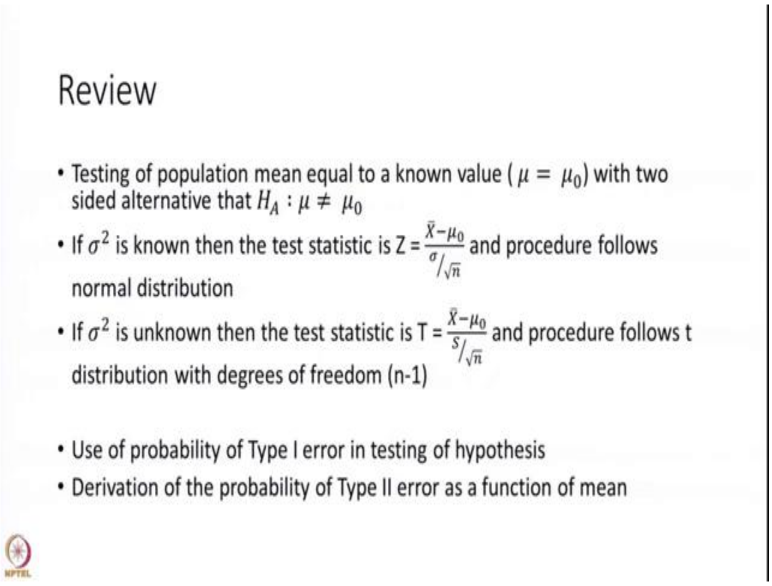
Dealing with Materials Data:
Collection, Analysis and Interpretation
Professor Hina A Gokhale
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay
Lecture 71
Hypothesis Testing IV

Hello, and welcome to the course on dealing with materials data. In the process of statistical inference, we are undergoing the sessions on hypothesis testing.

(Refer Slide Time: 00:34)



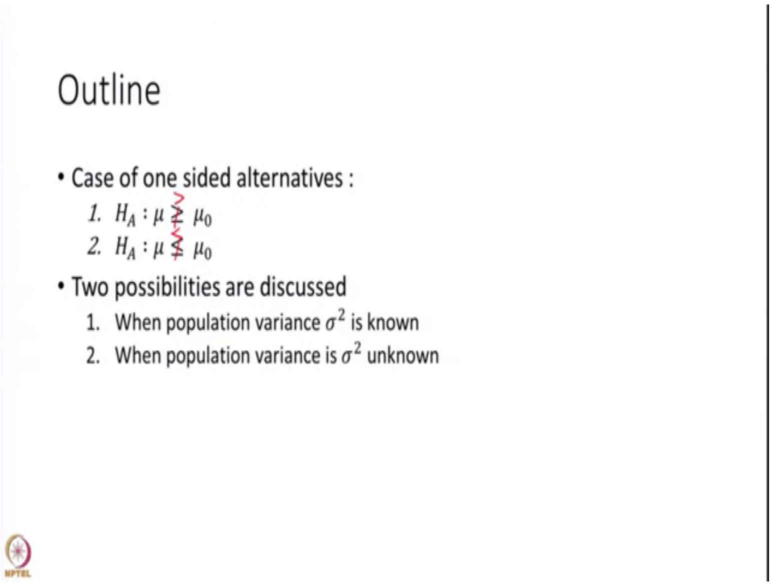
So far, we have considered the testing of hypothesis with the two sided alternative that hypothesis that

$$H_0 : \mu = \mu_0 \ \ vs. \ H_A : \mu \neq \mu_0$$

We found that when the population variance is known, the test statistic is Z which is

Z = $\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ If it follows, if the population is normal this follows a normal distribution. If σ^2 is unknown then the test statistic is T = $\frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ and procedure follows t distribution with degrees of freedom (n-1) . Then we also found that the six step of classical hypothesis testing procedure can also be carried out by working out the probability of type 1 error. And we also went through, how to derive the type two error as a function of alternate hypothesis mean Mu when it is not Mu 0.

