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

Lecture - 19
Two - sample t-test

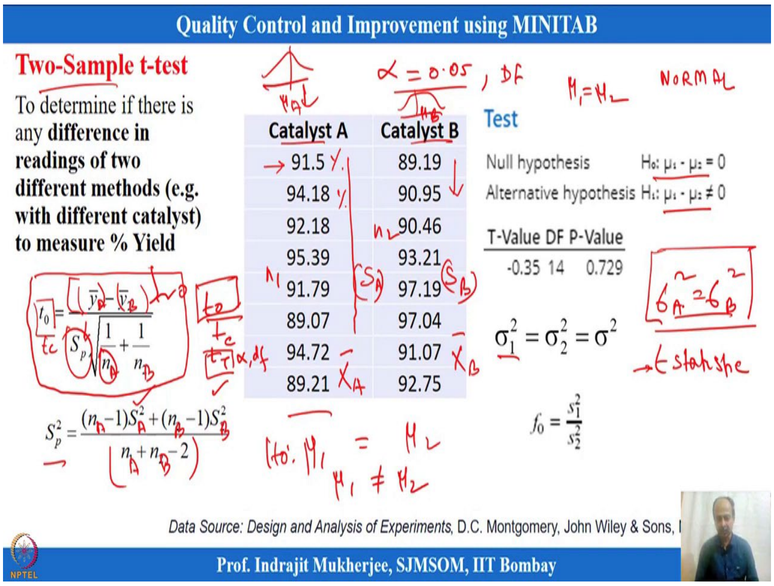
Hello and welcome to session 19 of our course on Quality Control and Improvement with MINITAB. So, I am Prof. Indrajit Mukherjee from Shailesh J. Mehta School of Management, IIT Bombay. So, earlier session what we have done is that we have talked about hypothesis testing.

And in that case, how to compare the sample observations with the population mean or some value. Whether we can compare that one, we can make a judgment out of that, whether the mean is this value or not, is it equal or not equals to conditions, that is the simplest way we can explain hypothesis testing.

And we have used a z-test and we have used a t-test condition to satisfy that one. And we have also seen that if underlying assumptions of normal distribution fails in that case, what is to be done, what conversion we can do like box-cox transformation we have seen. And then we do the hypothesis testing or the converted data or otherwise we go for non-parametric test. And that considers median and ranking concept.

And based on that, we get the p-values and we make interpretation, $p < 0.05$, we go for the alternate hypothesis. And if $p \geq 0.05$ we cannot reject the null. So, that is the condition based on which we make analysis.

(Refer Slide Time: 01:41)



And so now what we will do is that, we will go to a specific test which is known as two-sample t-test which is very relevant to our quality concept, and that can be considered as the starting point of experimentation ok.

So, this concept I will give over here the statistical concept that is used to do this testing, and how it is to be done in MINITAB. So, here this problem that I am highlighting over here is that there are 2 different catalyst that is used.

And the experimenter is interested in to know that whether to use catalyst A or catalyst B which improves the yield. So, I have a data observation over here which is from catalyst A, when I have used catalyst A, what was the yield.

So, this is the percentage yield that was reported like that. And this is catalyst B which was when I have used catalyst B and the sample observations that I have, and what is the yield percentage that is measured over here.

I want to check whether the in populations this $\overline{\mu_1}$ that we are getting over here and $\overline{\mu_2}$ average – whether they are same, whether they are different like that. So, I have $\overline{X}_A$, $\overline{X}_B$ the average value that we are getting out of the samples over here.

We can have standard deviation of A (σ_A), we can also get standard deviation of B (σ_B) that information also we can get from this data observation that we have ok. Now, I have to make a judgment:

$$H_o : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

So, while doing this test, so what we do is that we use as a, we use a specific test statistic that is known as t test statistic over here. So, we will make a calculated value that will be calculated as:

$$t_0 = \frac{\overline{y}_A - \overline{y}_B}{S_P\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

So, if $\overline{y}_A - \overline{y}_B$, is close to 0, we can assume they are close in that case they are not much different.

So, my analysis should show the result should come out that hypothesis testing whenever I am doing that, if it is close to 0, in that case we expect that the null hypothesis cannot be rejected. If it is very different from 0, what is expected is that we should reject the null hypothesis.

So, that is the basic interpretation I am trying to make over here. But the test statistics that is used over here, this difference is calculated, and then a pooled standard deviation is calculated. n_A and n_B is the number of observations for catalyst A and catalyst B, respectively. This can be same or this can be different .

And then we calculate a pool standard deviation for the analysis over here which is given as,

$$S_p^2 = \frac{(n_A - 1)S_B^2 + (n_A - 1)S_B^2}{(n_A + n_B - 2)}$$

So, we can calculate this pooled standard deviation because all information is known over here. This pooled standard deviation will be placed over here. And based on that, we will calculate a t_0 value which is known as we can we can also say t calculated value like that.

Then T-tabulated value can also be seen because the degree of freedom is given over here. And we are assuming α let us say 0.05 or 5 percent is the level of significance that we are assuming over here. And in this case, what we can do is that with a given level of α and degree of freedom over here that is $n_A + n_B - 2$ over here, what we can do is that we can get the value of tabulated calculated value of t_0 over here. And this can be compared with the tabulated value like that which depends on α and degree of freedom. So, that can be compared and based on that we can make a judgment. So, that was the old way.

And nowadays what we are doing is that we are seeing the p-value what I told because of software interface we are getting p-values, any of the software will report the p-values. And based on that, we can reject or accept the null hypothesis like that, based on the outcomes of p-values like that ok.

So, here also when I am doing these two-sample t-test to compare the means over here, $\mu_1 = \mu_2$ or $\mu_1 \neq \mu_2$, some assumptions are there. So, over here basic assumption is that the data that is coming from catalyst A and catalyst B both should follow normal distribution like that.

So, normality assumption is there. So, normal assumptions is required for each individual observations that each individual data sets that we are getting. So, this should follow normal distribution. And also we are assuming that the data are independent.

So, set of data over here with catalyst A that we have collected over here has nothing to do with the data set that is in catalyst B. They are independent with each other. So, there is no correlation that should exist between data set A and data set B like that, or catalyst A with catalyst A and catalyst B like that.

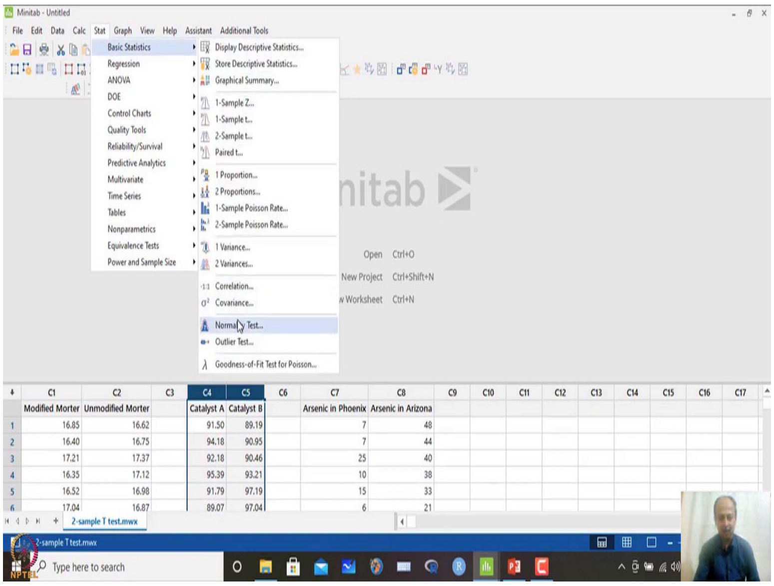
So, that check is also needs to be ensured over here. And then what we can do is the third check that is required is that whether the variance over here so the S_A that we have calculated over here can we estimate whether the variance of A is same as variance of B like that.

So, this will dictate what type of t statistics that we will use t statistics that we will use. And in case they are same, one t statistics will be used; if they are different, another t statistics will be used with a given degree of freedom like that ok.

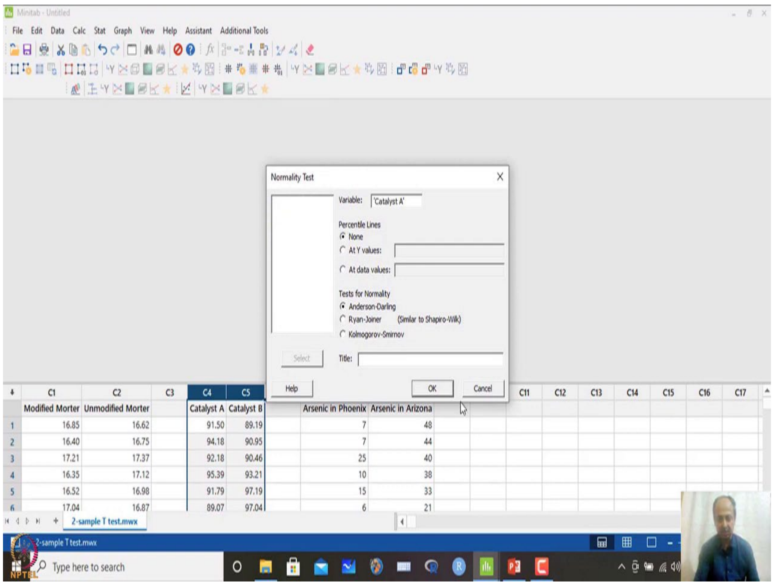
So, that is given condition and you can see any books to understand the basics on this two-sample t-test and the statistical that is t statistics that is used like that. So, degree of freedom will change in case this varies like that. So, in case this is not, this is not same in that case some different test with different degree of freedom is required like that ok.

And so for that we need to do these checks. So, normality one independent testing that the data sets are independent and the variance is same or not that will also considered over here. So, this data set I have in MINITAB. So, I will try to do the test as per the requirements like that.

(Refer Slide Time: 08:17)

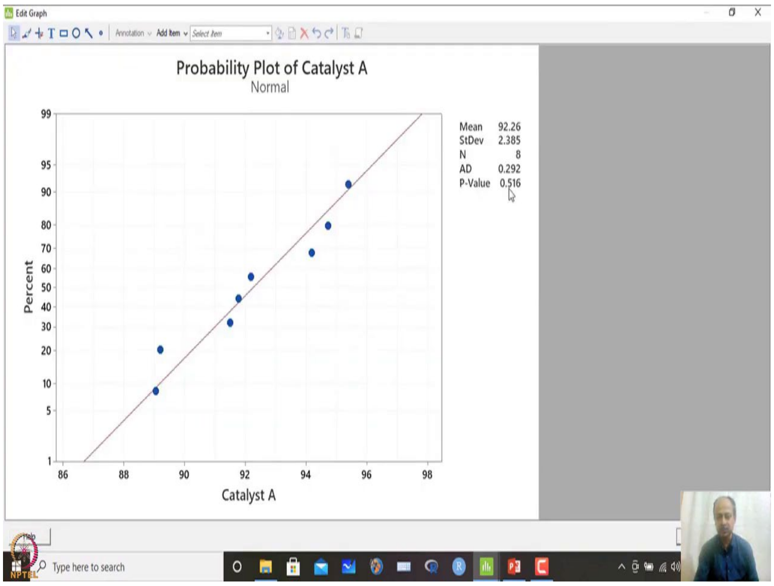


(Refer Slide Time: 08:26)

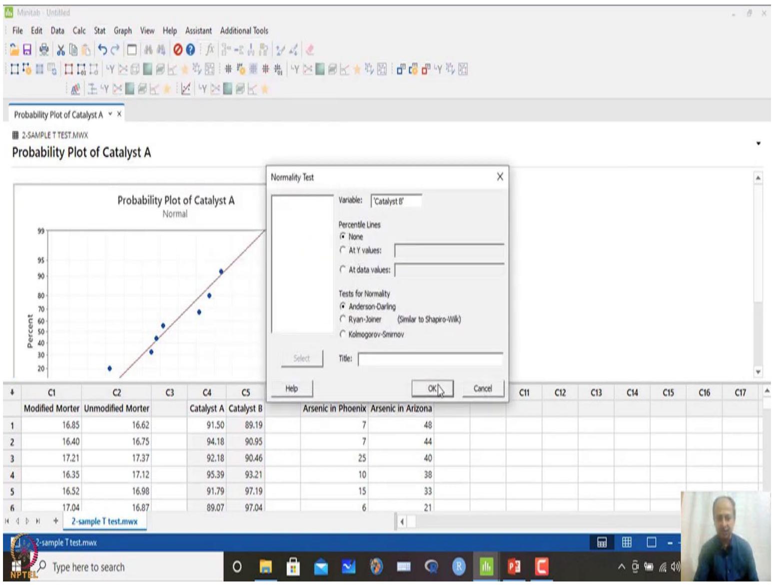


So, what I will do is that here you see that data set is in C4 and C5 column, what I will do is that I will go to stat and basic stat over here I want to check normality first. So, whether I want to sure whether catalyst A is normal or not data set that I am having. I am using Anderson-Darling test again. For the data set A, and I will click ok over here.

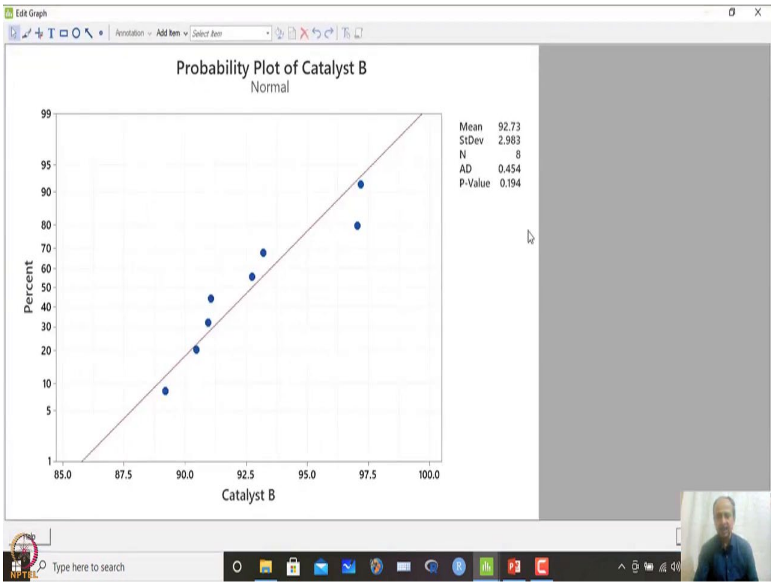
(Refer Slide Time: 08:36)



(Refer Slide Time: 08:51)



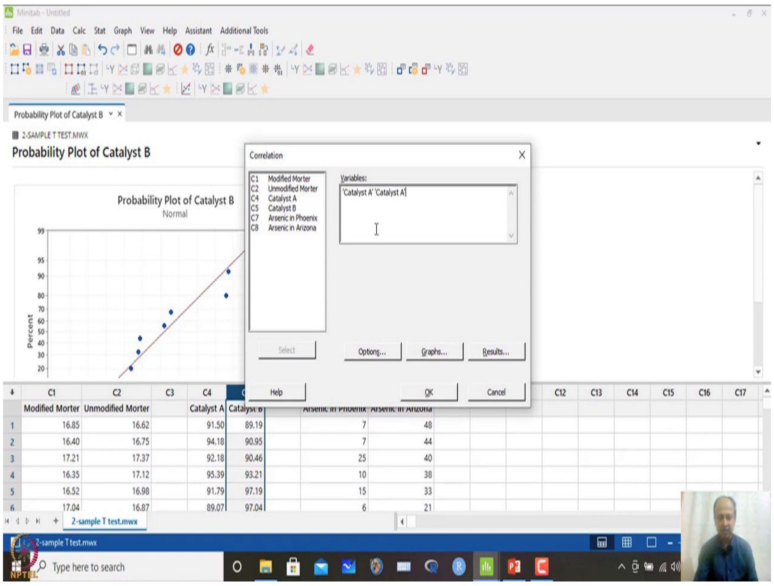
(Refer Slide Time: 08:55)



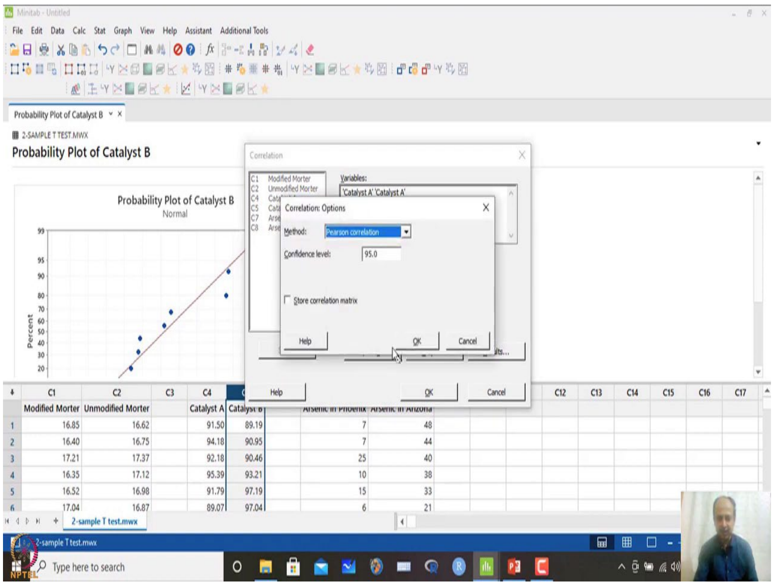
So, when I give ok, what I see is that the p-value that is reported over here is around 0.516. So, data seems to be normal. So, there is no problem. So, I can close this one. And similarly I can do for second for normality test for data set B, catalyst B over here. And I do the same testing over here.

