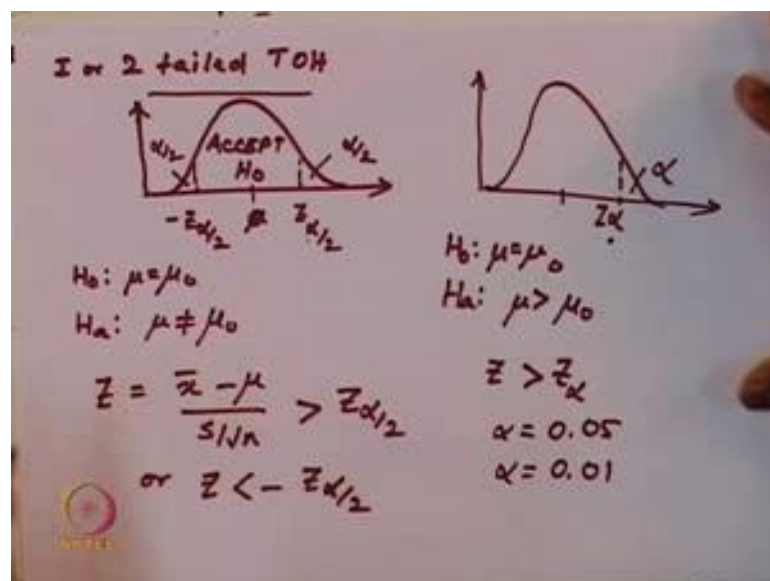


Introduction to Biostatistics
Prof. Shamik Sen
Department of Bioscience and Bioengineering
Indian Institute of Technology, Bombay

Lecture – 35
T-test

Hello and welcome to today's lecture. So, yesterday in last days lecture we have discussed about 1 or 2 tailed test of hypothesis.

(Refer Slide Time: 00:28)



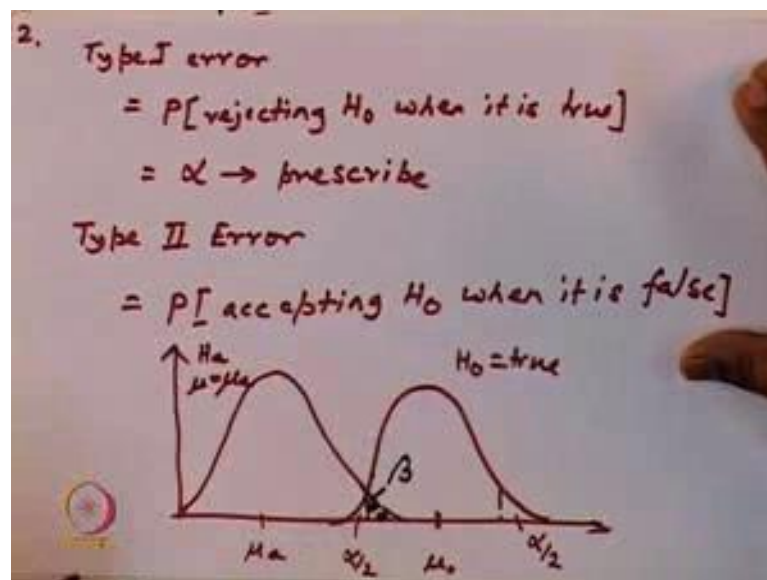
So, in a 2 tailed test of hypothesis depending on your confidence interval alpha. So, you determine this value of Z which is alpha, Z alpha by 2 this area under the curve is alpha by 2, this is alpha by 2 and accordingly you accept. So, this is the acceptance region for H naught whatever is your null hypothesis. Let us say your H naught is mu equal to mu naught, so then this value becomes 0, after you normalize for a standard normal variable. And your alternative hypothesis H a is mu not equal to mu naught. So, that is why you consider both the tails the left tail and the right tail and that is why it is called a 2 tailed test of hypothesis.