(Refer Slide Time: 02:06)



In this session, we are going to consider the case of two-sided alternatives that can be two of them. Mu is greater than or it

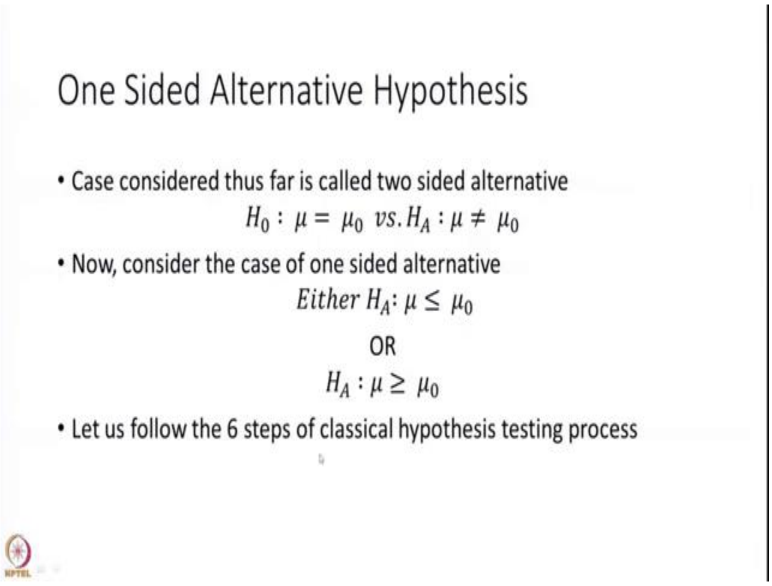
Case of one sided alternatives:

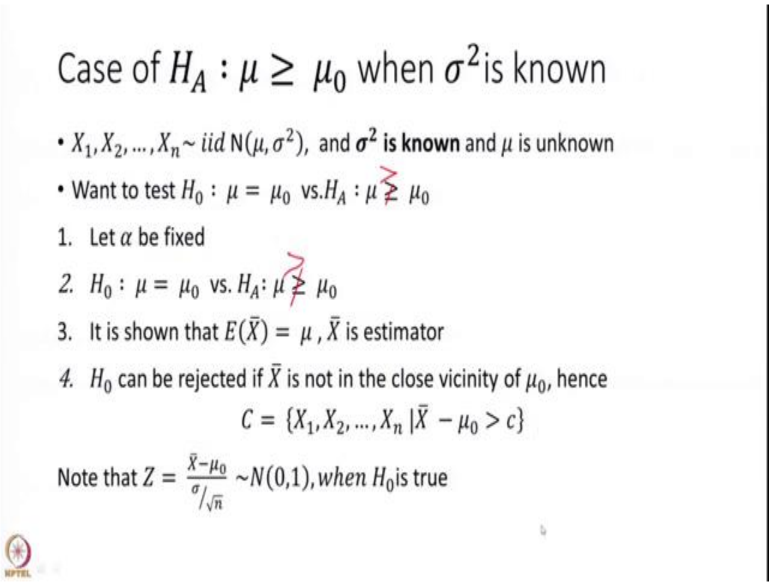
1.$H_A : \mu > \mu_0$

2.$H_A : \mu < \mu_0$

So, there are again we are taking the two possibilities under normal population assumption that variance is known and variances unknown.

(Refer Slide Time: 02:37)





One-sided alternative hypothesis you have either this or this and we will follow the six steps of classical hypothesis testing. So, let us take the case when sigma square is known Mu is unknown and we are assuming normality that is the population is normal with mean Mu and variance sigma square.

