the difference x bar minus mu, maybe I would also want to normalize based on the variance.

And that is what like we are defining z, which is the difference divided by the variance. Now, let us ask the question, what fraction of the time the difference is going to be small? If the fraction of the time this difference is going to be small then maybe it is good, your test is good. And if the fraction of the time if this difference is going to be large, maybe you are not actually capturing the true population mean.

So, using this quantity, what we are going to, so using this quantity what we are now going to call it as a statistic in this case, we can ask the question what is the probability that my z is going to be greater than or equals to z? So, notice that this z is what you have computed based on your sample. If this z happens to be large, let us say somewhere here, let us say now

I am going to look into a gaussian distribution with the mean 0 and variance 1, and now this is mean 0.

And if this happens to be small, let us say this z happens to be here that means that this probability is going to be probability z, like I will be basically covering the probability of all this. So, if this z is small, my probability is going to be higher. And the other way if this z happens to be, let us say here, then this probability is going to be smaller. So, based on that, we can think that if this probability is large, and I am going to call this value as my p-value, p is equals to probability, sorry probability that, p is equals to probability that z is going to greater than or equals to z.

And if the probability of this happening is let us say large that means basically my z is small that means it is actually good. So, maybe I would like to accept, maybe like, if large I would like to accept the null hypothesis. So, null hypothesis I mean to say that so your test is what your test is basically now trying to check whether these samples are coming from a distribution like my null hypothesis here probability that my, let us say my theta is mu and my null hypothesis, alternate hypothesis is that your theta is not equals to mu or maybe, yeah.

So, if this quantity is large, it is likely that the samples you are observing and by the way this probability is under your, assuming that your samples are coming from your true parameter that you want to test against, if this quantity is large, then you want to accept the null hypothesis saying that yes, indeed they are coming from mu. And if small, if not or small then maybe it is possible that your samples are not coming from this parameter mu, that is, because in this case, that z is large, and that is why this probability is going to be small.

So, if not or small then you may want to consider the alternate hypothesis. Or it is equivalent to saying that you want to consider an alternate hypothesis maybe like in this case possibly your null hypothesis is not true and this is kind of a providing a evidence that maybe your claimed that it is to be null hypothesis is in current. And if large maybe actually what we will do is, let us put it slightly differently.

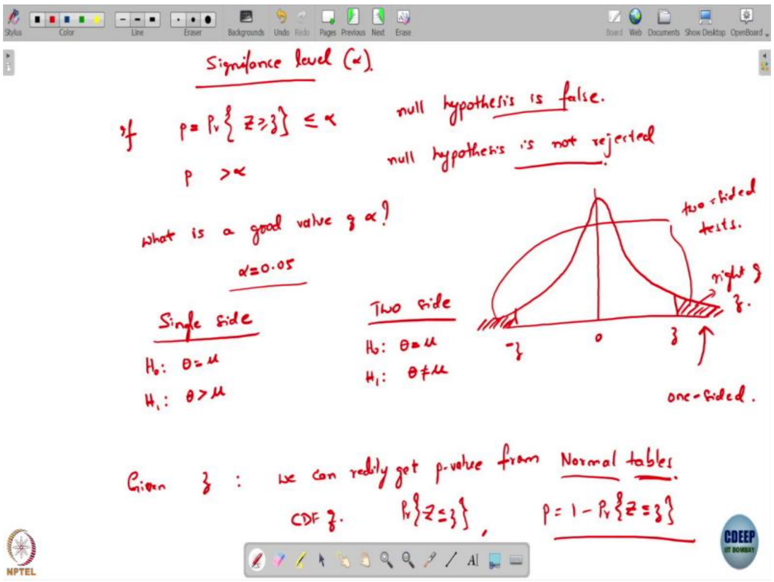
So, in this setup, what we are saying. Let us say you having sample and you have this hypothesis, to check. One is you are going to check whether these are coming from with the population mean mu and alternate hypothesis, what we are going to call is something challenger, your claim is that this is going to be a null hypothesis with parameter mu and somebody is going to challenge you saying no, no, this is not the case it is something else.

