

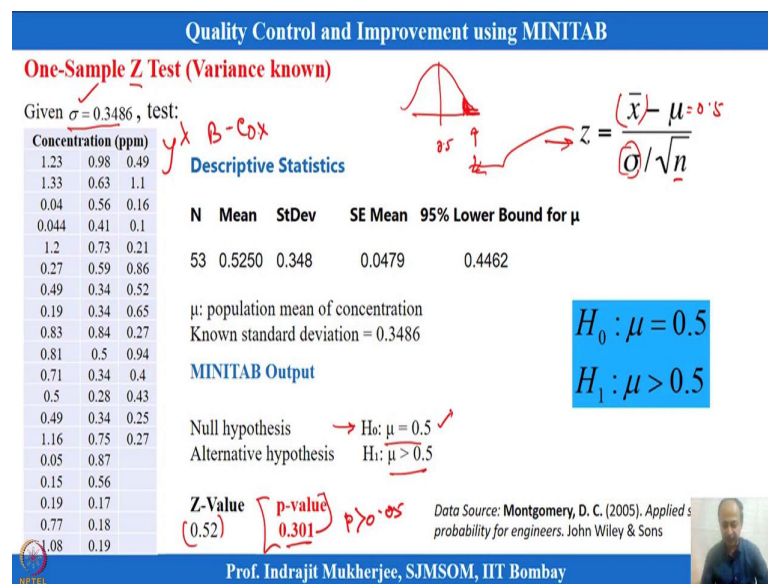
**Quality Control and Improvement with MINITAB**  
**Prof. Indrajit Mukherjee**  
**Shailesh J. Mehta School of Management**  
**Indian Institute of Technology, Bombay**

**Lecture - 18**  
**One sample t-Test**

Hello everyone and welcome to session 18 on Quality Control and Improvement with MINITAB. And I am Professor Indrajit Mukherjee from Shailesh J Mehta, School of Management, IIT Bombay.

So, previous session what we are doing is that we are talking about hypothesis testing and we are talking about One sample Z test that we are doing. And in that case, what we have mentioned is that one example we have taken and we will just try to highlight and then we will go ahead with that, ok.

(Refer Slide Time: 00:46)



So, this was the concentration in ppm which was monitored. And the hypothesis and  $\sigma$  was known over here. So, in this case  $\sigma$  was known. So, whenever population  $\sigma$  is known in that case that is why we are using Z test over here. And what we have seen is that we want to test the hypothesis that whether is  $\mu$  equals to 0.5 ,  $\mu$  greater than 0.5 over here.

And the Z statistic that was used was  $\bar{x}$  which is derived from the observation over here, and  $\sigma$  is given over here and n is the total number of observations that is there. And this  $\mu$

will be equals to 0.5 that we have assumed and so, I can calculate the Z statistics over here.

So, in this case; and then what we see is that whether the Z-value that is; so, this is 0.5 region over here whatever will be the Z-value. So, this is the rejection zone that we have defined based on one-sided test over here. And if my values is falling somewhere over here Z-value is falling that is Z statistics is falling somewhere over here in the rejection zone in that case we will go for the alternate hypothesis. We will go for the alternate hypothesis.

So, in this case what we have seen is that data because data is non-normal and we have transformed the data. So, we initially we have assumed that data is normal. Then also we have seen that we cannot reject the null hypothesis because my Z-value is around 0.52, and the corresponding p-value over here is 0.301. So, this values indicate that p is greater than 0.05.

So, I have to I cannot reject the null over here. So, now we go by the null hypothesis over here. So, we cannot reject or we do not have enough evidence to prove that  $\mu_0$  equals to  $\mu$  is greater than 0.5 like that, ok. And also, we have seen that in case this normality assumption is not true. We can convert the data by using lambda transformation over here y to the power lambda that box-cox transformation what we have used.

And what we have seen is that after conversion, then we converted  $\sigma$  and then what we did is that after conversion we have applied the same logic of Z test over here and we have seen that, that also satisfies the condition; that means, we cannot reject the null over here, ok. And so, that was the example that we have taken. So, that was the example that we have taken we can take some more examples over here and try to see that what happens in other scenarios like that.

(Refer Slide Time: 03:14)



**Quality Control and Improvement using MINITAB**

**P-value :**

P-value : **'Smallest' level of significance** that would lead to rejection of Null Hypothesis. (or **'observed' level of significance**)

- Compare the p-value with the selected value of  $\alpha$
- If the **p-value is less than or equal** to the **level of significance**, we judge the p-value to be small enough to reject the null hypothesis.
- If the **p-value is greater than level of significance**, we **do not reject the null hypothesis**.

$\alpha = 0.025$   
 $z_c = 1.96$   
 $z_c > z_{table}$   
 $p \leq 0.05$   $H_1$   
 $p > 0.05$   $H_0$

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

So, before we go into that we need some brief idea about that what is this p-value that we are talking about which is used for hypothesis testing and any other analysis on statistical analysis when we talk about level of significance like that; what is this p-value? So, normality test we have used this one p-value concept like that. So, what is this p-value?

So, when I am doing this test on normality test, when I am doing this hypothesis testing what is assumed is that the  $\alpha$  level of significance is around 0.025 let us say, 0.025. And on this side; if I am doing both-sided test what I am assuming is that on both sides it is 0.025 over here. And I am observing that this is the new condition and this is the  $\bar{x}$  condition where it falls? If it is falling within this region, I will reject the null hypothesis basically, ok.

So, MINITAB gives you an option to calculate the corresponding p-value. So, I am using a Z statistic over here and based on the Z statistic value that is calculated value. And then I compare this calculated value, X can be converted into Z statistics calculated value over here and this Z. So, this will be converted into Z on this axis and this will be  $z_c$ .

So, this is the 0 point over here, this will be 0 condition over; this will be converted into 0. The mean will be equals to 0 when I convert into Z. So, then what I do is that? Z statistics I am trying to derive that Z statistics over here. And based on this Z-value and with a level of significance of 0.025 on this side, so if my Z-value gives me this gives me a corresponding value.

So, here the Z statistics over here will be compared with a tabulated value, earlier days what you used to do is that this is calculated and tabulated value depends on this  $\alpha$  level of significance what I have defined over here. So, calculated value will be compared with tabulated value like that, from table Z table we will get.

And if this is calculated value is greater than then we will tabulated value, then we will reject the null hypothesis that was the earlier concept. So, I can do hypothesis testing in two ways. So, one is that using tables and we compare that values and go by that, otherwise we go by the p-value concept over here; what is used by MINITAB, and we and many people goes by p-value concept like that.

So, I do not predefine this  $\alpha$  level of significance, but I have in mind that if  $\alpha$  value if I get if this is the region of region of acceptance zone over here, so then what due to computational nowadays availability of computational software's like that. What can be done is that whenever I am doing this test or Z-values that I am calculating, so this will be 0 and this will be calculated value over here.

For a given calculated value that we have calculated. So, we can also calculate that what is the probability or the area which is more than Z. So, probability of Z greater than Z c like that. So, this area can be defined; this can be calculated over here and MINITAB does it automatically for you.

So, it gives you a smallest level of significance at which the null hypothesis can be that leads to rejection of the null hypothesis. That means earlier what we are assuming this is strict region that we are considering over here and this is the acceptance region we are considering over here.

Now, what we are what we will do in a reverse way we are calculating the  $\alpha$  values like that. This is observed level of significance, the  $\bar{x}$  condition that the  $\bar{x}$  values from the sample statistics give me a location of Zc over here and this Zc corresponds to an  $\alpha$  value which is different from the  $\alpha$  that we have assumed.

So, it is a calculated  $\alpha$  value for given value of  $\bar{x}$  like that. So, it is a representation that if this is the  $\bar{x}$  then what is the  $\alpha$  level, at what level of  $\alpha$  I can reject the null hypothesis? That is the smallest level of significance that is known as the P-value over here.

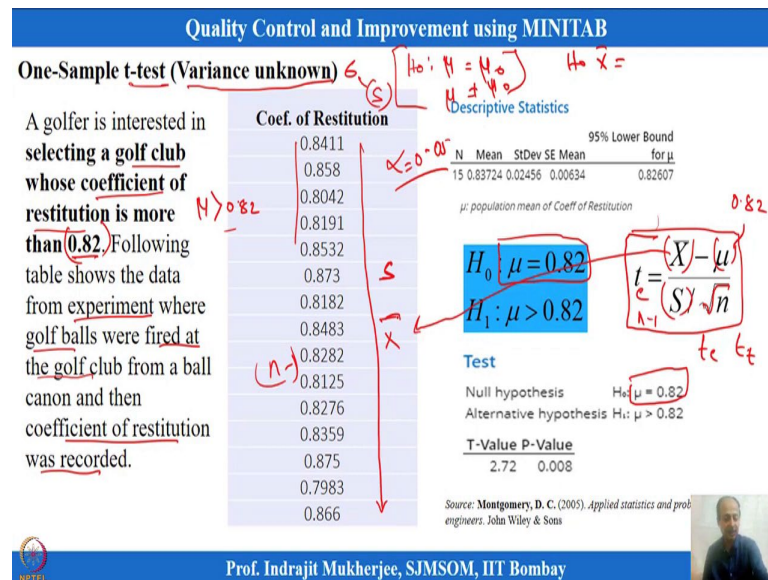
So, if P-value is less than 0.05 we need to reject the null hypothesis, so we need to reject the null hypothesis. So, this is the condition that we will apply or less than equals to condition we can apply. And P is more than 0.05 what we will do is that we will accept the null, so we will go by the null. And if there is an equals to condition we will go by  $H_0$  or alternate hypothesis like that. So, this is the condition that is used.

So, earlier days we used to see the tables and we compare that one because P-value calculation is quite complicated for various distributions like that. Nowadays, for computational we have so many software's like that immediately software will give you the p-value, corresponding p-value for a specific scenarios of either I am doing one sample Z test or any other test hypothesis testing.

So, p-value which is the actual level of significance at what point I can reject the null hypothesis that will be given. So, in when you have to interpret what you have to; what you have to see is that if P is less than equals to 0.05 or not if there is a condition then a null has to go in that case we will go by the alternate hypothesis. We have to go by the alternative hypothesis over here, ok.

So, that is the interpretation of p-value. So, P is less than equals to 0.05, we will reject the null hypothesis, we will reject the null hypothesis. If it is more than that the reverse condition, if it is greater than that we will assume that null has to be accepted. So, in that case, I cannot reject the null and we do not have evidence to reject the null like that. So, this is the condition. And let us take one more example.