$H_0 : \mu = \mu_0$ vs. $H_A : \mu > \mu_0$

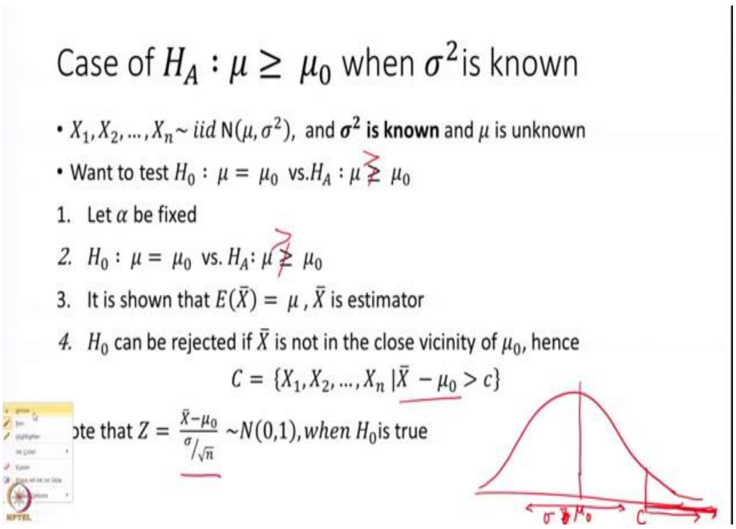
It is shown that $E(\bar{X}) = \mu$, $\bar{X}$ is estimator

H_0 can be rejected if $\bar{X}$ is not in the close vicinity of μ_0, hence

$$C = \{X_1, X_2, \ldots, X_n \mid \bar{X} - \mu_0 > c\}$$

Note that $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1),\ when\ H_0$ is true

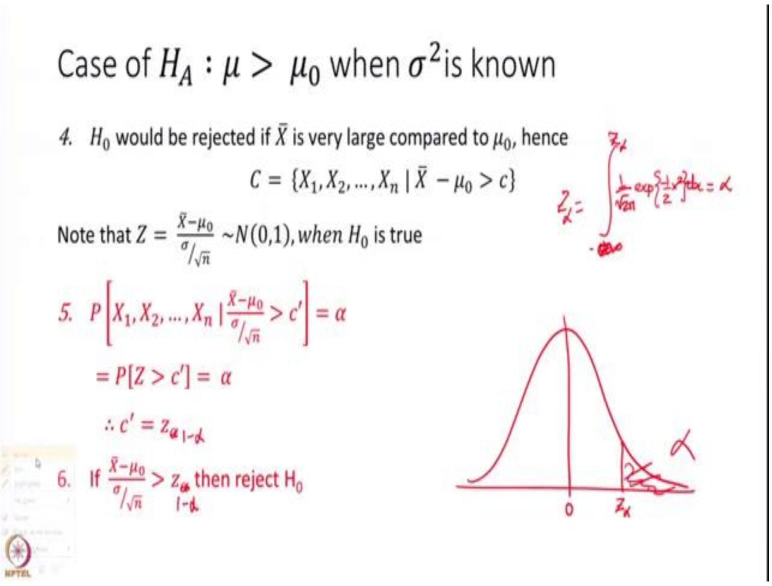
(Refer Slide Time: 03:56)



So, if you look at the sorry. If you look at the normal population, this is your Mu 0. And there is some sigma here, some value sigma is there. What we are saying is that the X bar has to be in the vicinity, but you have to reject it, if it is very far away on the positive side. So, actually if it is anywhere far away from Mu 0, you have to reject it, but to some extent we are going to accept it. So, we are going to make some limit if it is very far away. And this value is what I call C.

If it is very far away, then this is the region where I am going to reject it. So, I am saying that X bar minus Mu 0 has to be greater than C. So, this is if I plot this X bar minus Mu 0, it has to be greater than somewhere here. So, if your X bar value itself is some larger than this value, then you are going to reject it. So, again our test statistic is standard normal deviate Z, which is X bar minus Mu 0 over standard deviation over square root n, which is a normal with a known population mean and standard deviation, when H naught is true.