And then for catalyst B what I get is that again I get a p-value which is more than 0.05. So, here also normality assumption is not violated. And then what we can do is that whether they are independent or not catalyst A and catalyst B what we can do is that we can see the correlation coefficient over there.

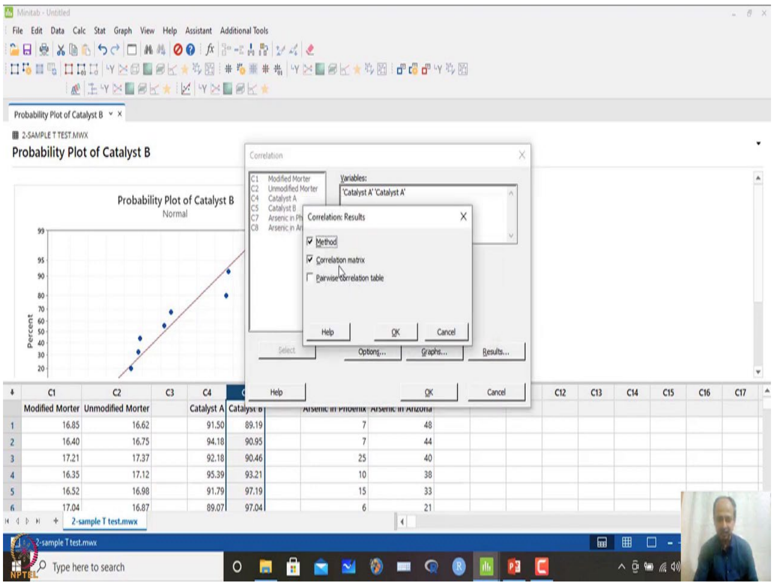
(Refer Slide Time: 09:11)



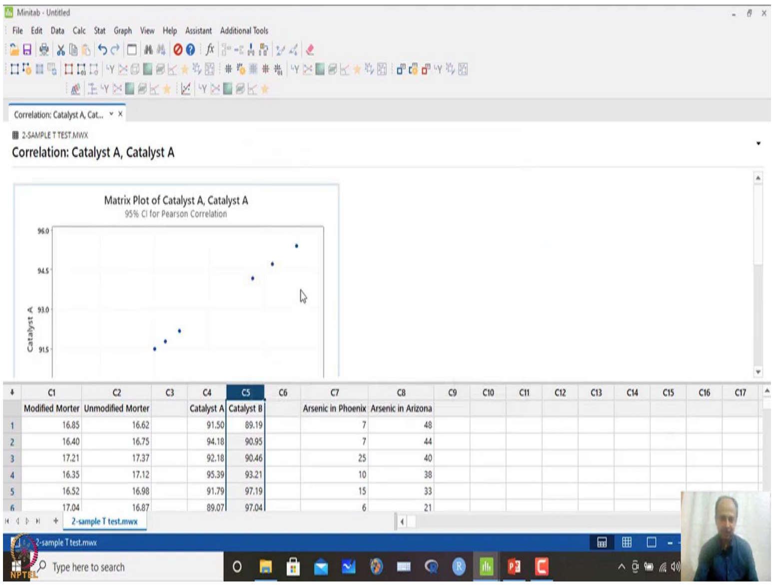
(Refer Slide Time: 09:15)



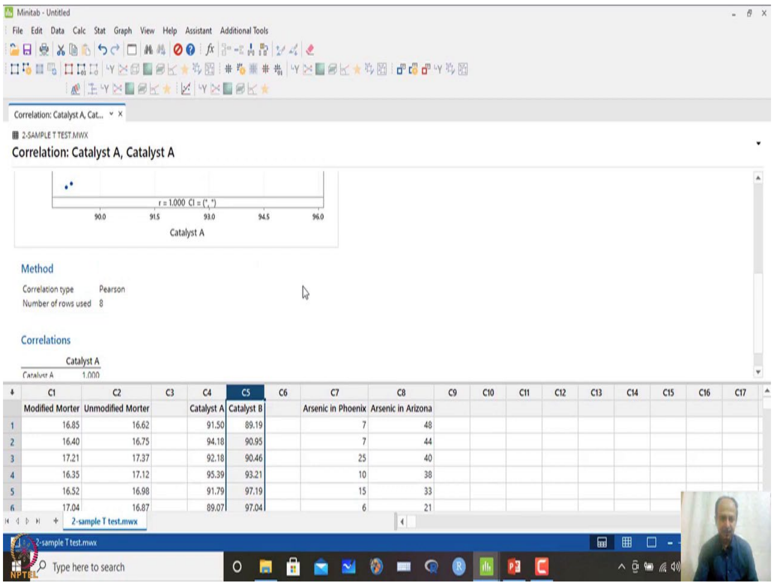
(Refer Slide Time: 09:20)



(Refer Slide Time: 09:26)

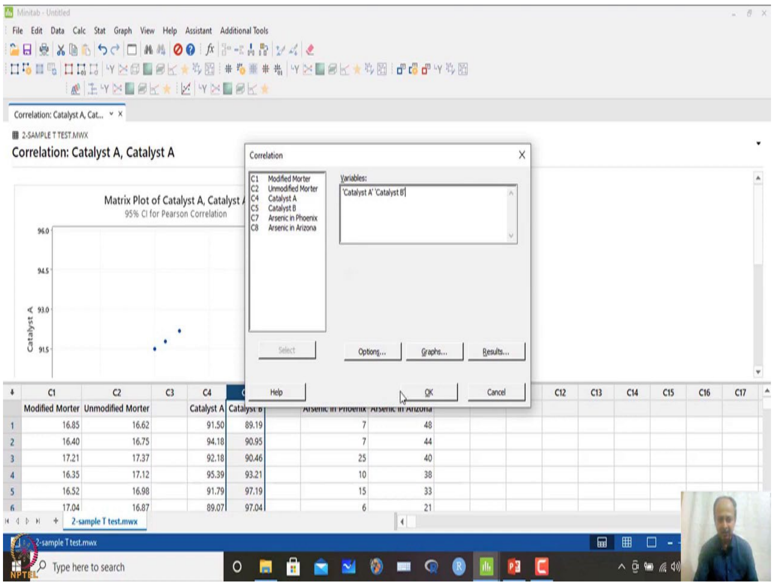


(Refer Slide Time: 09:27)

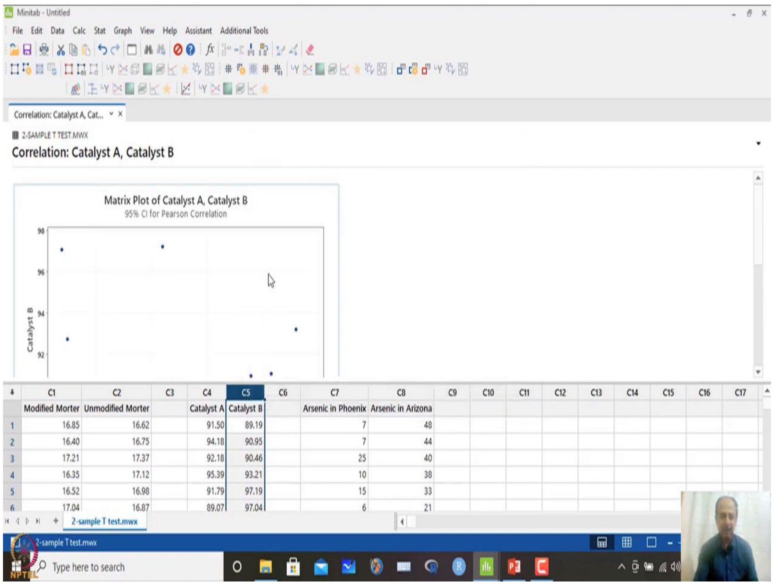


So, correlation can be checked over here for catalyst A and catalyst B over here. And options is that I can use Pearson correlation, and then results what I can do is that correlation matrix. So, this can be reported over here, and I click ok. So, in this case what we get is that near perfect relationship what we are getting over here.

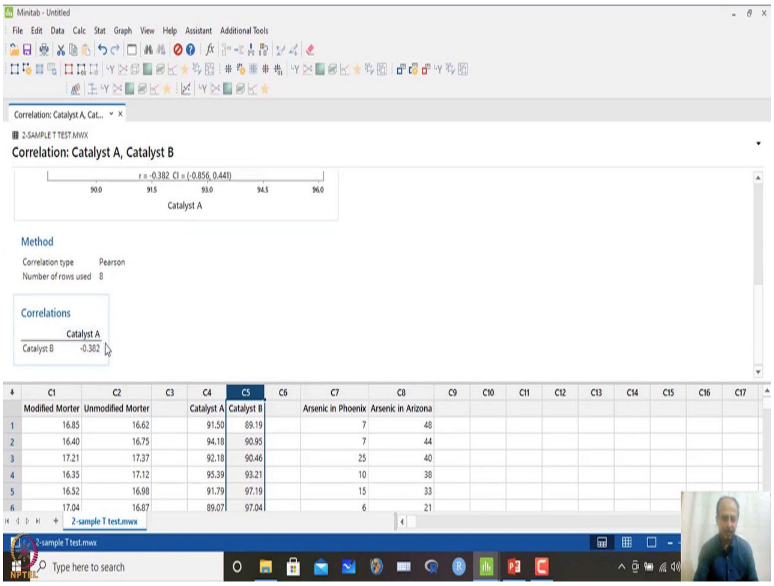
(Refer Slide Time: 09:42)



(Refer Slide Time: 09:45)

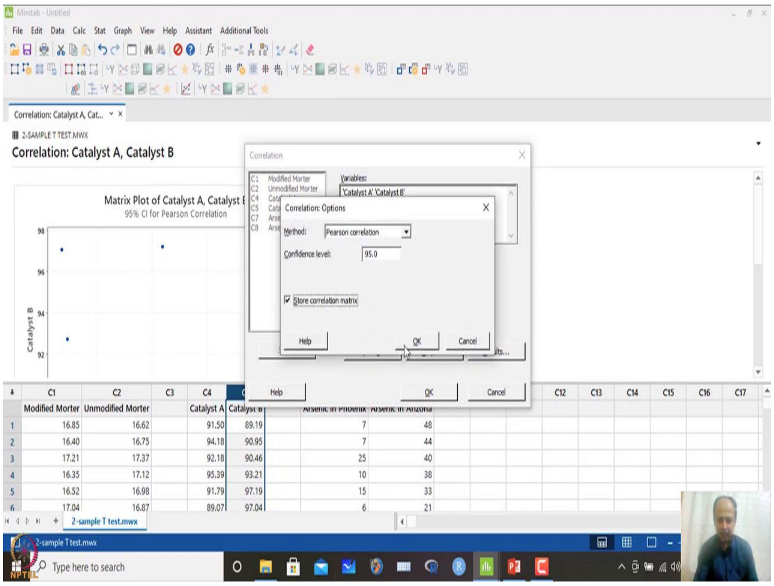


(Refer Slide Time: 09:46)

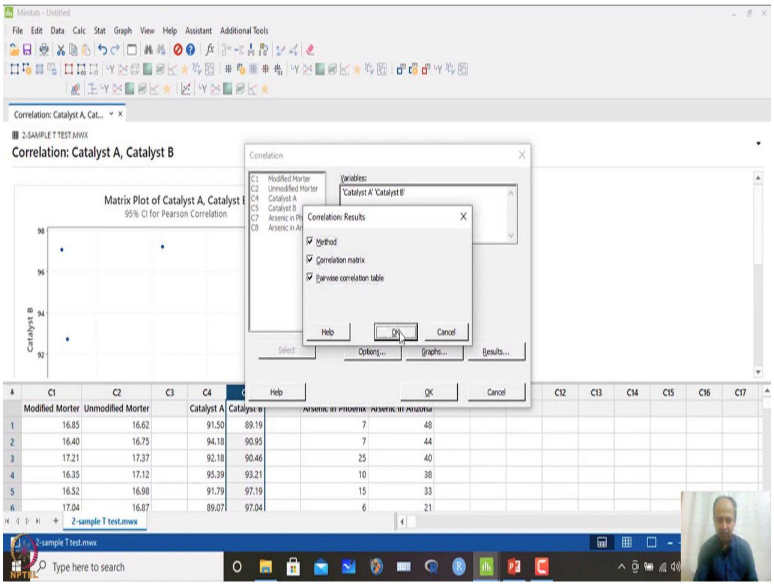


So, these values are 0.3 approximately, 0.3 negative, negative correlation what it is showing. And also check the P-values for this. So, we can go to stat and again basic statistics like that correlation analysis over here.

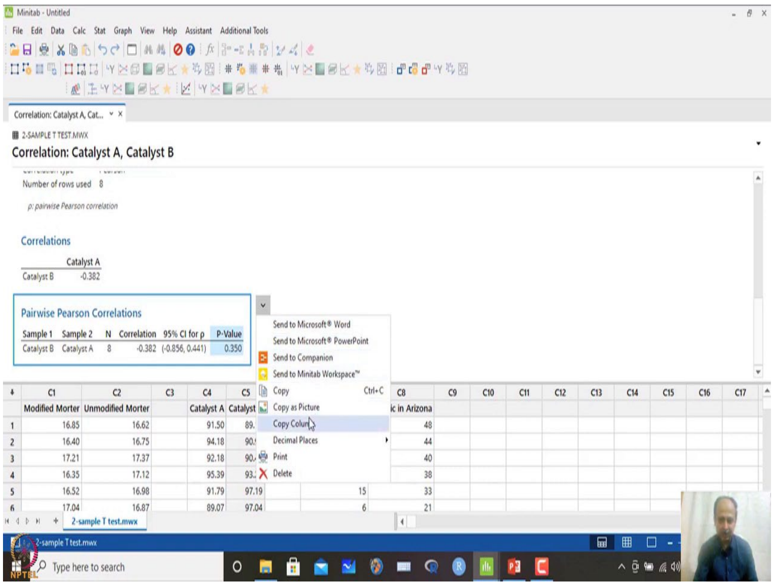
(Refer Slide Time: 10:09)



(Refer Slide Time: 10:15)

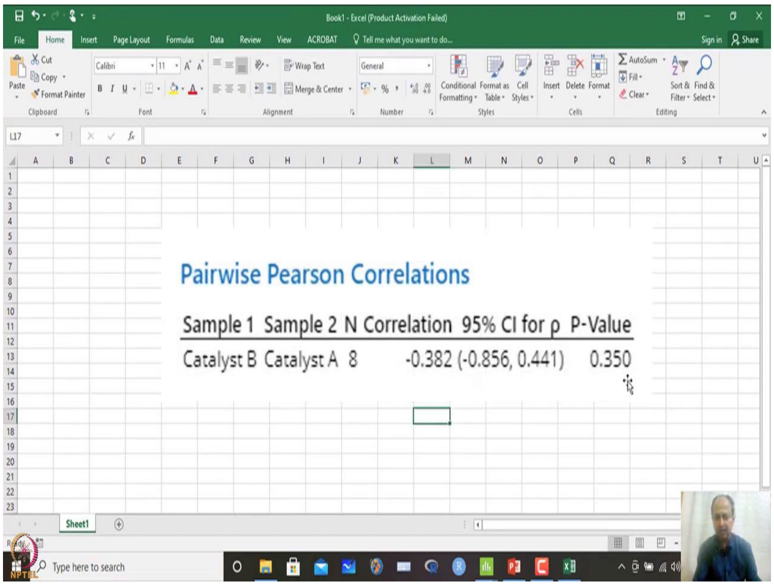


(Refer Slide Time: 10:21)



So, options what we can give is that store correlation matrix, and graphs results. So, we click pairwise correlation matrix and then we can say ok. And then what we can see is that p-value will be reported over here. So, if you see the p-values over here and I am copying as a picture.

