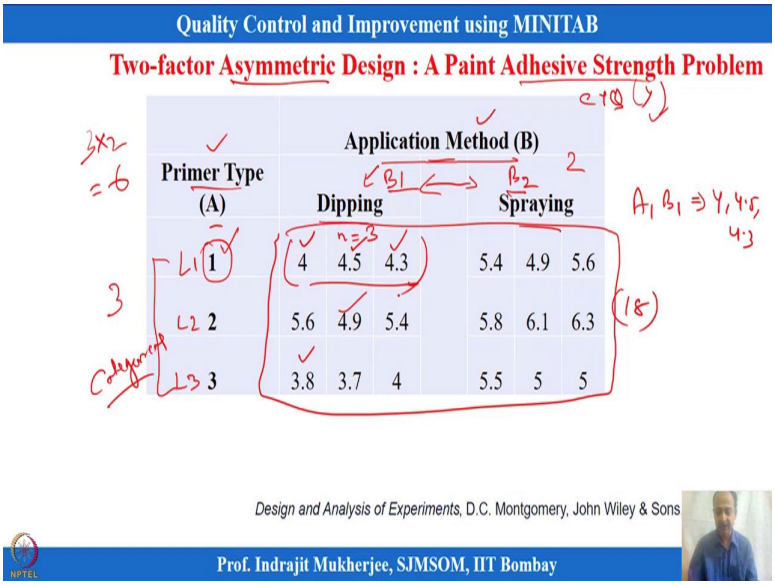
Quality Control and Improvement with MINITAB
Prof. Indrajit Mukherjee
Shailesh J. Mehta School of Management
Indian Institute of Technology, Bombay

Lecture - 29
Two-factor asymmetric Design, Symmetric Factorial Design, Two-way ANOVA

Hello and welcome to our session 29 on Quality Control and Improvement with MINITAB. I am Professor Indrajit Mukherjee from Shailesh J Mehta School of Management, IIT Bombay.

(Refer Slide Time: 00:35)



So, we are discussing about experimentation with two factors and we have taken one example, where we want to maximize the adhesive strength and that example is taken from design and analysis of experiment by Montgomery. And the experiment is a combination of two factors over here, one is primer type that is factor A and another one is application method that is factor B. Earlier we were dealing with one factor and changing that factor at different levels, what we have seen one way analysis of variance.

Now, we will adopt two way analysis of variance because primarily there are two factors; primer type and application method. Primer types are at 3 levels and application methods are at 2 level. One is dipping method. So, if you go to a paint shop, you will find that there are different ways to do paintings. One is dipping and another is spraying; this is the condition that is given.

Two options the industry is having and the three different primer types are generally taken from suppliers, type 1, 2 and 3. So, the experimenter want to figure out that what is the best combination of primer type and application method that will give me will maximize my adhesive strength, which is the CTQ.

So, experiment was conducted like this way. Primer type (A) has 3 levels and application method (B) has 3 levels. So, total 6 combinations are possible, A1-B1, A1-B2, A2-B1, A2-B2, A3-B1, A3-B3.

So, an each of this combination each of this combination is run 3 times over here. So, combination A1-B1 we can think and for this combination we have taken 3 observations for this, 4, 4.5, 4.3 and all with different samples. So, replicates (n) equal to 3, what we can think of in this experimentation.

And we wanted to also ensure the randomization aspects of that in experimentation and while taking this data what was followed. So, randomly we create either B1 or B2 and primer type A1 and A2 like that and I have number of samples. So, this observation may have may be the first observation, but this may be the second observation; this may be the second observation, again this may be the third observation or experiment that was run like that. So, everything is randomized; the total data that is generated over here.

So, you can see $6 \times 3 = 18$, 18 data points are generated over here and completely randomized over here. So, that we do not know if there is third factor hidden factor. So, to so, to minimize the effect of the hidden factors like that, we have randomized the experimentation over here.

This is the CTQ or adhesive strength that was measured over here 4, 4.5 and 4.3 like this. So, all these are measurements of the adhesive strength and this is the complete experimental setups of all possible combination. This is known as asymmetric design. Because one is at 3 level 1 is at 2 levels like that.

When the levels are equal in that case it becomes a symmetric design. Here what we are looking at all possible all possible combination for different levels of A and B.

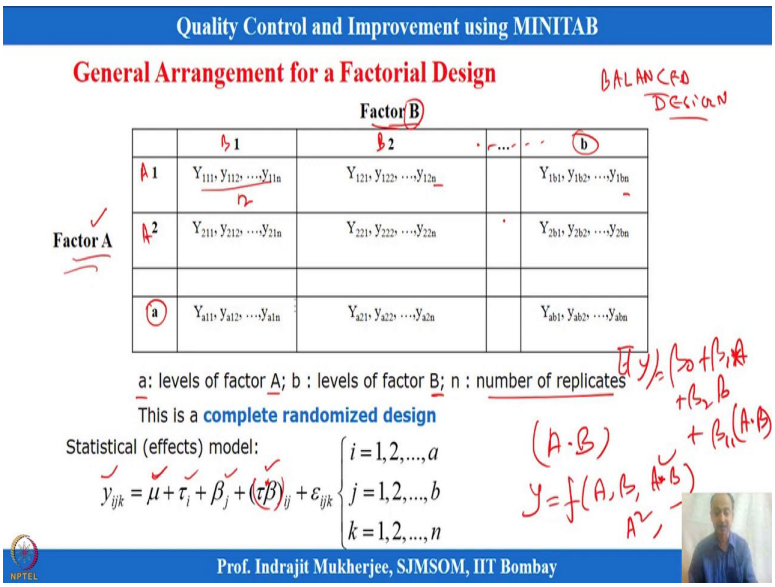
So, this is asymmetric design and we will try to see how to analyse asymmetric design. And also please remember that this adhesive force over here, which is y characteristics

which is a continuous variable over here, and the other factors this can be categorical this can be assumed to be categorical; that means, there is no sequence we can place them so, any of the primer types 1, 2, 3. So, this is a categorical variable also an application method is also a categorical variable that you can see.

So, factors that we have selected is a categorical variable, but the response that is coming out of the process; that means, the adhesive strength is a continuous variable and you have to remember that in design of experiment, y should be continuous. Then we can do this analysis of variance.

Factors can be continuous also and factors can be mixture of continuous variable and categorical variable that is also possible. Here we have two variables, which are categorically specifically primer type and application method. Application methods has 2 levels; dipping and spraying. Primer type has 3 options; we are having primer type 1, 2, 3.

(Refer Slide Time: 05:39)



So, this is the general tabular we collect the data. So, we have seen A1-B1 combination, B1-A2 combination. Like this we can generalize up to B levels; we can have b levels of factor B and a levels of factor A over here. So, this can be generalized, we can write a generalized form for data we are collecting.

And each is having n replicates over here. This is a balanced design. Theory supports that we should balance the design as it gives you better estimation.

That is one of one of the assumptions that generally people try to do in design of experiment. Number of samples that is taken for any combination is generally equal. Although analysis can be done if it is not equal, but we are assuming a balanced design over here.

So, in this case factor A has total a levels and factor B has total b levels and n is the number of replicates that we have considered. So, mathematical model that is for each of this variable is written as

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

y is modelled with overall mean over here and the effect of factor A and effect of factor B and also we have taken another combination of this, because in single factor this is not coming. Whenever more than one factor what will happen is that there is a possibility of interaction, which is known as A multiplied by B over here.

So, sometimes and to understand interaction what we can simply think of that when we write a function, let us say A and B are continuous variable over here. So, in this case y can be a function of A, y can be a function of B, y can also be function of A multiplied by B. So, this is the or when we write polynomial equations like that.

Curvature in the in the response surface is because of this interaction; if it is present at all. So, in this case or higher order terminologies it can be a square also in the models. So, when we develop the regression equation:

$$y = \beta_0 + \beta_1 A + \beta_2 B + \beta_{12} AB$$

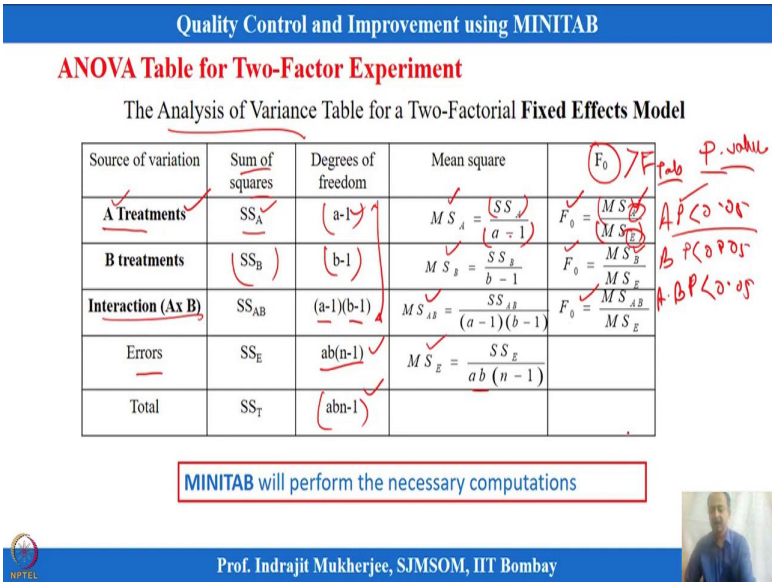
In simple terminology, we can think of that as an interaction over here. So, in the models it has to be considered. Now, we have to check whether the interaction is significant or not and based on that only, best combination of A and B can be derived ok. MINITAB gives you option to estimate the interactions also.