(Refer Slide Time: 08:38)



So, in this case, one more example what I am seeing is variance is unknown that is given over here. A golfer is interested in selecting a golf club whose coefficient of restitution is more than this one. So, my condition over here is the  $\mu$  value should be greater than then only I will accept the golf club like that. So, golf club means which is used to hit the ball basically.

So, there is a design of the golf club. So, I have to select a golf club with coefficient of restitution is more than this one. Following table shows the data for experiments where golf. So, experimentation comes over here where golf balls were fired at the golf club from a ball canon and then coefficient of restitution was recorded, ok.

So, these are the values that we are getting over here. So, these are the sample observations. So, from here what we can do is that we can calculate average; we can calculate standard deviation of the data like that, ok. And population variance is not known. So, when we are doing hypothesis testing remember that we are doing hypothesis testing on the population.

So, whether  $\mu$  equals to  $\mu_0$  or  $\mu_0$  equals to  $\mu$ , we are not testing whether this  $\bar{x}$  is certain values like that. So, this is not the way we are testing over here. So, this is all about hypothesis testing, all about population not about sample over here. Sample observation will lead to reject or acceptance of the null hypothesis like that, ok.

So, here what we can get is that we can get  $\bar{x}$  over here and the condition is that we will consider select that golf club which is having a coefficient of restitution of more. So, one of the; one of the club was taken and in that case it was observation was given over here, so this will have a mean and standard deviation.

So, in this case, I can when variance is unknown for the population of  $\sigma$  is unknown. So, I can make an estimation with  $S$  over here and I can replace the estimation of  $\sigma$  by  $S$  which is an unbiased estimator of  $\sigma$  and in that case what will happen is that I have a test statistic which is known as  $t$  test over here.

And the test statistics that you see over here is this one where  $t$  is equals to; so calculated value will be based on what is the  $\bar{x}$  value which will come from here. And then  $\mu$  over here is taken as 0.82 over here which is the condition that is given it should be more than this. So, null will be equals to condition over here,  $\mu$  equals to 0.82 that is also considered over here you can see, ok.

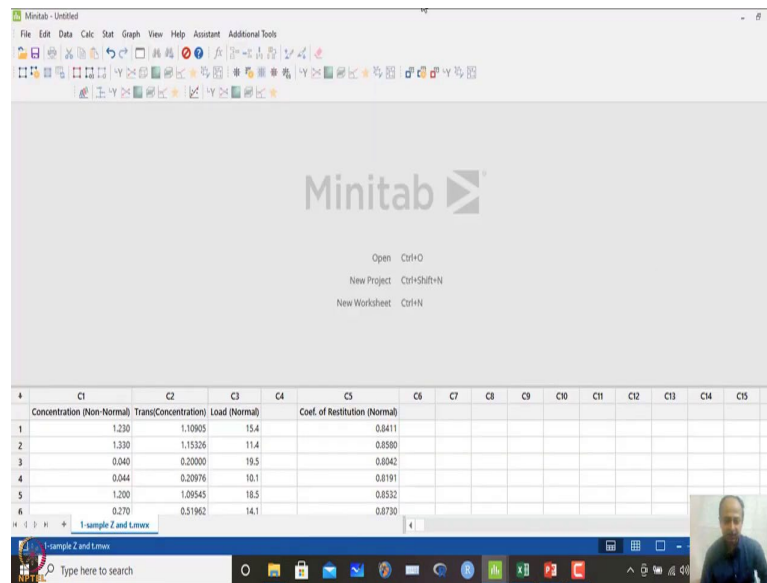
And standard deviation calculation will be given. So, instead of  $\sigma$ , I am taking  $S$  over here and the square root of  $n$  that is the condition in also in sampling distribution we have seen that this is the this is; so, variance follows the  $\sigma$  by root  $n$  condition. So, here also we are taking instead of  $\sigma$  we are assuming  $S$  by root  $n$  over here.

And that will follow it is  $t$  distribution and this  $t$  distribution. So, some calculated value of  $t$  will be calculated and that will be compared with a tabulated value like that. So, that will be compared. So, whenever I am talking of  $t$  distribution in that case degree of freedom is important, so  $n$  minus 1.

So, here I have  $n$  observation  $n$  minus 1 is the degree of freedom that will be considered for the analysis. And  $\alpha$  level of significance let us say we have we are assuming a 5 percent level of significance. So,  $\alpha$  equals to 0.05 at that level we want to test this one.

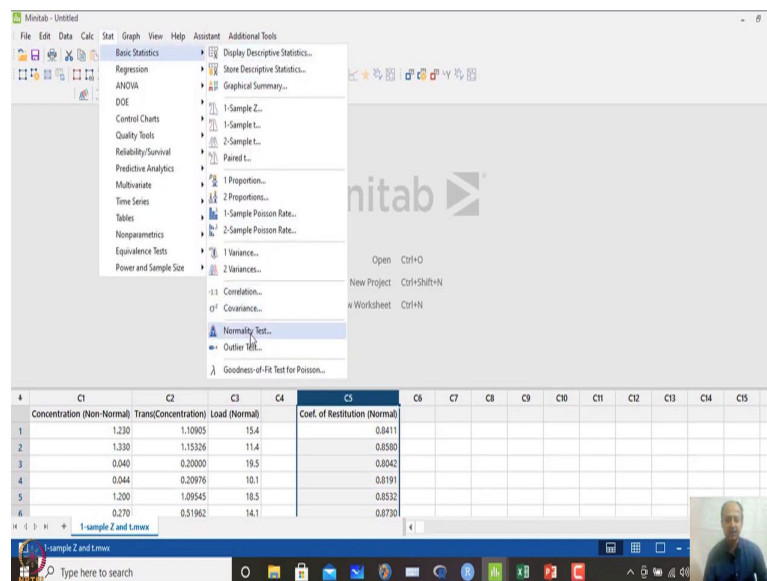
So, whenever I am using a  $t$  test the underlying assumption is that data follows normal distributions like that. So, data follows normal distribution that has to be checked and then only we can go for  $t$  test like that when variance is up. So, that is the condition, we will see that one.

(Refer Slide Time: 12:02)



So, we will go to the Minitab file and try to see how this analysis is done.

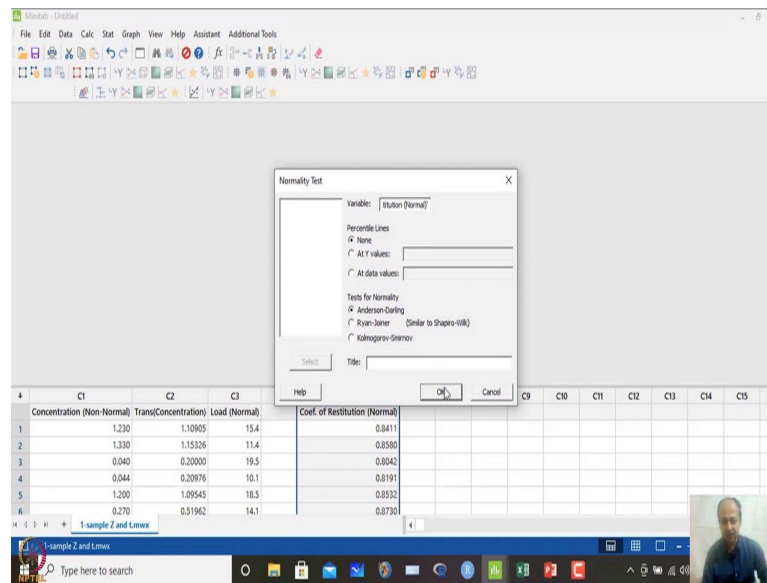
(Refer Slide Time: 12:08)



So, coefficient of restitution, so this is the data set that I am having. So, I will go to basic statistics and I will try to confirm whether it follows normal distribution or not.

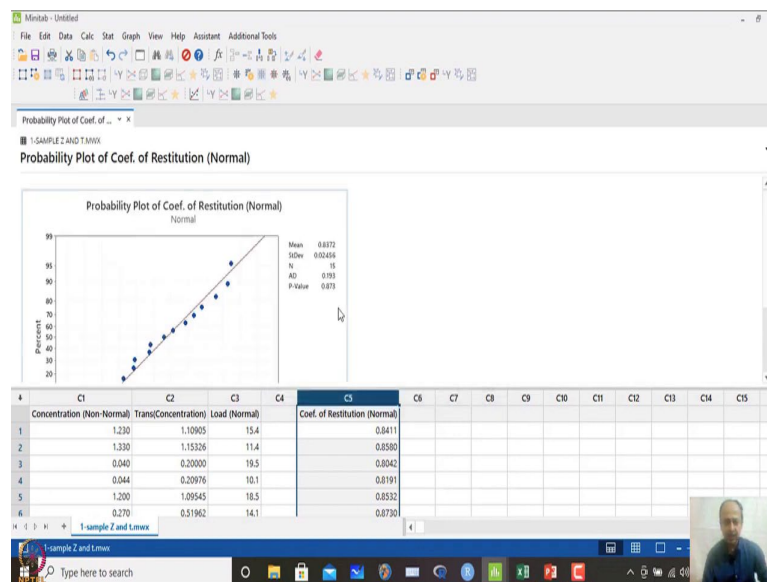


(Refer Slide Time: 12:11)

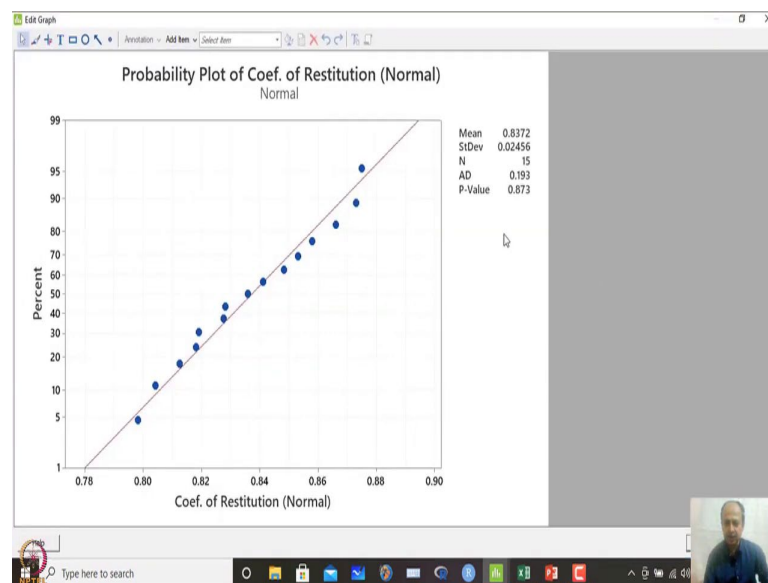


So, in this case, coefficient of restitution I am doing. So, in this case I am doing Anderson-Darling test again. So, I will click ok.