(Refer Slide Time: 10:44)

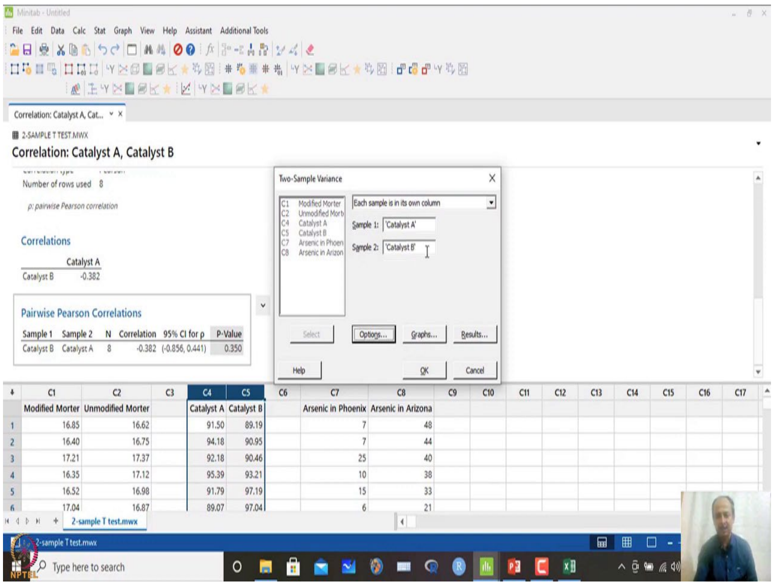


And I will paste it let us say in excel, so that it is visible also. So, let me just paste this one to show you what is the correlation that is coming out to be between catalyst A and catalyst B. So, let me just click this one. And let us paste the whatever information we have got. So, over here what you see is that p-value is greater than 0.05, so that means, catalyst A and catalyst B there is no significant correlation that exists between catalyst A and B.

The data set that I have got in sample 1 and sample 2, so they are not. So, correlation also can be checked by Pearson correlation we have discussed like that. And that whether they are significantly correlated or they are not significantly correlated that also can be seen with the p-values that is reported over here.

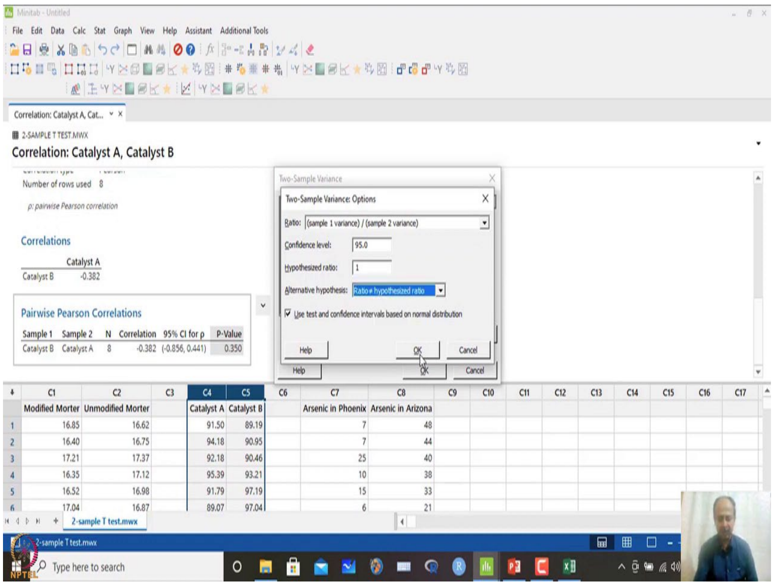
And it says that p-value is more than 0.05 will indicate that there is no, there is no statistically significant correlation that exists. If it is less than 0.05, significant correlation will be then we can say that there is a significant correlation. So, data seems to be independent over here. So, the second condition also holds. So, in this case, third condition, what we have to do is that whether the variance is same or not. So, I will do a 2 variance test like that 2 variance test over here.

(Refer Slide Time: 11:42)



So, both samples are in different columns. So, sample number 1, I will give catalyst A, and sample number 2 is catalyst B like that.

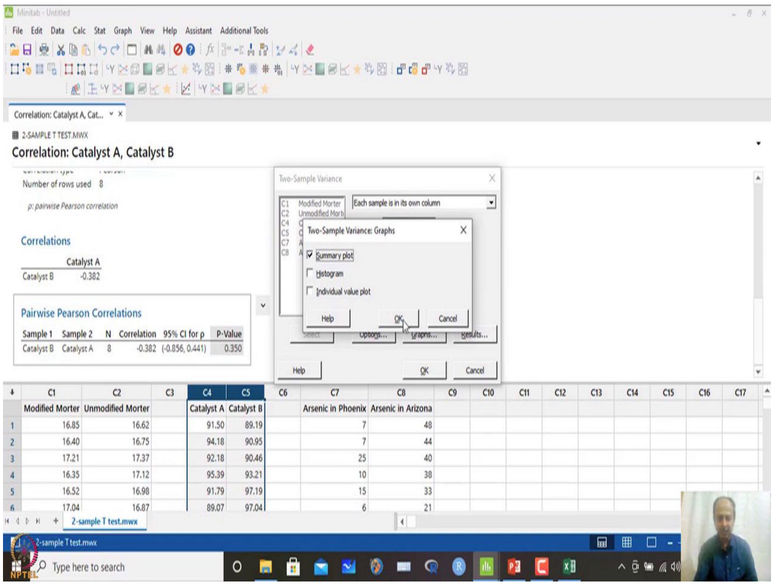
(Refer Slide Time: 11:58)



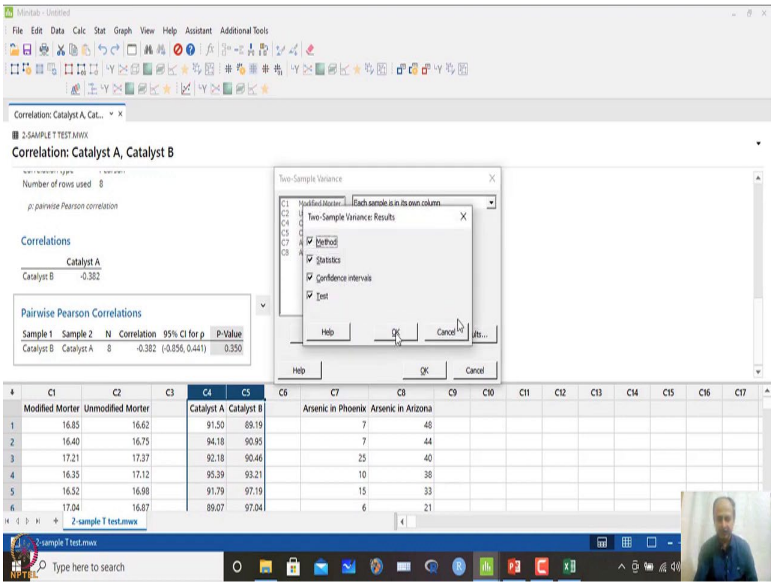
So, then I go to options over here. So, because I have checked normality, I will use test and confidence interval based on normal distribution over here. So, what I want to check whether the variance is same. So, I can use variance testing over here. I can also use standard deviation testing over here. I am using variance test let us say.

So, in this case, whether the ratio whether the variance is same or not same that I want to check over here. So, ratio not equals to hypothesis ratio over here, so whether there so both sided test I am doing over here is assuming normal distribution. So, in this case if I give ok.

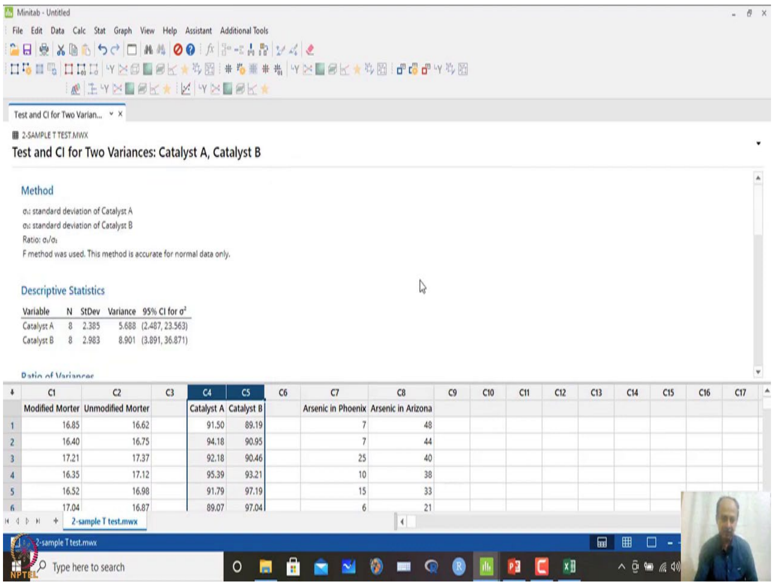
(Refer Slide Time: 12:31)



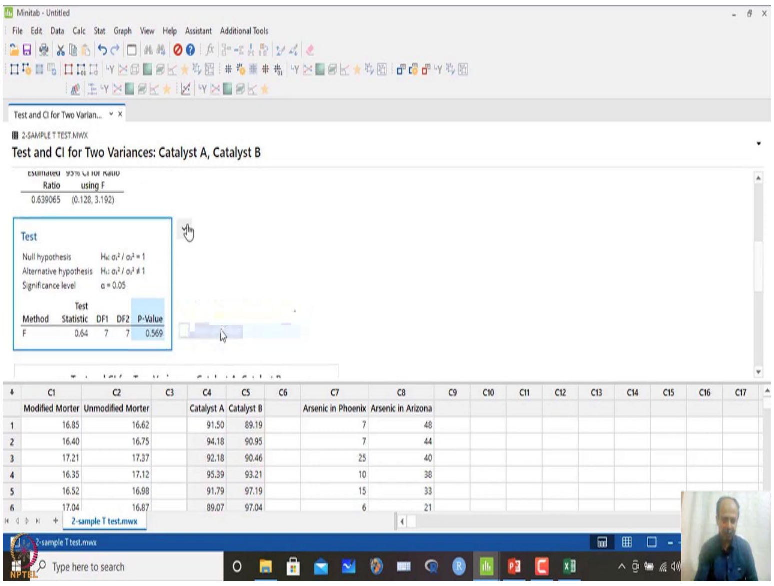
(Refer Slide Time: 12:36)



(Refer Slide Time: 12:40)

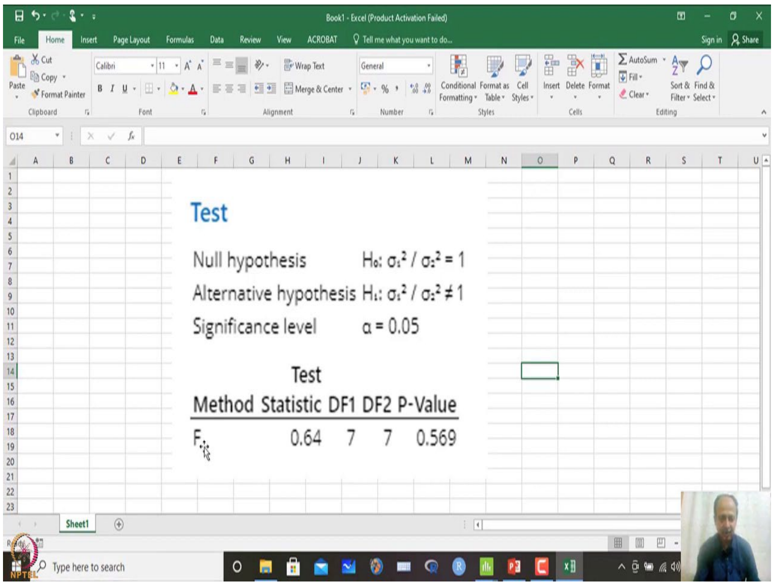


(Refer Slide Time: 12:45)



And then what I can see is that I can see summary plots like that, I can also see results like that. All the results are given over here. So, this is by default and I click ok over here. And what I get is that I will get a F statistics over here.

(Refer Slide Time: 12:56)

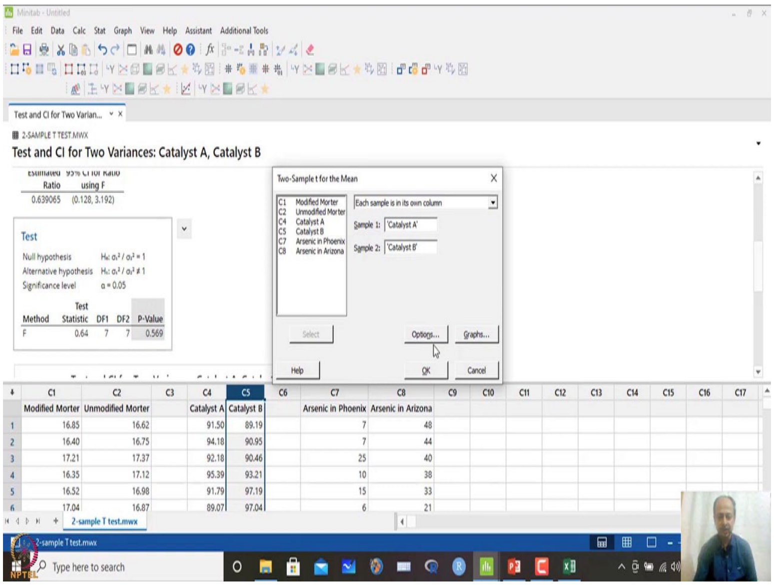


So, I can remove this one now. And I can show you this one what is the results outcome over here. So, if you see the results over here, Fstatistic was used over here method is Fthat is mentioned over here, and the statistic value is 0.64.

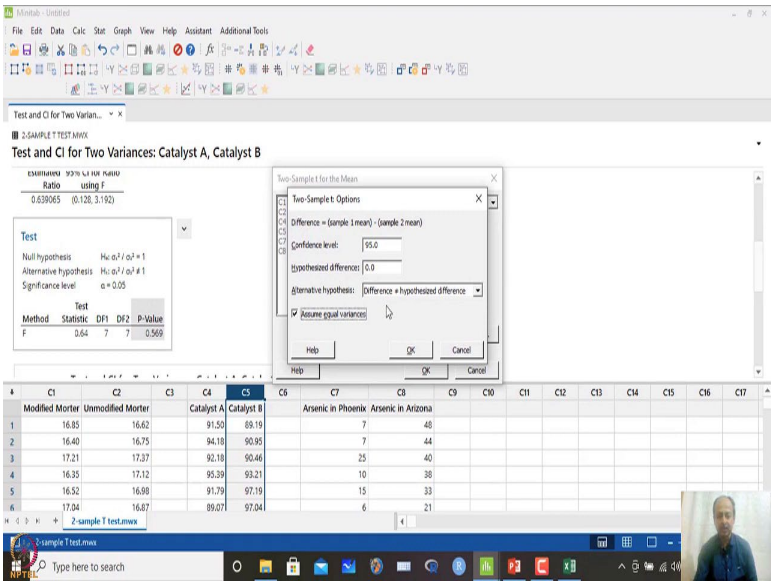
And in this case, if the $\frac{\sigma_1^2}{\sigma_2^2}$ is very close to 1, And based on that, what comes out to is 0.64 which is close to 1 basically and that is confirmed also that p-value is not significant over here. So, if P-value is not significant, that means, the variance is same basically. When variance is same, I will go for that condition while doing the hypothesis testing or 2-sample t-test like that.

