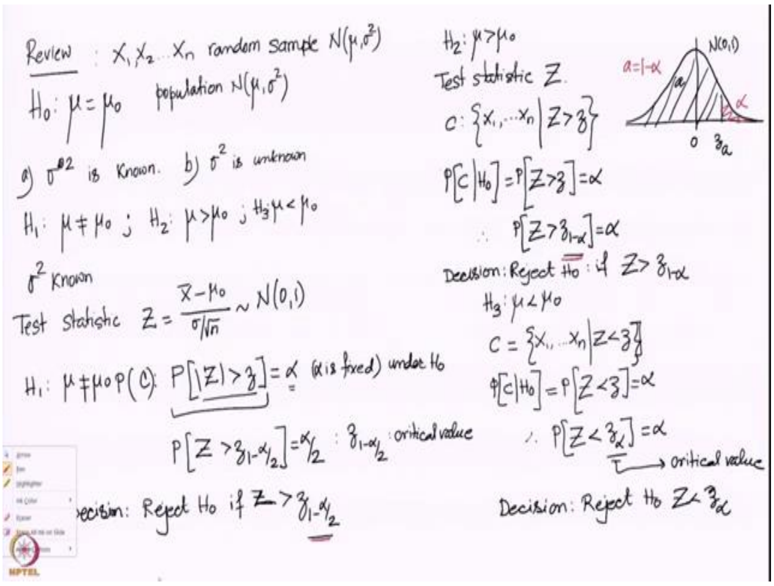
Dealing with Materials Data:
Collection, Analysis and Interpretation
Professor Hina A Gokhale
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay
Lecture 72
Hypothesis Testing V

Hello and welcome to the course on dealing with materials data. Presently we are going through the sessions on hypothesis testing.

(Refer Slide Time: 00:35)



We have so far considered the sessions way which deals with so if we quickly write down what we have seen, we have dealt with the hypothesis $H_0 : \mu = \mu_0$

We have assumed that population is normal. This mean Mu and variance sigma square. And we consider two possibilities, σ^2 is known and σ^2 is unknown .

In either case we would like to discuss or we have already discussed the three alternates. The first alternative we talk is a two-sided alternative,

$$H_A : \mu \neq \mu_0$$

Second alternative is that a $H_A : \mu > \mu_0$ and the third alternative we have discussed is $H_A : \mu < \mu_0$ under both the cases that sigma square is known and sigma square is unknown.

We found when sigma square is known, we found that test statistic is Z which is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Here we have assumed that we have a sample X1, X2, X3, XN is a random sample from normal population with mean Mu and variance sigma square. So, X bar is an average of X1, X2, X3, XN and we found that this is distributed as normal (0,1) it has no unknown parameters, it is completely known distribution.

And therefore, if we consider the alternative that Mu is not equal to Mu 0, we found that the critical region can be defined as probability or the probability of critical region can be defined as Z in absolute value is greater than some value Z which should be alpha, where alpha is fixed is the type one error. This is under H0 that is when the null hypothesis is true, we call this as a, this is the critical region and this is the probability of critical region, which is equal to alpha.

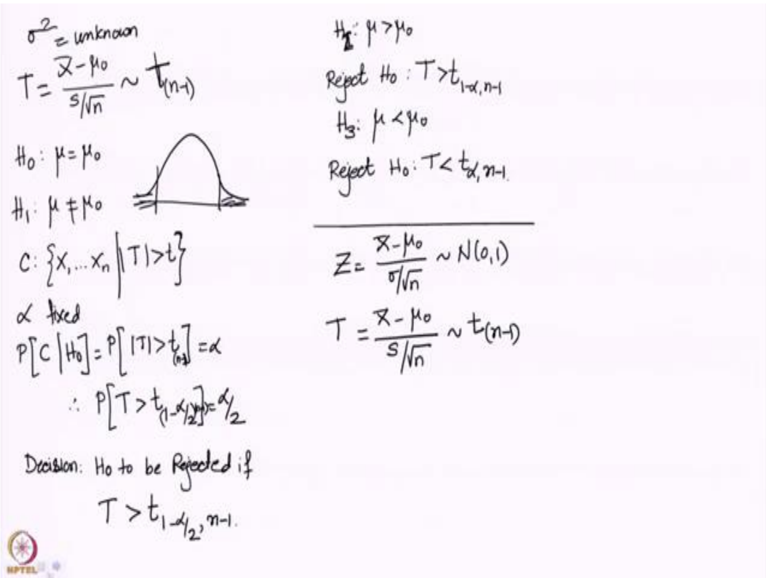
And then we find that actually in that case probably to of Z greater than small z 1 minus alpha by 2 is equal to alpha by 2. And this is the, in that case, Z one minus alpha by 2 is a critical value means that the decision is going to be reject H0, if this statistic Z is greater than Z 1 one minus alpha by 2. If we take the alternate of Mu greater than Mu 0, then again, the test statistic is same is Z and the critical region let us properly defined it is X1, X2, XN such that Z is greater than some value Z, small z. This is your critical region under H0.