So, effect of factor A on the CTQ, effect of factor B on the mean of CTQ and what is the effect of interactions of A B on the expected value of y or CTQs we want to analyze.

So, all these 3 components can be estimated if you have done all combinations; if you have taken all combinations of A and B that is possible and MINITAB gives you a estimation of each of these factor effects clearly. And ANOVA analysis will also show, what is the effect of A whether it is significant or not B is significant.

When I change the level of A whether it is influencing the expected value of y; when I change the level of B whether is it impacting the value of expected value of y or when I change both is it impacting the expected value of y.

(Refer Slide Time: 09:29)



So, these are the general terminologies and you can see books to see the derivations of this.

This is 2 factor ANOVA, we can think of analysis of variance table. So, in this case, when we have done all combinations and we can estimate the effect of A and that is source of variation; that means. So, sum of square of variation can be calculated which is SS_A.

As we have a levels of A so, $a-1$ is degree of freedom. B has b levels, so $b-1$ is the degree of freedom and interaction effects what we are talking about A multiplied by B has $(a-1)(b-1)$ degrees of freedom.

Then total degree of freedom will be total number of observations that we have taken minus 1 and then error degree of freedom can be subtraction of this minus all of this. So, if we subtract this one, we will get this formulation $abn-1$

So, I can calculate the mean square of A, mean square of B and I can also calculate mean square of AB and mean square of error can also be calculated based on the degree of freedom that is that already we have, but it requires some degree of freedom. So, I need error degree of freedom also to calculate mean square error over here. So, it cannot be 0 or we do not have.

So, based on which we can derive all this mean square and then what we can calculate is F values of for this effect of A. Effect of A can be calculated by effect of A whether it is significant or not how do we check that?

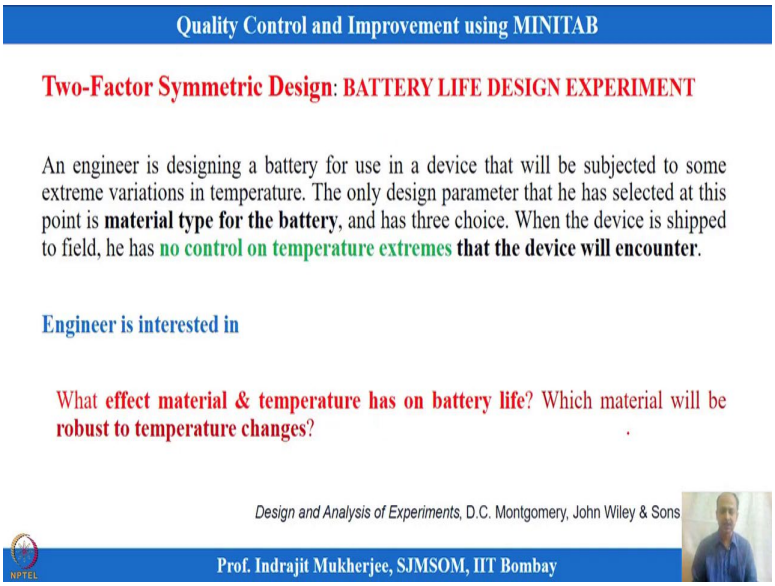
We take the mean square error divided by means divided by mean square error mean square factor A divided by mean square error over here; will give me a F value. Similarly for B also we get a F value and similarly for AB interaction, we get a F value. Then this F value will be will be whether this is greater than tabulated value.

So, F_0 whether it is greater than this one, this will generate the p values over here and we can also calculate the p values. So, p values will be generated by MINITAB and if p is less than 0.5, what we will say is that factor a influences the expected, when I change the factor A it is influencing the expected value of CTQ like that.

So at least there is 2 levels when I change from one level to the other that is the interpretation in one way analysis also. Similarly for B also we can calculate, we can see whether the p value is less than 0.5.

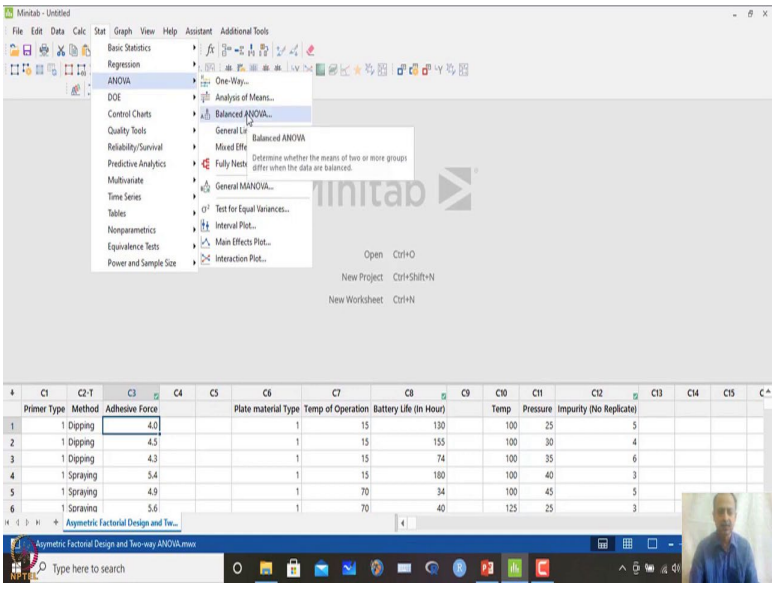
So, everywhere p value can be calculated and MINITAB will do it automatically for you and to say that whether A is significant, B is significant or AB is significant like that. So, we can interpret that way and MINITAB does this complete calculation. If you fit the data set and it will give you all information's like that ok. So, that is the interpretation.

(Refer Slide Time: 12:36)



So, let us try to see the examples that we have taken and try to see how to analyse the data and represent the data in MINITAB. So, I am taking the same examples that with primer type and dipping and spraying method.

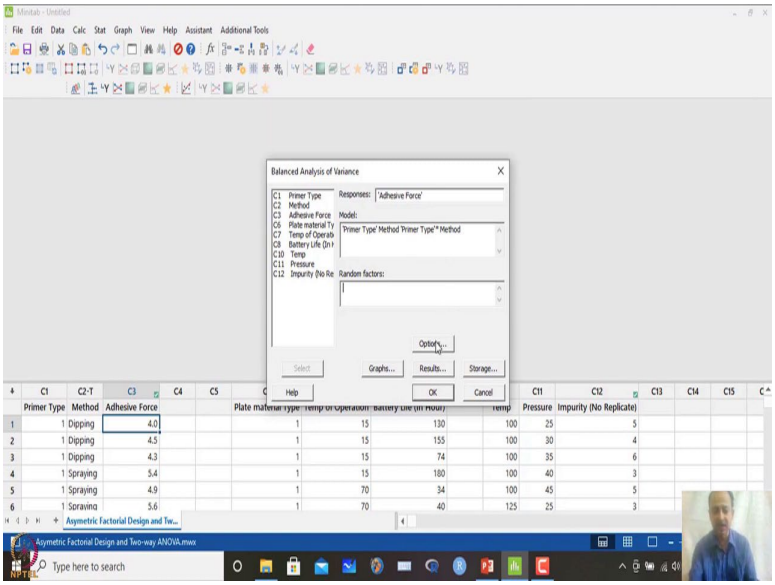
(Refer Slide Time: 12:54)



So, the data is located over here in C1, C2 and C3 column. So, primer type methods and adhesive force over here and this experiment was carried out. So, how to analyse this one? What we have to do is that, we have to go to *stat* and then go to *ANOVA* and there

is a *balanced ANOVA* information over here. There is a balanced ANOVA information over here.

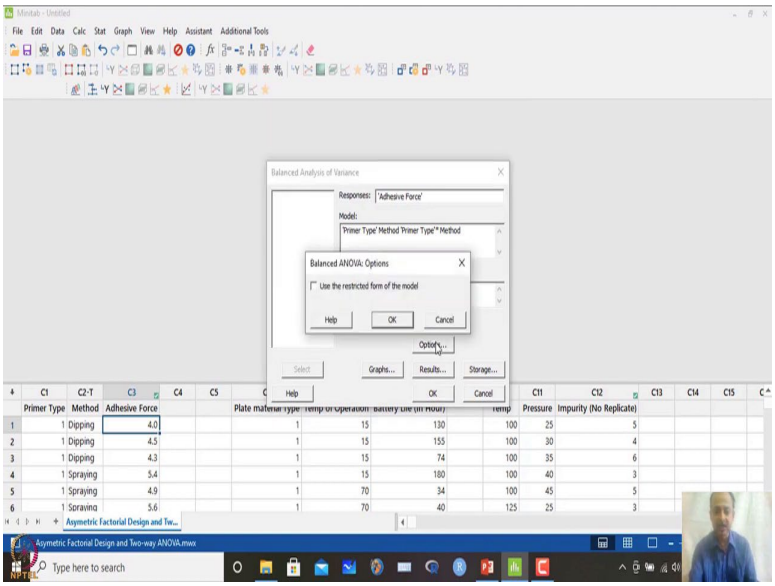
(Refer Slide Time: 13:15)



