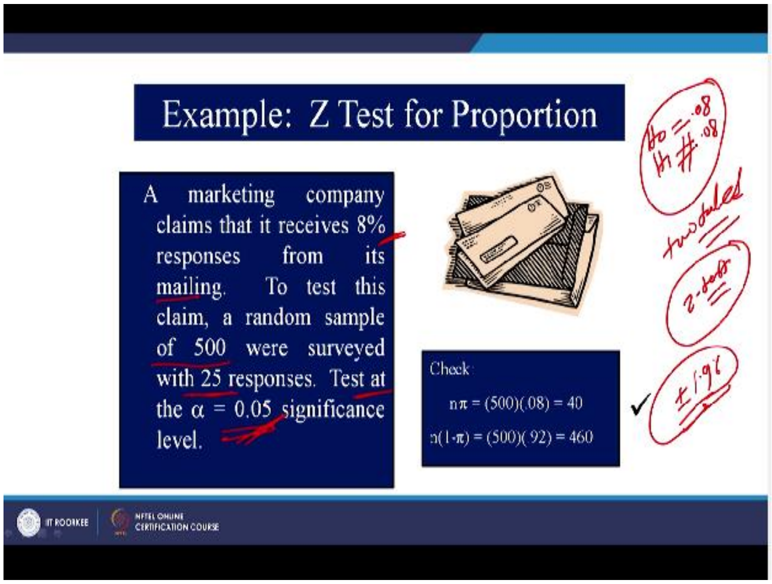
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology - Roorkee

Lecture – 35
Hypothesis Testing of Proportions - II

Hello friends. I welcome you all in this session. As you are aware in previous session we discussed questions on one sample one tailed t-test. We have seen questions on hypothesis testing of proportion where in we have used z-test and it was a two-tailed test, right. Let us look at one more example on hypothesis testing of proportion and the question is like this.

(Refer Slide Time: 00:57)



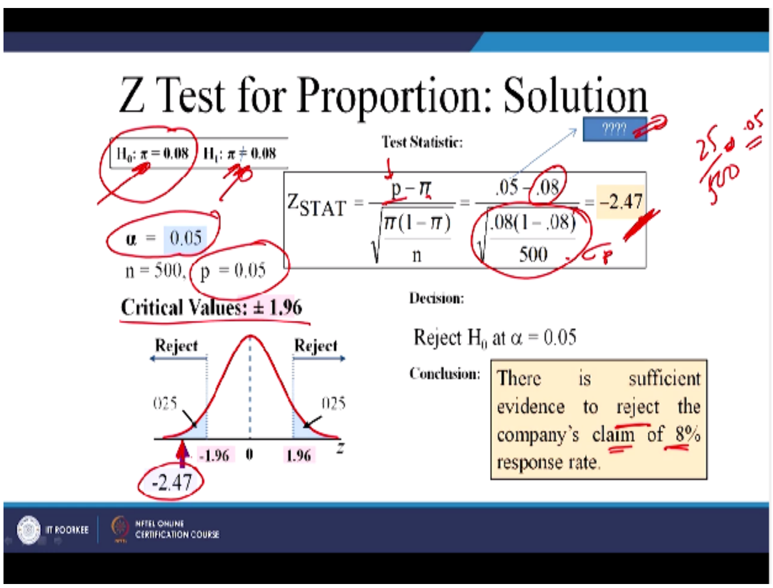
A marketing company claims that it receives 8% responses from its mailing. So this is the claim of the company, 8% responses the company receives from its mailing, right. Now you want to test whether this claim is correct or not. So what you have said, what you have done, you just collect a sample of 500 respondents or samples and you found that only 25 responses. So you need to test this hypothesis z, 0.05 significance level.

So how would you solve this question? A marketing company claims that it receives 8% of responses from its mailing. To test this claim, a random sample of 500 were surveyed with 25 responses. Test this hypothesis 0.05 significance level. So you have to take couple of decisions over here for solving this question. So you need to first find what is your null hypothesis and alternative hypothesis, right and then what?

Whether you would be using t-test of z-test, right, okay. So the null hypothesis here is that the company's claim is that it receives 8% responses. So it is 0.08, right, 0.08 is null hypothesis and alternative hypothesis is that it does not receives this much responses from mailing list or from mailing. Is not it? So this is a case of two-tailed test. Two-tailed, what? Whether it is t-test as a test.

Do you think there are two conditions you are fulfilling of applying t-test? If you do not fulfil those two conditions, then use z-test. So here, you would be using which test? Z-test, right and the significance level is 0.05. Since we are using z-test, what would be the critical limits, critical table limits? It would be 1.96. Is not it? I think by this time you would have understood that 0.05 significance level z-values are 1.96 both sides, right.