(Refer Slide Time: 05:30)



And therefore, we come to the test statistic, this is where the differences so,

$$P\left[X_1,\ X_2,\ \ldots,\ X_n \mid \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > c'\right] = \alpha$$

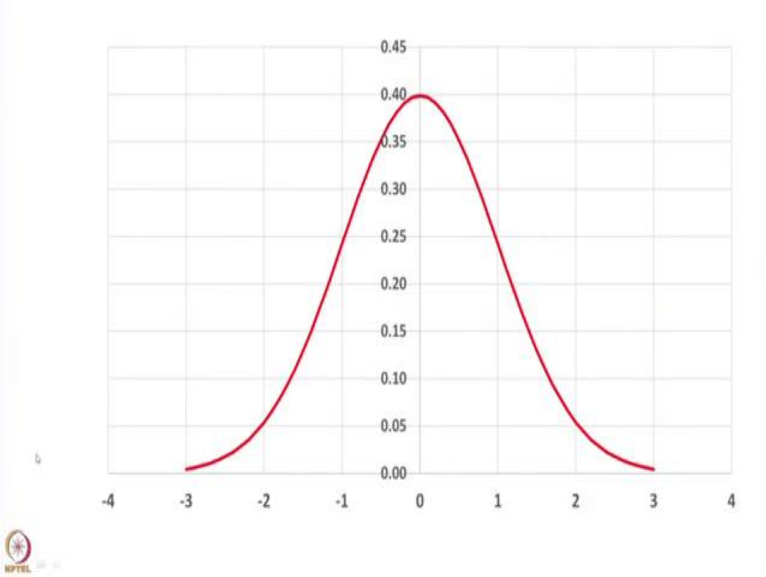
$$= P[Z > c'] = \alpha$$

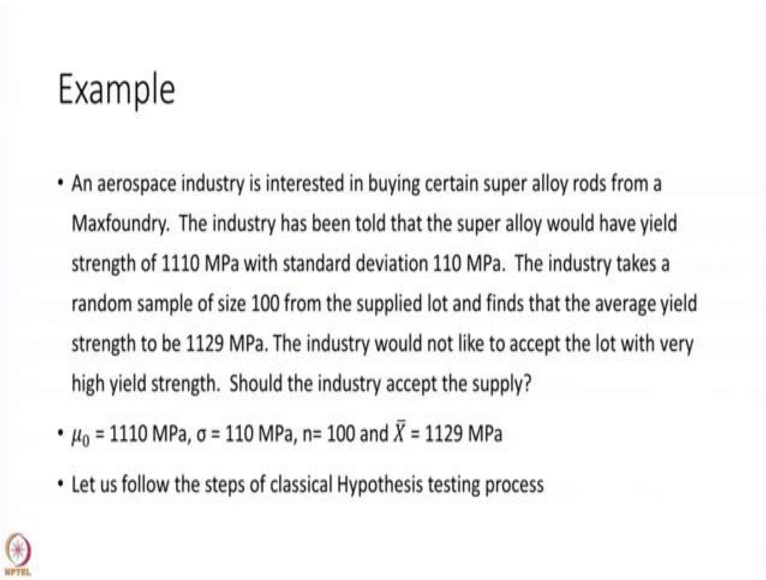
$$\therefore c' = z_{1-\alpha}$$

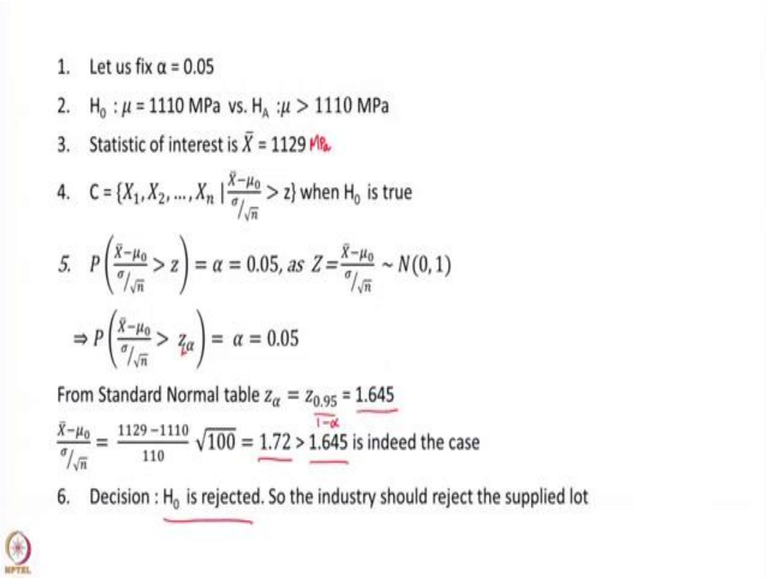
$$\text{Actually, If } \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha}$$

Then you reject the null hypothesis please make this correction.