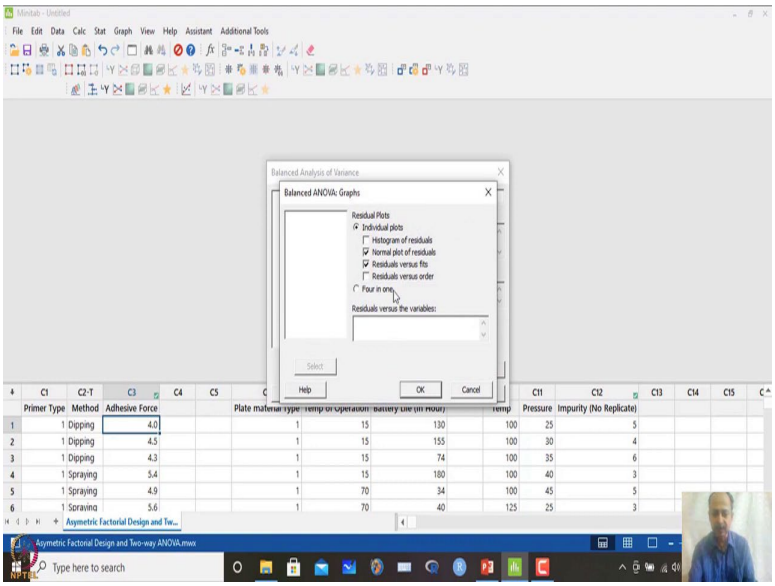
So, what you do is that *ANOVA*, *balanced ANOVA* and you click that one and then what you will do is that you have to identify, which is the response that you have to analyse. This is adhesive force which we want to maximize let us say and primer type and the method are the 2 factors that we have selected in one is in C2 column one is in C1 column like that. And to understand we have to also incorporate primer type multiplied by methods over here to understand the interaction effect is prominent or not.

We do not have any random factor. This has fixed effect model. We are we are selecting that this is a fixed effect model and there is no random factor as such.

(Refer Slide Time: 13:52)



(Refer Slide Time: 13:59)

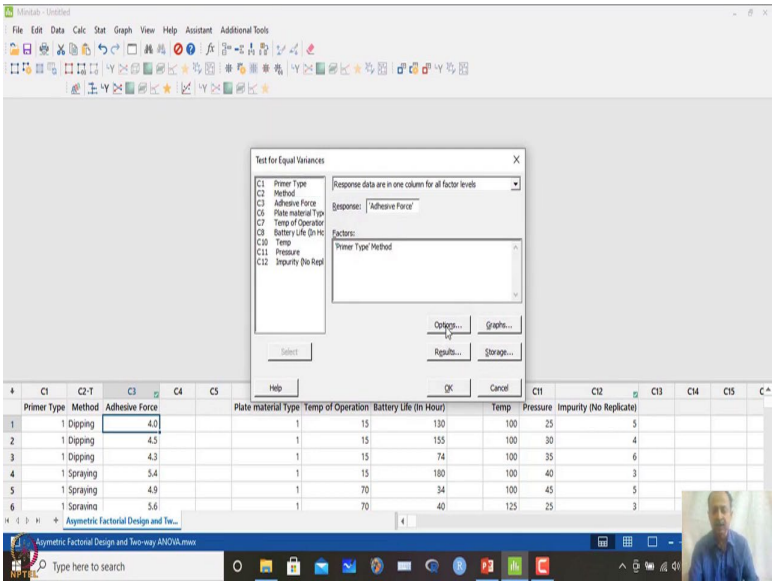


So, in this case what we will do is that in options we will not use we will not take this one. So, so in graph what we will do is that we have the same assumptions like in regression, here also assumptions remain same.

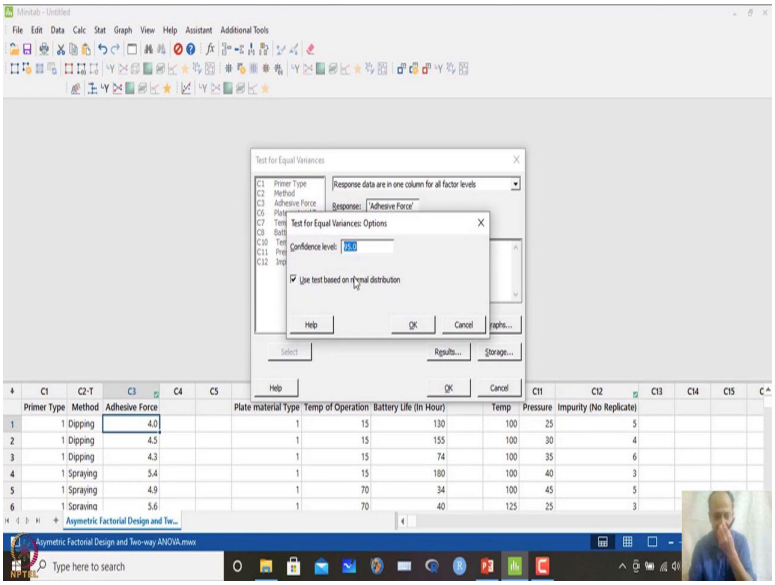
So, residual should be normal, it should be there should not be any heteroscedasticity like that and residuals versus order like that. So, this can also be verified, when we are doing design of experiments like this 2 way analysis of variance. So, what we can do is

that we can also see whether the variance is same or not because of change of these factors over here.

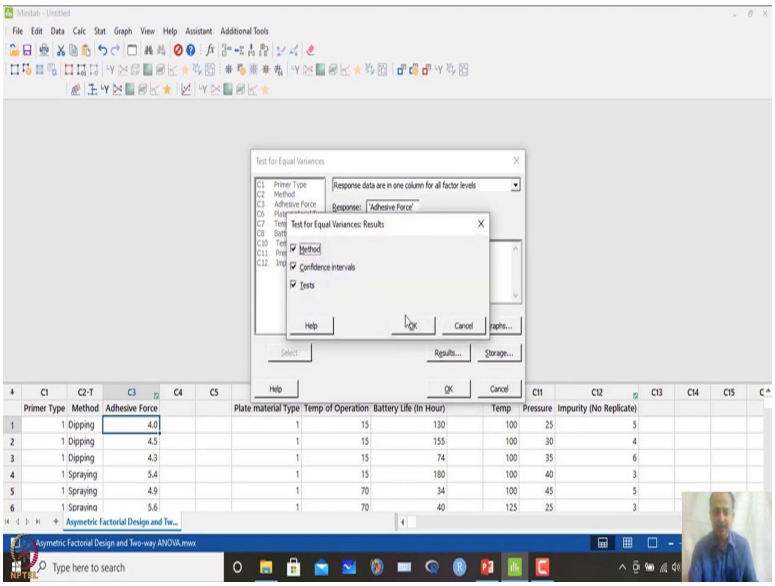
(Refer Slide Time: 14:25)



(Refer Slide Time: 14:35)



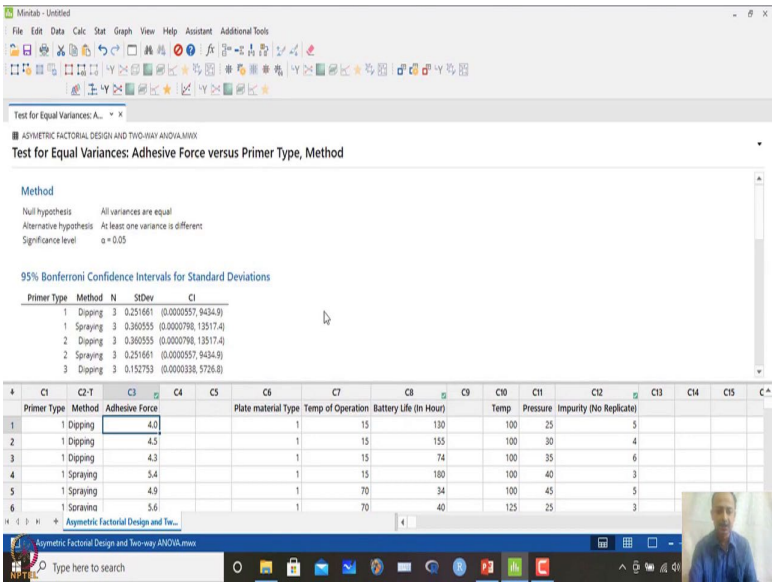
(Refer Slide Time: 14:45)



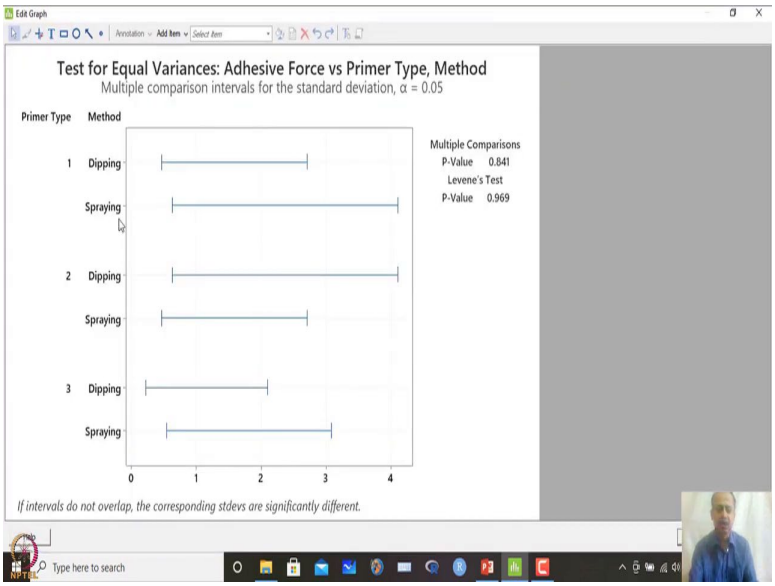
So, if you go to ANOVA, you will find test of equal variance whether the variance is same for different combinations like that. So, I have given that adhesive force is the is the variable or response over here and which are the two factors primer type and method over here.