(Refer Slide Time: 03:53)

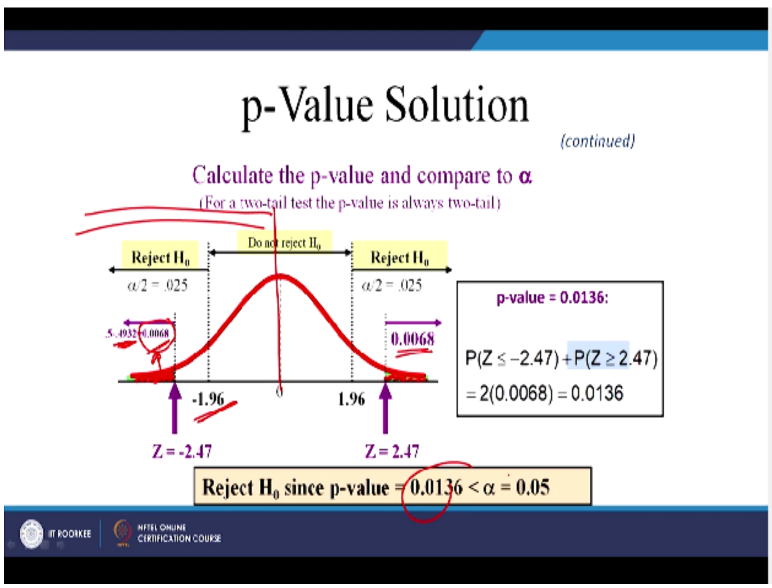


So this is your null hypothesis, operation proportion is 0.08 and it is not 0.08, right. Now this is your sample proportion, right and how did you get this p value? How should you be getting it? Since π is known, right is 0.08. It is 8%, right, it is there is question itself. So this is 0.06 alpha and this sample proportion is 0.05. How did you get this? Because in a sample of 500 which you surveyed, you just had 25 responses, right.

So it is 25/500, right. So it becomes 0.05. Is not it? So that is why p is equal to this, pi is equal to this and this is standard error of proportion. This is what is there. So you should calculate z statistics, s - 2.47 and of course, you critical values are 1.96, right, + and -. So compare your calculated z value with critical values. So this is, this end rejection region, right, this -2.47 is rejection region.

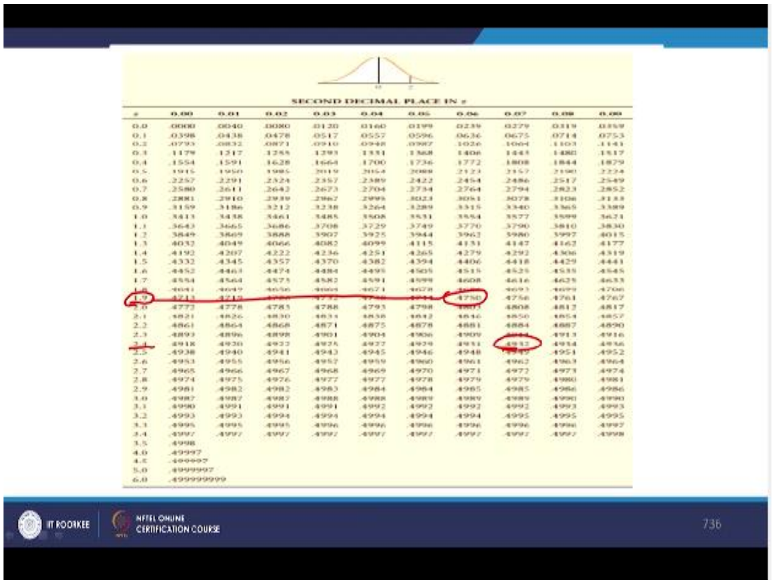
So you will reject the null hypothesis. So there is a sufficient evidence to reject the company's claim of 8% response rate. So we will say that the response rate is not 8%, right. We are saying that this is not accepted. In other words, we have rejected this and we have accepted the alternative hypothesis, right. It means the response rate is less than 8% or more than 8%, right. So let us look at one more example on hypothesis testing of proportion.

(Refer Slide Time: 06:16)



In fact, the same question you can solve using p value approaches well. So this is the case wherein z value is 1.96. Now at 1.96, you need to look at the probability values. For that, you need z table. So let us look at what are those values there in z table?