So, if your Z value for your test statistics let us say $\bar{x} - \mu$ by s/\sqrt{n} it is greater than Z of alpha by 2 or Z is less than minus Z alpha by 2, you know that your test of I your H naught can be rejected otherwise it you cannot reject H naught. And for a singled tailed test of hypothesis, for a single tail test of hypothesis you are interested in

knowing whether your μ , your H_0 is $\mu = \mu_0$ and alternative hypothesis is $\mu > \mu_0$. So, you want to know whether this value is greater than the confidence level. So, if your test statistic is greater than Z_{α} . So, then you will reject your hypothesis and you can accept your alternate hypothesis.

Now, alpha corresponds to the confidence level, you can have alpha equal to 4-5 percent confidence level, you have alpha equal to 0.05 for 1 percent confidence 99 percent confidence level you choose alpha equal to 0.01 so on and so forth. Now yesterday, while testing of hypothesis we discussed the possibility of 2 types of errors which can come up.

(Refer Slide Time: 02:58)



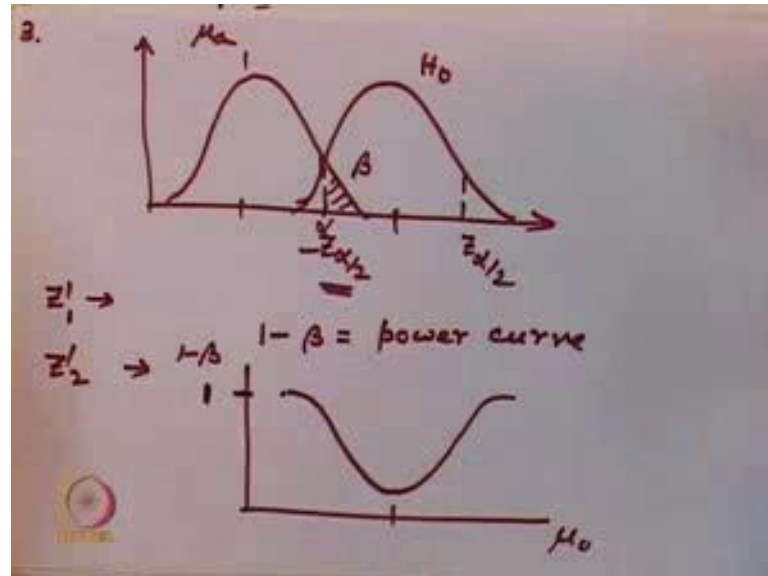
You have a type I error is a probability of rejecting H_0 when it is true and this is nothing but this alpha and this you prescribed.

However you can have another type of error or type II error which is the probability of accepting H_0 when it is false. So, if I draw let us say this is your H_0 equal to true, this is your mean and this corresponds to μ_0 . So, you have for a 2 tailed test of hypothesis you have this is your alpha by 2 this is your alpha by 2.

Now, let us say your sample gives you a mean of μ_a which is different. So, then the distribution let us say you get, so this is your alternative hypothesis which is $\mu = \mu_a$, this is distinct that μ_0 . So, what you see is this area, this area in this part of

the curve this you call as beta and this is what your probability of accepting H_0 when it is false.

(Refer Slide Time: 04:58)



So, how do you calculate it, how do you calculate beta? So, for your H_0 you calculate Z of alpha by 2 minus Z of alpha by 2 and Z of alpha x 2 and you know what is the range, what is the confidence interval. Corresponding to these values of Z for here you have to calculate Z prime, let us say this corresponds to μ_0 . So, you have to calculate Z prime using these values of Z of alpha by 2, using these values of Z of alpha by 2 you will get Z prime 1 for this value 1 and Z prime of 2 and using this you can calculate what is this area under the curve which is your beta.

So, what you can clearly see from the way we have drawn this if this curve shifts to the left then your beta decreases for the same value of alpha your beta decreases if this curve shifts to the left. How will this curve shift to the left? If you see increase your sample size which means that your deviation of this population keeps on decreasing. So, the curve $1 - \beta$ it is called the power curve and you can plot $1 - \beta$ as a function of let us say μ_0 and if let us say this was your value of μ_0 corresponding to H_0 is true and you plot your $1 - \beta$ curve it will look something like this, where far from it you will get a value which is close to 1, which is close to 1.

