Introduction to Biostatistics

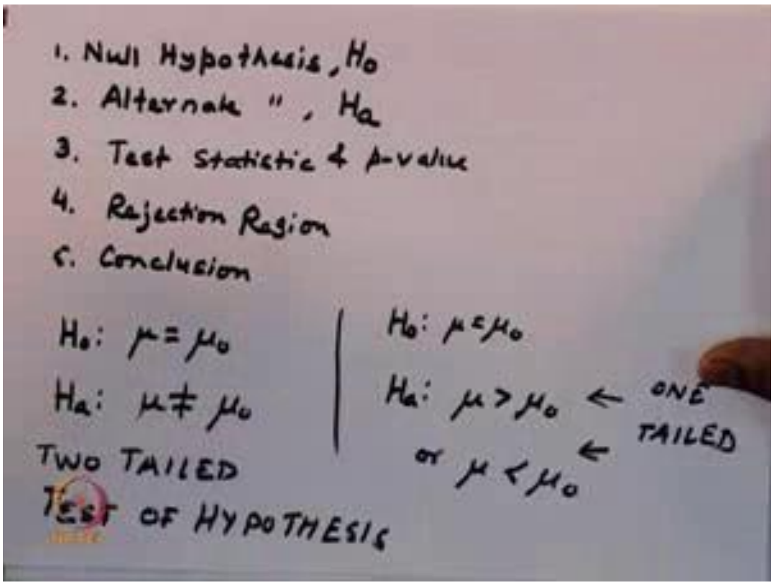
Prof. Shamik Sen
Department of Bioscience and Bioengineering
Indian Institute of Technology, Bombay

Lecture - 33
Test of Hypothesis - 3 (1 tailed and two tailed test of Hypothesis, p-value)

Hello and welcome to today's lecture. So, today we will do a brief recap of our discussions of test of hypothesis from last lecture and then we continue to discuss few other examples or few other concepts in this testing of hypothesis. So, what is the test of hypothesis and why is it important. So, we know that in practical cases you as a statistician might have to estimate some parameter for a population or to make a decision when you are given that parameter.

So, for all these cases you have a hypothesis. So, for a test of hypothesis there are 5 components for this test and first one is the null hypothesis which is written as H naught.

(Refer Slide Time: 01:02)



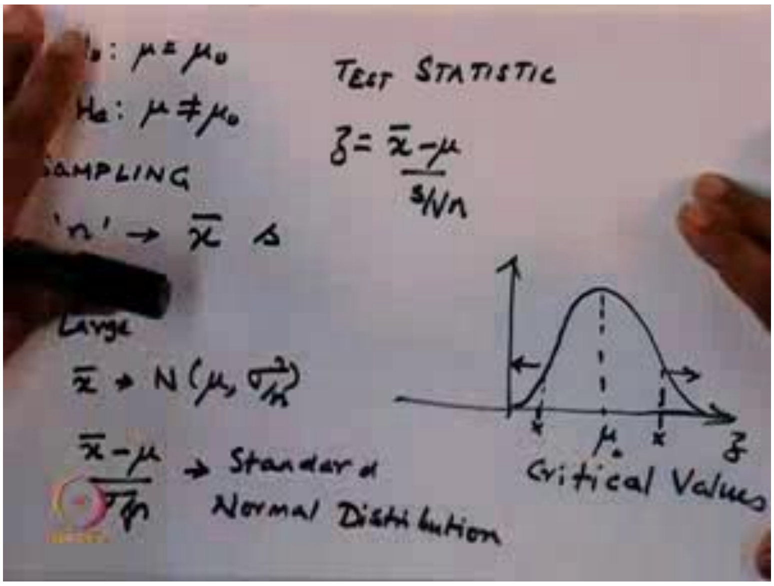
Second one is the alternated alternate hypothesis, hypothesis it is referred to as H a, you make these decisions as to either your null hypothesis is true or the alternate hypothesis is true. So, you either accept the null hypothesis and reject the alternate hypothesis or you reject the null hypothesis and accept the alternates. So, based on this, this is you calculate, you decide based on some test statistics and a p-value which is represented representative of the confidence you have in taking in measuring this test statistic.