So, in this case, in catalyst what we have seen is that it follows individual follows normal, then their data is independent like that. And also the variance of A and variance of B are same in population basically. So, then what I will do is that stat, I will go to stat basic stat and 2-sample t-test. Now I will apply the 2-sample t-test over here.

(Refer Slide Time: 14:01)



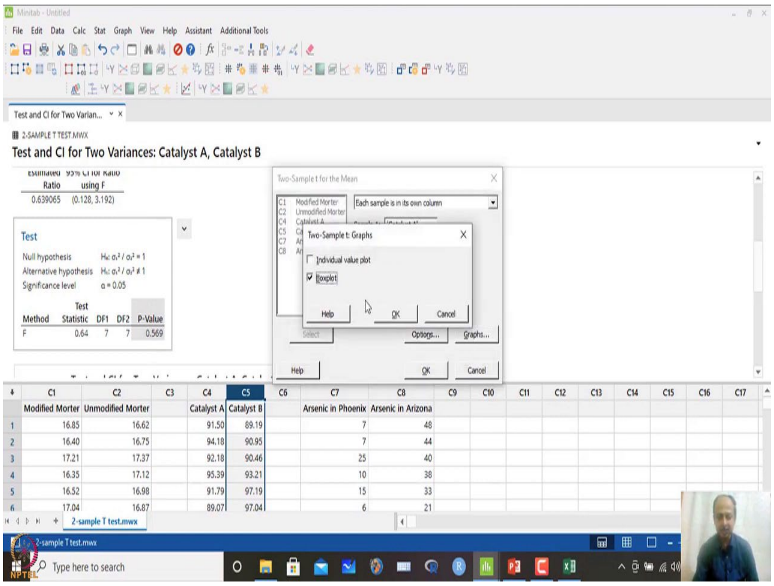
(Refer Slide Time: 14:12)



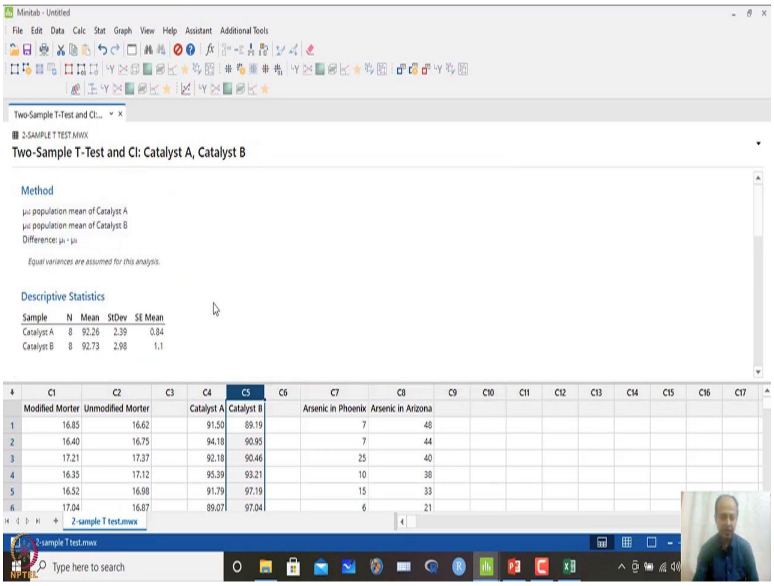
So, when I apply 2-sample t-test like that, it will ask are they in one column or they are in different column, they are in own column we have mentioned. So, all are in different columns. So, catalyst A and catalyst B, we are trying to do.

And then you go to options and then you assume equal variance. So, this you have to click over here assume equal variance. 95 percent is the confidence band that we are using every time, so that is by default and difference not equals to condition. We want to check catalyst A or catalyst B whichever is giving me is the means are different or not.

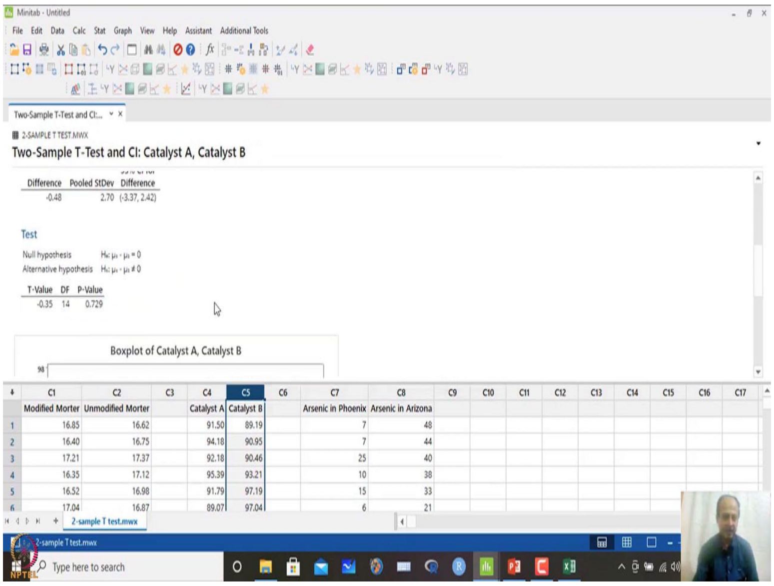
(Refer Slide Time: 14:34)



(Refer Slide Time: 14:38)

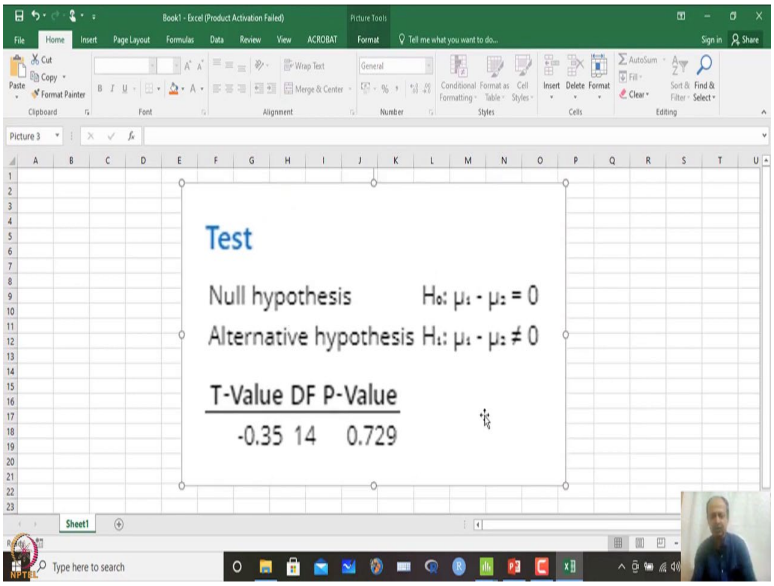


(Refer Slide Time: 14:39)



So, both sided test I am doing over here. So, in this case, I will give ok. And then box plot can also be seen over here box plot over here. And I click ok. What will happen is that I will get some values over here, and I will get a corresponding values over here. So, if I copy this one.

(Refer Slide Time: 14:53)



And then in that case what will happen is that I will I can paste this one over here. So, this is already done. So, we can just enhance this visibility for enhancing the visibility.

So, in this case, when we have done the two-sample t-test, what the value that we are getting is 0.729 is the p-value.

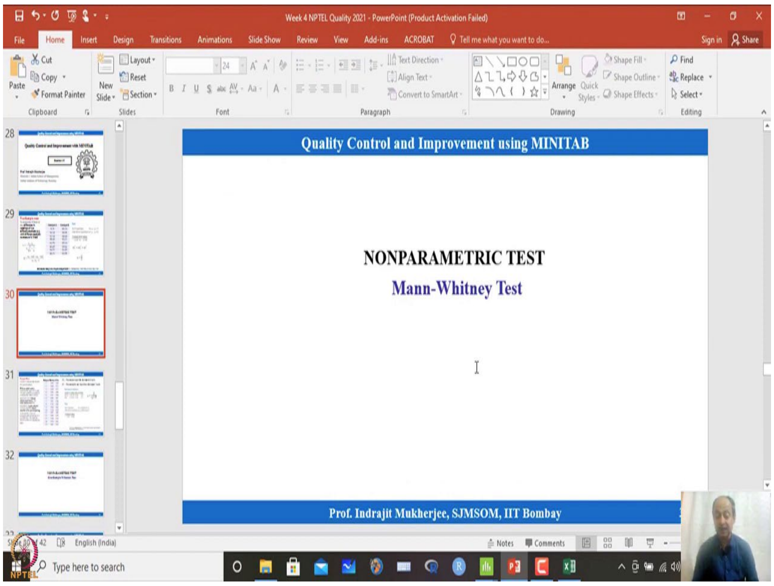
And that is more than what is what is recommended value of 0.05. So, I cannot reject the null hypothesis. So, in this case $\mu_1 = \mu_2$; that means when I am using catalyst A and catalyst B means are not differing much statistically basically ok. So, whichever catalyst you use the overall mean is remain same.

So, in this case, whatever improvement may be with a new catalyst that you have developed that is not giving you higher yield as compared to catalyst A. So, we can retain the catalyst A like that until unless we have enough evidence that this is effective, we will catalyst some new catalyst is effective, we will not go ahead and say that we will not make claims like that catalyst B is effective as compared to A.

So, that is this type of testing like a drug is effective or not like that, so with the original drug.

We do the testing and we calculate what is the effectiveness of the drug with some measures like that, and with a second drug which is the new drug that you have developed whether it is giving you better efficiency as compared to drug a like that that can be tested using this type of two-sample t-test like that ok. So, what is required when I am doing a two-sample t-test is that now all these assumptions needs to be satisfied over here. All these assumptions needs to be satisfied.

(Refer Slide Time: 16:17)

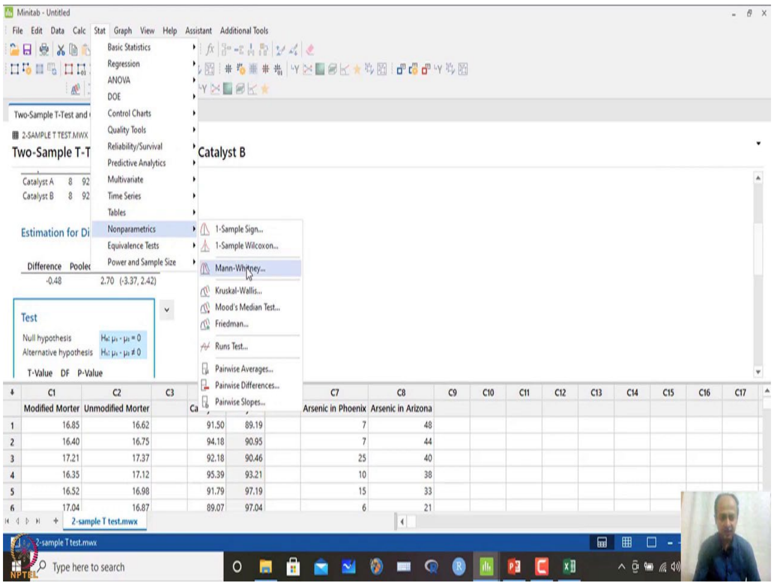


And, but if this assumption fails, one option is that conversion of the data set like what we have mentioned like that. We convert it to normality and then do the testing. And both the data set has to be converted. So, in that case, you have to keep in mind like that because to make a fair comparison between the 2 data set like that ok. So, and another option is that another option is that non-parametric testing.

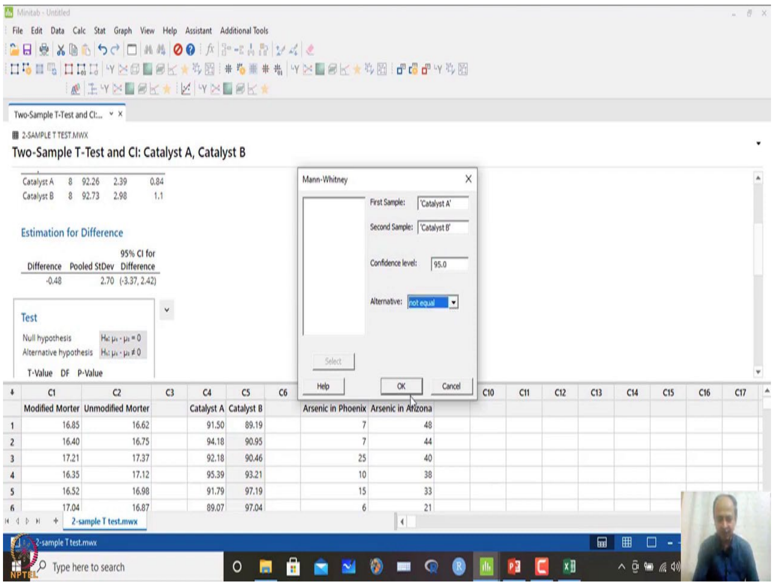
So, I have a non-parametric option which is known as Mann-Whitney test in case you do not want to go about it because the assumption fails. So, I want to assure big. But this test are so robust I can assure you that even if some deviation happens small deviation to moderate deviation happens.

In that case, the conclusion will be more or less same with non-parametric like that. So, I will use a Mann-Whitney test which is provided by MINITAB and that is the recommended one when the assumption fails. And you do not want to go with the assumption. So, in that case, I will go directly to non-parametric testing over here.

(Refer Slide Time: 17:19)

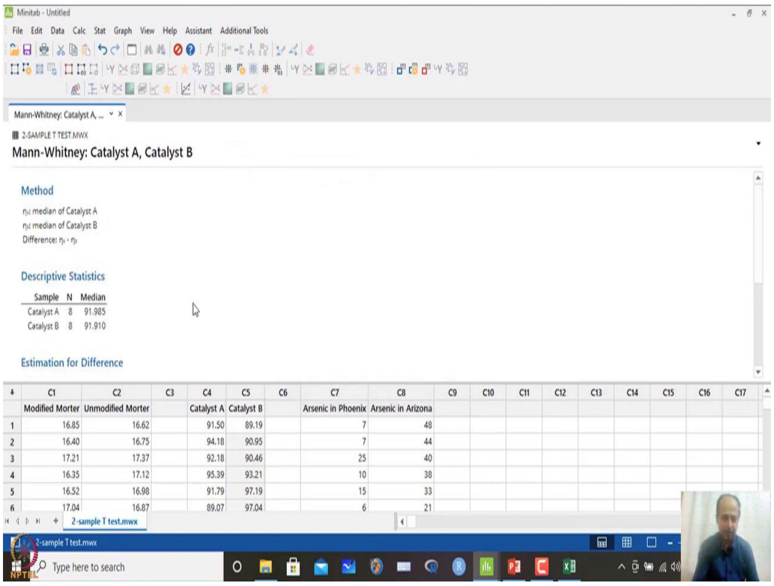


(Refer Slide Time: 17:20)

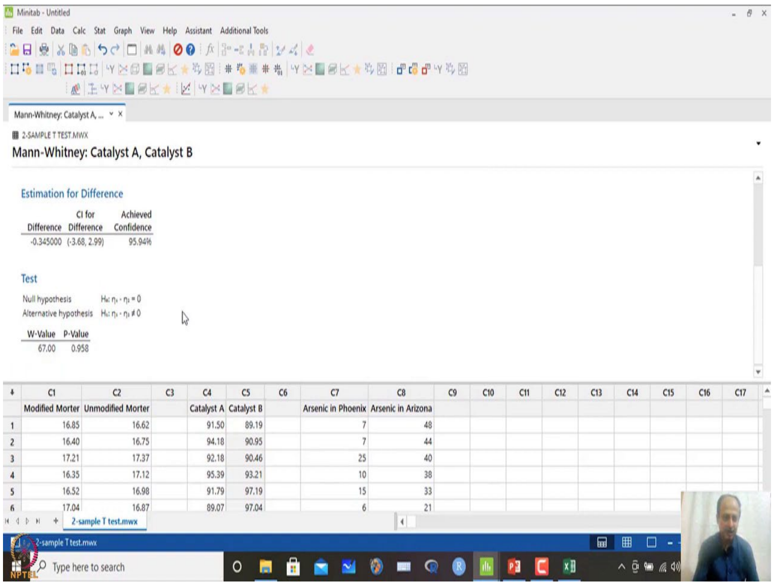


So, Mann-Whitney test I will do. And in this case, median again, the median whenever I am going for non-parametric median value is recommended as compared to mean over here. So, first sample is in catalyst A and second sample is catalyst B. Let us assume that our distribution fails, assumption fails.