Why? Because when mu is far from this value then you will never, it is unlikely that you will make this mistake your beta will 0. If your curves are shifted are separated from each other, if your curves are shifted separated from each other then you will get a value of beta which is equal to 0 and then 1 minus beta will approach 1. So, at two end points this 1 minus beta curve approaches 1.

(Refer Slide Time: 07:23)

4. Binomial proportions
 $SE = \sqrt{pq/n}$ means of
 We compare 2 Binomial proportions
 $H_0: p_1 = p_2$
 $H_a: p_1 \neq p_2$
 $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_1 \hat{q}_1 / n_1 + \hat{p}_2 \hat{q}_2 / n_2}}$

So, we had discussed one more case for the case of binomial proportions in this case the standard error is given by root is equal to root of pq by n, but if we want to compare the means, if you want to compare two binomial proportions. How would we do it? So, in that case let us say our null hypothesis, we want to compare the means of 2 binomial proportions. In this case my H naught is going to be proportion p 1 equal to p naught p 2 and H a is going to be p 1 naught equal to p 2. So, your test statistic Z is calculated as p 1 minus p 2 by root of, so this is experimental variable by root of p hat q hat by n 1 plus p 2 hat q 2 hat by n 2.

So, let us take a sample example to see how we can do solve this problem.

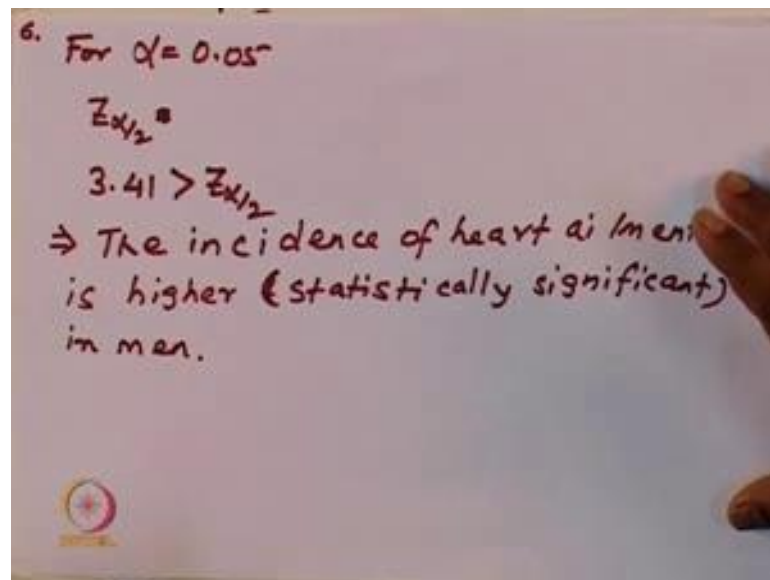
(Refer Slide Time: 09:00)

5. HOSPITAL RECORDS
52 men out of 1000
23 women " " "
were admitted with heart ailments.
IS THERE A SIGNIFICANT Diff. in heart
DISEASE among MEN & WOMEN
 $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{52 + 23}{1000 + 1000} = 0.0375$
 $p_1 = \frac{x_1}{n_1}$ $p_2 = \frac{x_2}{n_2}$
 $Z = \frac{0.052 - 0.023}{\sqrt{0.0375 \times 0.9625 \left(\frac{1}{1000} + \frac{1}{1000}\right)}} = 3.41$

So, imagine in hospital records of a given hospital. Suggest that 52 men out of 1000 and 23 women out of 1000 were admitted with heart disease, heart ailments. So, we want to know is there a significant difference, difference in heart disease among men and women. So, what we want to calculate see if I give you again this expression, our H naught remains the same as before that we want to calculate, we want to test the hypothesis that p_1 is equal to p_2 or the alternate hypothesis p_1 is not equal to p_2 .