(Refer Slide Time: 12:21)

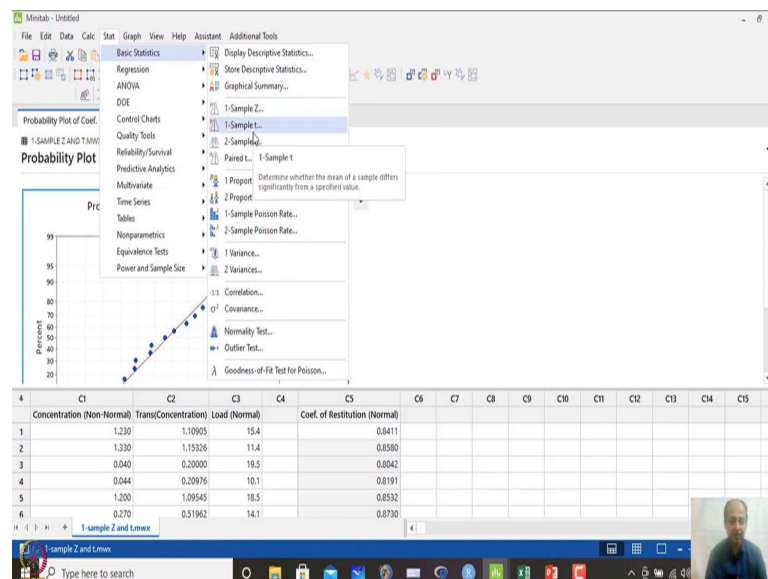


(Refer Slide Time: 12:23)



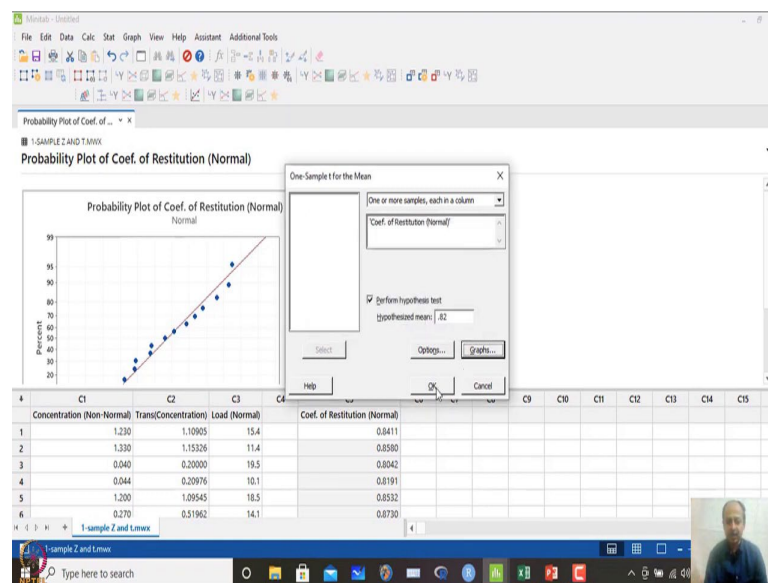
And when I do the analysis of normal distribution test what I found is that? P-value is around 0.873. So, by P-value interpretation what we know is that P is more than 0.05; in that case null hypothesis over here is the data is normal. So, in this case, we cannot reject the null. So, in this case data is normal. So, that is proved over here.

(Refer Slide Time: 12:45)



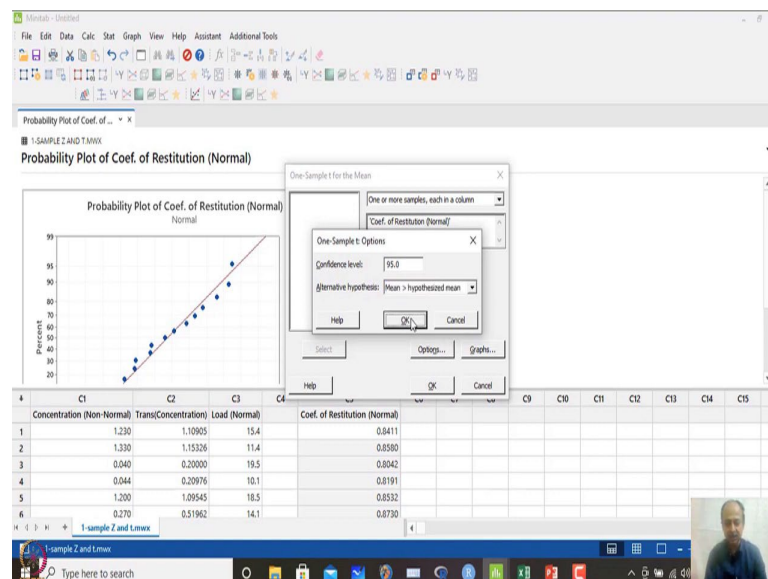
So, when the data is normal in that case we can apply the t-test, one-sample t-test. So, what you have to do is that go to stat, basic stat one sample t. So, one sample Z we have done, when  $\sigma$  is known one sample t we are doing over here.

(Refer Slide Time: 12:53)



So, in this case again, we will just reload the different set of data which is coefficient of restitution. So, in this case perform hypothesis and when we are doing that and the condition 0.82 which was given like that we want to test like that and that is the condition. So, this is one-sided test,  $\mu$  greater than 0.82.

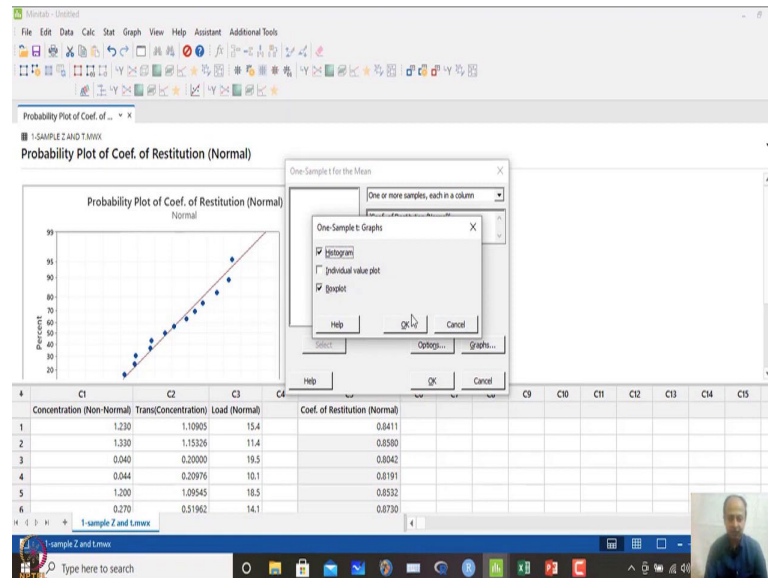
(Refer Slide Time: 13:17)



So, what we will do is that we will just mention that one in the options and we go by the options, and then 95 percent is there. So, I will write mean greater than hypothesized

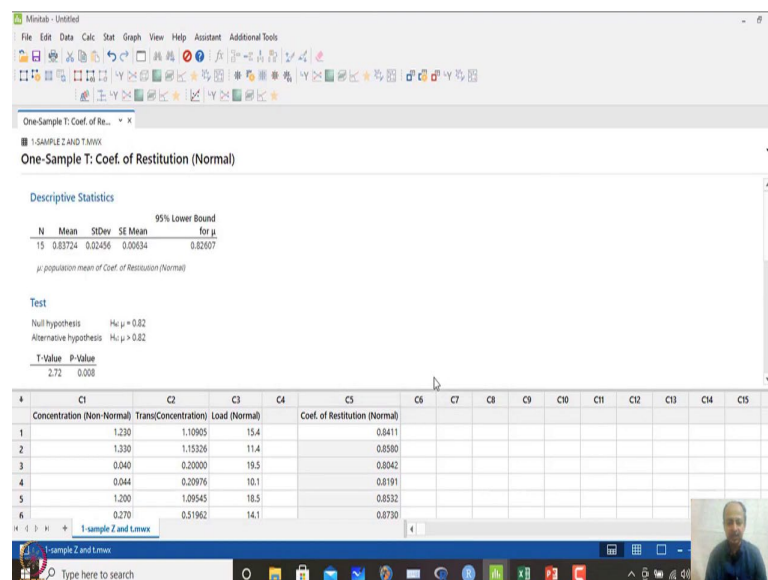
mean. So, if that is alternate hypothesis that I am considering over here greater than condition.

(Refer Slide Time: 13:31)



So, then in graph what we can do is that we can see histograms of the data set, also box plot of the data set like that what we can see.

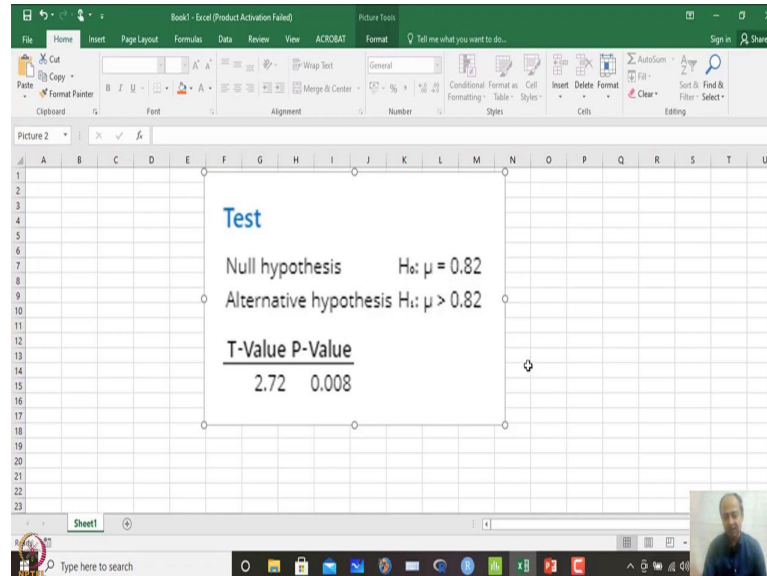
(Refer Slide Time: 13:39)



So, if I click ok and I click ok of the data set what will happen is that. So, this is the data summary descriptive statistics what you see mean and standard deviation is calculated and confidence interval is also given. So, if you go down what will happen is that this is

the test statistics that you can see. So, I can paste it over here to have an enhanced view, ok.