You have some rejection criteria, rejection criteria and then based on this rejection; based on this rejection criteria or rejection region you draw your conclusion. So, the null hypothesis let us say our agenda is to see whether the productivity of a line or a company has increased or decreased or even to test what it is.

So, in that case my null hypothesis can mean let us say mu is equal to mu naught. So, in other words what I am saying that the productivity of this company is this. I can have an alternate hypothesis either in the form of mean not equal to mu naught or in another situation you can have H naught as mu is equal to mu naught and H a as mu either greater than mu naught or mu less than mu naught. So, the difference between these 2 tests of hypothesis these in this condition you are not concerned whether mu is greater than mu naught or mu is less than mu naught or you want to ascertain is whether mu is significantly different from that of mu naught. So, this test is called a two tailed test is called a two tailed test of hypothesis versus all these both these cases are one tailed test of hypothesis.

So, when we talk about two tailed test of hypothesis or one tailed test of hypothesis what do we mean?

(Refer Slide Time: 04:01)

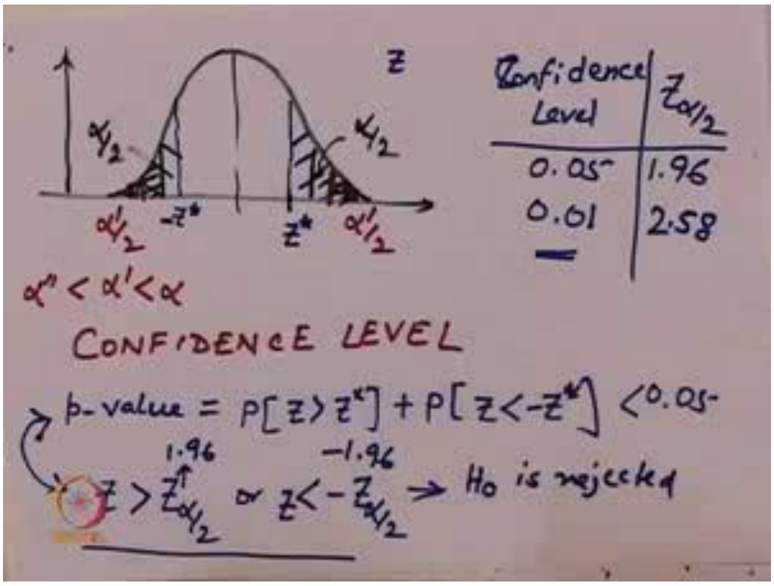


Once again your H naught says some let us say mu is equal to mu naught in this case and your H a is mu not equal to mu naught, how will you ascertain this? You will do some sampling let us say from the population you draw a sample of size n and based on this sample you calculate the sample mean x bar and the sample standard of deviation s.

Now, what you know from central limit theorem is if your sample size is large if this is large then x bar follows normal distribution with mean of mu and variance of sigma square if this was the population. So, I can compute a test statistic. So, if x bar follows this then x bar minus mu by sigma should follow a standard normal variable. So, this, if your x bar then this is x sigma square by n and this is sigma by root of n it should follow a standard normal distribution.

So, since I only have the sample I do not know exactly sigma, but what I have calculated is s. So, my test statistic is calculated by this value Z defined as x bar minus mu by s by root n. So, now, what do we mean by 2 tails? That is say this is your mu value or mu naught value and you will have a distribution this is roughly normal distribution, you will have a distribution which is on both sides of this mu naught and let us say you have to decide how do you come to the conclusion that your hypothesis is either right, that is you accept your hypothesis or it is wrong and you will accept the alternate hypothesis. So, you decide based on what is called a rejection region. Let us say and this for a two tailed test you can have the rejection region is when your value of Z is either greater than this, this particular threshold or it is less than this threshold. So, these values are called the critical values.

(Refer Slide Time: 06:52)



Now, if I draw this curve again. So, you can theoretically draw or reject once this threshold is crossed or you can reject once this threshold is crossed. So, what we have done is for the black so far, so you can compute the probability that you will reject which is this curve, this is the area under this curve. Let us say this is alpha this is alpha by 2 and this is alpha by 2 because this curve is symmetry. For the red, accordingly you see this area is alpha prime by 2 or alpha prime by 2 and the way I have drawn it is alpha prime is less than alpha.