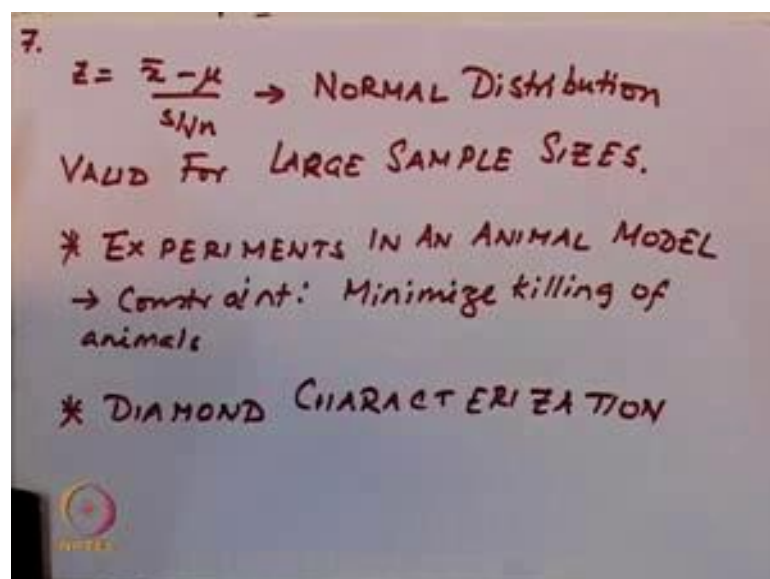
Now, we want to calculate this test statistic Z equal to p_1 hat minus p_2 hat by this; however, we do not know, so we do not know what value of p to put for 2 different populations. So, what we can do is we can estimate p hat as an average if you can estimate p hat. So, you can have p_1 as x_1 by n_1 and p_2 as x_2 by n_2 you can put in the value of p hat as x_1 plus x_2 by n_1 plus n_2 and this will give you a value of 52 plus 23 by 1000 plus 1000 which is equal to 0.0375. So, based on this p hat value I can calculate my z. So, based on this p hat value I can calculate my z which is p_1 hat which I know p_1 hat which I know to be 0.052 minus 0.023 by root of I use this common value in to q which is 0.9625 into 1 by 1000 plus 1 by 1000. So, if you plug in these values you will get a value of 3.41, we will get to a value of 3.41.

(Refer Slide Time: 12:27)



So, for 95 percent for alpha equal to 0.05 we know for a 2 tailed test this value is much greater, so Z of alpha by 2 is equal to 1.96. So, I know that 3.41 is much greater than Z of alpha by 2. So, I can say that the incidence of heart ailments is higher that is statistically significant in men. So, this brings us to our understanding of test of hypothesis for reasonably large samples. But, what if you are handling much smaller samples?

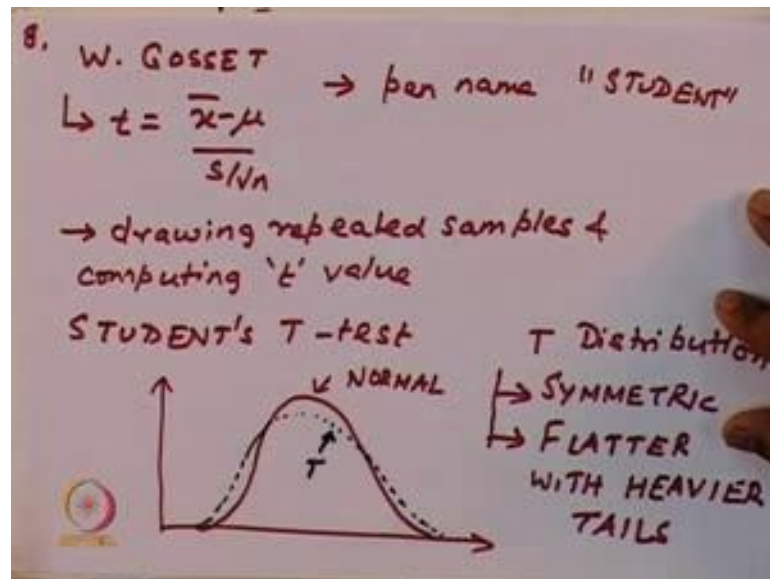
(Refer Slide Time: 13:56)



So, the entire test statistics Z equal to $\bar{x} - \mu$ by s by root n being a normal distribution is only valid for large sample sizes, for large sample sizes. Imagine the following scenarios; you want to do some experiments in an animal model. So, what is your constraint? The constraint is to minimize killing of animals.