(Refer Slide Time: 13:58)



So, this is showing me a value of 2.72 and P-value is 0.008, P-value is 0.008. So, when P is less than 0.05, we will go for the alternate hypothesis, we will go for the alternate hypothesis; that means, the average value of the population will be greater than 0.82 so, that is confirmed over here using this test.

So, with one sample we are doing this and we are finding out that the P-value is less than 0.05. So, null has to go over here. So, null has to go over here and the P-value indicates that we will go for the alternate hypothesis over here. And if this is true, so in that case anything which is more than that, so we will select this.

So, we will select this golf club because it is giving me a average which is on population we expect that it should be greater than 0.82. A level of significance over here P-value is considered as 0.05 and in certain situation P-value can be considered as 0.01; sorry  $\alpha$  value can be selected as 0.01 like that.

So, 90 percent confidence band we are considering acceptance zone we are considering and. So, in that case; so, scenarios it depends on the experimenter what value of  $\alpha$  we will select. So,  $\alpha$  is generally taken as 10 percent or 5 percent like that, so that is the general condition, ok.

And if you want to be very sure, so 99 percent also acceptance band we can take. So, we do not want to take any chance like that. So, if it is critical, decision is very critical in that case may be the 99 percent band we will consider and in that case only one percent possibility that my conclusion will go wrong.

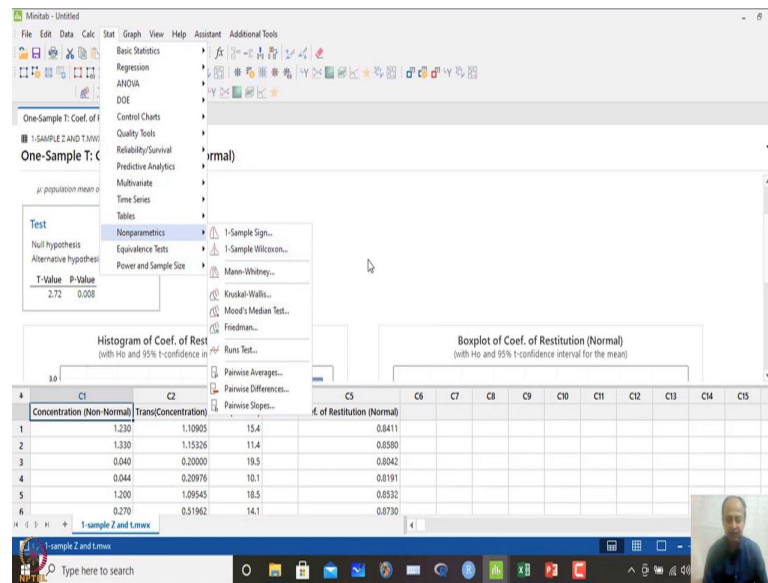
So, in that case  $\alpha$  value will be 0.01; 0.001 basically. So, that will be the condition that we can think of sorry sorry 0.01, sorry 0.01 will be the condition that we can adopt like that. So, it depends on the scenarios what  $\alpha$  level we will select like that, ok. So, this data set shows that normality assumption is there and the t value, P-value is, P-value is less than. So, significance level is reached. So, we can go for that, ok.

Now, in this case normality we have transformed the data. So, one data set that was given that concentration data set, and in that case what we have seen is that it is non-normal, so conversion has worked. So, in this case scenario is favorable for us, so box-cox transformation has worked.

So, after transformation we have cross checked and we have seen that whatever conclusion that we have drawn assuming normal that is also if I can transform the data and our conclusion is also correct. So, second time when we converted the data also our analysis shows that we cannot reject the null, so it is approximately. So, p-value is again less than that one.

So, in case conversion is not happening it is a skew distribution or something conversion is not working, so in that case what we can do is that? There is another type of test which is known as non-parametric test, which uses the median as a reference and uses rank concept like that.

(Refer Slide Time: 16:51)



So, when you go to stat you go to this non-parametric options that you have, so in this case you will find two options, one sample Wilcoxon and one for sample sign test like that. So, what is to be; which scenario will be preferred, which one to be preferred like that, so that we can highlight over here.

(Refer Slide Time: 17:12)

**Quality Control and Improvement using MINITAB**

**NONPARAMETRIC TEST**

→ **One-Sample Wilcoxon Test**

(Assumes Symmetric Distribution, e.g. uniform)

→ **1-sample Sign Test**

(Does not assume Population distribution Symmetry)

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

So, in case data is non-normal, so I want to test, so distribution assumption is I want to avoid that one. So, in this case, there are options like that non-parametric test. So, non-parametric test does not assure a strong decision over here, but what we can say is that

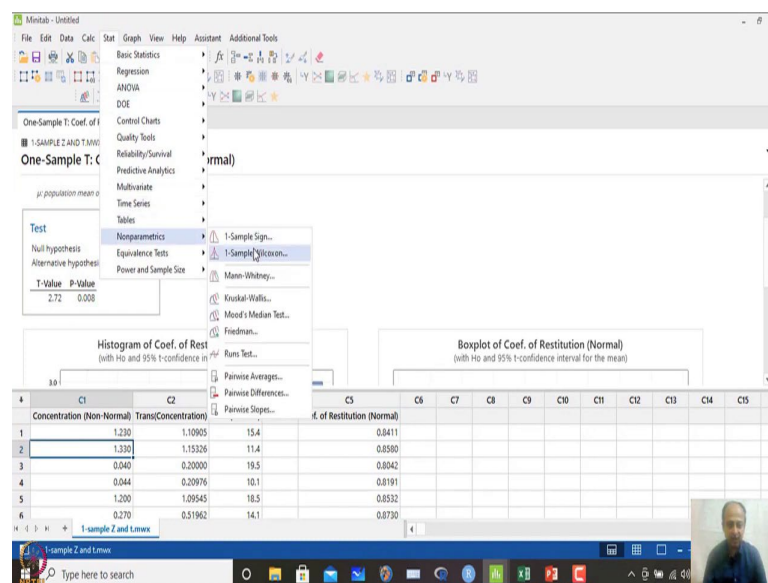


this is the alternative that we have in case everything fails. So, we have this alternative, but we cannot be 100 percent sure that this test works 100 percent of the time. So, it will give you correct decisions like that.

So, in this case, One-sample Wilcoxon test some guideline over here Assume Symmetric Distribution like normal distribution what we have. When asymmetry exist, so in that case 1-sample Sign Test can be used seeing the distribution, based on that I can use either Wilcoxon test. And if it is does not assume population, this does not assume symmetry of the distribution, so this is this we can use, this also we can use, so one-sample or otherwise we can use one sample Wilcoxon test.

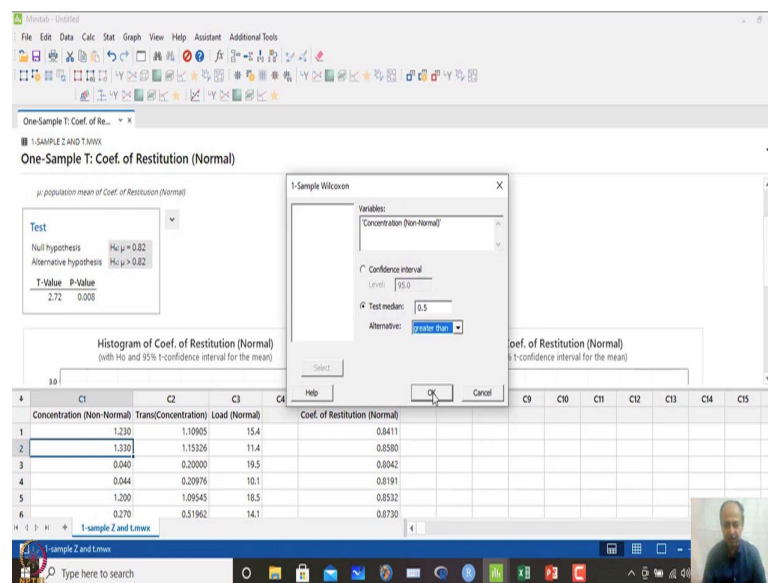
So, how do we see that one? So, when we are using this one, we can go back to the data set over here which is non-normal. And in this case, what we can do is that go to stat and go to non-parametric, go to one sample Wilcoxon assuming the distribution more or less symmetry over here.

(Refer Slide Time: 18:15)



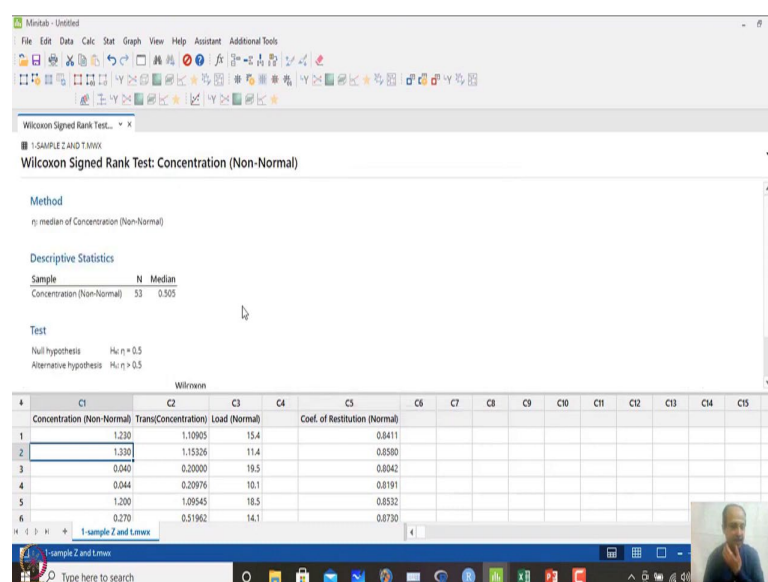


(Refer Slide Time: 18:17)



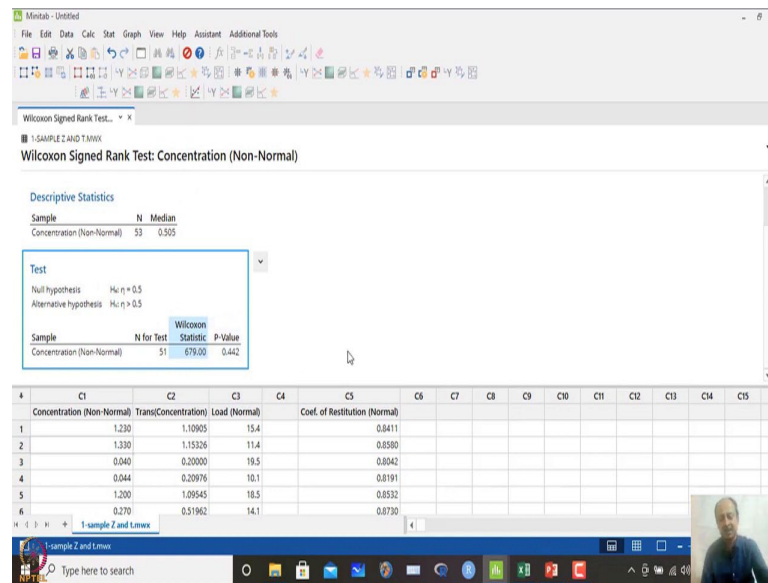
So, in this case, what I will do is that I will go for this concentration nonnormal data. And here we are testing the medians basically. So, in this case again, so we do median testing over here, so not mean. So, you have to remember this one. So, median is assuming to be 0.5 over here we are doing this test, and greater than condition remains same. So, this condition we are testing.