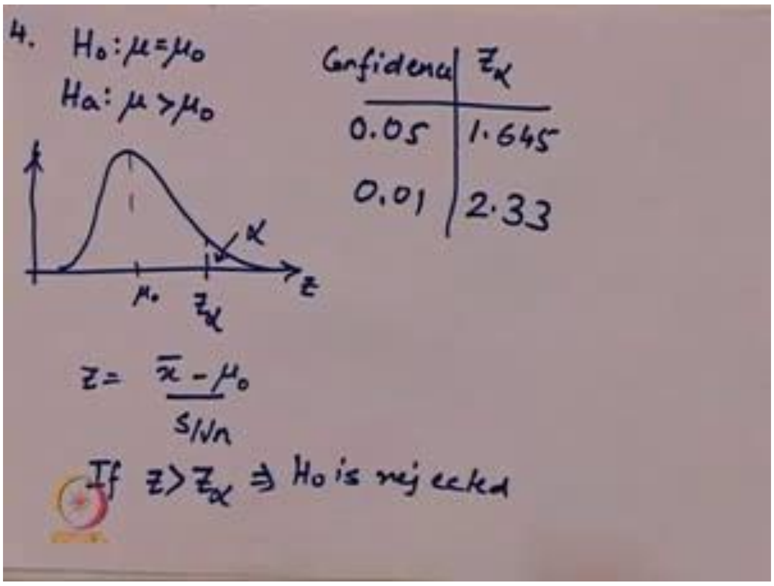
So, as you decrease alpha you can go to alpha prime or you go to alpha double prime you can define an alpha double prime which is even smaller, you can define an alpha double prime which is even smaller. So, what you are changing is essentially the confidence level. What is p-value is let us say you have calculated a particular value of, you have calculated this value of Z statistics test statistics are let us says Z is this one.

So, then for a two tailed test this should be minus Z, for a two tailed test the area under this whole curve the area under this curve is p-value, p-value in other words if I draw for this particular case let us say if I label this as Z star then p-value is the probability that my Z is either greater than Z star or the probability that Z is less than minus Z star. So, you can have various levels of confidence for the two tailed test you can have, so this value I call for a two tailed test we call it as Z alpha by 2.

So, for the confidence level you can have Z of alpha by 2. So, if you want a confidence level of 0.05 which means that 95 percent certainty you want to reject H naught. So, this particular value of Z alpha by 2 is 1.96. If you want 0.01 then this value becomes 2.58. So, in other words you want to calculate probability Z greater than 2.58 or probability Z less than minus 2.58.

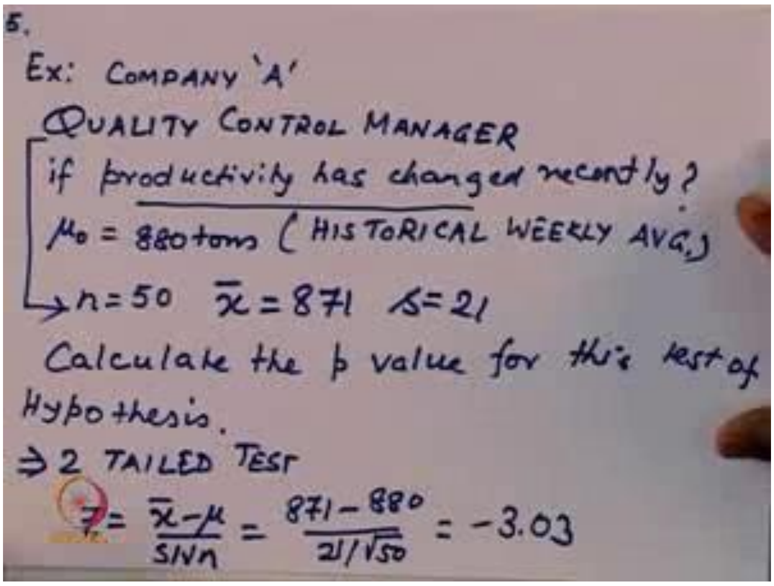
Now, instead of calculating the area under the curve since this value is nothing, but the area under the curve if you calculate this value of Z. If you compute Z is either greater than Z alpha by 2 or less than minus Z alpha by 2 then you can automatically draw the conclusion that. So, if this criteria is full file then you can say H naught is rejected for a particular case. So, this is an equivalent statement. So, if your p-value is less than 0.05 is equivalent to saying Z is greater than in this particular case Z is greater than 1.96 or Z is less than minus 1.96.