(Refer Slide Time: 07:09)







So, anyway I have already shown in the plot. So, if we take the same example again, exactly the same example as before and now we do the 1 sided testing in which we take V is equal to Mu 0 is that is the hypothesis is that the population V is 1110 MPa what says alternative mean is greater than 1110 MPa. And at present we have found our, from our sample, the mean value, the sample mean is 1129 MPa, sorry for these mistakes in the slide. So, this is MPa. So, accordingly we work out procedure.

1. Let us fix α = 0.05
2. H_0 : μ = 1110 MPa vs. H_A :$\mu > 1110$ MPa
3. Statistic of interest is $\bar{X}$ = 1129
4. C = $\{X_1, X_2, \ldots, X_n \mid \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > z\}$ when H_0 is true
5. $P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > z\right) = \alpha = 0.05, as\ Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$

$$\Rightarrow P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha}\right) = \alpha = 0.05$$

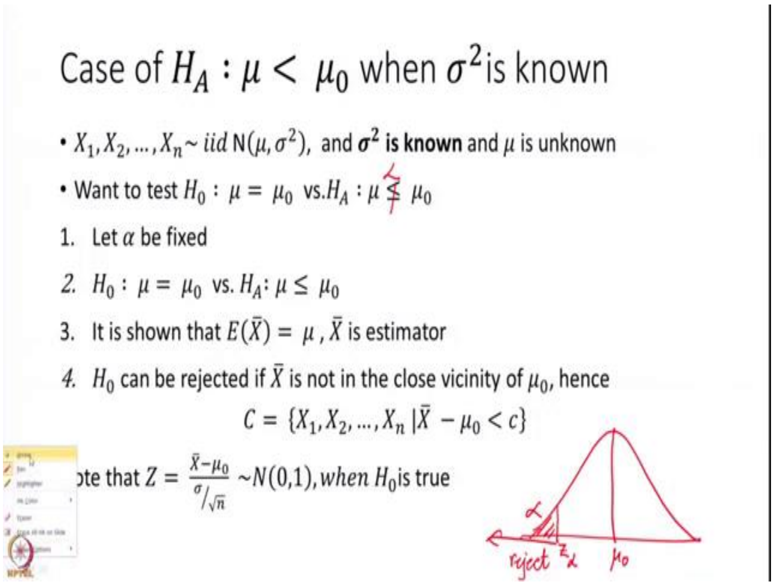
From Standard Normal table $z_{1-\alpha} = z_{0.95}$ = 1.645

$$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{1129-1110}{110}\sqrt{100} = 1.72 > 1.645 \text{ is indeed the case}$$

6. Decision: H_0 is rejected. So the industry should reject the supplied lot

I think the mean value has also been changed, this value X bar has also been changed.

Refer Slide Time: 09:28)



So, let us take the other alternate I think you would understand in the other alternate what has to happen before going into the detail if you are looking at the population distribution

- $X_1, X_2, \dots, X_n \sim iid$ N(μ, σ^2), and $\boldsymbol{\sigma^2}$ **is known** and μ is unknown
- Want to test $H_0 : \mu = \mu_0$ vs.$H_A : \mu < \mu_0$

1. Let α be fixed
2. $H_0 : \mu = \mu_0$ vs. $H_A : \mu < \mu_0$
3. It is shown that $E(\bar{X}) = \mu$, $\bar{X}$ is estimator
4. H_0 can be rejected if $\bar{X}$ is not in the close vicinity of μ_0, hence

$$C = \{X_1, X_2, \dots, X_n \,|\bar{X} - \mu_0 < c\}$$

Note that $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1),\ when\ H_0$is true

5. $P\left[X_1, X_2, \dots, X_n \,|\, \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < c'\right] = \alpha$