So, so you can say that probability of critical region under H0 is equal to probability that Z is greater than small z should be alpha. And therefore, what we find is that it has to be probability that Z greater than Z1 minus alpha is equal to alpha. If this 1 minus alpha by 2 and 1 minus alpha is confusing you, I would like to remind you, that you take a standard normal population, this is normal 0, 1 this is your mean 0. Then Z of Z sub A is defined as this probability, so that this probability is A.

So, in this case, if we want this probability to be alpha, then A has to be 1 minus alpha. And therefore, we are taking 1 minus alpha or a similarly, we are taking this because this is two-sided. So, then when we come to the next possibility is that alternate of H that Mu is less than Mu naught. Then our critical region is going to be X1, X2, XN such that Z is less than small z, sorry it should be this way curved bracket. So, probability of C under H0 is equal to probability of Z less than small z, which should be alpha.

And therefore, we find that probability of Z less than Z alpha is equal to alpha and therefore the critical value, this is the critical value, this is the critical value. In this case the decision is, in the previous case the decision is if reject H0, if Z is greater than Z, small z1 minus alpha and in this case the decision would be reject H0. If Z is smaller than small Z alpha. So, this is what we have done so far.

(Refer Slide Time: 09:48)



Next thing what we have made change is that suppose we are taking the sigma square as unknown then please recall that the Z statistic turns into a T statistic, which is same as X bar minus mu 0 divided by S over square root N, where S is the sample standard deviation. And now this is distributed as T with N minus 1 degrees of freedom. So, again there is no parametric dependence on it, we already know N, so, we know N minus 1. So, we know the distribution, this should be small t sorry, it is a distribution that we are already aware of.

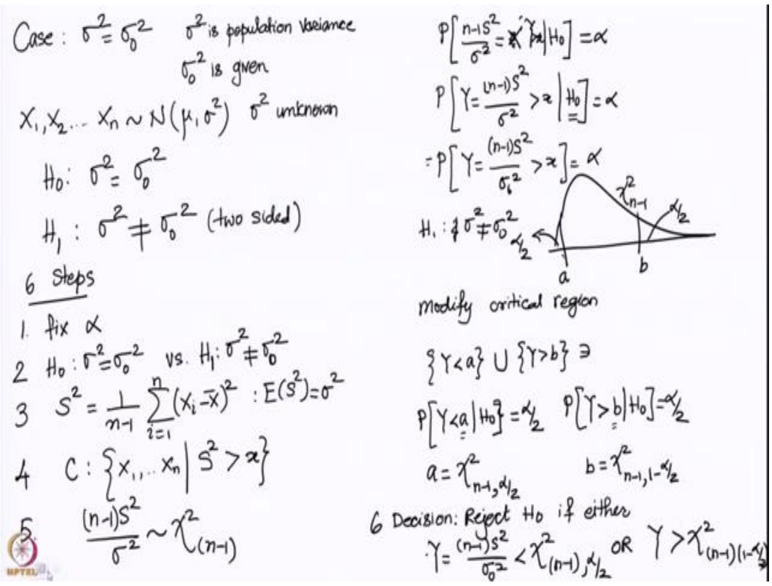
And therefore, now you can start that your H0, which says that Mu is equal to Mu 0 under H1, that Mu is not equal to Mu 0 your critical region is going to be all X1, X2, X3, XN such that the absolute value of T is greater than some t. That is, if you take a T distribution, which is a symmetric distribution, and we say that if T lies in any of this area, it is our rejection region within this area we accept it.

So, when with a fixed alpha from the beginning, we would like to have probability of critical region given H0, which is probability of absolute T greater than small t, which has to be alpha. And therefore, just as we did in the previous case, capital T greater than 1 minus alpha by 2 has to be alpha by 2 and this becomes your criteria for decision which you will say that H0 to be rejected, sorry rejected if T is greater than small t1 minus alpha by 2, sorry here I should try it n minus 1 and here also I should write n minus 1 because that specifies is degrees of freedom.

So, on the same line, quickly we can say that if we consider the testing of hypothesis H1, I mean the sorry, the alternate of H2, which says that Mu is greater than Mu 0. Then do I need to give much explanation please recall, our decision is reject H0 if the same T is greater than small T, 1 minus alpha with N minus 1 degree of freedom. And in the case of the third hypothesis, Mu less than Mu 0, third alternate hypothesis, the decision making is going to be reject H0, if T is less than t alpha N minus 1, n minus 1 is degrees of freedom.