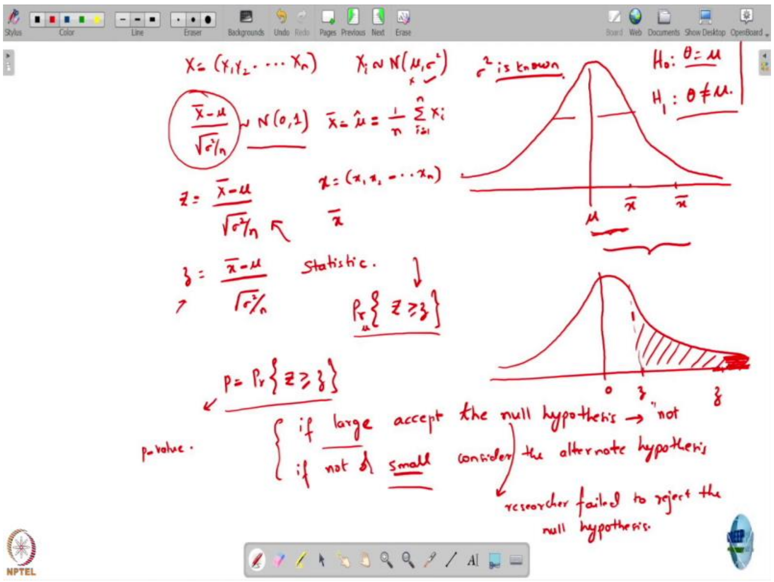
Now, in that case, if this quantity happens to be small, then it is kind of indication that your null hypothesis may not be true and alternate hypothesis is true. And on the other hand, if this probability is large, it is in favor of null hypothesis it is basically saying that, it is saying that not basically you can say that let me see what is the correct word here. You can say that. Here, you can say that researcher failed to reject the null hypothesis.

So, in this setup, we are basically saying that whatever the tests you have provided to verify your null hypothesis whether to accept or not and you have a challenger here somebody is challenging you and if you could show that this probability or this p-value in this case is large then you can say that researcher failed to reject your null hypothesis.

And on the other hand, this is small then you are basically validating that or basically the researchers validating that whatever your claim on the null hypothesis is incorrect. So, then the question arises here is what is this large and small, how to quantify this, is there a way and that is where we need to put some threshold here.

(Refer Slide Time: 16:15)





And significance level come to picture. So, you want to say that the p-value you have let us say Z is greater than or equals to z if this is less than r equals to c, then you can say that, let me see what is the exact term I can use, then we can say that your null hypothesis is false. On the other hand, if this p-value you have is greater than alpha then null hypothesis is not rejected. And then question comes what is the good value of this alpha?

Usually, alpha is taken to be 0.05. If your p happens to be smaller than 0.05 then you are claimed that the null hypothesis is true is not accepted or it is your null hypothesis is false. On the other hand, if p happens to be larger than 0.05 then your opponent who was challenging you is not providing sufficient evidence to say that your hypothesis is false or in other words, the null hypothesis is not rejected, you are still in the business.

So, because of that, if you are making a hypothesis and the p-value for that hypothesis, the p-value computed under that null hypothesis happens to be larger than 0.05 then your hypothesis is not rejected. And so, you should always make sure that if you are making a claim on a null hypothesis the p-value should be as large as possible, or at least it should be more than a 0.05.

And this 0.05 just like kind of a thumb rule people take it is not necessarily that one has to stick to 0.05. Depending on your application, if you feel that my application is such that I do not need to be too stringent or I need to be relaxed. Depending on that you can set your alpha to be maybe any value between maybe let us say 0 to half, whichever is more applicable. Now, when you are doing this test there are two possible ways, like when you are making the hypothesis.

One is let us call single side and another is like a two side. In the single side like let us say your hypothesis is like let us say theta is some mu and your alternate hypothesis is just saying that maybe just like your theta is greater than mu. They are just saying that no, like somebody is claiming that the true parameter is going to be mu and you are challenging by saying that no, the true parameter is simply be larger than mu just like one sided.

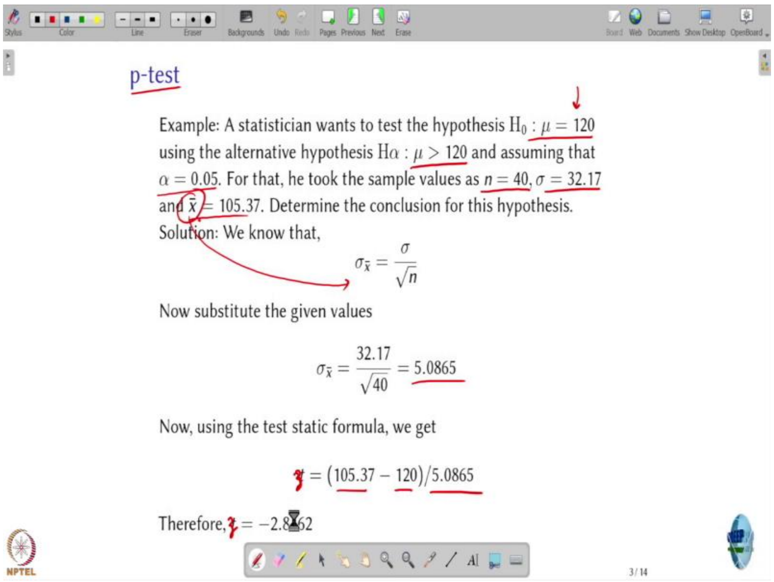
On the other hand, the two sided say that, you are going to say that, the true null hypothesis is simply parameter mu, but your challenger may say that no, he simply say that no, this is not mu it is something else it could be larger or smaller that is why it is called two sided. But in both the cases like this is like standard normal. So, if it is like one sided you have z here, you will look into this, this is like a one-sided case.