(Refer Slide Time: 10:46)



So, for this one tailed test, when you calculate the one tailed test in this case let us say my H naught is mu equal to mu naught and H a is mu greater than mu naught. In that case, I will just label it as Z alpha where alpha is the area under this curve. So, what will you do in this case? So, this is your mu naught you will calculate Z as x bar minus mu naught by s by root n, if Z is greater than Z alpha then H naught is rejected. So, the corresponding values of Z alpha for a one tailed test, so you have the confidence interval you can have 0.05 and 0.01, this is 1.645 and this is 2.33. So, these are the one tailed and the two tailed test.

(Refer Slide Time: 12:06)

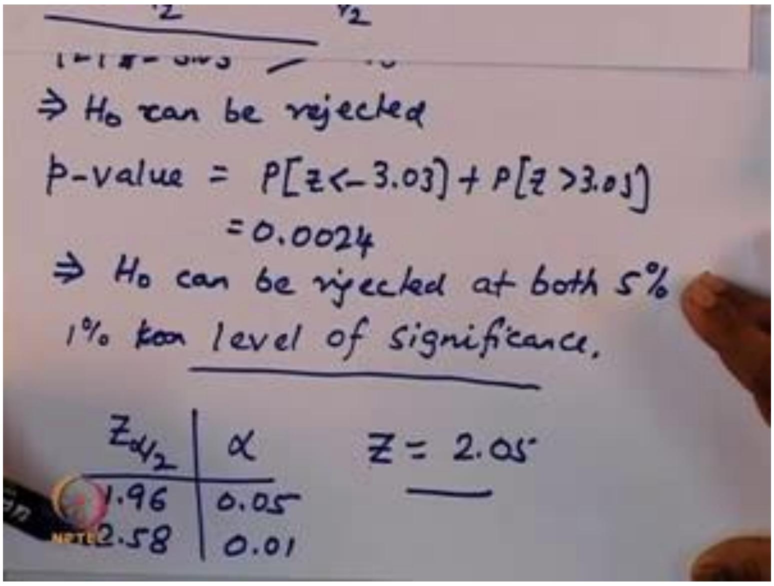


Let us take a (Refer Time: 12:06) example. So, imagine you have a company A and the quality control manager say, control manager wants to evaluate, wants to evaluate if the productivity. So, what does we want to evaluate? If productivity has changed recently, what are you given? The average productivity or mu naught is equal to 880 tons. So, this has been the historical rate, historical weekly average. So, what the quality control manager does he draw the sample of 50 products and for this he calculates x bar as 871 and s is 21. So, we want to calculate the p-value for this test of hypothesis.

So, what you know is the quality control manager wants to evaluate his productivity has changed which means that he does not want to see whether it has increased or whether it is decreased. So, this implies this is a two tailed test. You can calculate the value of Z which is going to be x bar minus mu by s by root n x bar is 871 minus mu is 880, s is 21 by root of 50 this gives me a value of minus 3.03, minus 3.03.

So, for the two tailed test, for the two tailed test if I draw this curve again what we have been, what I have shown what we know is add the 5 percent confidence level or 5 percent level of significance where Z alpha by 2 is 1.96 now your Z value here is 3.03 which is much greater than 1.96.