And in options I am using not if I do not use normal distribution assumptions for the data set and we will get by Levene's test, we can get that one. And results we have given all possibilities over here. If you want to store that is a possible, but I will click ok let us say let us assume.

(Refer Slide Time: 14:51)



(Refer Slide Time: 14:53)

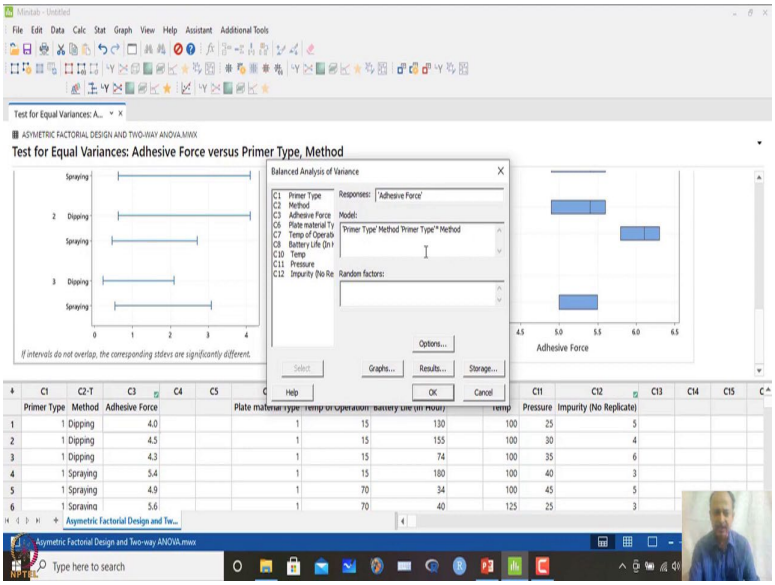


And in this case, what happens is that I get a p value over here. So, different combinations of primer type and methods over here what you are seeing is that and the p value of Levene's test what is important for me. And Levene's test indicates that p value is more than 0.5.

So, in this case this indicates that there is a at least no heteroscedasticity when I change the combinations of primer type. And methods like that, more or less all the variance is

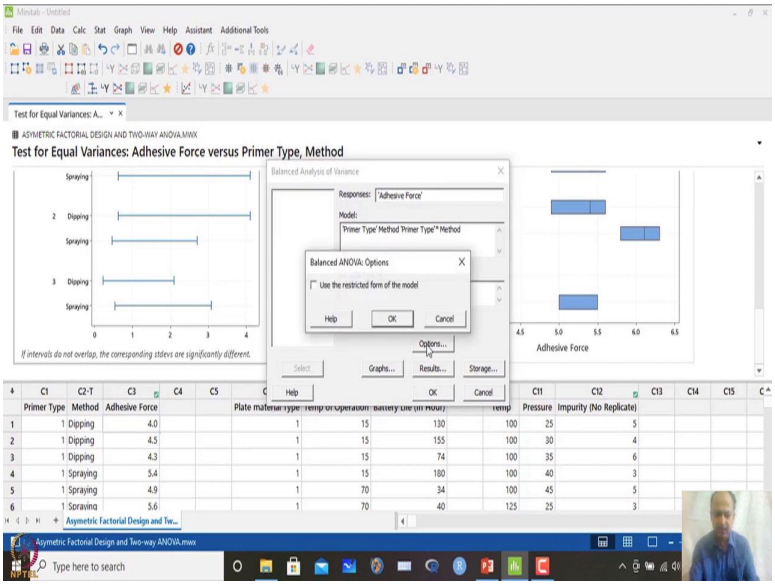
same and that is not different. So, in this case, we can adopt this one. So, first this assumptions is verified over here.

(Refer Slide Time: 15:37)

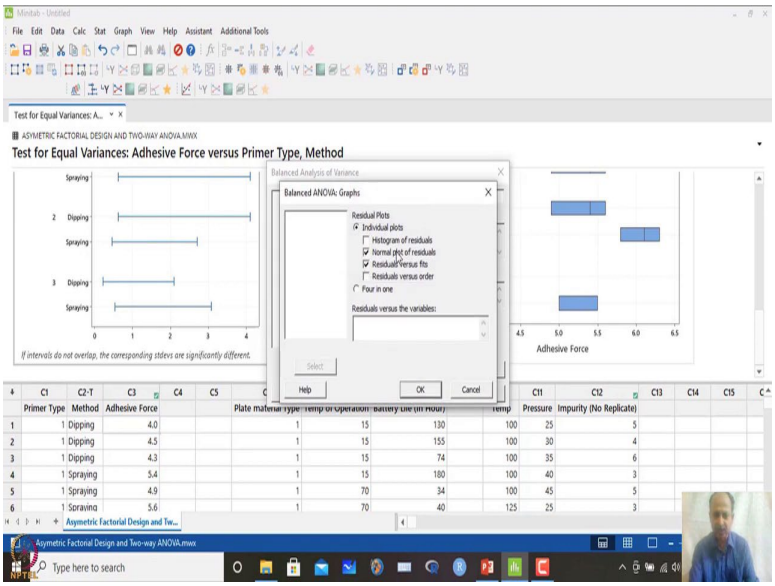


Now, what we can do is that, we can go to stat ANOVA; balanced ANOVA let us say an adhesive force is a factor that is taken over here. Adhesive force, you can just click this one and this is the interaction that we have taken.

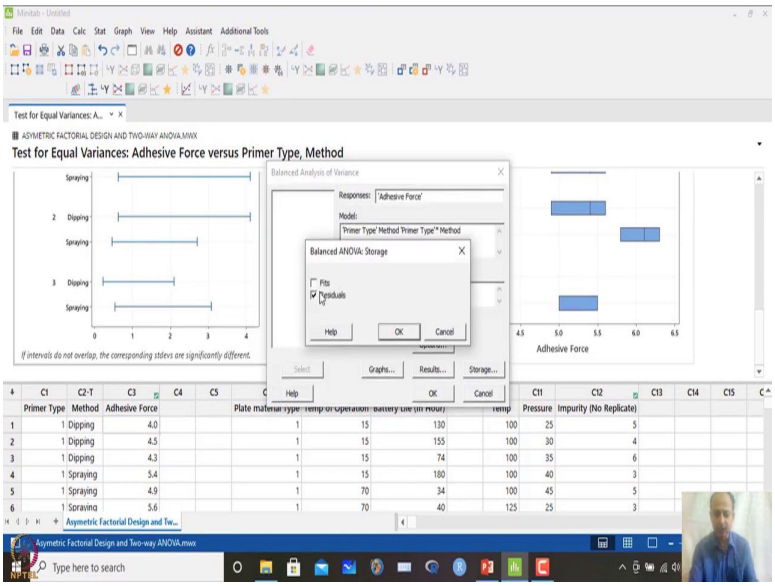
(Refer Slide Time: 15:40)



(Refer Slide Time: 15:42)

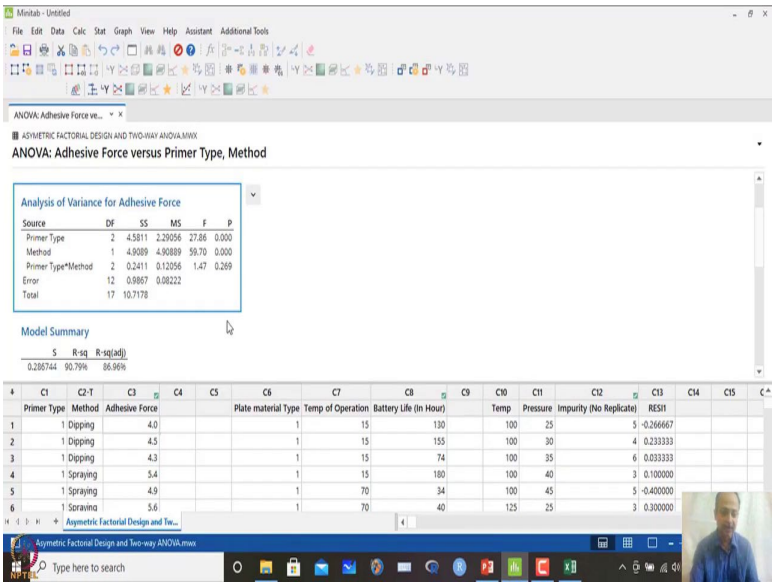


(Refer Slide Time: 15:51)



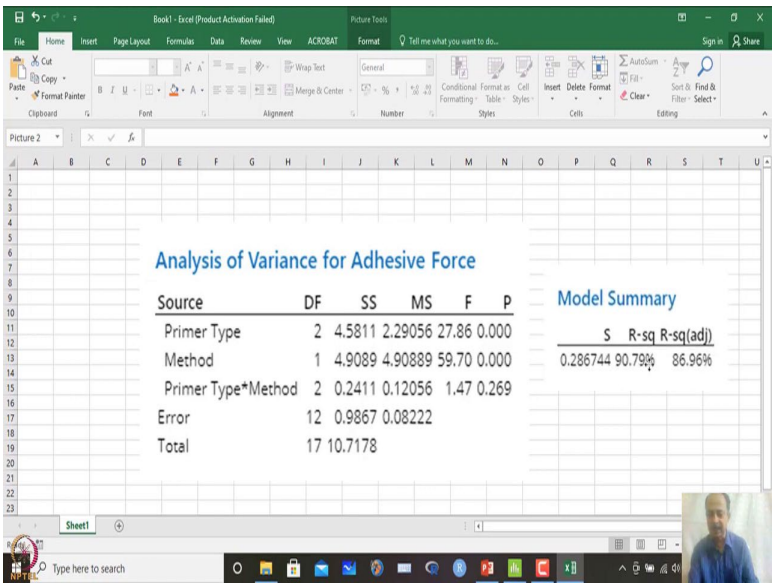
And in options, we will do normal probability plot residual versus we to see if heteroscedasticity is still there in the error like that, we want to check. And if you want to store the residual, you can store the residual also to see the normal distribution assumptions like that. Then what you do is that you click ok.