(Refer Slide Time: 17:39)

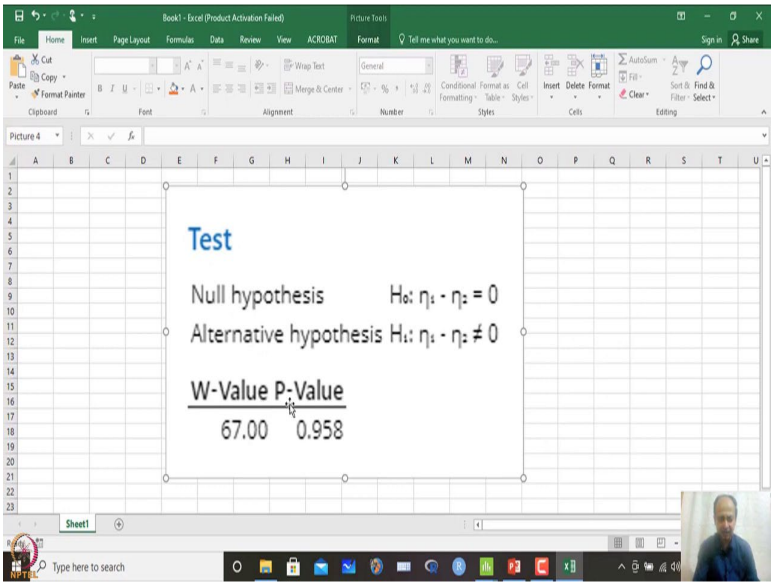


(Refer Slide Time: 17:40)



And I am taking not equals to condition over here. So, I am doing and if I click ok over here. So, what will happen is that you will get a W statistics over here and you will get a value of p-value over here, so that will be reported like that. Here also p-value is reported. So, here also we will get some p-values like that ok.

(Refer Slide Time: 17:50)



W is the statistic that is Mann-Whitney statistics that we are getting over here. And this p-value indicates that p is more than 0.05. So, indicates that medians are not different over here. So, these are the difference of medians like that; medians are not different over here.

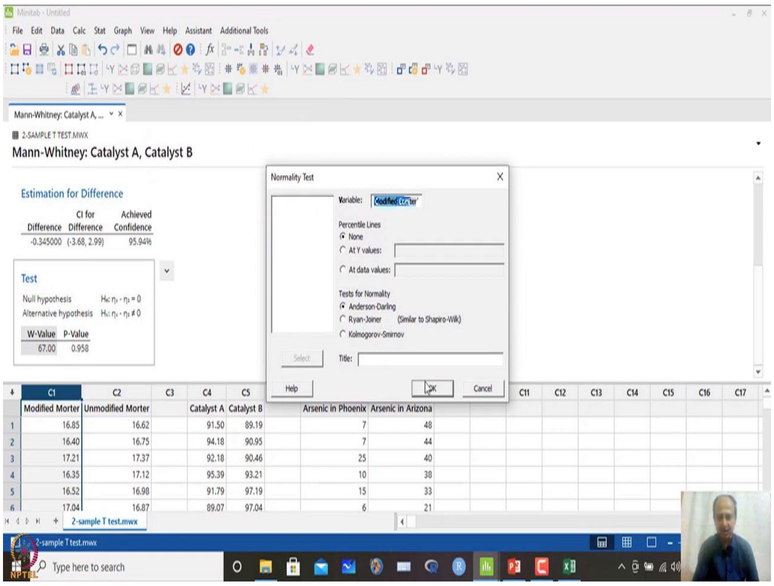
So, basically conclusion is that both the drugs are equally effective like that. There is no one cannot give higher yield as compared to the other one. So, percentage yield improvement is not much significant statistically significant what we have done. So, catalyst A yield is same as in population is same as catalyst B.

So, with one sample information, let us say over here what we have. So, we have 1, 2, 3, 4, 5, 6, 7, 8 observations like that for a given catalyst, and another eight observations on this side. So, at the degree of freedom, you see 14. So, $n_1 + n_2 - 2$ that will be 14 , so that degree of freedom is used over here; and the corresponding p-value says that there is no difference between catalyst A and catalyst B like that ok.

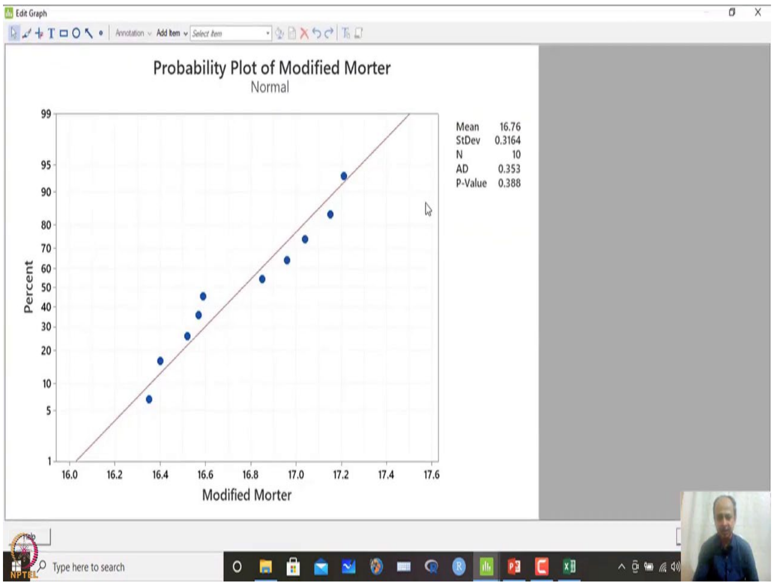
So, we can have different examples over here like this is one example taken again from Montgomery's Design of Experiment Book, where some cement formulation is checked. So, one is modified one and one is unmodified. Unmodified is the original formulation; modifier is the new formulation of cement that is there. And we want to check and do the two-sample t-test and want to confirm that whether they are different or whether they are same like that.

So, what we will do is that again we will do the basic statistics testing over here. So, I want to check two-sided test let us say. So, in this case, what I will do is that first I will check the normality. So, I will check for modified morter, is it normal. So, and the data set says that when I do the first testing over here, P-value is 0.388.

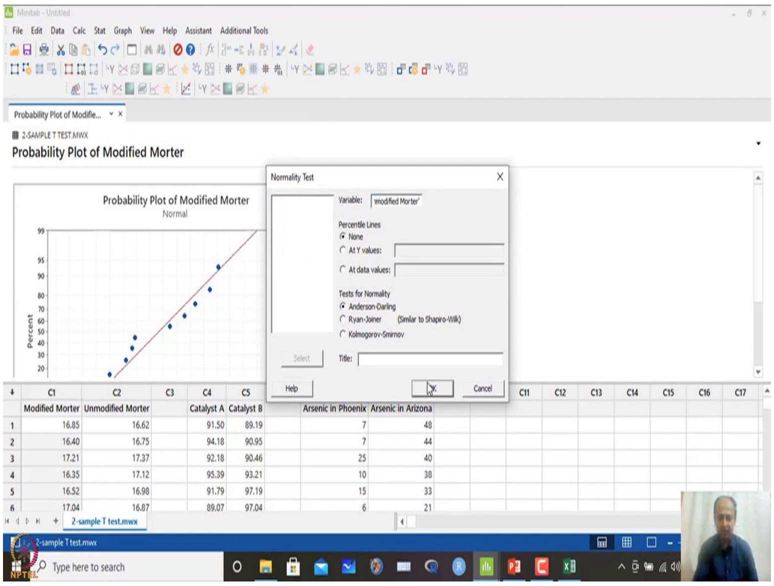
(Refer Slide Time: 19:28)



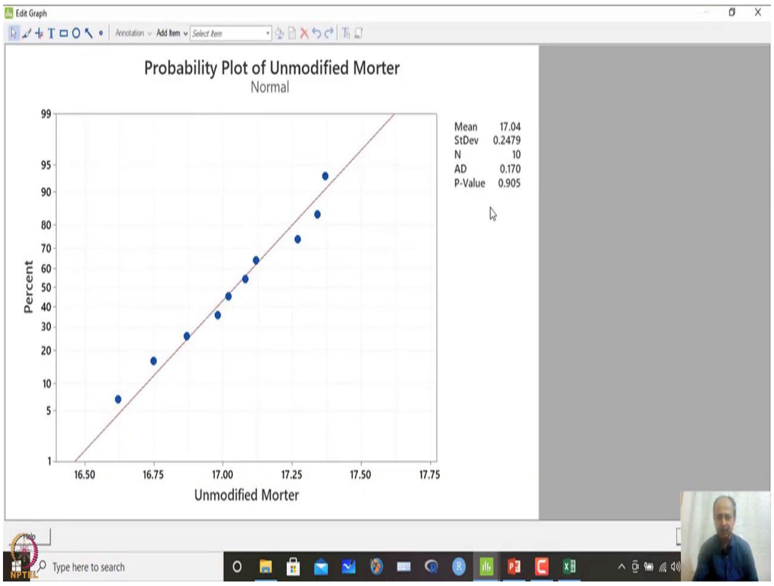
(Refer Slide Time: 19:32)



(Refer Slide Time: 19:45)

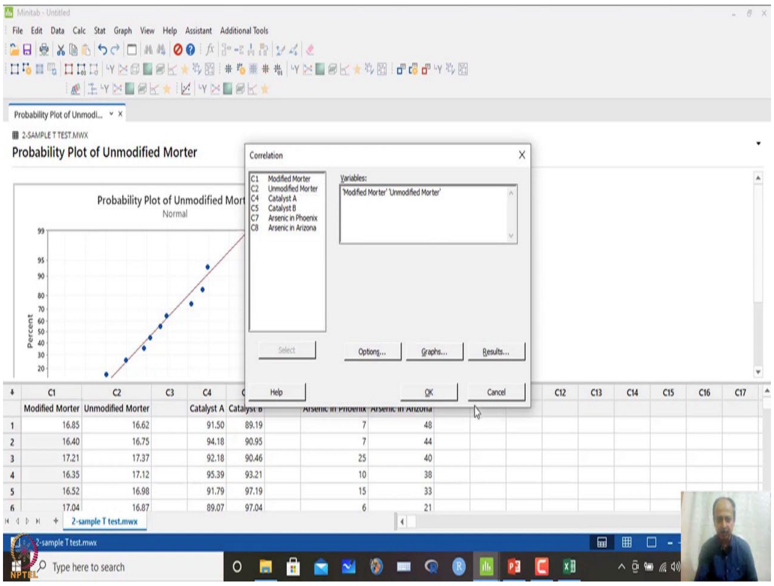


(Refer Slide Time: 19:46)

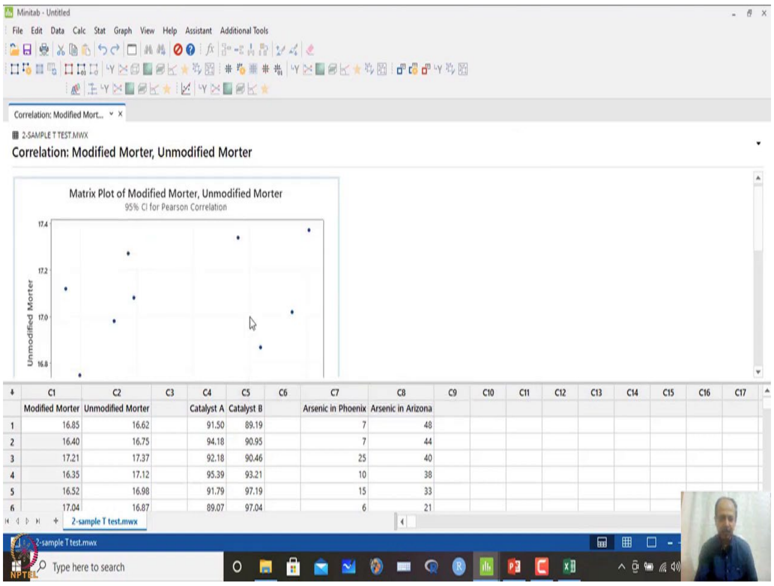


So, it is normal. So, second data set what I will do is that second we will do the same normality testing. And instead of modified, I will use unmodified over here and check the normality measures over here. And what I am observing over here p-value is 0.9 approximately that also indicates that unmodified morter data that we have is also in population will be normal. And then what we can do is that so first assumption is constant.

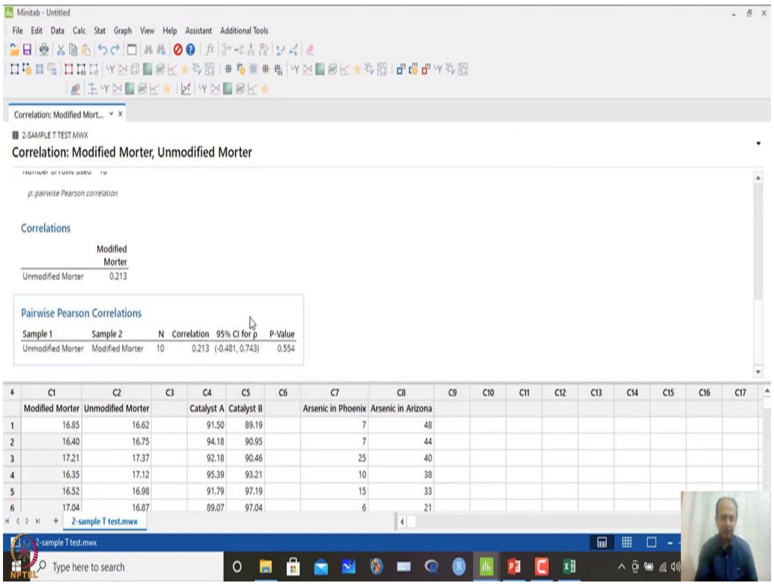
(Refer Slide Time: 20:06)



(Refer Slide Time: 20:10)



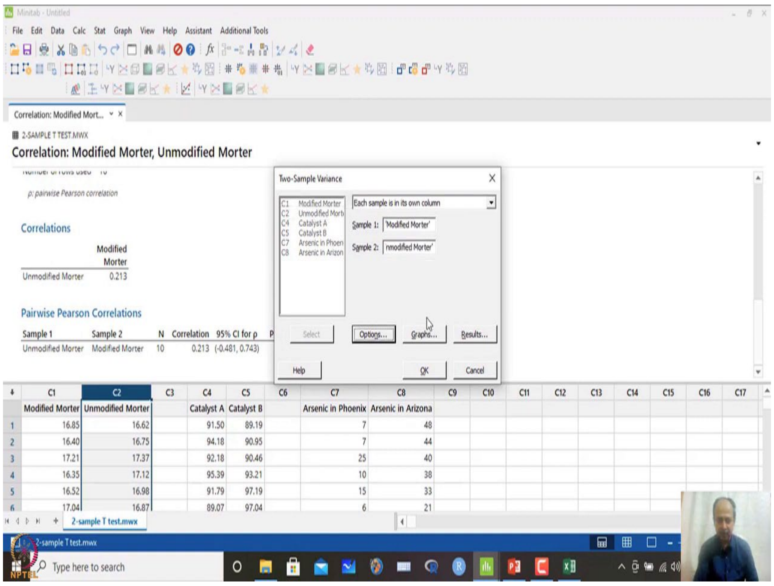
(Refer Slide Time: 20:11)



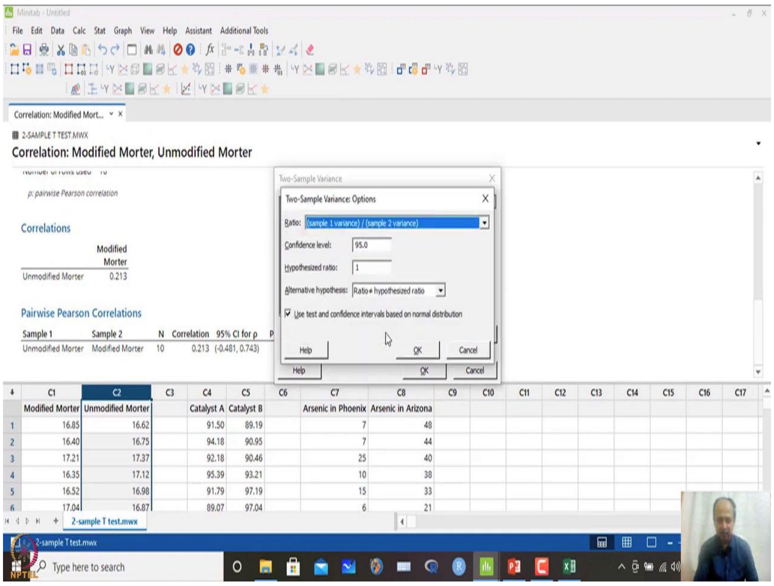
Then second assumption what we can do is that basic statistics in that case correlation we can check. So, modified one and unmodified one. So, I will click ok over here. And I will go the p-value that is coming over here. And the p-value observed is 0.554. So, I am not converting I am not taking the, so this also indicates that modified and unmodified the correlation is about 0.2. So, I told you that thumb rule is that approximately more than 0.7 will be significantly different significant correlation exist like that.