(Refer Slide Time: 06:45)



So at 1.96, this 1.9 and this is the one, right, 0.475, right. So we are interested in this area, right. This is 0.49, that is not the value. In fact, yes, at z, let us look at, at z 2.47, 2.47, 2.4 and 7 is 0.4932, right. So z value, which we have calculated is 2.4, right, 2.47, right. So at 2.47, the area under curve is 0.4932. This 0.4932, this one, this area is in fact 0.062 because this area is 50%, right, 0.5 and we are interested in this area, right.

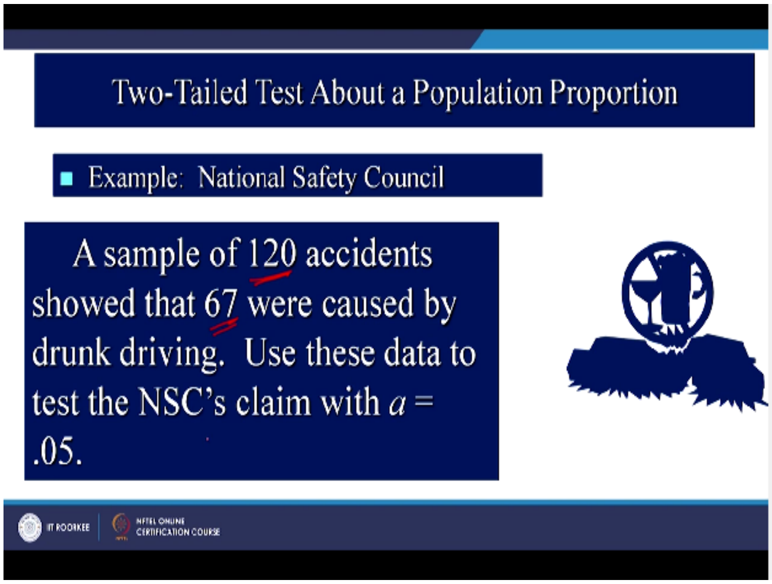
So 0.5-0.4932 and this is the area. This side and similarly on right side, you will get the same value, same area, right. So you just add these 2, you will get value or probability value. So this is 0.01362. Is p value less than alpha? Yes, and no. Is p value less than alpha? So p value is less than alpha we will reject, null hypothesis, right. So this is what you should take decision using p value approach, right.

So using critical value approach, we rejected null hypothesis. Is not it? Similarly, using p value approach, we have rejected hypothesis and you can use in fact the confidence interval approach as well to confirm this particular finding of rejecting this null hypothesis, right. So let us look at one more example. So this question is on hypothesis testing of population proportion using two-tailed test, right.

(Refer Slide Time: 09:32)

So there is National Safety Council. Now for Christmas and New Year's week, the National Safety Council estimated that 500 people would be killed and 25,000 would be injured on the nation's roads, right. This is what the National Safety Council estimated, right. The NSC claimed that 50% of the accidents would be caused by drunk driving. The council claims that the 50% of the accidents would be caused by drunk driving.

(Refer Slide Time: 10:21)

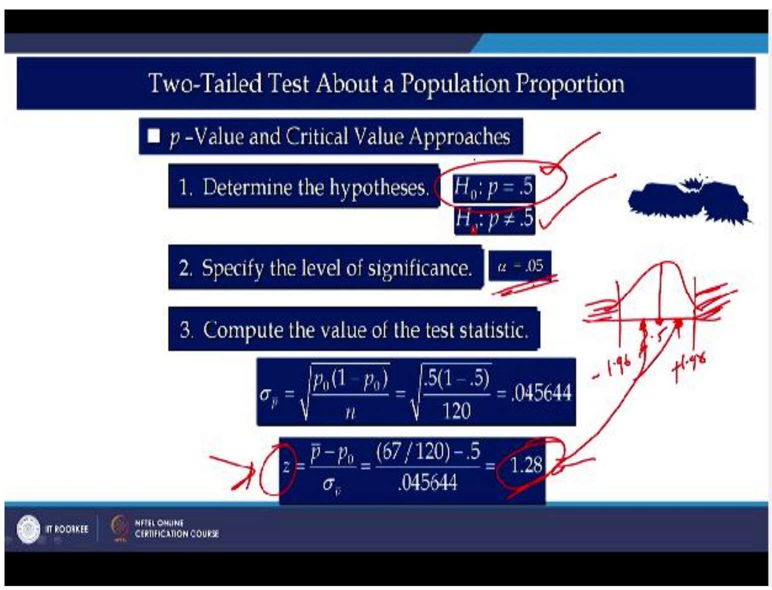


The council takes a sample of 120 accidents and showed that 67 were caused by drunk driving. Use these data to test National Safety Council's claim. And what is the claim? Let, what is the claim, that 50% of the accidents would be by drunk driving. But when they took a sample of 120 accidents, there were only 67 cases of drunk driving. So what would be null and alternative hypothesis here? Just look at this question once again.