(Refer Slide Time: 18:42)



And in this case what will happen is that you will get some observation median values over here.

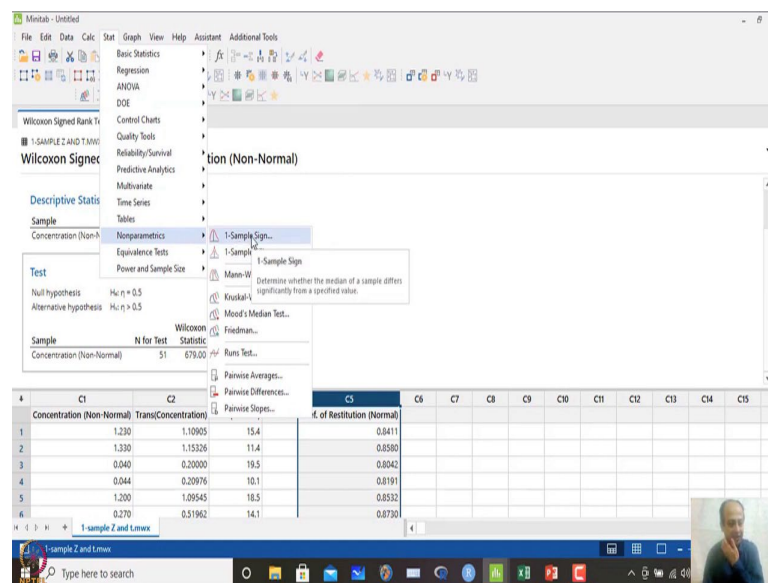
(Refer Slide Time: 18:46)



And the testing is done on the median value which is what you see over here median is equals to 0.5, median not equals to 0.5. So this Wilcoxon statistics is given 679, this is the like t statistic, Z statistics this is one of the statistic that is calculated. And corresponding P-value is given over here what do you see 0.442 like that, ok. So, 0.442 indicates that P-value is more than 0.05, so I cannot reject the null over here.

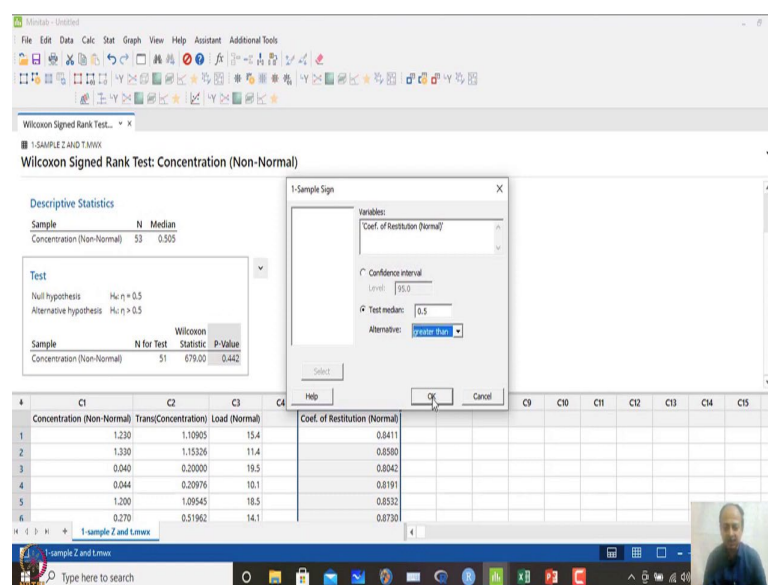
Similar conclusion was done when we have converted the data and without conversion also the conclusion was same, so but everywhere it may not happen. So, conclusion based on classical approach, may not be same as non-conventional approach which is over here or non-parametric approach which does not assume distribution like that, ok.

(Refer Slide Time: 19:39)



So, this is true for this, and if I go for this coefficient of restitution over here again assuming that the distribution we are not going by any distribution assumptions. We can also do 1 sample sign test over here.

(Refer Slide Time: 19:47)



So, this 1-sample sign test also you mentioned that this is the coefficient of restitution I want to do. Confidence interval is this, and test median is around 0.5 we want to test this one, sorry 0.5 over here, and I want to test whether it is greater than condition is satisfied or not. I click ok.

(Refer Slide Time: 20:02)

Sign Test for Median: Coef. of Restitution (Normal)

Method  
n: median of Coef. of Restitution (Normal)

Descriptive Statistics

Sample	N	Median
Coef. of Restitution (Normal)	15	0.8359

Test

Null hypothesis  $H_0: n = 0.5$   
Alternative hypothesis  $H_a: n > 0.5$

Sample	Number < 0.5	Number = 0.5	Number > 0.5	P-Value
Coef. of Restitution (Normal)	0	0	15	0.000

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
Concentration (Non-Normal)	1.230	1.10905	15.4		0.8411										
Trans(Concentration)	1.330	1.15326	11.4		0.8580										
Load (Normal)	0.040	0.20000	19.5		0.8042										
	0.044	0.20976	10.1		0.8191										
	1.200	1.09545	18.5		0.8532										
	0.270	0.51962	14.1		0.8730										

(Refer Slide Time: 20:05)

Sign Test for Median: Coef. of Restitution (Normal)

Method  
n: median of Coef. of Restitution (Normal)

Descriptive Statistics

Sample	N	Median
Coef. of Restitution (Normal)	15	0.8359

Test

Null hypothesis  $H_0: n = 0.5$   
Alternative hypothesis  $H_a: n > 0.5$

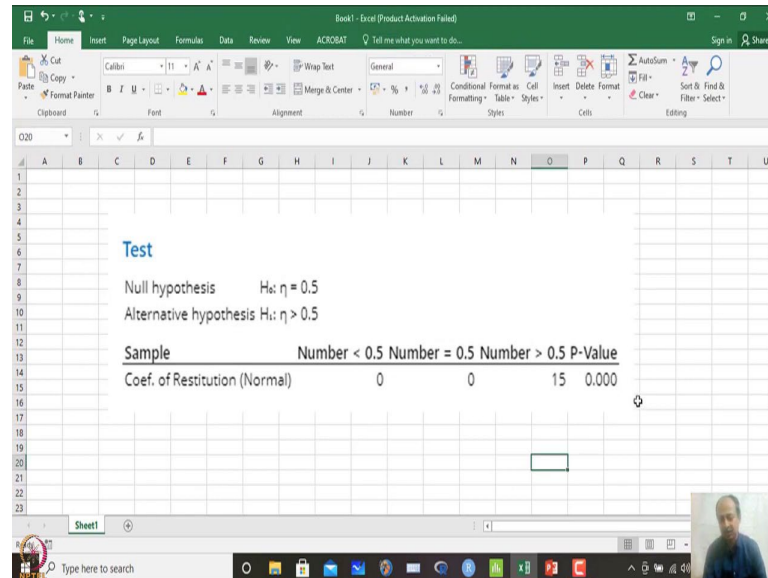
Sample	Number < 0.5	Number = 0.5	Number > 0.5	P-Value
Coef. of Restitution (Normal)	0	0	15	0.000

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
Concentration (Non-Normal)	1.230	1.10905	15.4		0.8411										
Trans(Concentration)	1.330	1.15326	11.4		0.8580										
Load (Normal)	0.040	0.20000	19.5		0.8042										
	0.044	0.20976	10.1		0.8191										
	1.200	1.09545	18.5		0.8532										
	0.270	0.51962	14.1		0.8730										

And then in that case sign test for median is given and also the conclusion over here, you go to the last P-values that you see over here, the P-value indicates significant significance over here. So, that we have to reject the null that is 0.5 basically, ok. So, all observations are median, whatever observation we have, all 15 observations are beyond this 0.5.

So, based on that some P-value calculation can be done and P-value is around 0. So, MINITAB will indicate, so whenever you see the output of MINITAB, so I am copying this one and I am pasting this one. So, I can delete this and let us paste this one.

(Refer Slide Time: 20:40)



Sample	Number < 0.5	Number = 0.5	Number > 0.5	P-Value
Coef. of Restitution (Normal)	0	0	15	0.000

So, observations that you see I am enhancing this one. So, what you will see is that P-value is approximately equals to 0000, up to third place of decimal , I will report the P-values like that.

It is not that P-value is exactly equals to 0 over here it indicates that up to three place of decimal this is shown and beyond three place of decimal MINITAB does not show P-values like that. It can be 00001 like that. So, we do not know. So, up to three place of decimal can be calculated over here.

If you are using some other software let us say r interface, you will find the P-value with some values and beyond up to 5 place of decimal you can see the P-values like that, ok. So, that is also possible in other software's like R which is freely available and we can we can also do this type of testing over there also, ok.