So, this is what we have learned so far in the hypothesis testing. The important two statistics we have found so far is Z which is X bar minus Mu 0 or sigma square root n distributed as a normal 0,1. And in the other case of our test statistic is T which is X bar minus Mu 0 over sample standard deviation square root n, which is distributed as a students three distribution with n minus 1 degrees of freedom.

(Refer Slide Time: 14:44)



In the todays case, what we would like to do is consider a case when you want to test that sigma square itself is equal to some pre decided value sigma 0.

$$Case\ \sigma^2 = {\sigma_0}^2$$

So, today what we want to consider is a case of sigma square is equal to sigma naught square, where sigma square is population variance.And sigma naught square is given, when this is the case, when this is the case we start out that we have a sample X1, X2, X3, XN from again a normal population with mean Mu and sigma square and we want to sigma square is unknown. And we want to test the hypothesis that

$$H_0 : \sigma^2 = {\sigma_0}^2$$

$$H_A : \sigma^2 \neq {\sigma_0}^2 \text{ (Two sided)}$$

sigma square is a given value sigma naught square; the situation can arise just as it would arise in a any testing of hypothesis for Mu is equal to Mu 0. Again, we consider the alternate hypothesis H1, first as sigma square is not equal to sigma naught square, it is a two-sided alternative.

Let us follow the six steps, let us follow the six steps of hypothesis testing.

1. Let α be fixed
2. $H_0 : \sigma^2 = {\sigma_0}^2 \; vs \; H_A : \sigma^2 \neq {\sigma_0}^2$
3. It is shown that $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \; : E(S^2) = \sigma^2$
4. Critical region $C = \{X_1, X_2, \ldots, X_n \,|S^2 > x\}$
5. $\frac{(n-1)*S^2}{\sigma^2} \sim {\chi^2}_{n-1}$
6. $P\left[Y = \frac{(n-1)*S^2}{\sigma^2} > \text{x}\right] = \alpha$

Now, because the alternatives on both sides, actually this is insufficient, we have to say that for alternate which is both sides that which says that sigma square is not equal to sigma naught square in a Chi square distribution, it should lie somewhere between two limits A and B, where this is also alpha by 2 and this is alpha by 2.

So, considering that we modify our critical region as,

$$\{Y < a\} \cup \{Y > b\}$$

$$P[Y < a|H_0] = \alpha/2 \text{ and } P[Y > b|H_0] = \alpha/2$$

$$a = {\chi^2}_{n-1,\alpha/2} \; ; b = {\chi^2}_{n-1,\alpha/2}$$

$$Decison\ of : Reject\ the\ H_0\ null\ hypothesis\ if$$

$$Y = \frac{(n-1)*S^2}{\sigma^2} < {\chi^2}_{n-1,\alpha/2} \;\; or \;\; \frac{(n-1)*S^2}{\sigma^2} > {\chi^2}_{n-1,\alpha/2}$$

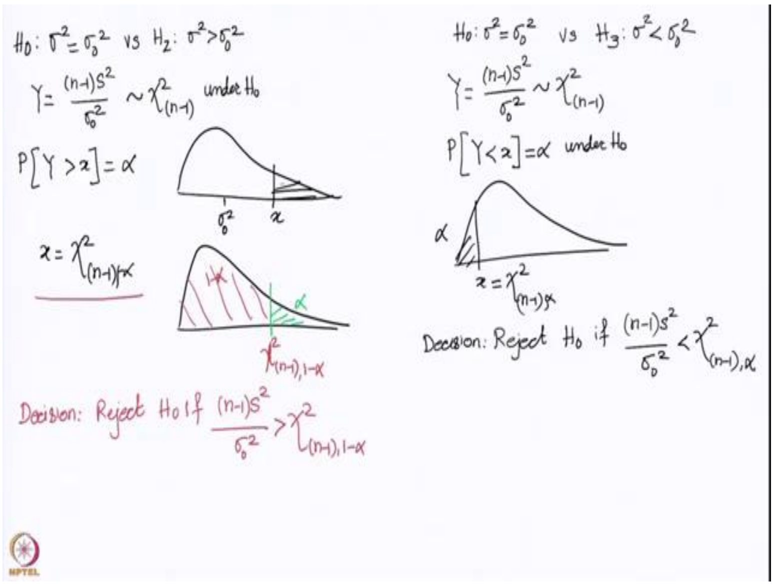
So, these are your critical value and your decision is going to be reject H0, if either P sorry, either Y which is n minus 1 S square over sigma naught square is less than Chi square n minus 1 alpha by 2 or Y is greater than Chi square n minus 1, 1 minus alpha by 2. I hope this is clear that we are what we find let us quickly go through it.

If we follow the six steps, this is the sixth step. So, if we follow the six steps, we fix the first alpha value. So first we fixed an alpha value. This is our null hypothesis and the alternate hypothesis, we find that S square is the correct statistic to take to find a test statistic for this or to find a critical region for this test.