Read this question carefully and try to formulate null hypothesis alternative hypothesis. So the NSC claimed that 50% of the accidents would be caused by drunk driving. So this would be your null hypothesis, right. This H0=0.5 and alternative hypothesis would be, it is not equal to 0.5. Is not it? So this is a case of two-tailed test, right. In the begin inning itself, we have said. But by reading the question and by analysing the situation, you should be able to frame null hypothesis as well as alternative hypothesis.

Now this is a case where in you have to apply either z-test or t-test, right. So how would you take the decision? It is a case of t-test or z-test? Let me look at; let me give you this part of the question as well. A sample of 120 showed that 67 were caused by drunk driving. So you get any hint or not. So this is a simple case of z-test, right, okay.

(Refer Slide Time: 12:43)

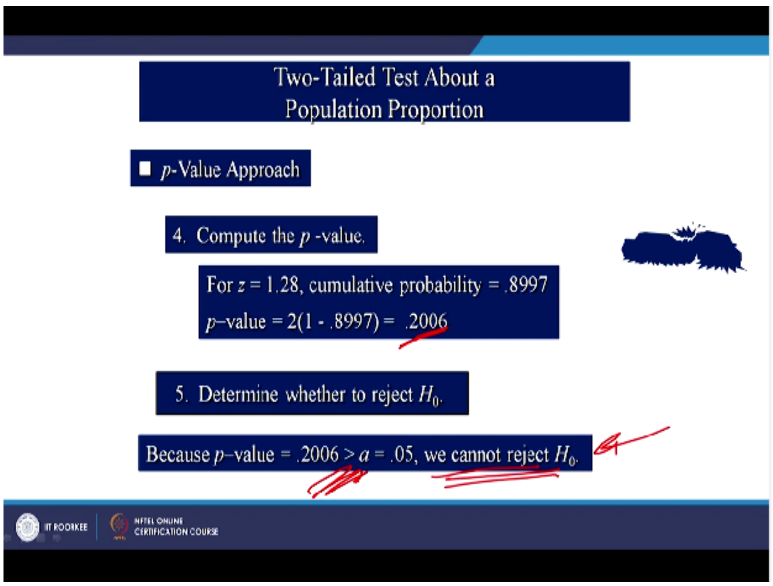


So you would be having, let us say, the first step is you have formulated null hypothesis and alternative hypothesis. This Ha, p is not equal to 0.5, level of significance 0.5. We can calculate the appropriate z statistics which is z here that is 1.28, right. Since this is a case of two-tailed test, right. So how the distribution would look like. Is not it? So it would be like this. Is not it? So you have got, these two are rejection region, right.

And hypothesized population parameter is 0.5. Is not it? z calculated is 1.28 and what will be the table value of z. Of course, it would be 1.96, is not it? Because significance level is 0.05, right. So this is -1.96, this is +1.96 and calculated value is somewhere here, right. This is somewhere here, right, 1.28, not here, somewhere here. Is not it? So you will reject null, not reject null hypothesis.

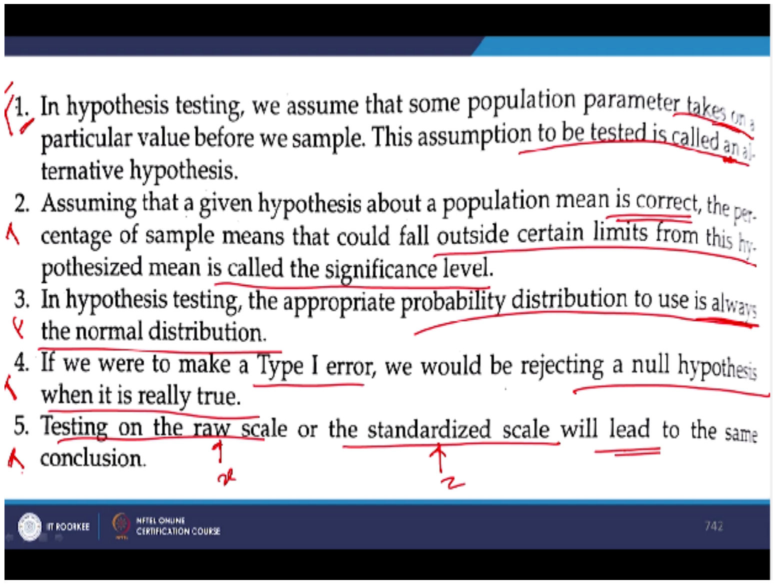
Since the calculated z value falls in critical range, so we will not reject null hypothesis. We will not reject null hypothesis and we will say that the 50% of accident would be by drunk driving, right. So this is how you can solve a question like this.