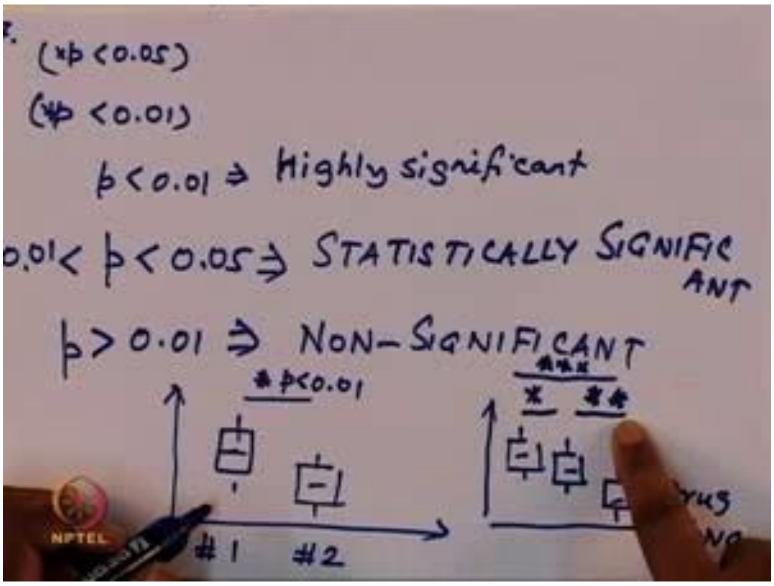
(Refer Slide Time: 15:07)



So, I can safely conclude, I can safely conclude that Z, mod of Z is equal to in our case 3.03 which is much greater than 1.96. So, this would mean that I can reject H naught can be rejected, if I calculate the p-value, so the p-value would be given by P of Z less than minus 3.03 plus P Z greater than 3.03 and you get a value of 0.0024. So, if you look at these values you have 5 percent level is 0.05 or 1 percent level is 0.01. So, this tells us, this tells us that H naught can be rejected at 5 percent levels at both 5 percent and 1 percent level of significance.

Now, you can imagine a situation imagine a situation I get a value of, so what I know if I plot Z alpha by 2 again and alpha for 1.96 I have 0.05, for 2.58 I have 0.01 right. Let us say in a particular case I get a value of Z of 2.05. So, what this means is I can accept it at the 5 percent confidence level, but I have to reject it at the 1 percent confidence level because this Z is not greater than 2.58, and this causes some level of ambiguity and this is the reason why we always report the p-value for the test, the p-value for the test. 4 5 6. This is the reason why we always report the p-value of the test.

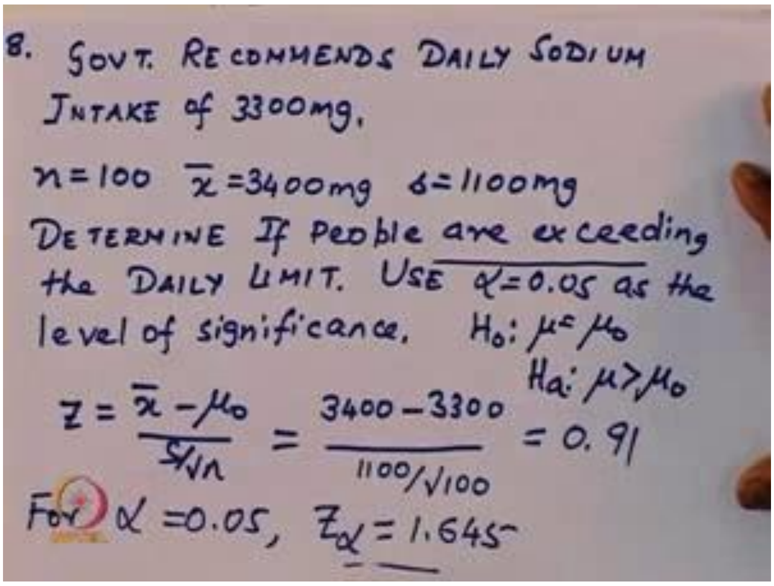
(Refer Slide Time: 17:54)



So, if you look at research papers you will see words like, so general you might see words like star p less than 0.05 or star p less than 0.01. Now what is accepted if your p is less than 0.01 you typically, so you typically you say that the results are highly significant, your results are highly significant. If your p is less than, so then you say these results are statistically significant. If your p-value lies between these 2 values you will say statistically significant, but if your p is greater than 0.01 then your results are non-significant.