(Refer Slide Time: 15:57)



What will happen is that, you will get a ANOVA analysis table like this. What we have just discussed. We can interpret this table.

(Refer Slide Time: 16:25)



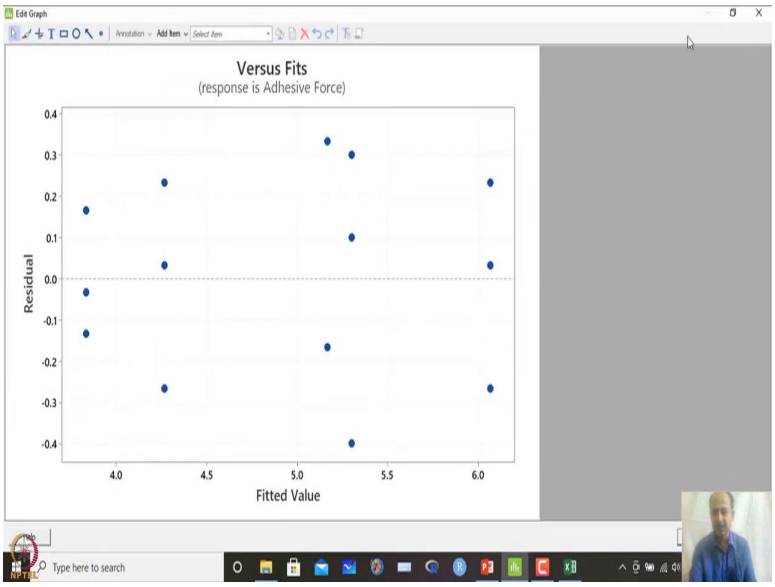
What I see primer type is the source of variation. Means when I change the primer type it is impacting the expected value of y because p value is less than 0.05. And when I change the method is it also significantly impacting the adhesive strength over here.

This primer type interaction between primer type and method is impacting the expected value? No, basically it is not, because p value is more than 0.05; this is 0.269.

So, this is prominent from this ANOVA analysis and it is also showing that the model how much adequate this model is. So, in this case if I paste this one and just show you the model adequacy checks that one of the checks that is R square adjusted value, because there are 2 variables over here.

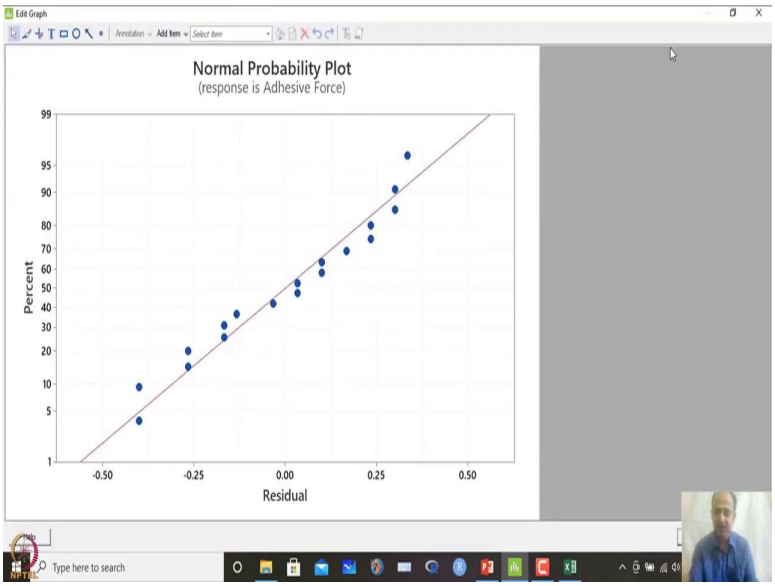
So, it is around 86 percent which is quite good enough. And this says that total variability of the y explained by these two factors (primer type and methods) is up to 86 percent. And this is the interpretation that we are getting over here.

(Refer Slide Time: 17:52)



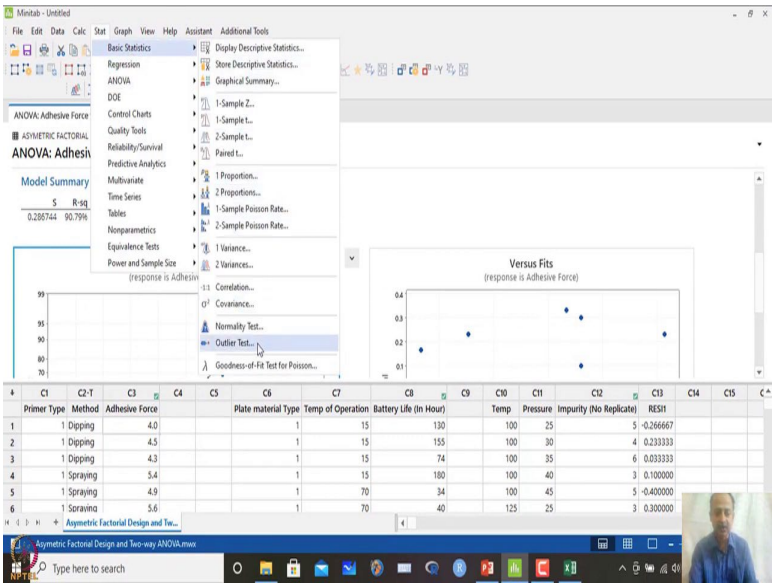
And normal probability plot is given over here and also the residual plot, which does not seems to be very there is no pattern as such. So, we can say the heteroscedasticity is not there which is also proved at the initial stage when we have done this Levene's test like that.

(Refer Slide Time: 18:01)

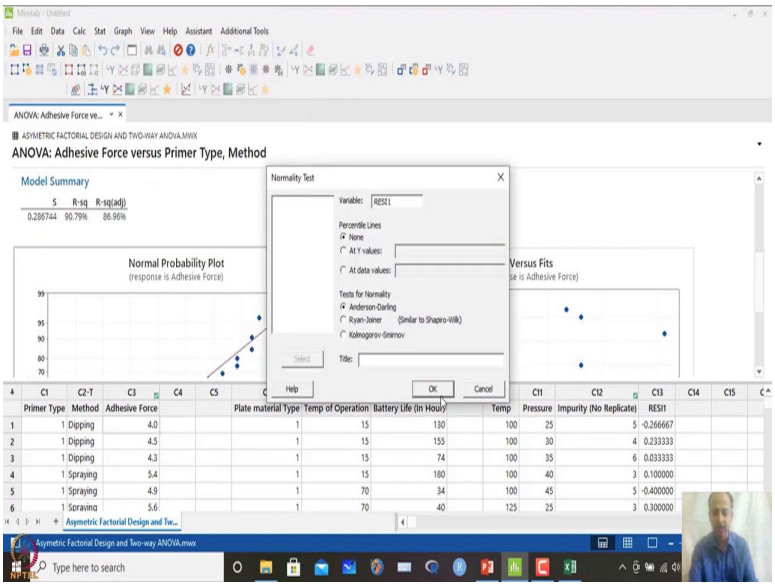


So, normal probability plot also does not show much deviation and we can check this one and whether it is adhering to the normality assumptions, because we want to make conclusions based on this.

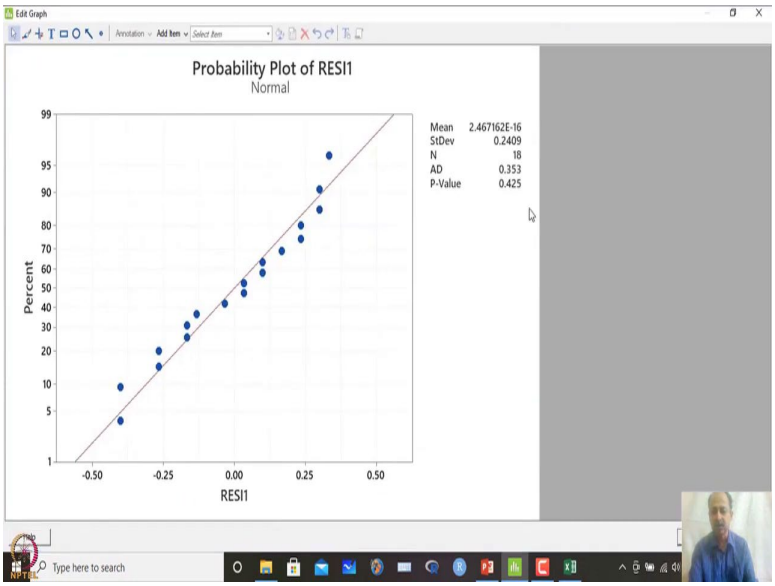
(Refer Slide Time: 18:09)



(Refer Slide Time: 18:13)



(Refer Slide Time: 18:14)



So, basic statistics what we can do is that, we can do the normality test at the end of the data set residual and if you do this one what will happen is that you will get a p value which is more than 0.05, 0.425 and that indicates that it is not deviating from normality assumption.