(Refer Slide Time: 14:40)



In fact, in this question, you will get the same answer using p value approach as well. So we will not reject null hypothesis because p value is more than alpha. Is not it? So we will not reject null hypothesis. The same is the decision we took here. We were not rejecting null hypothesis because calculated z value is in none rejection. So we will not reject null hypothesis and same is the answer using p approach or probability approach.

(Refer Slide Time: 15:18)



So let us look at couple of exercise on hypothesis testing of one sample test and using t-test or z-test using one-tailed test or two-tailed test, right. So we have worked out lots of examples on this and now we will move on to exercise, right. So let us look at first question. And you have to tell me whether this is true or false. In hypothesis testing, we assume that population parameter takes on a particular value before we sample.

This assumption to be tested is called an alternative hypothesis. What do you think? In hypothesis testing, we assume that some population parameter takes on a particular value before we sample. This assumption is to be tested is called alternative hypothesis or some other hypothesis? So this is called null hypothesis, right. So this statement is false, right. Let us move on to question number 2.

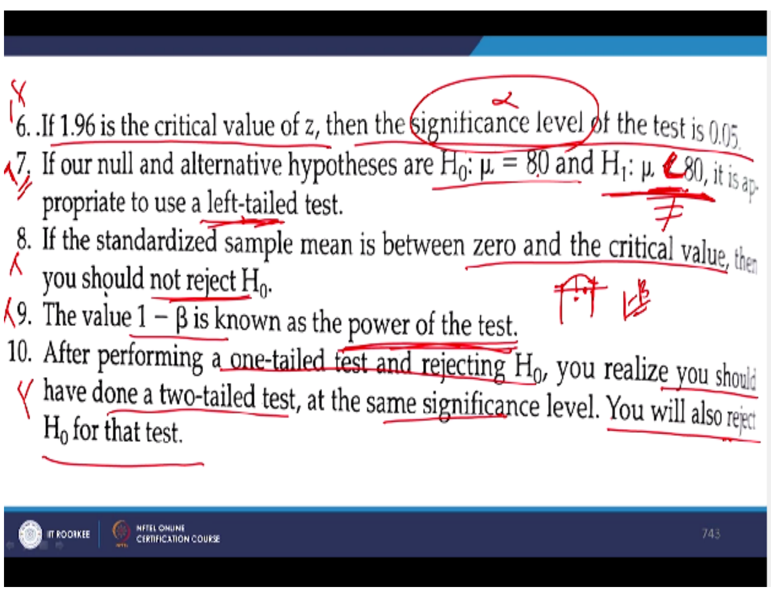
Assuming that a given hypothesis about a population mean is correct. So we know that the given hypothesis, our population mean is correct. The percentage of sample means that could fall outside certain limits from this hypothesized mean is called the significance level. Is this correct? The percentage of sample means that could fall outside certain limits from this hypothesized mean is called significance level.

Yes, this is called significance level. So this is true. So let us look at next one. In hypothesis testing, the appropriate probability distribution to use is always the normal distribution. Just read this sentence carefully. In hypothesis testing, the appropriate probability distribution to use is always the normal distribution. Is this correct? Do we always use normal distribution or z distribution?

No. We sometimes use two distribution as well, right. So this statement is false, right. Let us look at the next one. If you were to take type I error, we would be rejecting a null hypothesis when it is really true. So what is type I error and type II error? Just think of it. If we were to make a type I error, we would be rejecting a null hypothesis when it is really true. Yes. When we reject a true hypothesis, it is called type I error, right.

So this is true, okay. Let us look at point number 5. Testing on the raw scale or standardized scale will lead to the same conclusion. Testing on the raw scale or z scale will lead to the same conclusion. So you should know what is, in fact, it is nothing but z scale, right, z values, right. Raw scale means the x values of that particular variable, right. So what do you think? Testing only raw scale or the standardized scale will lead the same conclusion, yes. It will lead the same conclusion, right. Let us move on to the next question.

(Refer Slide Time: 19:40)



If 1.96 is the critical value of z, then the significance level of the test is 0.05. Significance level means alpha. Is this true? What do you think? If 1.96 is the critical value of z, the significance level of the test is 0.05. This is false statement. In fact, you should know how to calculate or how

to find out probabilities for given z value and how to convert, let us say, z values into probability, right.