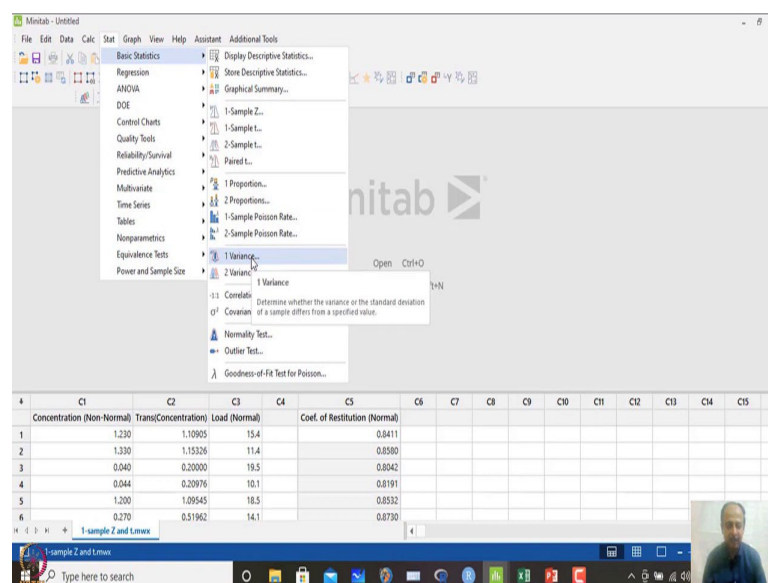
So, what we are showing over here is that this indicates that when I am doing that, I am not assuming any distribution over here symmetry we are not assuming over here. We can go for this type of analysis and we will go for a conventional approach whenever I

know the distribution follows, it is following normal distribution like that. And we want to test whether the mean is equals to some values or not.

We can also do variance testing over here, we can do 1- sample variance test over here. So, if some example exist over here we can see that.

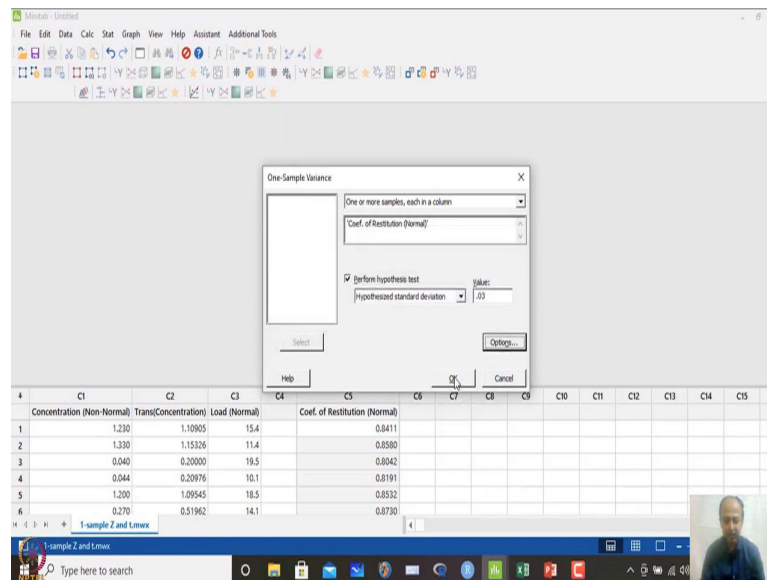
So, we can test the variance whether it is equals to. So, our variance that has come out, what we can see over here standard deviation is 0.02, 0.02 is the average that we are getting. So, in this data set, let us say I want to check the variance over here and I want to do the testing on variance like that, ok whether variance is equals to some value.

(Refer Slide Time: 23:20)



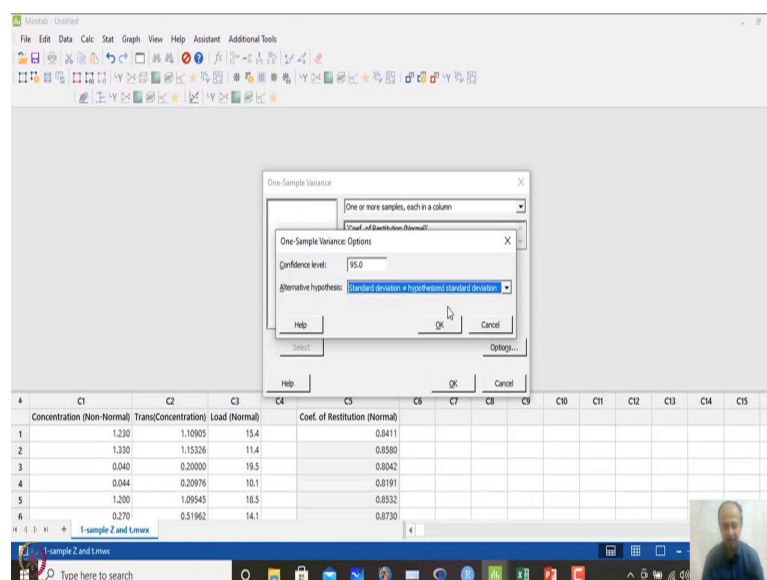
So, if I go to stat and I go to basic stat over here and I want to do a variance test, there is an option one variance test like that, instead of mean I am testing the variance over here.

(Refer Slide Time: 22:28)



Coefficient of restitution, so I want to perform the hypothesis value over here, let us say I want to see whether it is equals to 0.03 something like that, so because the analysis says it is close to 0.02 So, if you see the standard deviation that is calculated over 0.245 like that, let us assume that whether it is close to 3 or not, like 0.03 or not we want to test that one. And hypothetically I am assuming 0.03 over here.

(Refer Slide Time: 23:00)



So, in options what we do is that 95 percent we do not change confidence level over here, and standard deviation I am doing both sided test over here let us say, not equals to



condition; whether it is equals to 0.03 or it is not equal to. So, two sided test I am doing over here.

(Refer Slide Time: 23:15)

**Test and CI for One Variance: Coef. of Restitution (Normal)**

**Method**  
is standard deviation of Coef. of Restitution (Normal)  
The Bonett method is valid for any continuous distribution.  
The chi-square method is valid only for the normal distribution.

**Descriptive Statistics**

N	StDev	Variance	95% CI for $\sigma$ using	
			Bonett	Chi-Square
15	0.0246	0.000603	(0.0194, 0.0358)	(0.0180, 0.0387)

**Test**

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
1	Concentration (Non-Normal)	Trans(Concentration)	Load (Normal)	Coef. of Restitution (Normal)											
2	1.230	1.10905	15.4	0.0411											
3	1.330	1.15326	11.4	0.0580											
4	0.040	0.20000	19.5	0.0042											
5	0.044	0.20976	10.1	0.0191											
6	1.200	1.09545	18.5	0.0532											
7	0.270	0.51962	14.1	0.0730											

So, in this case, I will click ok over here, and I have a chi-square test for this because we have also seen that confidence interval when we are doing the confidence interval case for  $\sigma$  also for population variance there also chi-square distribution was used like that.

(Refer Slide Time: 23:33)

**Test and CI for One Variance: Coef. of Restitution (Normal)**

**Method**  
is standard deviation of Coef. of Restitution (Normal)  
The Bonett method is valid for any continuous distribution.  
The chi-square method is valid only for the normal distribution.

**Descriptive Statistics**

N	StDev	Variance	95% CI for $\sigma$ using	
			Bonett	Chi-Square
15	0.0246	0.000603	(0.0194, 0.0358)	(0.0180, 0.0387)

**Test**

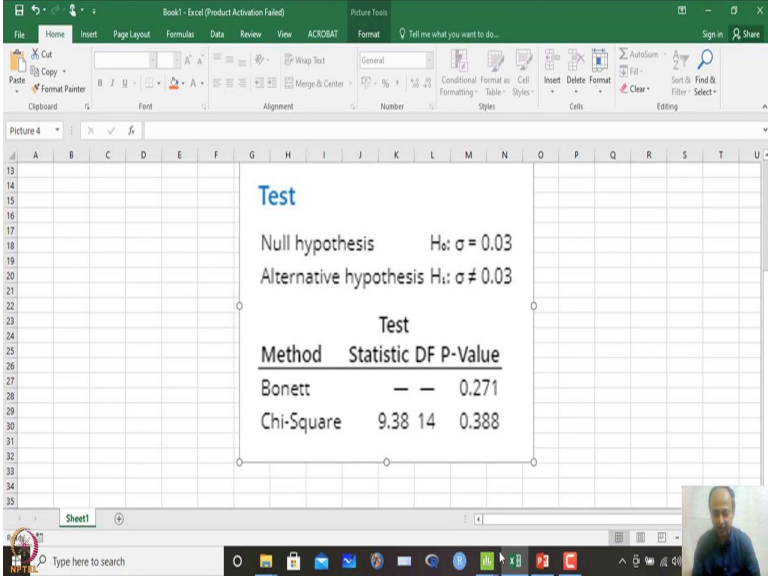
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
1	Concentration (Non-Normal)	Trans(Concentration)	Load (Normal)	Coef. of Restitution (Normal)											
2	1.230	1.10905	15.4	0.0411											
3	1.330	1.15326	11.4	0.0580											
4	0.040	0.20000	19.5	0.0042											
5	0.044	0.20976	10.1	0.0191											
6	1.200	1.09545	18.5	0.0532											
7	0.270	0.51962	14.1	0.0730											

Here also chi-square distribution is used to define this test statistic over here. And it will give you an idea of the P-value concept over here. So, you have to see this, the chi-



square test, the test statistics over here with degree of freedom that is given. So, I can just copy this one and paste it in excel, so that you can see.

(Refer Slide Time: 23:49)



Test

Null hypothesis  $H_0: \sigma = 0.03$   
 Alternative hypothesis  $H_1: \sigma \neq 0.03$

Method	Statistic	DF	P-Value
Bonett	—	—	0.271
Chi-Square	9.38	14	0.388

So, I will enlarge the test statistics what we see over here. So, what we will see is the chi-square test that we will see over here. Test statistics is 9.38, degree of freedom is 14, information that we have and this is 0.388, so that indicates that we cannot reject the null hypothesis which is  $\sigma$  equals to 0.03 because our value also shows that that is close to 0.03 over here.

So, what we are; what we are saying over here is that, that we can do test for, and underlying we are assuming that the it follows normal distributions like that it follows. Normal distribution that is the underlying assumption, but variance can also be tested. So, we have an option in MINITAB where we can do variance testing also.

So, mean testing what we are doing, Z test, t test. Similarly, when we are testing for one sample variance like that in that case. We have a chi-square distribution to define that one and based on that we can define the test and P-value will indicate whether to reject or accept the null hypothesis like that.