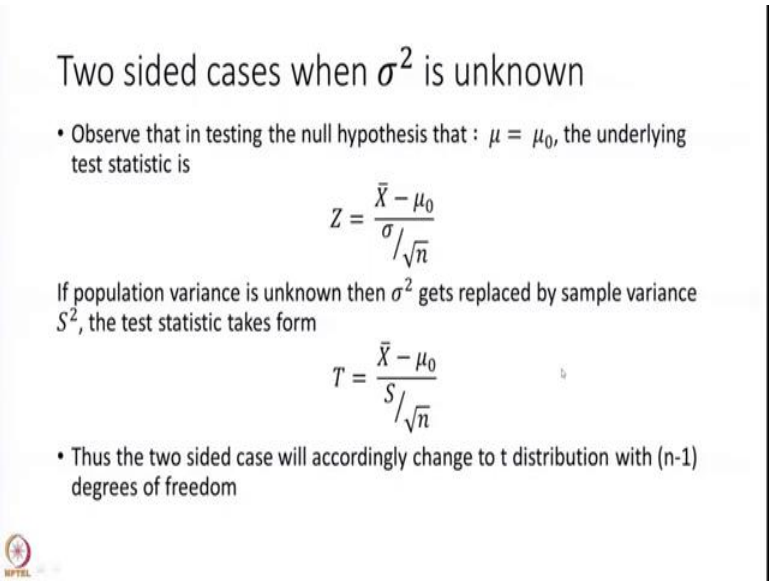
$= P[Z < c'] = \alpha$

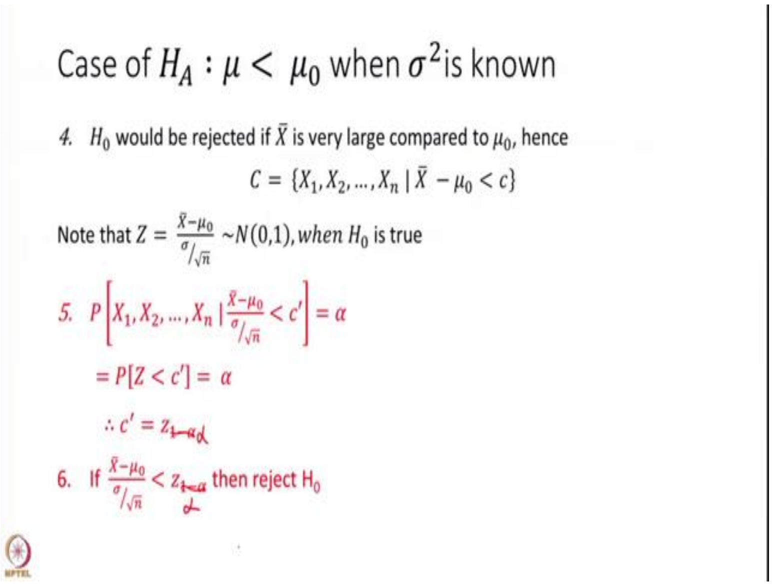
$\therefore c' = z_{\alpha}$

6. If $\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < z_{\alpha}$ then reject H_0

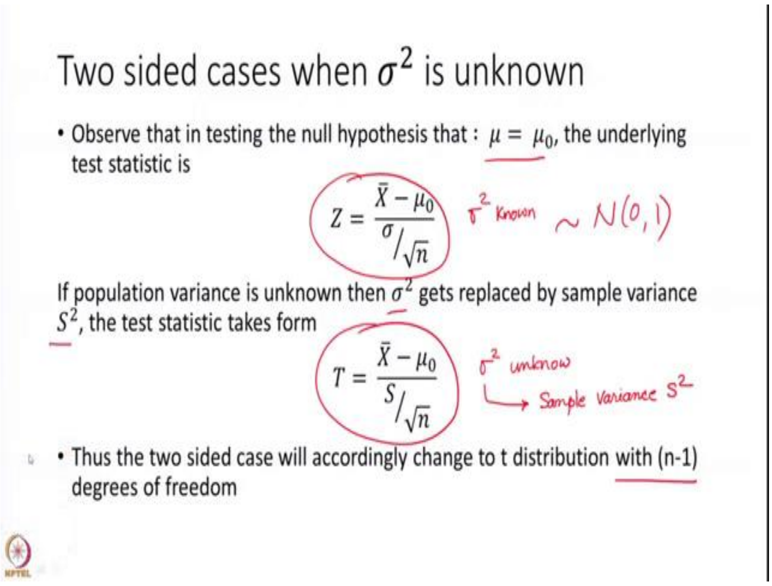
And if you take the example, you can work out that it is also, you can work out the example and you will find whether you take meanwhile you to be say 1108 which is smaller than 1110 and then see if the hypothesis, how the hypothesis procedure works and you reject the null hypothesis or the accept it.