So you need to know how to look at table carefully. So sixth number is false, right. If our null and alternative hypothesis are these, null hypothesis is this and alternative is less than that. It is appropriate to use left-tailed test because this sign is less than type. So we will use left-tailed test, right. So it is appropriate to use. So seventh is true. Is not it? Had it been like this, greater than 80, then you would have used right-tailed test.

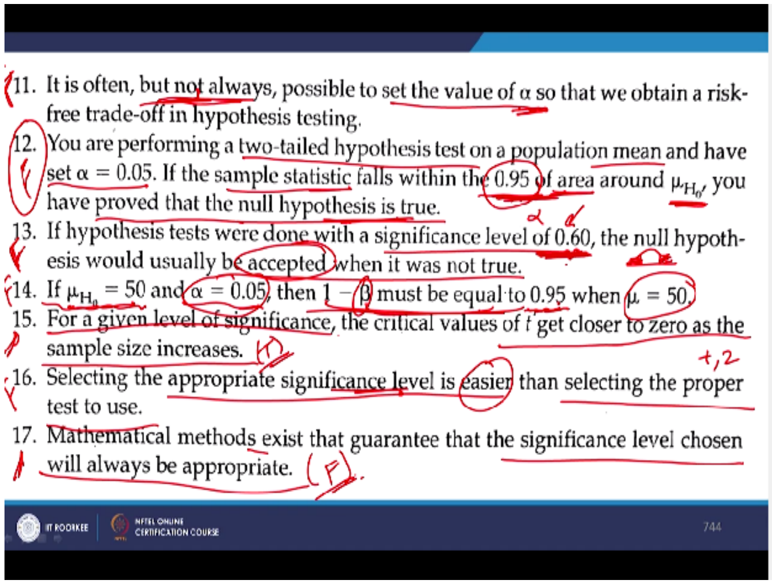
If had it been not equal to, then we would have used two-tailed test, right. Let us look at eight one. If the standardized sample mean is between 0 and critical value, then you should not reject null hypothesis. If the standardized sample mean is between 0 and critical value, then you should not hypothesis. It means there is a situation like this. So this is your 0 value, right and let us say, so this is your critical value and the standardized sample mean is between this, right.

So then you should not reject null hypothesis. Yes, you should not reject null hypothesis. This is true, right. Let us say, this is one more critical value, right. So these are critical values and your sample mean is in this range. 1-beta is known as power of test. Is it correct? In fact, we have seen relationship between type I and type II error wherein we have said that 1-beta is power of test, right.

What does it mean? What do you mean by power of test? Power of test is the probability of rejecting a null hypothesis when it is false. So that is power of test. So this statement is true. Let us look at the next one. After performing a one-tailed test and rejecting null hypothesis, you realize you should have done a two-tailed test, at the same significance level. You will also reject null hypothesis for that test.

Keep in mind, you wanted to have a 2 sample, a two-tailed test. But you just used one-tailed test, right. So do you think that the result would remain same or the result would be different? Would you still like to reject null hypothesis? This is false. Your answer will change. Is not it? Because you will have 2 rejection regions in this case, right. So your answer might change. Is not it? But it will not, it may, in fact, it will change. Is not it? So this is false, right.

(Refer Slide Time: 24:46)



Let us look at the next one. It is often, but not always, possible to set the values of alpha so that we obtain a risk-free trade-off in hypothesis testing. It is often but not always possible to set the value of alpha. Generally, we set the value of alpha in second state, right. Before deciding on whether you would be using t-test or z-test, right. So it is often but not always. In fact, it is always possible to set the value of alpha.

So this statement is false. Is not it? So you set the value of alpha a priori, right, before calculating t statistics of z statistics. Let us look at next one. You are performing a two-tailed hypothesis test on population mean and have set alpha=0.05. if the sample statistics fall within 0.95 of area around s, around hypothesized population mean. You have proved that the null hypothesis is true. Is it correct?

What do you think? You are performing a two-tailed hypothesis test on a population mean and have set alpha is equal to this, if the sample statistic falls within 0.95 of area around, 0.95 area around this. You have proved that null hypothesis is true. No this is completely false. In fact, when you say; in fact, you can prepare distribution for this question and then you can draw your rejection region and non-rejection region, right.

In fact, it is not 0.95 of area, but it is from, the sample statistics has to fall within critical limits, right. Then you would have said, you would have not rejected the null hypothesis and you would have said the null hypothesis is true. Is not it? So you need to look at critical values. Let us look at the next one. If hypothesis tests were done with a significance level of, let us say, 0.60, right, significance level is 0.60.