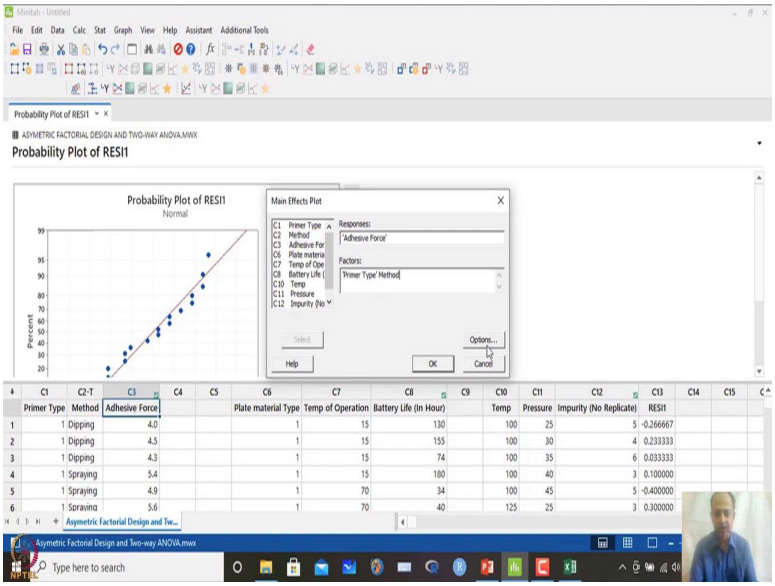
So, model adequacy check is an important aspects, when even if I am doing two factor analysis of variance. So, in that case also, the error assumptions that we have taken in

regulation is also applicable over here and we have to adhere to that; if it is not again transformation and all these things will come ok.

Anyhow, so, this is the primer type methods and adhesive force. Now, which is different from which one multiple comparison tests are also possible, but let us try to see one more thing important aspects over here, which is known as interaction plots like that. We want to see that plot plotting is also possible over here analysis of variance.

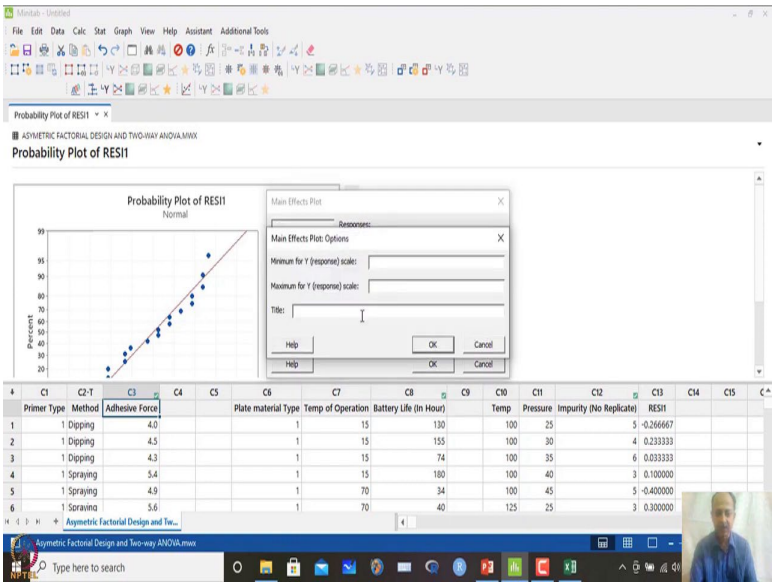
So, in this case there are two plot options that we will see; main effect plot and interaction plot over here.

(Refer Slide Time: 19:05)

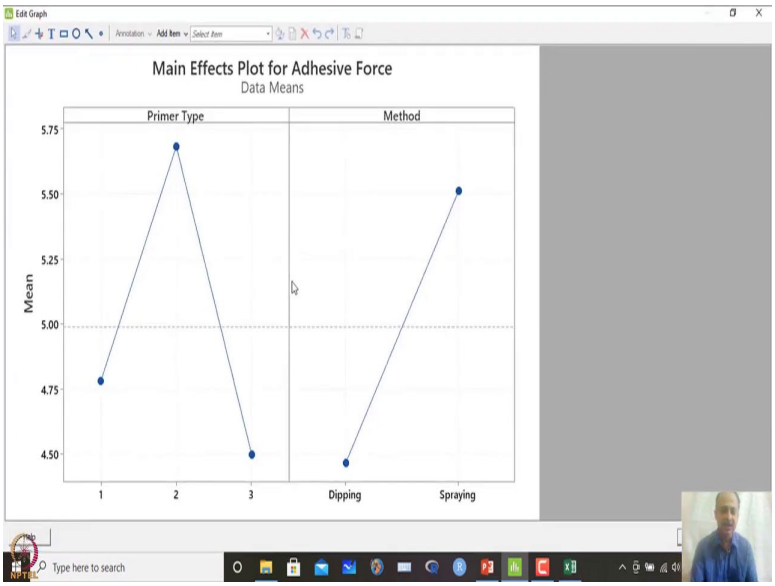


Let us assume let us try to see what is main effect plot. So, main effect plot we have to draw this one and try to see. Adhesive force is a response over here and factors that we have considered primer type and method type over here.

(Refer Slide Time: 19:14)



(Refer Slide Time: 19:21)



So you will get the main effect plot like this. What does it indicate basically? When primer type is 1, what is the average value of this experimental data set that we have got. When the primer type is 2, what is the average value that we are getting of adhesive strength that this is the point that we are seeing over here. Similarly, for when the primer type is 3, what is the adhesive strength? So, if you have to select over here, which type of primer type. I will select that will maximize the adhesive strength immediately I can say 2 is the 2 is the primer type that I should adopt over here.

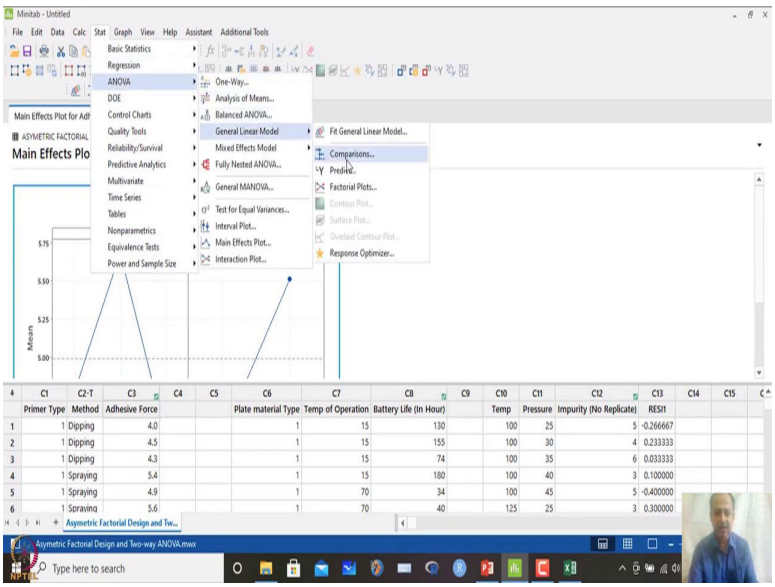
Similarly, methods for dipping and spraying over here what you are observing over here? Dipping is giving a lower mean as compared to spraying, when I am using spraying method over here. So, if you are going by method selection, I will go by spraying always over here ok. I will go by spraying over here and primer type 2 over here.

When there is no interaction, I will go by the main effect plot and I can get the best combination based on this main effect plot over here which is also we can do by seeing the interaction plots.

So, this is the interpretation. So, when we have main effect plot, there is no interaction just do the main effect plot and see the best combination. And to find out the best combination what you can do is that primer type 2 and methods over here spraying can be adopted over here.

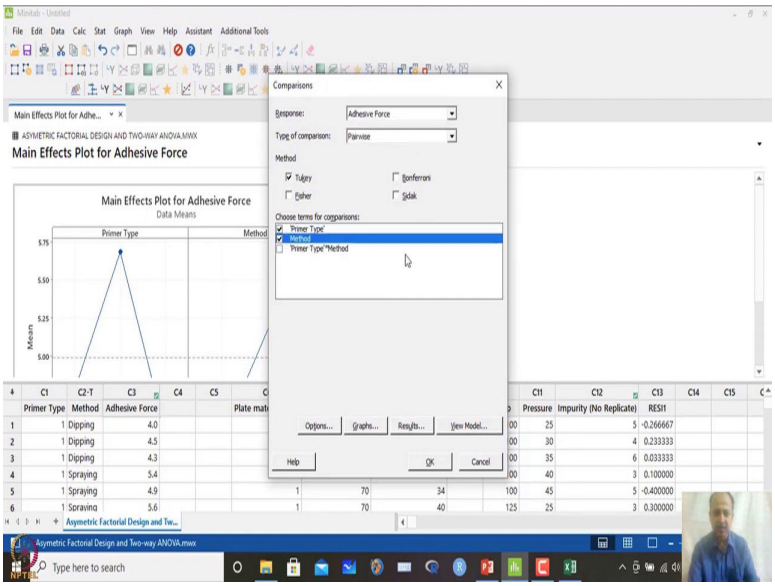
And multiple comparison tests can be seen whether primer types 2 is very different from 1 and 3 like that whether spraying methods is very different from dipping methods like that; that is also possible to be done.