But we have a p-value testing for that. So, we have a p-value hypothesis testing measures over here. So, that will confirm that there is no correlation between the two data set that we are having modified and unmodified. They are independent data observation over here ok.

(Refer Slide Time: 20:53)



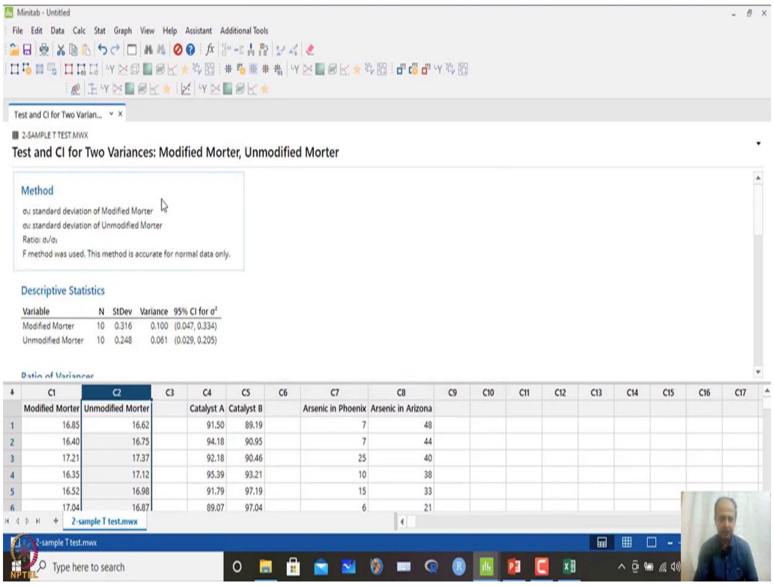
(Refer Slide Time: 20:58)



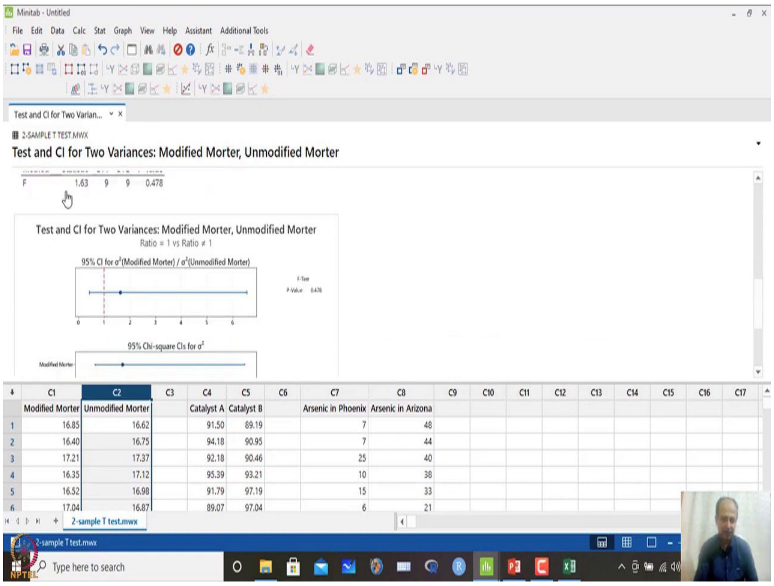
So, the final testing that we have to do is that whether they are same, whether the variance is same or not. So, I do a two variance test. And I will go for this 2 variance test over here. So, one is modified; one is unmodified.

So, I will go to options over here. 95 percent confidence level we will take over here. And this is sample 1 variance and sample 2 variance. So, I am doing two-sided test, and I will click this option because normally assumptions we are making.

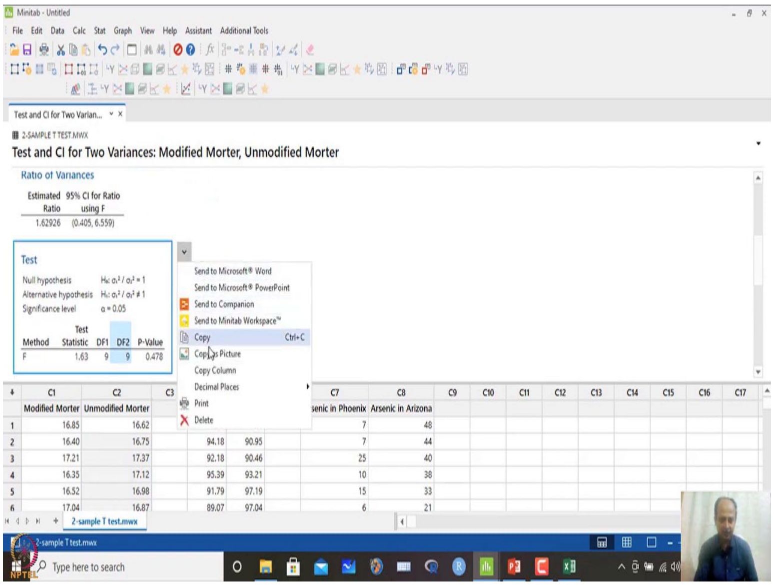
(Refer Slide Time: 21:14)



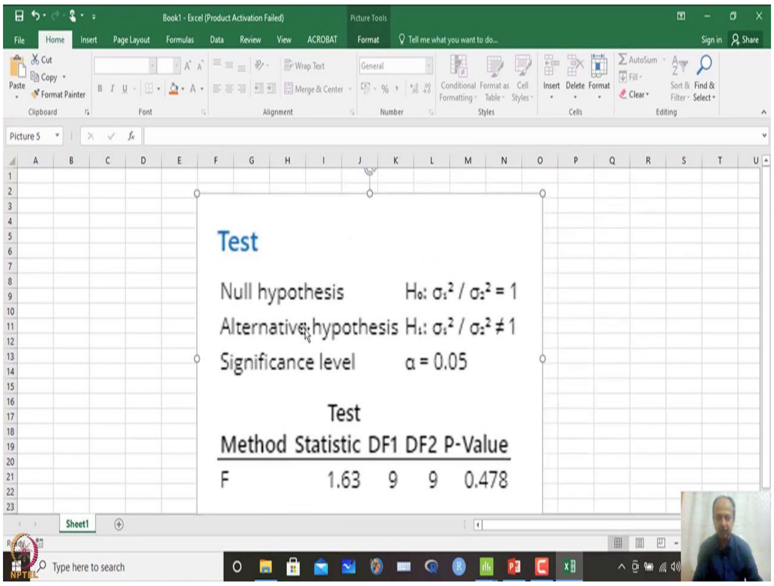
(Refer Slide Time: 21:16)



(Refer Slide Time: 21:20)



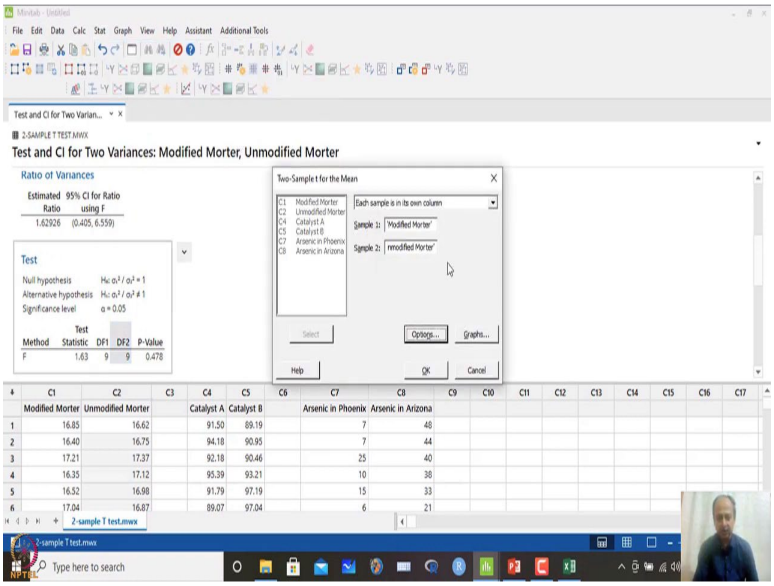
(Refer Slide Time: 21:27)



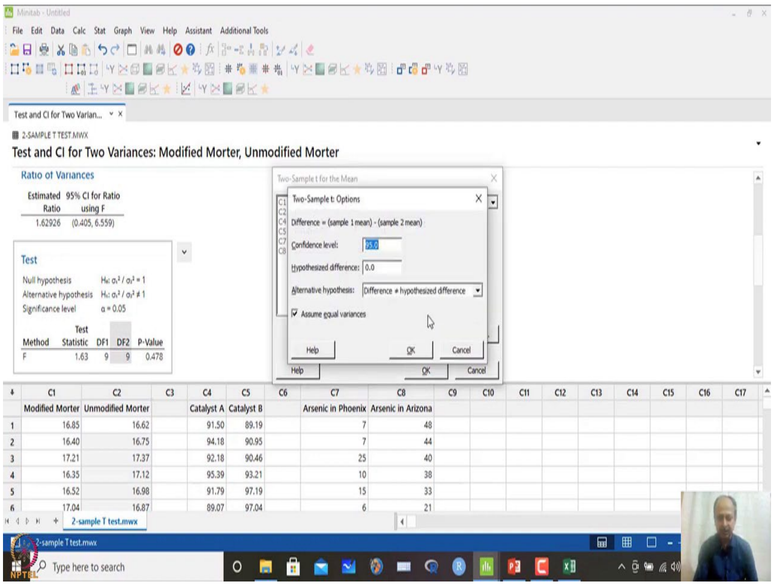
And we have also cross check that one. So, I will click ok. And then I will click ok over here. So, what will happen is that I will get a variance test over here, and F test indicates that there is no difference. So, F test if I can take this one, and I take it to excel. So, the values that we are getting over here is F test for the modified and unmodified cement, and that means, variance testing that we are doing for the two data set. So, p-value is 0.478.

And that indicates that basically the variance is same or the ratio of variance is close to 1, and that is not different from one like that. So, that can be confirmed from here. And so we can go, so every check is done over here. So, first test, second test and third is satisfactory.

(Refer Slide Time: 21:57)

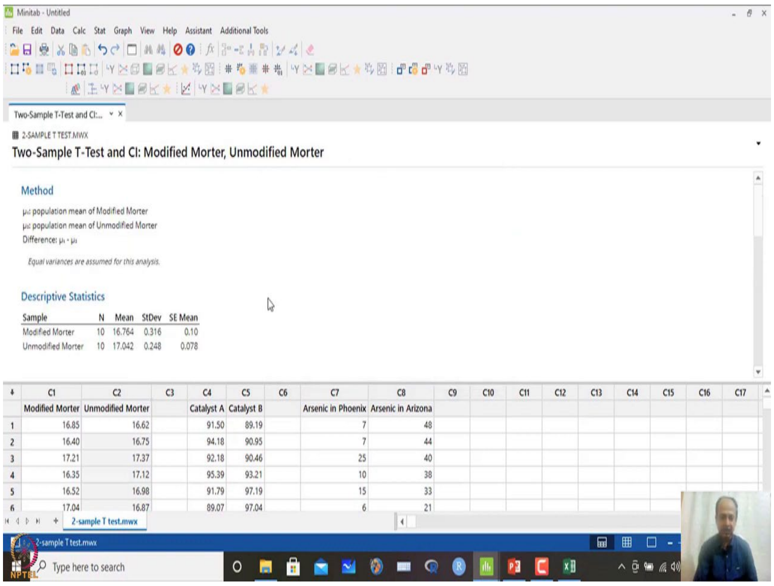


(Refer Slide Time: 22:04)

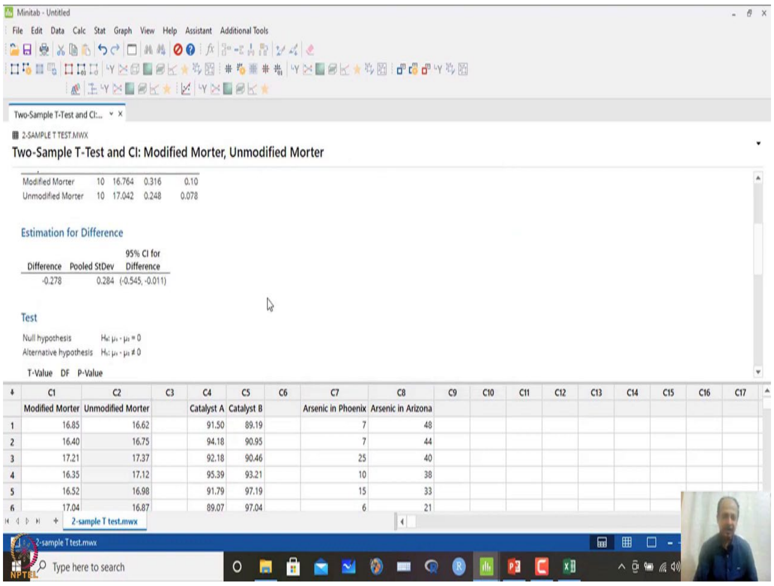


Then what we can do is that we go to basic statistics to sample t-test, and then one want to check whether modified cement is different from unmodified one. So, we can also do one-sided test like that. So, if I change the condition we can do which is higher than which one. So, that if it is different that can also be checked by the results itself. So, I, if I take equals to condition also we can do that so if they are significantly different like that.