So this is your alpha value. Is not it? So alpha value is this. It means this is 30%, this is 30%, right and remaining is 40%, right. The null hypothesis would usually be accepted when it was not true. What do you think? Why do we test, why we do not test hypothesis at high significance level? In fact, in one of the slides, I have taught you about why we do not test hypothesis a high significance level.

Because there is a probability of rejecting a null hypothesis which is true. So this statement is false, right. In fact, the null hypothesis would usually be rejected when it was not true, okay. So at high significance, the null hypothesis would be rejected if it is true, right, okay. So this is a false statement, right. Let us look at fourteenth one. If hypothesized population mean is 50, alpha is 0.05, then 1-beta must be equal to 0.95 when mean is 50.

What do you think? Is this true statement or false statement? Fourteenth, if hypothesized population mean is 50, alpha is this, then 1-beta must be equal to this. In fact, this is not 1-beta, this is 1-alpha. So this statement is false one. Is not it? So 1-alpha would be equal to this, right. So with alpha and beta, you can find out whether this statement is true or false. So let us move on to fifteenth one.

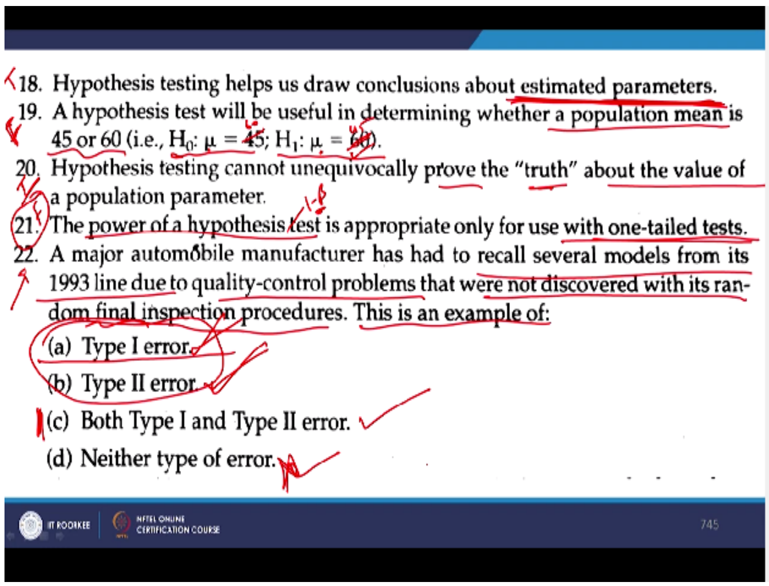
For a given level of significance, the critical value of t get closer to 0 as sample size increases. What do you think? For a given level of significance, the critical values of t get closer to 0 as sample size increases. Yes, it happens. So this statement is true one. Let us look at the sixteenth one. Selecting the appropriate significance level is easier than selecting the proper test to use. What do you think?

Selecting the proper significance level is easier than which test is to be used or to be applied.

What do you think? Is it correct statement? In fact, it is very difficult to decide the significance level. But it is easy to use, it is easy to decide whether you want to use t-test or z-test, right. So this statement is not correct. This statement is false. Let us look at the next one. Mathematical methods exist that guarantee that the significance level chosen will always be appropriate.

So do you have mathematical methods which would tell you that significance level chosen or the chosen significance level was a correct one. In fact, you do not have any such mathematical method, right. So this statement is false, right. Seventeen number is false. Let us look at the next one.

(Refer Slide Time: 32:17)



Hypothesis testing helps us draw conclusion about estimated parameter. Is it correct? Is it correct, eighteenth one? Hypothesis testing helps us draw conclusion about estimated parameters, yes. About, we generally estimate population parameter first or we assume population parameter and only basis of that sample, we prove or disprove our hypothesis, right. Let us look at nineteenth one.

A hypothesis test will be useful in determining whether a population mean is 45 or 60. So null hypothesis is this and alternative hypothesis is 60. Is this correct? Hypothesis test will be useful in determining whether population mean is 45 or 60, given null hypothesis is this and alternative hypothesis is this. No, it is not possible. Is not it? So you can have either 45 over here, is not it?

If not 45 over here, and if we keep 60 over here, then you need to keep 60 over here, right.

So this statement is false, right. So let us look at one more. Hypothesis testing cannot unequivocally prove that truth about value of population parameter. Hypothesis testing cannot unequivocally prove the truth about value of population parameter. So what about this? Yes, it can prove in fact. Hypothesis testing is a procedure which can unequivocally prove the truth about the value of population parameter, right.