But if it is a two sided you need to consider this probability as well. So, both this together will give you the probability the p-value for the two-sided tests. And the good thing about this p-values is you can readily compute them by looking into the tables for your standard normal distribution. Like I mean, if I give you z, you know already given z we can readily get p-value from gaussian tables or maybe just I can just say normal tables.

One point to note here is this normal tables, they usually are given for probability like z is equals to less than or equals to z that is basically CDF of your normal distribution. But notice that the p-values are the complement of this. So, the p-values, we need to get the complement of this by simply taking the negation of that. And by doing this, you will end up finding this value which are to the right of z value.

So, notice that basically, this all worked well, when your samples are gaussian distributed or this quantity x bar minus mu normalized by square root of sigma square by n is normal distributed. And here another crucial thing we notice this the sigma square is known. Now, the next question happens is how does this change when the sigma square is unknown, and which could be often the case. We will study that by looking into our t-tests.

(Refer Slide Time: 24:25)

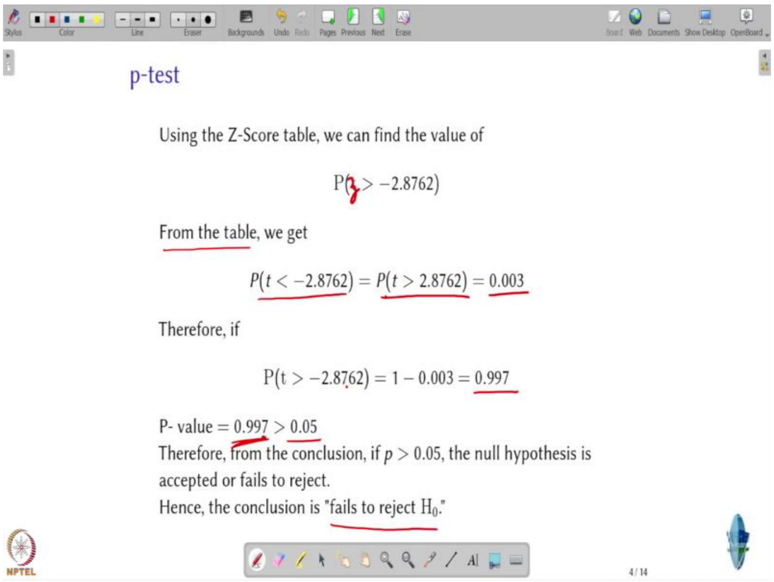


So, we actually, so far computed the p-value. Now, before I jump into the t-test, I want to just conclude what this p-test is, by looking into an example. Suppose let us say you have a, you are in a case where a statistician wants to test the hypothesis that the true population mean is 120. And somebody is challenging him saying that the alternate hypothesis is the one-sided value of mu being greater than 120.

And the threshold the significance level is being set to 0.05 here and the number of samples we have is 40, the variance is taken to be 37.5 and from the samples it has been computed that the sample mean is 105.37. Now, using the p-value, we want to conclude whether this x bar is statistically significant or not. So, now, how to do this? First, we are going to compute the statistics, for that I need the variance of my estimator.

So, I have this x bar and I know that it is variance is going to be sigma by square root n and sigma has been told to me as 32.17, n is 40. So, I got this value 5.0865. Now, I will compute my, I have been denoting this as z value. So, the z value here is simply x bar minus 120 and divided by. So, notice that here you will actually know this 120 because that is the null hypothesis against which you are testing and here you will end up with a z value which is minus 2.8762.

(Refer Slide Time: 26:32)



Now, for this using z score table you can calculate z being larger than minus 2.8762 in the following fashion, we know that from the gaussian tables probability that t is less than or equals to this because of the symmetry this is also equals to t greater than 2.8762 is 0.003. And now, probability that t being greater than minus 2872 is 0.997.

So, the p-value we have obtained it to be 0.997 which happens to be significantly larger than my significance level of 0.05. So, therefore, in this case, the p-test says that my, the p-test leads to the conclusion that we have failed to reject the null hypothesis that means null hypothesis claimed by the statistician is still valid and there is no evidence to provide that this is not true that is basically we have failed to reject the null hypothesis.