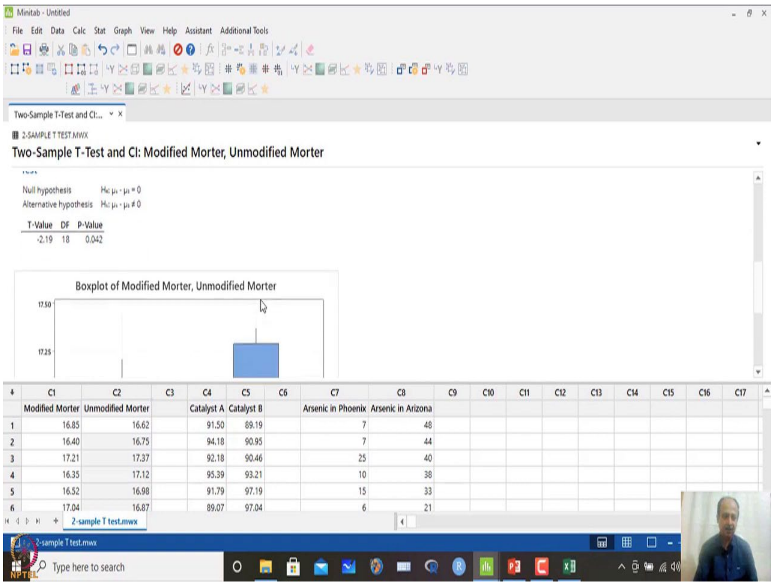
(Refer Slide Time: 22:18)



(Refer Slide Time: 22:20)

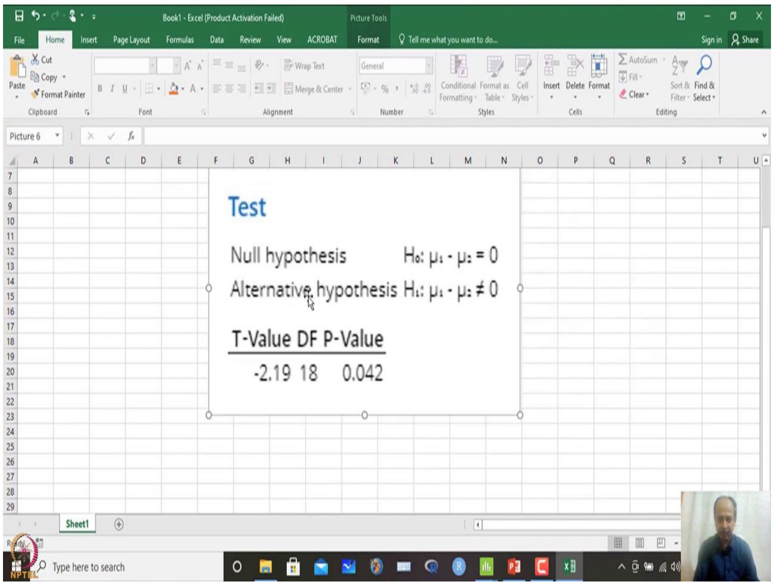


(Refer Slide Time: 22:21)



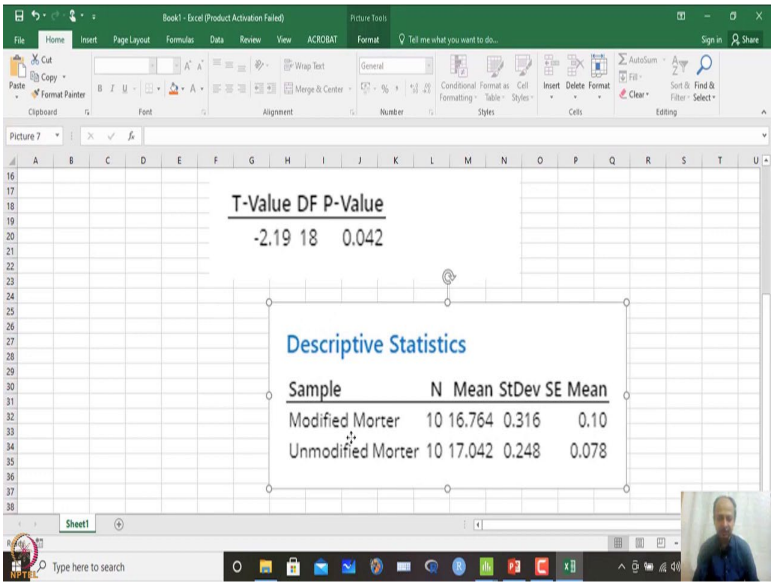
So, in this case, what we will do I will click ok. And after the test statistic, what we are getting over here. So, this can be seen same test statistics what we have got over here. So, this is catalyst 1 and catalyst 2. And for this what we are getting is that let us go back to this analysis over here. So, this I can copy this and place it over here.

(Refer Slide Time: 22:44)



So, let me place it over here. And we can see whether the modified and unmodified are different or not. So, here what we see is that p-value is 0.042, that means, what p-value is less than 0.05, that means, there is a difference between these two values that we are we are getting over here. So, that means, one mean is different from the other mean. So, modified morter and unmodified morter formulation. So, let us go to the mean which is different from which one.

(Refer Slide Time: 23:14)



So, let us go to this and we have confirmed that there is significant difference between these two. So, we can place it like that and paste it over here. And we can see enlarge this one and see which is different from which one. So, we have seen that P-value indicates that they are different.

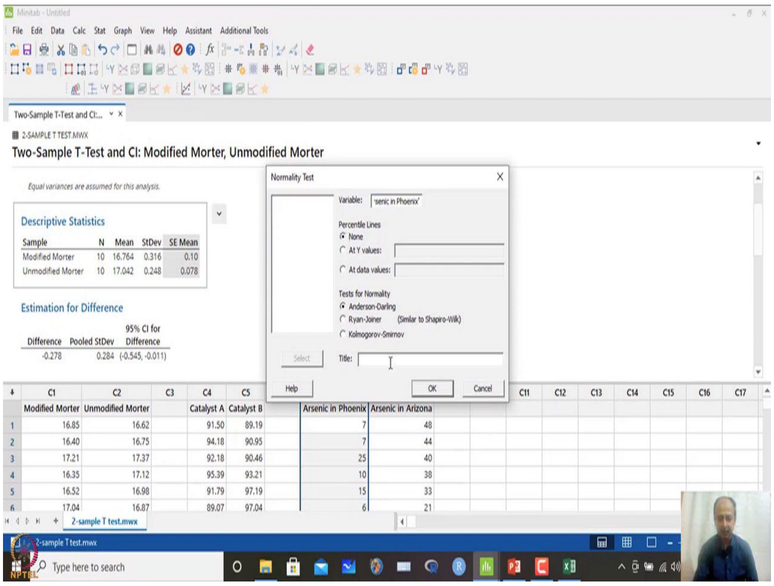
Now, modified morter is giving me a strength of 16.76, and unmodified is giving me a mean strength of 17.04 like that. So, unmodified is giving you higher strength as compared to modified morter over here, and they are significantly different statistically over here ok.

So, if you have to choose between modified morter and unmodified morter, So a tensile strength let us say tensile strength is measured over here which is the given condition. So, in this case whichever strength is higher, I will go for that.

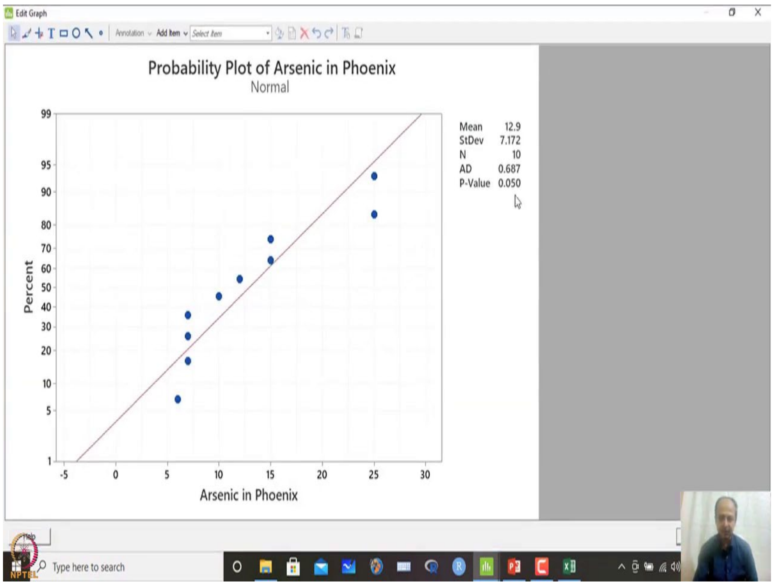
But the results indicate that unmodified is giving me average which is higher than the modified one. And they are statistically different basically what which was proved by this p-value over here. So, if you have to implement which formulation to be adopted, in this case, I will go for the unmodified or old formulation. I will not go for the modified formulation of the morter that is next one because that is giving me lower strength as compared to the unmodified morter.

So, in a population, at population level, what we can see is that unmodified one is giving me higher strength as compared to modified morter over here ok. So, that is the physical interpretation we can take out of this. We can take another example like what is given over here. Arsenic content in phoenix whether it is different from Arizona Arsenic content like that is another example where we can see whether which test to adopt like that.

(Refer Slide Time: 24:50)

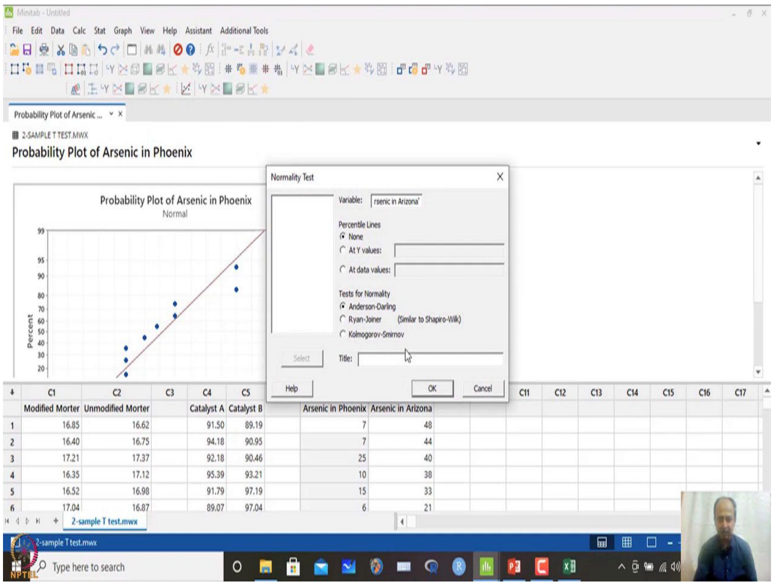


(Refer Slide Time: 24:53)

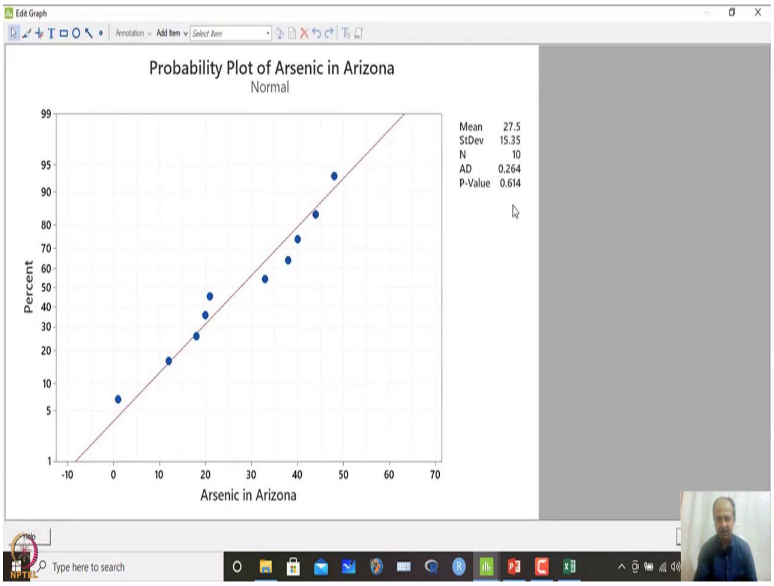


So, in this case, what we can see is that whether the basic assumptions holds true or not. So, in this case we will go for Phoenix and let us try to do the testing over here. So, it is just on the line that means, you see p-value is exactly equals to 0.05. So, we may consider that this is not greater than this one. So, this assumptions holds like that.

(Refer Slide Time: 25:14)

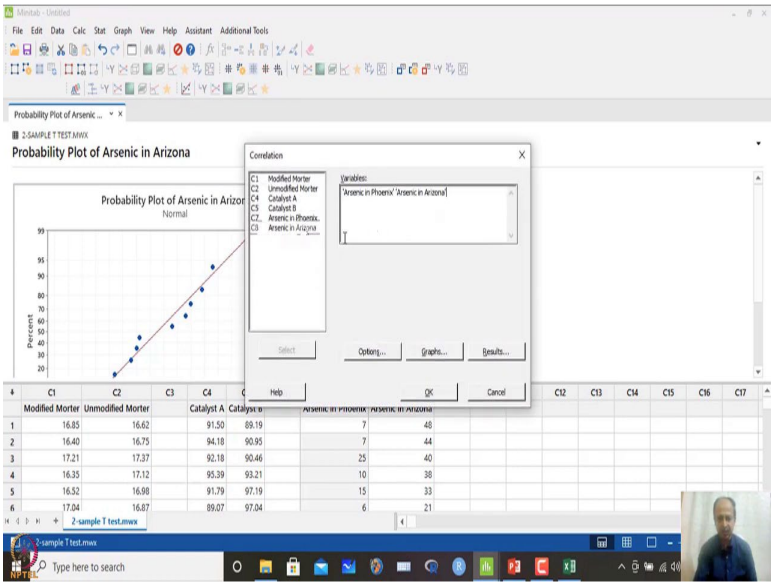


(Refer Slide Time: 25:18)

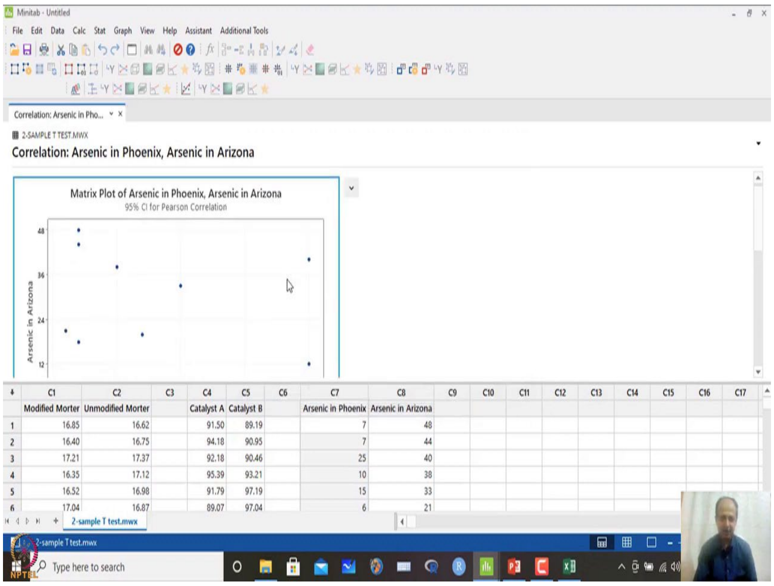


If I go to Arsenic in other condition like that, so I will go to basic status normal normality test. And in this case, I will take the Arsenic in Arizona and try to do this test over here. So, here also we see that the conditions are satisfactory over here. And we can assume that normality assumption is not violated as such.