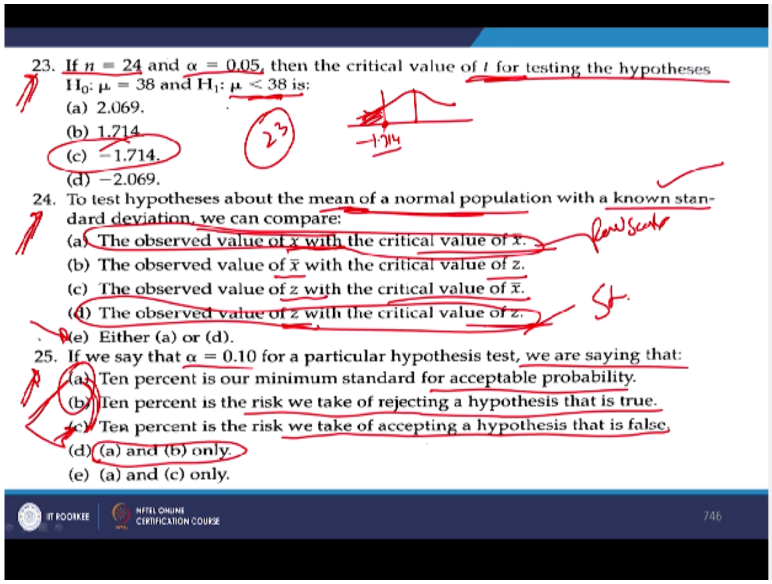
So this statement is true. So we will look at these 2 questions as well. The power of hypothesis test is appropriate only for use with one-tailed test. The power of hypothesis test is appropriate only for use with one-tailed test, is this correct? What is power of test? This 1-beta, is not it? Is appropriate only for use with one-tailed test or can be used with two-tailed test as well. Just think over it.

Yes, this is false. We can use it for two-tailed test as well, right. So this is false, right, 21 number is false, okay. Just look at this question, 22nd. A major automobile manufacturer has had to recall several models from its 1993 line due to quality-control problems that were not discovered with its random final inspection procedure. This is an example of type I error, type II error, both type I and type II error, or neither type of error. What would you do?

Let me tell you the answer cannot be C. is not it? Because you cannot have both the errors simultaneously. So either you are committing type I error or type II error. And of course, this also cannot be answer. Because the answer is somewhere, is either A or either B. So let us recall what is type I error. Type I error is an error when you reject a null hypothesis when it is true, right. So this is a case of type I error.

In fact, we have seen couple of examples on type I and type II error. You can always refer that lecture on type I and type II error. Now let us move on to the next one.

(Refer Slide Time: 37:05)



If n=, let us say, 24 and alpha=0.05 and the critical value of t for testing the hypothesis, for less than type, right. So this is a case of low tailed test, right or left-tailed test where n=24, right. So how, what would be the t value. You need to look at t table for this question. To answer this question, you need to refer to t table. So at what degrees of freedom, you should be looking at? You should be looking at 23 degrees of freedom.

Is not it and under which column? So you need to look at 0.90 column, right. So the t value would be for this question, this. So this is in fact a left-tailed test, right. So this -1.714, this value, right, okay. So this is a very simple question. You can easily find out the answer to this one. Let us look at the twenty fourth one. To test hypothesis about the mean of normal population with a known standard deviation, we can compute.

So standard deviation is known and we are testing a hypothesis testing of population mean, right. Now the observed value of x with the x, so we can compare. We can compare what? The observed value of x with critical value of x bar, the observed value of x bar with critical value of z, the observed value of z with critical value of x bar, the observed value of z with critical value of z.

So what is the correct answer? To test hypothesis about the population, mean of a normal population with a known standard deviation, we can compare what? In fact, this is the right answer, a or d, right. So of course, observed the value of z with the critical value of z. Then only we will take decision whether to reject or not to reject or you can have this situation as a, is not it?

So either this raw scale and this is standardized scale or z scale. Let us look at the next one, twenty fifth. If we say that alpha = this for a particular hypothesis test, we are saying that, what does it mean? The 10% is our minimum standard error for acceptable probability, 10% is the risk we take of rejecting a hypothesis that is true. What is alpha? Alpha is also known as type I error, right, okay.

Ten percent is the risk we take of accepting a hypothesis when it is false, a and b, means these two or a and c, right, these two. What is correct? Yes, please? So the answer is a and b only. So 10% alpha is nothing but we are taking a risk that we reject the hypothesis when it is true or accept when it is false. Is not it? So in this session, we discussed about different questions on hypothesis testing of one-tailed test, two-tailed test, t-test and z-test. We will continue with some more questions in next session. Thank you.