So, if you look at plots you might see the plot being represented as follows. Let us imagine this is the data for one particular condition. So, this is condition number 1. And for another condition this is your condition number 2 and you will see. So, this is condition number 2 and you have some matrix here and what will be shown like this is star p less than 0.01. Alternatively, you might also see let us say if you have 3 conditions, imagine this is three different drug concentrations, concentrations and you have measured some output. So, between these in many papers you might see this as written as star this as written as star star and this as written as star star star. What we showed is again that the level of difference between these 2 populations or these 2 measurements is significant, but this is statistically significant more of all three of them are significant, but the level of significance varies. So, in the text you will probably see the magnitude being reported.

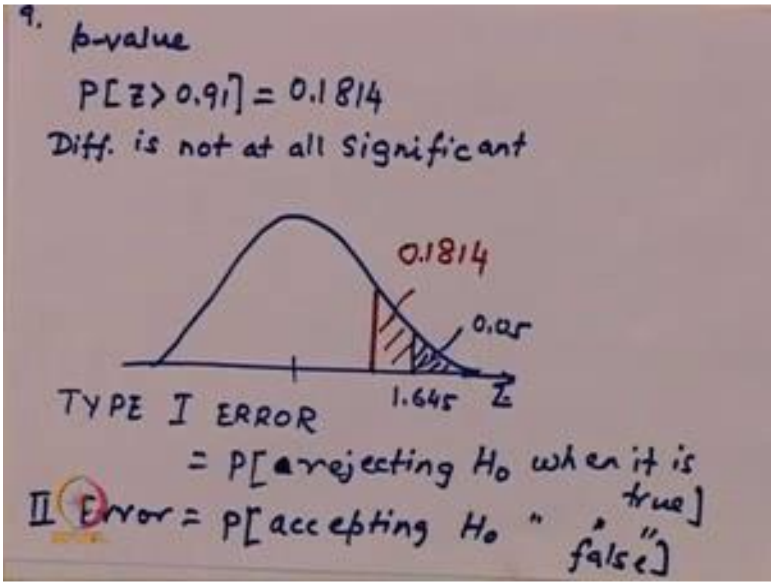
(Refer Slide Time: 20:50)



Let us discuss another example. So, imagine the government recommends daily sodium intake of 3300 milligrams. So, for a sample a random sample of 100 measurements these values turned out to be 3400 milligrams with a standard deviation of 1100 milligrams. So, the question we want to (Refer Time: 21:33) determine if people are exceeding the daily limit; use alpha equal to 0.05 as the level of significance. So, as before what we have to do? We have to calculate the test statistic which is given by x bar minus mu naught by s by root n, x bar is 3400 milligrams minus mu naught is 3300 milligrams, s is 1100 by root of 100. So, you get a value of 0.91.

Now for alpha equal to 0.05, Z alpha is 1.645. So, why do we use a value of Z alpha equal to 1.645? Because we are asked the question if people are exceeding the daily limit, so for this case my H naught is mu equal to mu naught and the alternate hypothesis is mu is greater than mu naught. So, I need to use the one tailed test distribution, one tailed test of hypothesis for which for alpha equal to 0.05 Z alpha equal to 1.645, but since the value of the test statistics is less than Z alpha which is 1.645 I can safely say that this is not statistically significant. This is not, one cannot clearly say whether it has increased or not. I can also calculate the p-value for this case, I can calculate the p-value and this comes out to be. So, I want to calculate probability of Z is greater than 0.91 which comes out to be 0.1814.

(Refer Slide Time: 23:51)



So, this tells you this probability is way too high. So, the difference is not at all significant. So, if I were to draw the same this is 1.645 of Z corresponding, this is 0.05 the area under this curve this area is 0.05, but what you are given what you are given, what you got is something like this and the area under this curve this whole curve came out to be 0.1814. So, this tells you that you cannot make the assertion that the daily intake has increased significantly.

So, most of the cases, most of the cases we say that it has increased or decreased, but when we make these predictions there are 2 types of errors which can accumulate. We call them either the type I error which is the probability of rejecting H naught when it is true or the type II error this is the probability of accepting H naught when it is false. I will stop here for today and in the next lecture we will discuss about these 2 types of errors.

Thank you for your attention.