(Refer Slide Time: 21:08)

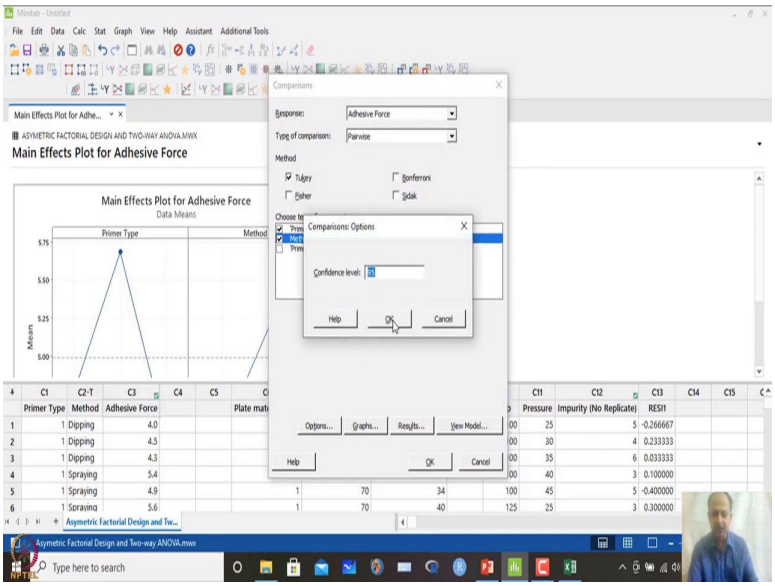


And that can be done when you go to stat over here, analysis of variance over here and maybe general linear model over here and go to comparison test over here.

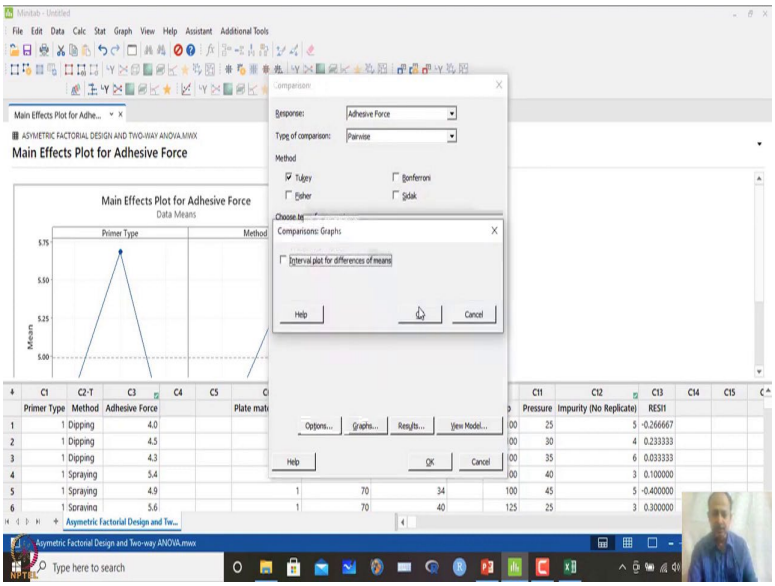
(Refer Slide Time: 21:12)



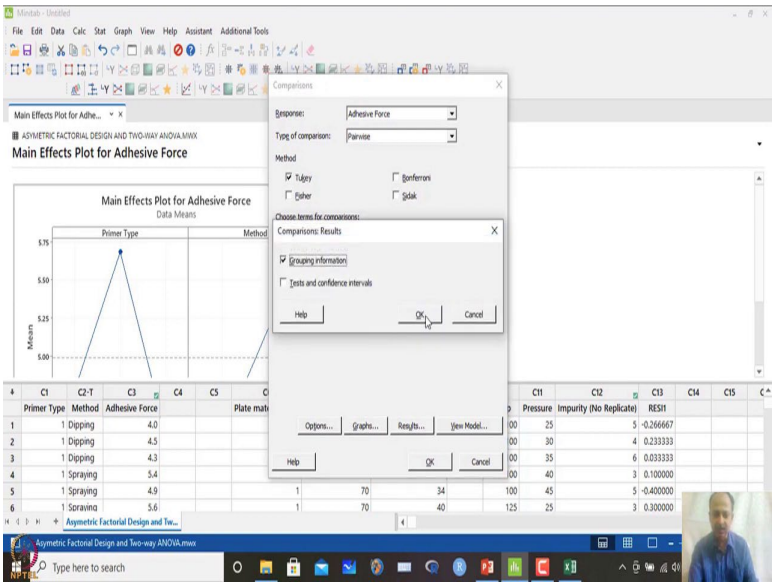
(Refer Slide Time: 21:22)



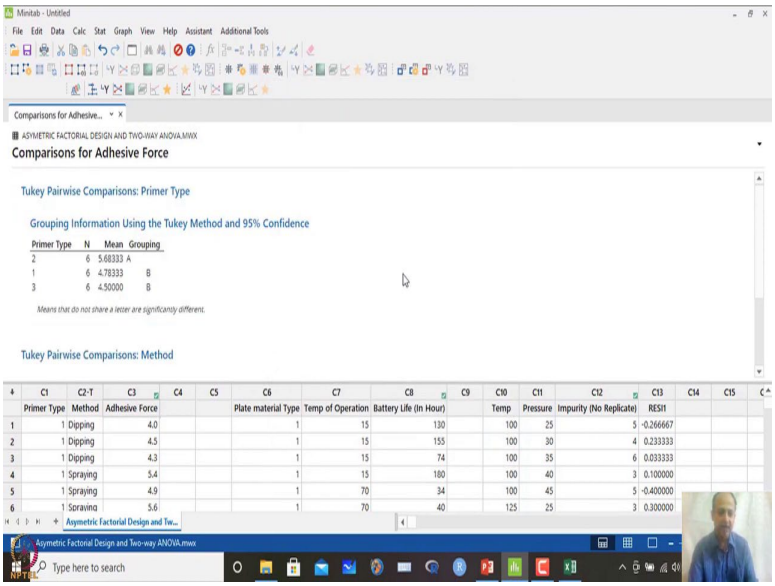
(Refer Slide Time: 21:24)



(Refer Slide Time: 21:27)



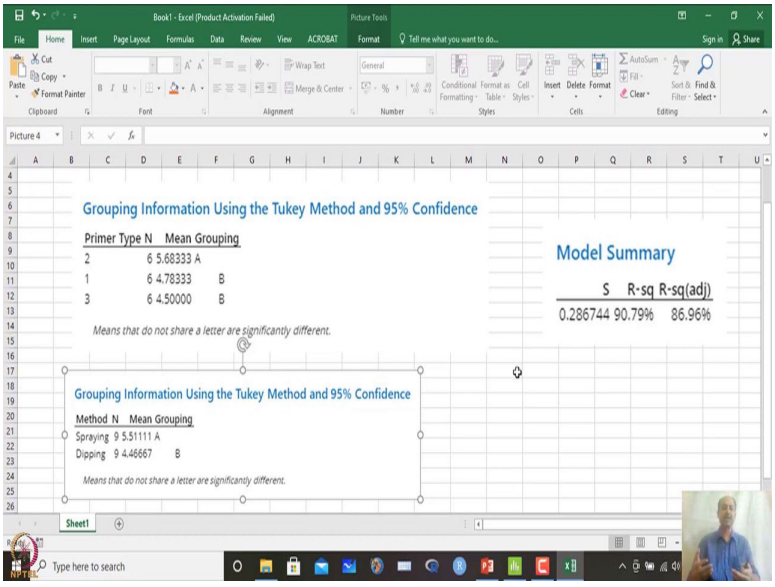
(Refer Slide Time: 21:31)



And we can use Tukey's comparison test for this, we will take the adhesive force over here and to case test and I want to see primer type of a method whether they are different like that.

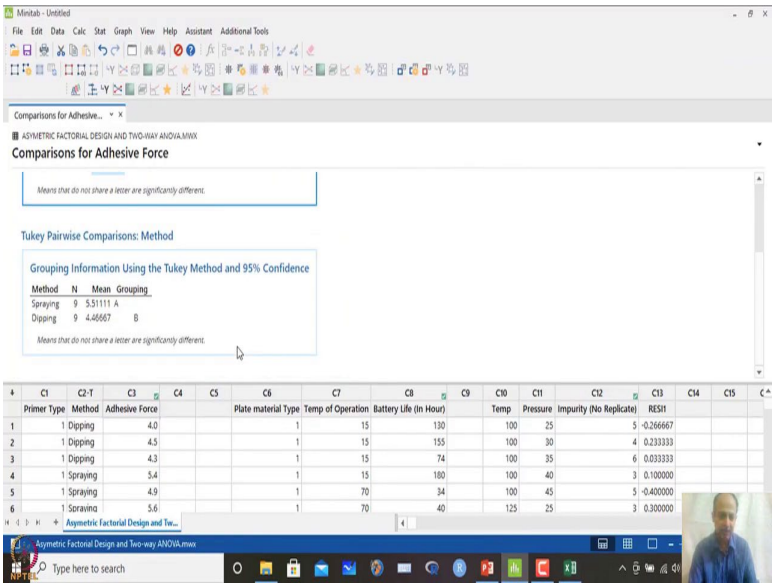
And in options we do not want to change anything over here, we do not want to see all these results and also grouping information is required, because Tukey's test is based on grouping. I go to grouping information. What I see is that primer type 2, if I can copy this one, you will be able to see copy as picture over here and I paste that one over here.