And therefore, this test statistic we chose as a sample variance because expected value of sample variance is the population variance, then we defined a critical region X1, X2, X3, XN such that S square should be greater than X. And then we take appropriate statistic which would be devoid of any unknown parameter.

So, we say that it is n minus 1, I better write a parenthesis here. So, it is n minus 1 S square over sigma square. Actually, here itself I should have used this kind of an arrangement, but we have actually continued with the same arrangement, but we find that when you simplify this and you want to find the critical value, we realized that this is a two-sided hypothesis. And therefore, we must have the two values, we must have the two values A and B. And therefore, we find that the critical region can be modified to this and finally, we come up with the test.

(Refer Slide Time: 26:12)



Now, I believe it is not very difficult to understand as to what would be the case if

$$H_0 : \sigma^2 = {\sigma_0}^2$$

$$H_2 : \sigma^2 > {\sigma_0}^2$$

Then naturally you will say that your test statistic is Y which is $Y = \frac{(n-1)*S^2}{\sigma^2} = {\chi^2}_{n-1}$under H_0.And therefore, a probability of

$$P[Y > x|H_0] = \alpha$$

Now we are correct because if you look at the region, we want the sigma square to be greater than sigma naught square, this is sigma naught square, then you would like to have it greater than some value, which is critical value, which will be a critical value which I am calling X.

But now, you can make out that the indirect case x is

$$x = {\chi^2}_{n-1,\alpha}$$

Please remember again, if you take the alpha Chi square distribution, and if you wish to have this value such that this is alpha in that case this area is 1 minus alpha.

And therefore, this value is I call it, sorry Chi square n minus 1, 1 minus alpha. So, that is the value which will come. So, the decision is going to be reject H0, if n minus 1 S square over sigma naught square is greater than Chi square with, value of Chi square with n minus 1 degrees of freedom at 1 minus alpha probability.

Similarly, you can easily show that if you want to test the hypothesis

$$H_0 : \sigma^2 = {\sigma_0}^2$$

$$H_2 : \sigma^2 < {\sigma_0}^2$$

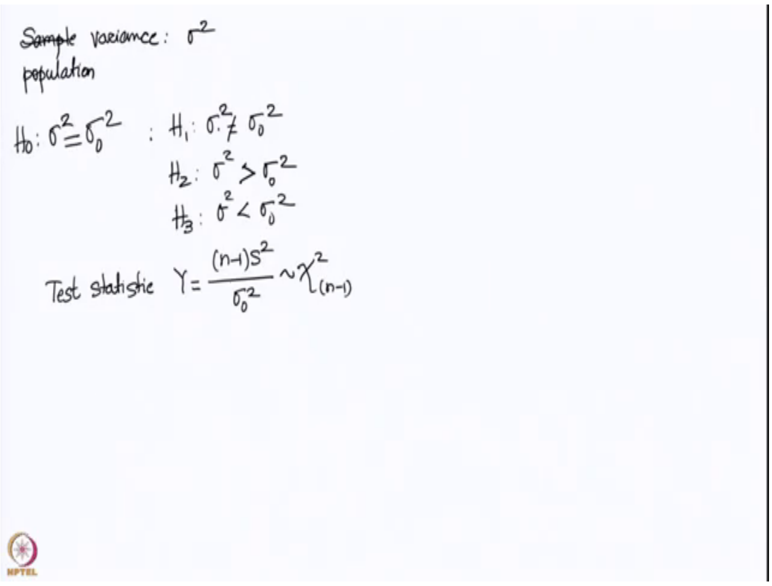
Then your test statistic is same, which is the test statistic is going to be $Y = \frac{(n-1)*S^2}{\sigma^2} = {\chi^2}_{n-1}$

And we are looking for a probability that

$$P[Y < x | H_0] = \alpha$$

So, if you look at the probability curve, you are looking at this where you have X, so that this is alpha and this is given by Chi square n minus 1 alpha. And therefore, your decision is going to be reject H0, if n minus 1 square over sigma naught square is less than Chi square with n minus 1 degree of freedom and alpha probability.

(Refer Slide Time: 31:13)



So, with this we conclude this session on testing of hypothesis, for testing of hypothesis for sample variance. So, sample variance, sorry not sample variance, it should be population variance sigma square, the hypothesis we tested is sigma square is a given value sigma naught square. And we considered three alternative one is a two-sided alternative, it says that sigma square is not equal to sigma naught square.

Second one say that sigma square is greater than sigma naught square. And the third hypothesis say that sigma square is less than sigma naught square. In all the cases we found that test statistic Y is nothing but n minus 1 sample variance divided by sigma naught square, which is distributed as Chi square with n minus 1 degree and then we derive the three tests. Thank you.