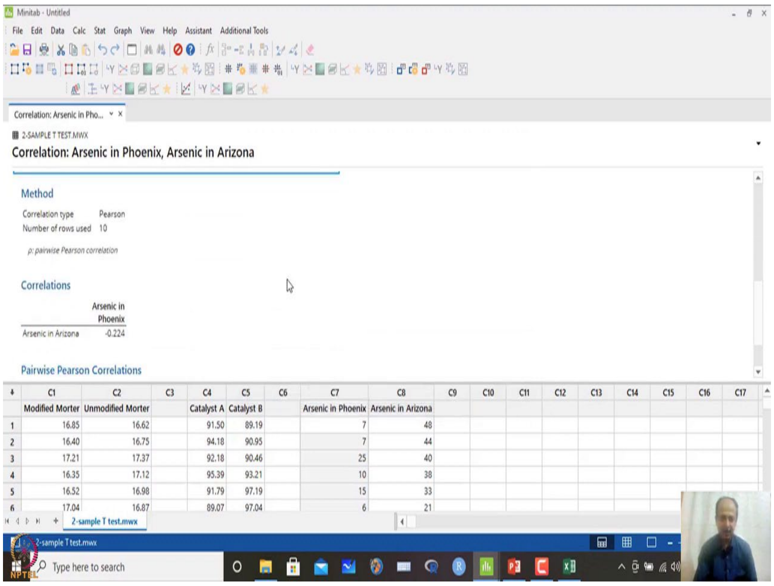
(Refer Slide Time: 25:31)



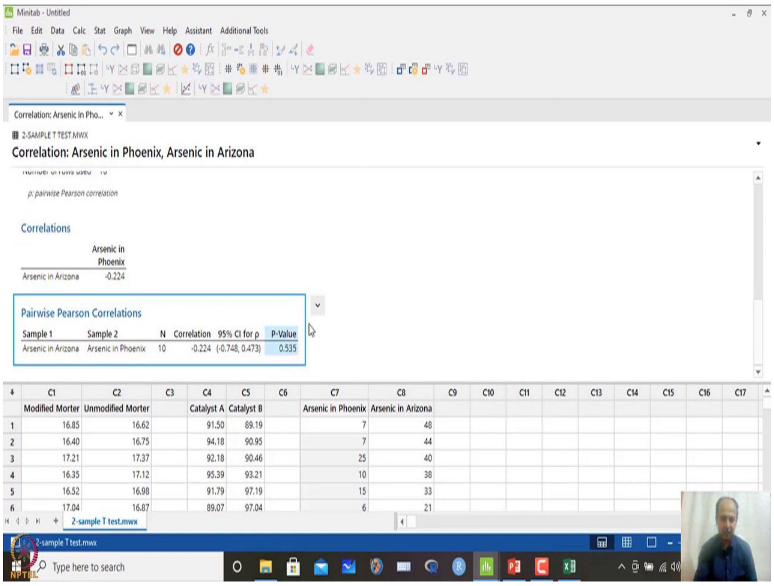
(Refer Slide Time: 25:38)



(Refer Slide Time: 25:39)



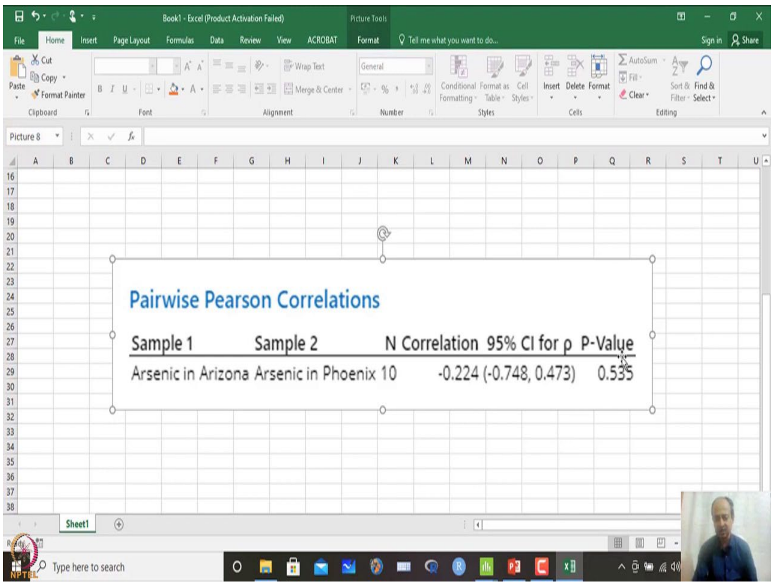
(Refer Slide Time: 25:40)



So, then we go to the basic statistics and correlation we can check between these 2 data set then we can test the p-values over here.

And what we see over here p-value is also not significant, that means, there is no high correlation between these 2 data set; that means, they are independent data sets.

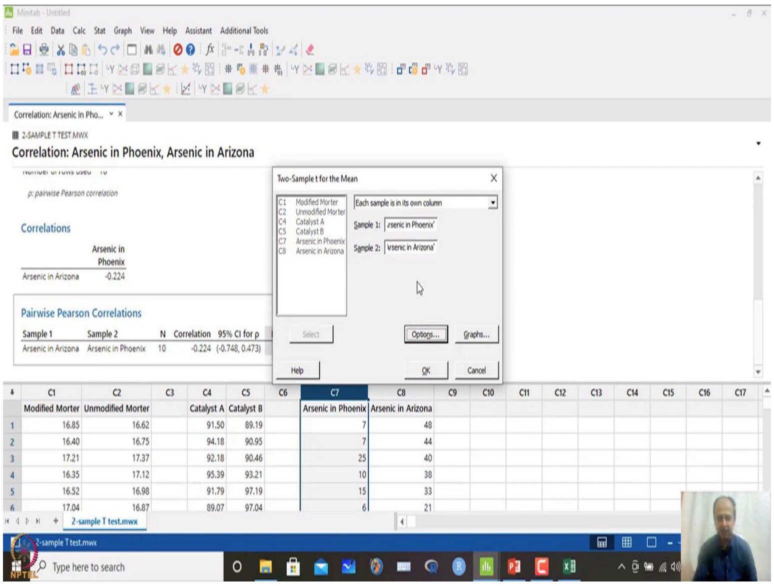
(Refer Slide Time: 25:58)



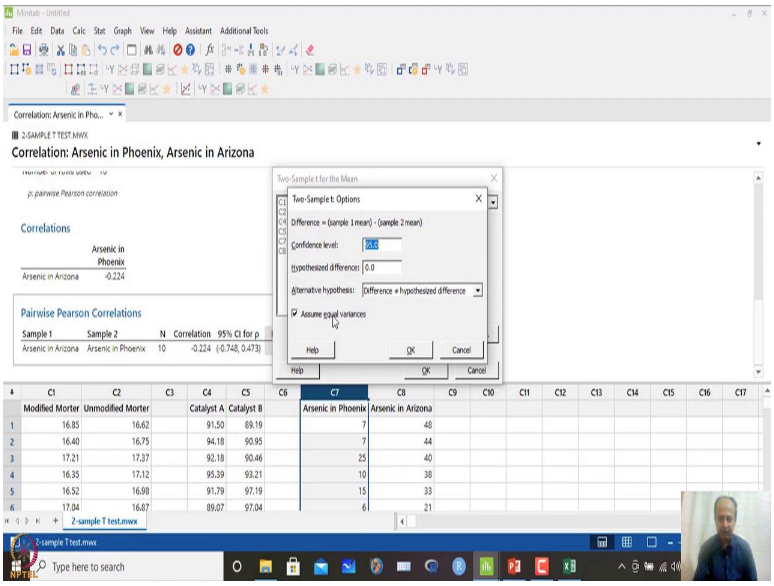
I see the Pearson correlation which is coming out to be 0.535. So, this paired Pearson correlation is coming out to be 0.535. p-value is more than 0.05. So, nothing significant correlation that. And it is correlation you can see is -0.224.

And anything more than 0.7, I told should be significant, but here -0.224. So, in this case, it should not come out to be significant over here. So, that is what is reflected over here in the p-value. So, 0.535 that is the value we are getting.

(Refer Slide Time: 26:49)

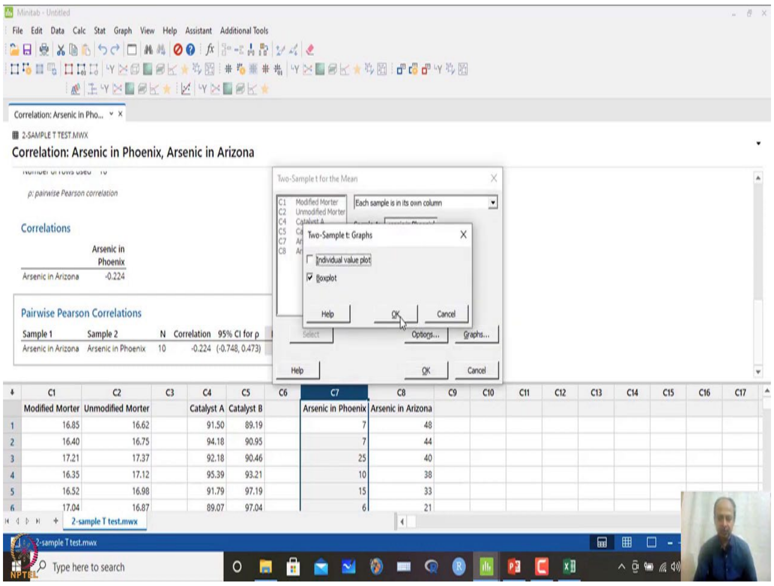


(Refer Slide Time: 26:57)

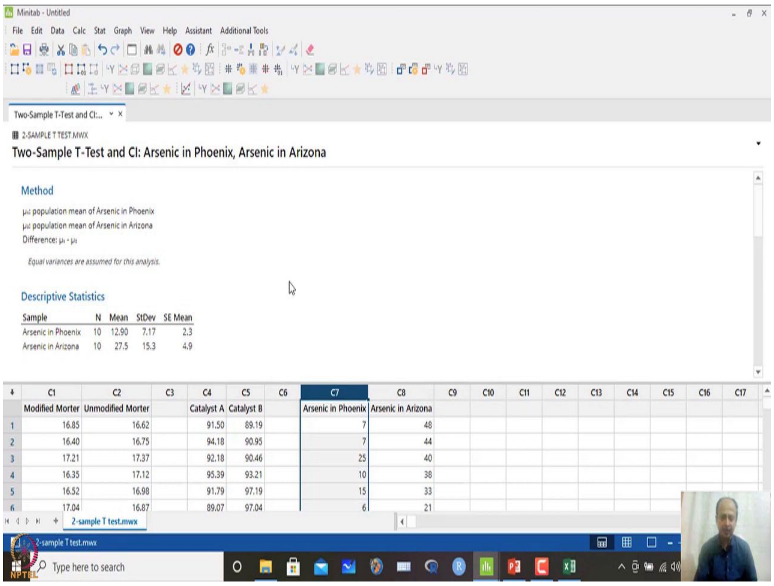


So, data sets are independent. So, in that case, the final test what we have to do is that all checks we have done. So, two-sample t-test we will do. And in this case only data set I will change. So, I will just take Arsenic, and then second one Arsenic is Arizona

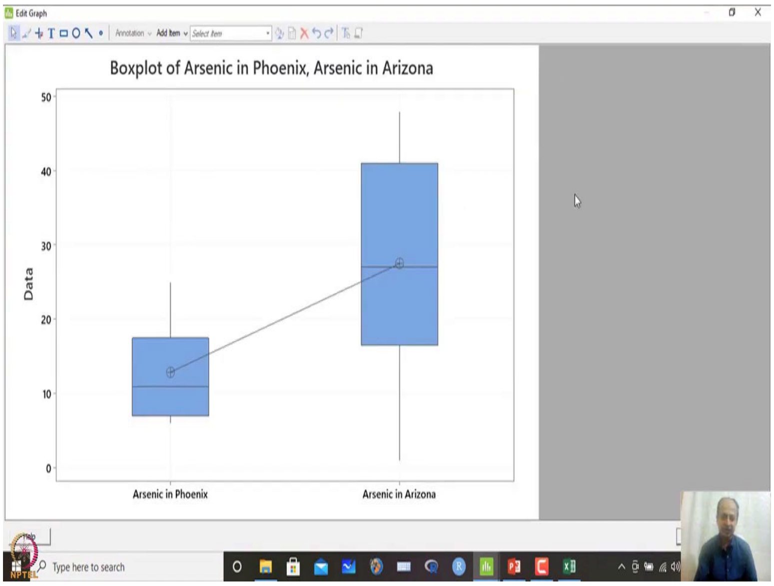
(Refer Slide Time: 27:09)



(Refer Slide Time: 27:13)



(Refer Slide Time: 27:15)

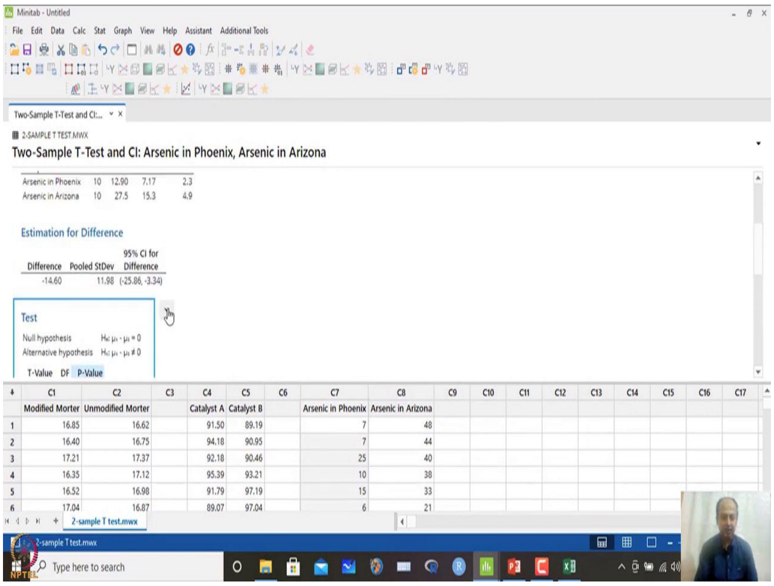


So, in options what we can do is that assume equal when that condition is satisfactory 95 percent that I have taken. The difference should be equals to 0 and not equals to condition will be the what I want to test. And graphically we can see the box plot. And I click ok over here all that. And in this case what happens is that you can see the box plot over here which will give you some idea of the data set.

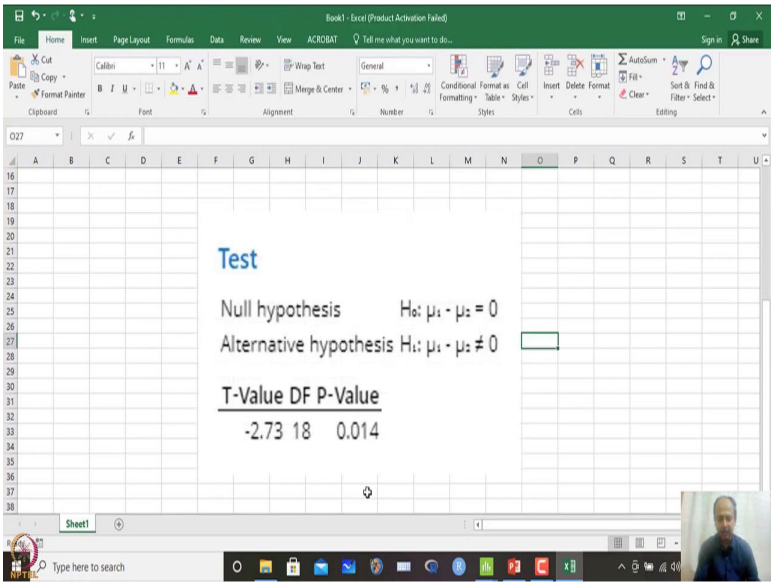
So, in this case what you see this is the average value, this is the average value what you get. So, mean value is 12.9 over here. And the mean over here what you see is around 27.5. So, there is a huge difference of slope you can understand that there is a huge slope between these 2 values over here. So, this should be prominent, that means, there is a significant difference between location of the mean of this and this over here. And the median value is also we can see a significant difference exist between these two.

So, our hypothesis testing should be able to identify this difference and it should reflect that there is a significant difference between the Arsenic content over here as compared to the Arizona content that we are getting. Box plot also reflects that fact.

(Refer Slide Time: 28:10)



(Refer Slide Time: 28:19)



So, to prove that one, we have to go to the what is the value that we have got in hypothesis testing p-values over here. So, I am just copy pasting the p-value analysis or final hypothesis testing that we have results that we have got. And I will paste it over here.

And I will just enhance this one, so that you can see. And what you see over here is the p-value is 0.014, that is less than 0.05. That indicates that and the T-value is around -2.73, and this is not close to 0, this is not close to 0.

So, in that case, our basic interpretation this p-value is less than 0.05. So, in that case what we can assure is that this two values are statistically different these two values are statistically different and in that case you need to be cautious when it is content of Arsenic. So, Arsenic content over here you see 27. So, Arizona it is about 27 mean average value. So, it is much higher as compared to the as compared to Phoenix. So, there is a statistical difference that exists between these two ok.

So, this is what we wanted to emphasize this is two-sample t-test where I can just compare before improvement and after improvement. So, before we have implemented some measures before we have done anything. So, what, what is the existing scenario? So, in quality what happens is that we try to see existing scenario. And then we do some improvement and then prove that this improvement has real effects and that is statistically different like that.

So, first phase of experimentation is that at one condition is it different from a second condition? So, to prove improvement happened or not happened, what required is that we need to do this type of hypothesis testing. And two-sample t-test is the most important hypothesis testing which is used in quality basically.

But that may be the starting point, but we need to know two-sample t-test and there are many other lectures where you can see two-sample t-test like that. And this is the way we do it in MINITAB, and the interpretation I have already told. So we will discuss about paired t-test also after this which is also very relevant where to use to sample t-test, where to use paired t-test like that that difference we should know and based on that.

We can go ahead with the experimentation where we do real improvements like that. So, we will discuss about analysis of variance concept like that which is the fundamental or pillars of design of experiments like that ok. We will stop over here and we will continue with paired t-test in our next lecture ok.

Thank you.