So, this way we can do one sample testing. So, most in quality we will find that some load information is given over here. Let us say this example was earlier taken I want to

test let us say this load is close to 15 or it is greater than that then only we will accept like the coefficient of restitution.

(Refer Slide Time: 25:04)

Test and CI for One Variance

Test

Null hypothesis  $H_0: \sigma = 0.03$

Alternative hypothesis  $H_1: \sigma > 0.03$

Test

Method Statistic DF P-Value

Bonett -- 0.271

Chi-Square 9.38 14 0.388

Normality Test

Goodness-of-Fit Test for Poisson...

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
1	Concentration (Non-Normal)	Trans(Concentration)	Load (Normal)		Coef. of Restitution (Normal)										
2	1.230	1.10905	15.4		0.0411										
3	1.330	1.15326	11.4		0.0580										
4	0.040	0.20000	19.5		0.0042										
5	0.044	0.20976	10.1		0.0191										
6	1.200	1.09545	18.5		0.0532										
7	0.270	0.51962	14.1		0.0730										

(Refer Slide Time: 25:07)

Normality Test

Variable: Load (Normal)

Tests for Normality

Anderson-Darling

Ryan-Joiner (Similar to Shapiro-Wilk)

Kolmogorov-Smirnov

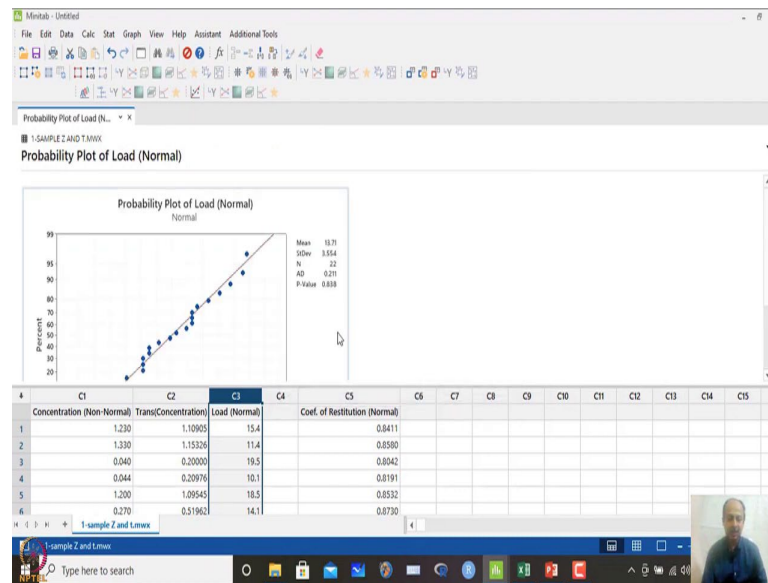
OK

Cancel

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
1	Concentration (Non-Normal)	Trans(Concentration)	Load (Normal)		Coef. of Restitution (Normal)										
2	1.230	1.10905	15.4		0.0411										
3	1.330	1.15326	11.4		0.0580										
4	0.040	0.20000	19.5		0.0042										
5	0.044	0.20976	10.1		0.0191										
6	1.200	1.09545	18.5		0.0532										
7	0.270	0.51962	14.1		0.0730										

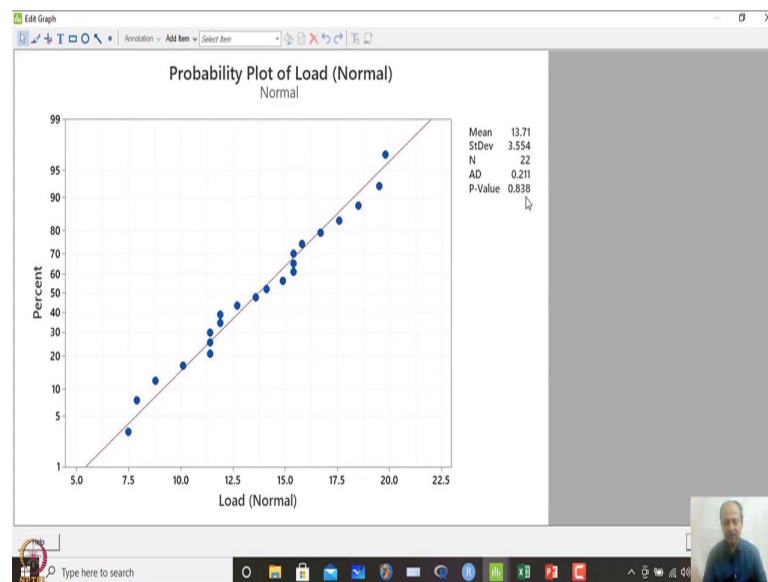
So, here also you can check whether the data set follows normal distribution like that. Again I am redoing this one. So, I will take the load information, load data information over here.

(Refer Slide Time: 25:14)



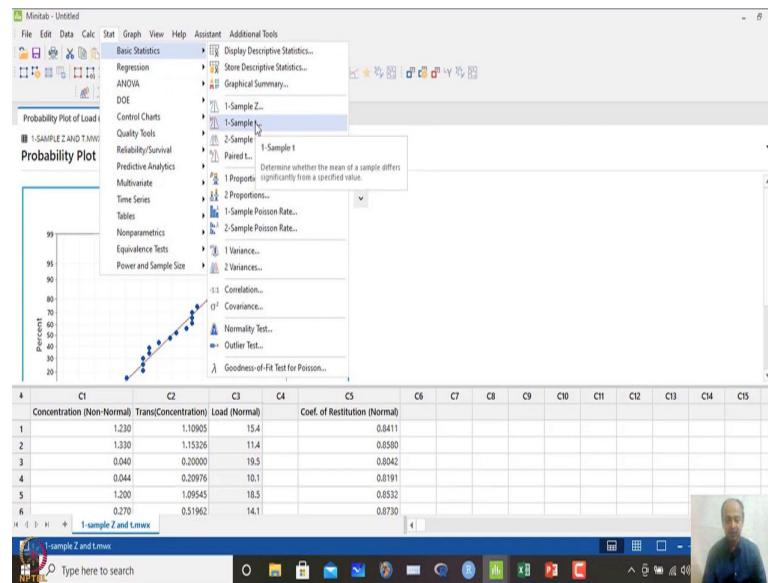
And I will do Anderson-Darling test, and I see that Anderson-darling test indicates; the data is falling normal because P-value is 0.838.

(Refer Slide Time: 25:17)



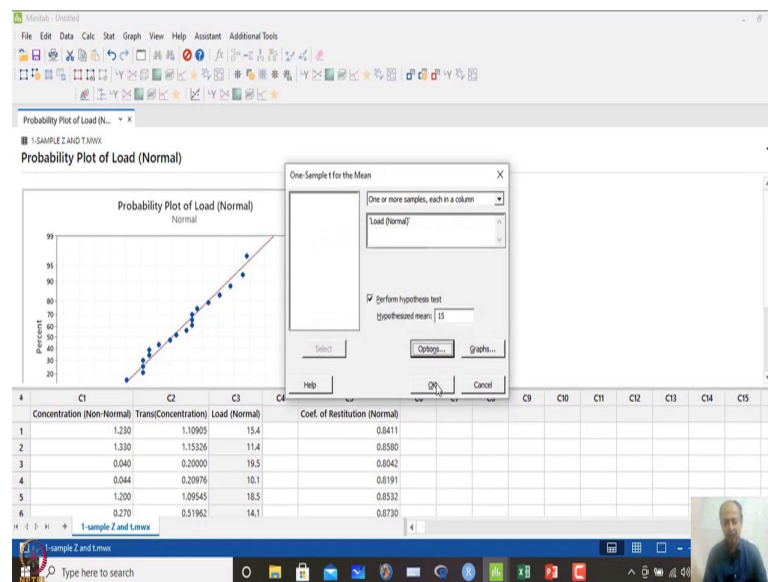
So, in this case there is no problem in normality assumptions.

(Refer Slide Time: 25:26)



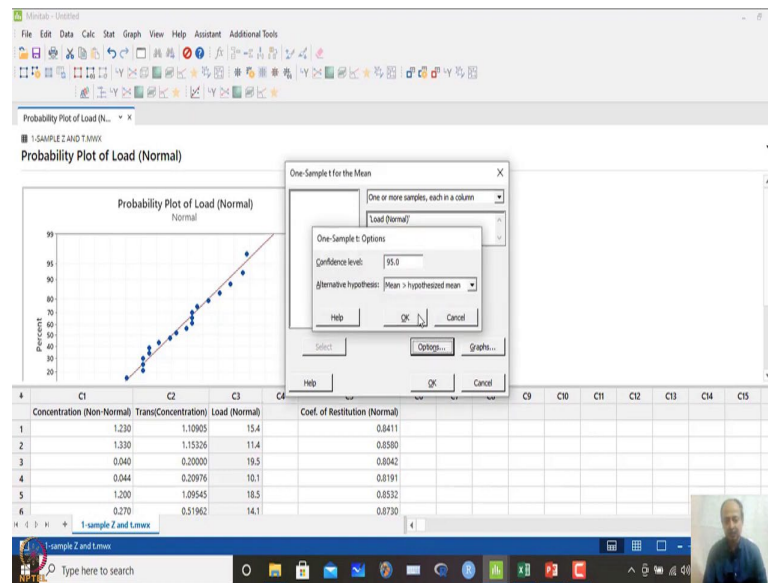
So, I can go directly for the testing of this. So, I will go for basic testing and one sample if  $\sigma$  is known,  $\sigma$  is unknown over here, so I will go by that.

(Refer Slide Time: 25:32)



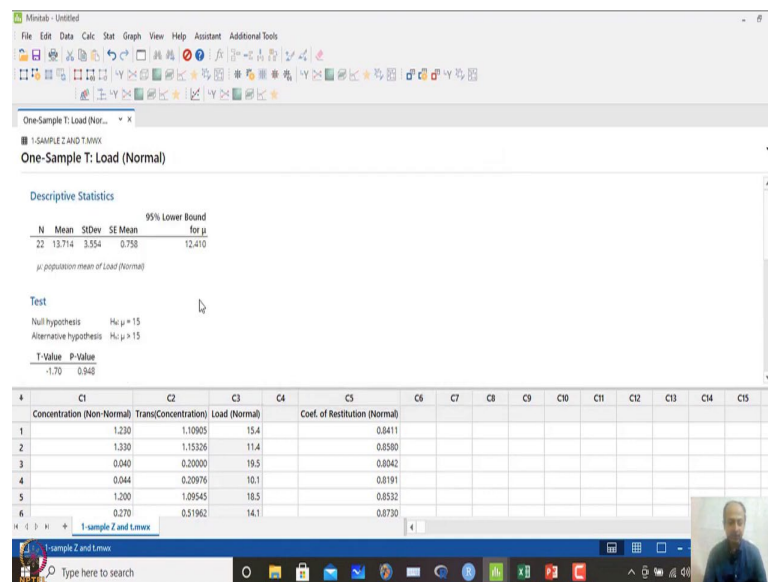
So, I will go for the load over here and I will perform the test whether it is more than 15 or something like that we can hypothesize mean over here.