Another example is say for example, you want to characterize the diamond weight some diamond characterization. So, because diamond is expensive you cannot afford to take a large sample size, you are constraint to take a much smaller sample size because of which your normal distribution assumption may not be valid.

(Refer Slide Time: 15:49)



So, this question has been addressed. So, this question has been addressed by Gosset and, so he published this study using he defined t as $\bar{x} - \mu$ by s by root n and the distribution he obtained. So, he calculated this t by repeatedly, so by drawing repeated samples and computing t value and he published this work under the pen name "STUDENT". So, this is why this particular test is called the student's T-tests, how does the t distribution look like? So, if your normal distribution is like this the T distribution is much flatter. So, the T distribution is also symmetric. So, this is your T and this is your normal.

So, what you can see for the T distribution this is also symmetric. So, the symmetry, but it is more flatter with heavier tails, what do you mean by heavier tails? You see in this portion of the curve these probabilities are higher and at that peak it is much suppressed.

So, this curve is on an average more flatter and if you increase your sample size or if as n tends to infinity T distribution merges with the normal distribution.

(Refer Slide Time: 18:49)

9. TEST OF HYPOTHESIS for 't' distribution

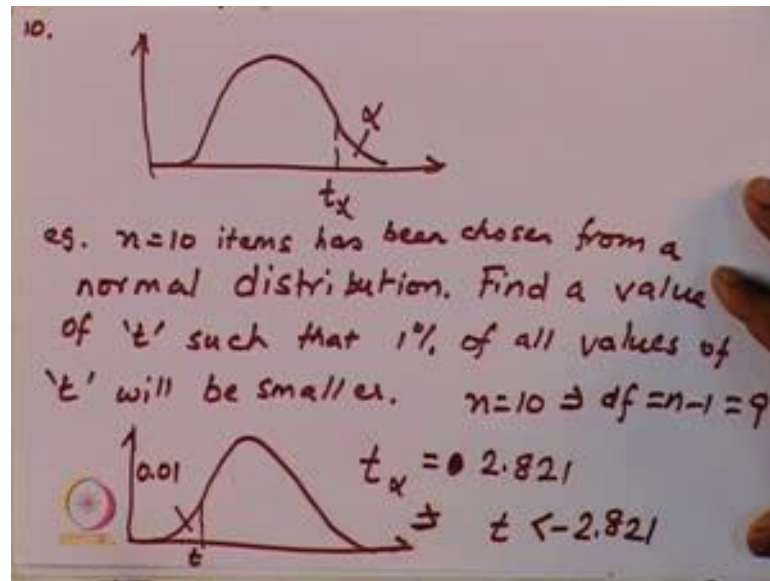
t- TABLE

df	$t_{0.1}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	df
1	3.078	-	-	-	
2	1.886	-	-	-	
3	
4	
⋮					

df = DEGREES OF FREEDOM
= $n - 1$
↳ SAMPLE SIZE

So, for large sample sizes you will get the same value whether you use the T test or the normal test normal distribution. So, for T distribution, how do we do test of hypothesis? So, what is provided in a t-table are the following, you have something called df or the degrees of freedom and you have t of point 0.1, t of 0.05, t of 0.025, t of 0.01 and get the degree of freedom and the degrees of freedom has values 1 2 3 4 and you have these values provided so on and so forth. So, df or the degree of freedom degrees of freedom is nothing, but n minus 1. So, if n is your sample size df is n minus 1.

(Refer Slide Time: 20:25)



So, as in previous cases, for the T distribution also you can calculate t of alpha where alpha is the area under this curve. So, for example, if n equal to 10 if a sample, so n equal to 10 items has been chosen from a normal distribution how will we find a value of t such that 1 percent of all values of t will be smaller. So, as per this question what we are given is n equal to 10 again implying the degrees of freedom is equal to n minus 1 equal to 9. So, if I draw this distribution. So, 1 percents it corresponds to 0.01. So, I want to find the value corresponding to, I want to find out the value corresponding to this particular value of t for n equal to for df equal to 9. So, I can as we have plotted for corresponding to 0.01 and df equal to let us say 9 we find out what is the value of t.