(Refer Slide Time: 10:15)





(Refer Slide Time: 11:04)



The two-sided case when sigma squared is unknown, I do not think we need to go through the whole procedure by now, because we have clearly seen that under the null hypothesis $\mu = \mu_0$,

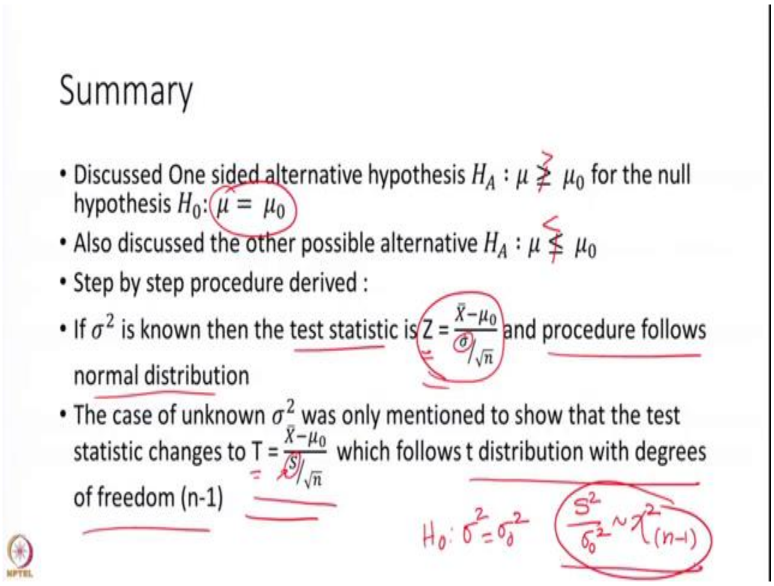
If population variance is unknown then σ^2 gets replaced by sample variance S^2, the test statistic takes form

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Thus the two sided case will accordingly change to t distribution with (n-1) degrees of freedom

While this becomes a normal distribution with 0 mean and 1 standard deviation. I do not think we need to repeat this in our totality because it is a very obvious case.

(Refer Slide Time: 12:24)



So, let us summarize it. We discussed 1 sided alternate of mean value to be $H_A : \mu > \mu_0$ or $H_A : \mu < \mu_0$. In both the cases we found that when sigma square is known, or rather in all the cases, we found that when sigma square is known, Z which is standard normal deviate, which is Z = $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ is the test statistic and the test will procedure will follow normal distribution. If sigma square is unknown, then the test statistic takes a value T = $\frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ because you replace sigma by sample standard deviation and it follows the T distribution with N minus 1 degrees of freedom.

Now, before we close the session, I would like to repeat or bring out 1 particular feature here. Remember that X bar that is the sample mean is a good estimator, it is an unbiased estimator of population mean. No matter whether the distribution is normal or non-normal or whatever. Similarly, sample standard deviation or sample variance is an unbiased estimator of a population variance well, it does not depend on the distribution. And therefore, these two statistics, the Z statistic, and the T statistic, play a central role when you want to test a hypothesis with respect to Mu.

Similarly, in the next session, we will say that if you want to have a hypothesis testing for the variance of the population, so you want to test a hypothesis, which is something like sigma square is equal to sigma naught square. And what do you think would be the central statistic? It has to be S square over sigma naught square. And it would be distributed as psi square with N minus 1 degrees of freedom. And just as the and Z this is going to be the test statistic. We will look into it in the next session. Thank you.