(Refer Slide Time: 25:41)

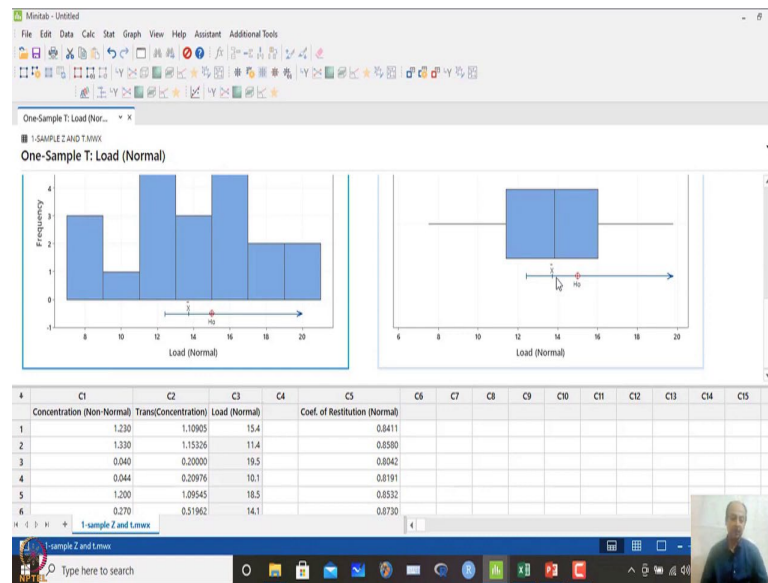


And in options, I will say that if it is greater than hypothesized means I will go for this thing.

(Refer Slide Time: 25:48)

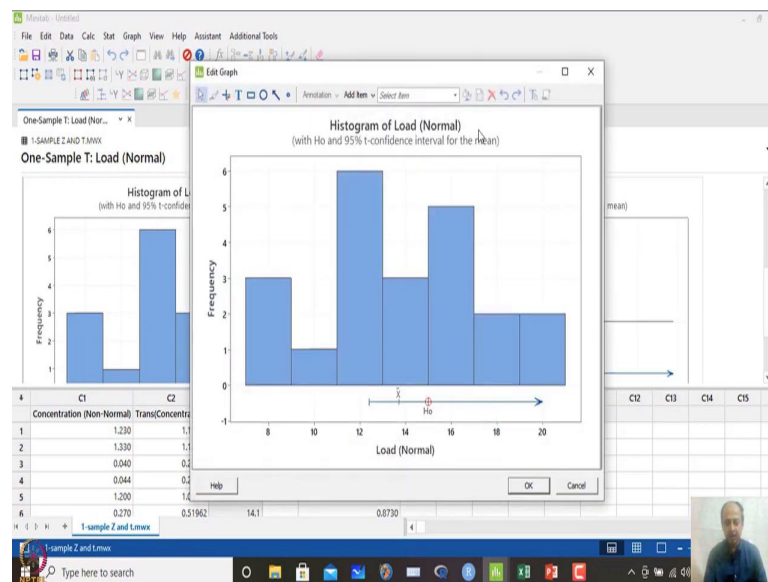


(Refer Slide Time: 25:50)



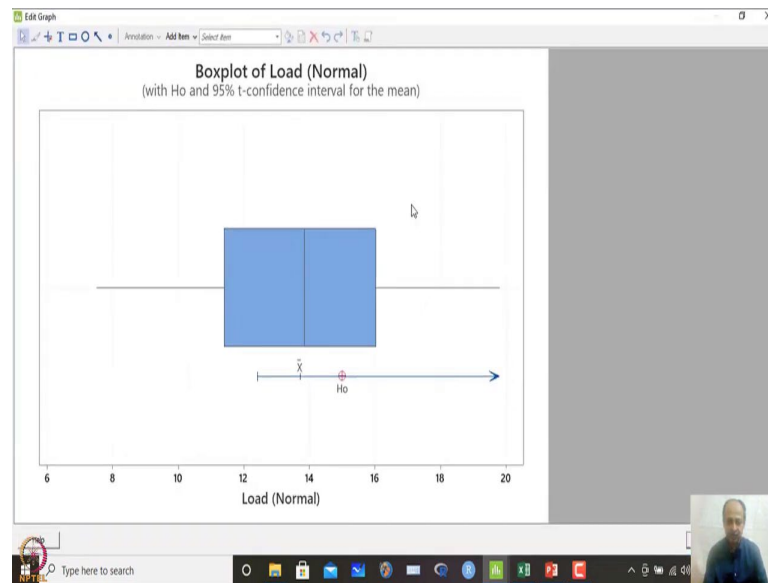
So, if I click, ok over here and what I see is the P-value is 0.948. Here also we have given options for the data to be disclosed in; we can draw the histogram over here.

(Refer Slide Time: 25:58)



There is a graph option that we have clicked earlier also. So, these things also you can see the distribution of the data over here. Normality assumption is following, so statistically we can say it is non-normal. So, this is following normal distribution. So, this is true.

(Refer Slide Time: 26:14)



And also you will see a box plot of this load over here; that means box plot with hypothesis location over here. Hypothesized mean is over here, location that plus sign that you see over here, red dots that you see over here this is the location of the null hypothesis and this is the average that we are getting, this is the average value  $\bar{x}$  location and this is the median value what is the 50 percentile, 50th percentile over here, in the box plot over here.

So, these values are, these values that  $\bar{x}$  we are getting is very close to the hypothesized mean and that is why what we are getting over here is what we are getting over here is p-value is coming out to be significantly higher than 0.05. That means, the hypothesized mean is very is not very different from the  $\bar{x}$  over here, so hypothesized mean value over here.

So, based on the data information, so this is making a conclusion the conclusion is based on not only the mean value that we are calculating also the variance of the spread of the data set that we are having. So, in that case, these two values are used over here to make a decision out of this.

So, t statistics include S and  $\bar{x}$  information over here. So, both the information's are used and based on that we are making a judgment over here whether the load is greater than 15 or it is less than or less than equals to 15 like that.

And based on that I will make a judgment over here. So, I have seen the data when it is normal what to do and if it is non-normal what options we have transformation options over here. We also have non-parametric test over here. We also have a non-parametric options for this. And non-parametric options that we have two options what we said is that, I can go directly to once one sample sign test over here.

And in case you find symmetry in the distribution when you plot the data set like that we can assume symmetry and uniformity in the distribution left-hand side and right-hand side of the distribution. In that case, what symmetric distribution like normal distribution is a symmetric distribution like that. In that case, non-parametric option what we can explore is Wilcoxon test one sample, Wilcoxon test like that, ok.

(Refer Slide Time: 28:26)

#### Quality Control and Improvement using MINITAB

### MINITAB 19 INTERFACE & ILLUSTRATION



Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay



So, what we will do is that; what we will do is that, next we will try to cover some illustration we have given over here.



(Refer Slide Time: 28:28)

**Quality Control and Improvement using MINITAB**

**Two-Sample t-test**

To determine if there is any **difference in readings of two different methods to measure % humidity**

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Catalyst A (X1)	Catalyst B (X2)
33	31.1
32.8	30.9
34.9	30.2
36.6	33.4
37.2	34.3
40.1	36.3
41	36.4
36.6	36.9
35.5	34.3
43.8	40.2
45.3	40.6
38.4	37.1
44.2	37.1
38.6	36

**Test**

Null hypothesis  $H_0: \mu_1 - \mu_2 = 0$   
 Alternative hypothesis  $H_a: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-0.35	14	0.729

$\sigma_1^2 = \sigma_2^2 = \sigma^2$

$f_0 = \frac{s_1^2}{s_2^2}$

Data Source: Design and Analysis of Experiments, D.C. Montgomery, John Wiley & Sons, I

Prof. Indrajit Mukherjee, SJMSOM, IIT Bombay

Then, we will discuss about in our next session what we will do is that, we will discuss about two sample t-test. This is the important starting point of experimentation. This is the starting point of experimentation and this is important. So, we have given a brief idea about confidence intervals.

We have given a brief idea of hypothesis testing, not that I have emphasized much about hypothesis testing and on confidence interval you can see some other lectures which gives you thorough idea about this and extensive lectures on that you can see.

But that is the basic idea that is required over here in this course. So, that is the foundation. So, I have given you some basic idea of the foundation where P-value is important for us based on which we are making a judgment. And we will do hypothesis testing everywhere. Most of the time experimentation is only hypothesis testing; that means whether to go for this decision or that decision like that.

So, we are into inferential statistics that we are doing over here in quality also. So, some amount of statistics is required understanding of that is required to go ahead with the experimentation concepts like that, what we will try to discuss about factorial design and then all other options that we have, ok.

So, everywhere some statistical analysis has to be done and based on that we have to make a decision. But always remember that it has to be a practical decision basically, ok.

So, sometimes P-value is less than 0.05, immediately I take a decision it is not like that. So, physical significance practicality of that whatever I have concluded that has to be taken into consideration when I am making a decision like that, ok. Values are very close although it is significantly different in that case, let us say computational time like that, with two algorithms you are checking that one.

So, computational time is very close like that and all those statistically significant, but practically it does not make any sense. So, you have to consider this P-value very very seriously when I am making a judgment like that; what is practical being an engineer or in a process you are knowledgeable of the process, so whether this difference is really significant.

So, that I say that this is really significant or the difference required is more then only we can say that this is significant like that. So, P-value, what is the value of P-value based on that we can make a judgment like that, ok. So, if P is very close to 0.5, but just less than 0.5 whether we will 0.05, whether we will go for that or we will we need more evidence or we need more things like that that we have to check when we are making a decisions like that, ok.

So, what we will do is that we will close here and we will start with the next session. Starting from here we will discuss about two sample t test which is extensively used in quality, which is extensively used in quality that we will start.

Thank you for listening.