So, since we need t to be smaller for all these other values. So, let us say this t value, t of alpha let us say comes out to be 2.8, so implying since all our other values have to be smaller than t's then t has to be less than minus 2.821. So, that is how you would do it.

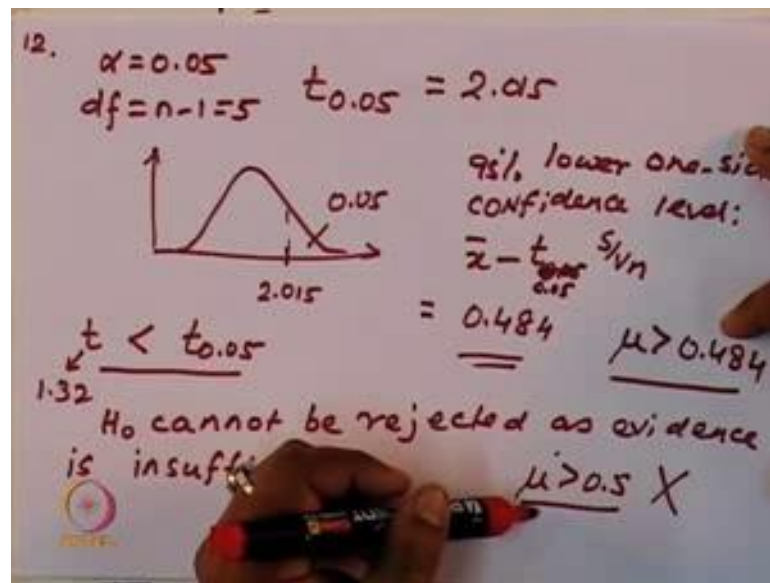
(Refer Slide Time: 22:57)

Ex: A new process for producing synthetic diamonds can be profitable if avg. wt. is > 0.5 karat.
 $n=6 \rightarrow 0.46, 0.61, 0.52, 0.48, 0.57, 0.54$
Is the avg. wt. > 0.5 karat.
 $H_0: \mu = 0.5$ $\bar{x} = 0.53$
 $H_a: \mu > 0.5$ $s = 0.056$
 $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.53 - 0.5}{0.0559/\sqrt{6}} = 1.32$

Let us solve a sample example. So, a new process for producing synthetic diamonds can be profitable if average weight is greater than 0.5 karat. Now you have taken n equal to 6 samples and their weights are respectively 0.46, 0.61, 0.52, 0.48, 0.57 and 0.54. So, you want to know, based on this is the average weight greater than 0.5 karat.

So, how will you solve this problem you begin with your null hypothesis which is μ equal to 0.5, your alternate hypothesis μ is equal to 0.5 is μ is greater than 0.5. So, for these values I can find calculate \bar{x} to be 0.53 and I can calculate their standard deviation which will turn out to be 0.056. With these values of \bar{x} and s , I can calculate the t statistic which is \bar{x} minus μ_0 by s by root n its 0.53 minus 0.5 by 0.0559 by root of 6 and I get a value of 1.32. Now corresponding, if I were to use, if I were to use α equal to 0.05 and corresponding to df equal to n minus 1 equal to 5 I can calculate the I can look up the t table to know that $t_{0.05}$ is 2.015.

(Refer Slide Time: 25:28)



So, if I were to draw this, corresponding to this you have t s 2.015. But our calculated t , so t is less than t of 0.05 right, t is less than this t value we have calculated to be 1.32, t is less than 0.05. So, I can conclude that H_0 cannot be rejected as evidence is insufficient, you can calculate the 95 percent lower confidence interval one sided confidence interval of a level, this comes out to be \bar{x} minus t of 0.05 into s by root n which is 0.484, sorry t of corresponding (Refer Time: 27:09) t into s by root n , so 0.25 only you calculate to 0.84. So, this tells us that μ is greater than 0.484. But we want our assertion was μ is greater than 0.5. So, this is not true because even a value of 0.5, 0.49 is not greater than 0.5.

With that I conclude our session for today. In the next class we will begin with some other, some little more discussion on T-test and then go into chi-square distribution.

Thank you for your attention.