(Refer Slide Time: 21:43)



So, what I will do is that I will just paste this one. And it says that primer to 2 is having a letter code of A which is very different from 1 and 3 like that. So, 2 we should select like that, because it is very different from the other one.

(Refer Slide Time: 21:57)

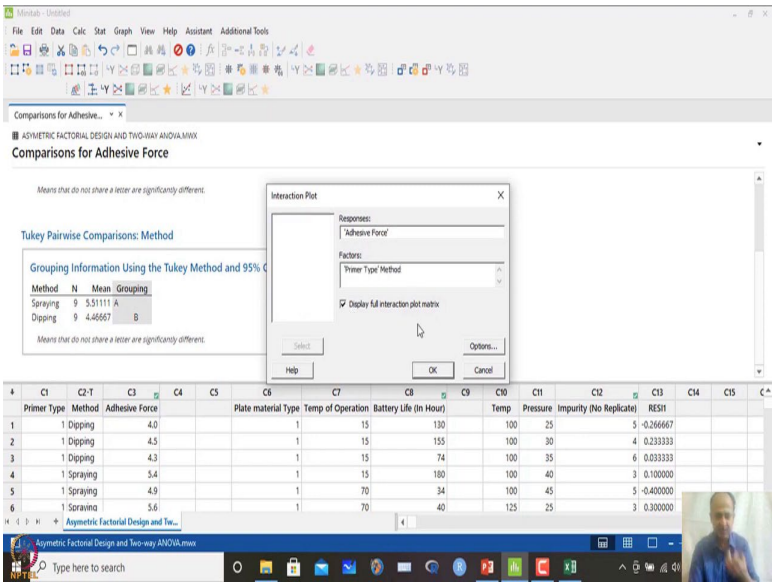


And then we can also see what happens with that spraying and dipping over here. In this case also we can copy as a picture and we can paste it over here to understand, because there is no interaction effects that is why we are seeing this grouping information of Tukey's test only on these individual factors. And what we are seeing is that spraying is very different, it is giving a higher mean as compared to the dipping methods over here.

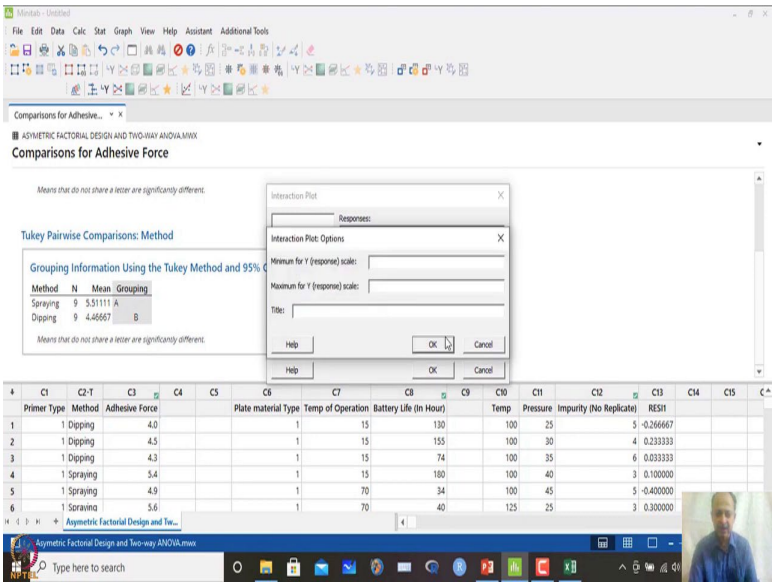
So, clearly I can identify that spraying should be adopted and primer type 2 should be adopted over here. So, that is the best combination, which is giving me a high expected value high expected value of the adhesive strength like that ok.

Now, we can also see this combination, best combination by seeing the interaction plot also. So, what we can do is that we can go to ANOVA analysis and we have an option of interaction plot also ok

(Refer Slide Time: 22:47)

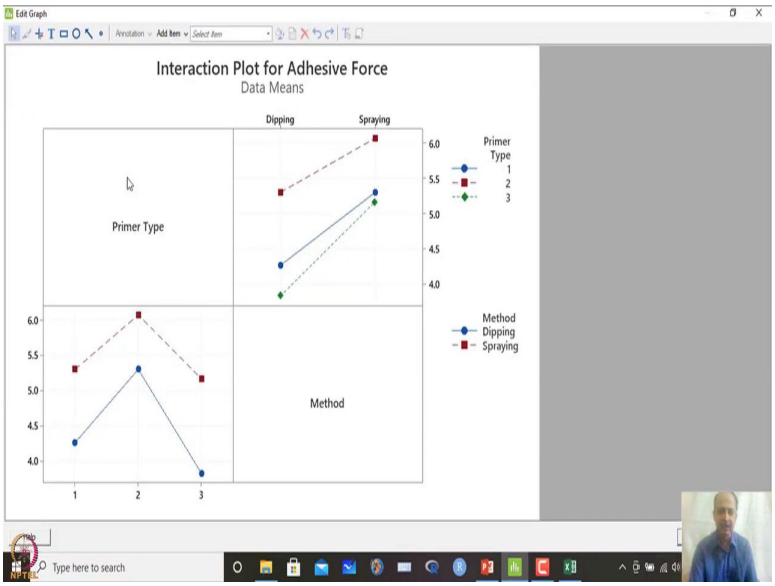


(Refer Slide Time: 22:58)



So, even if interaction is not prominent we can see the interaction plot. And how do we do that? Adhesive force and factor is primer type and method over here and I have clicked this display full interaction plot over here. And in this case, options I am not doing anything; I will click ok over here.

(Refer Slide Time: 23:00)



And when I click this one, I will get a full interaction plot over here. So, in this case you see one of the diagram over here I am taking the lower left hand side diagram over here. So, in this case what you see is that this blue line indicates that this is the dipping method and this is the spraying method. So, individual points over here this, I assuming this is the first point that we are locating over here.

So, when the primer type is 1 and we have adopted dipping method; what is the average expected value of adhesive strength that we have noted down, what is the average value of adhesive strength? So, this is the first point that we are getting. Similarly second point when combination is primer type 2 and dipping method what was the average values of the strength information adhesive strength like that.

Similarly, this point is generated like that and similarly, these are the on top what you see is that spraying method average strength that is reported over here. So, in this complete figure, what is what we can observe is that, this is the highest point, this is the highest point that we are seeing 6.066 that value we are getting over here.

So, this indicates that primer type 2 is the primer that we shield we should select and dipping or sorry and spraying is the method that we should adopt. So, because if I take these two combination the we are getting a higher adhesive strength, we are getting a higher adhesive strength over here.

So, either we can see from this side also and you can see also this diagram, interpretation remain same over here. So, we will go by primer type 2 and the this is the best combination; so, this top value that you see over here. So, spraying is the combination with primer type 2.

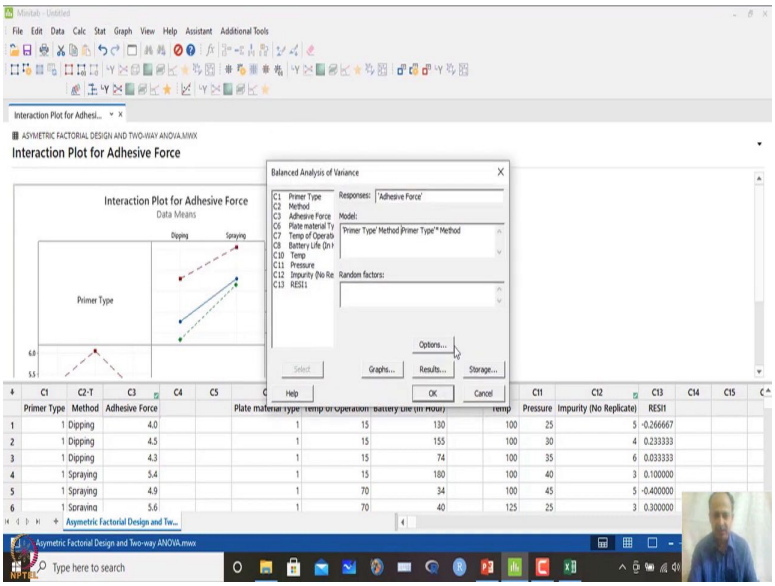
This way we can find out which is the best combination. And what we can what we can also see over from this is that if I cannot control let us say primer type over here 1, 2, 3 in actual manufacturing process whatever; if I cannot control this one, but I can control spraying over here. So, I will always freeze to spraying method.

I will always try to add a spraying method, because irrespective of the primer type that I am adopting over here 1 2 or 3. I will always by spraying method I will always get a higher value as compared to dipping method over here. So, if I cannot control this one, I will go by that one.

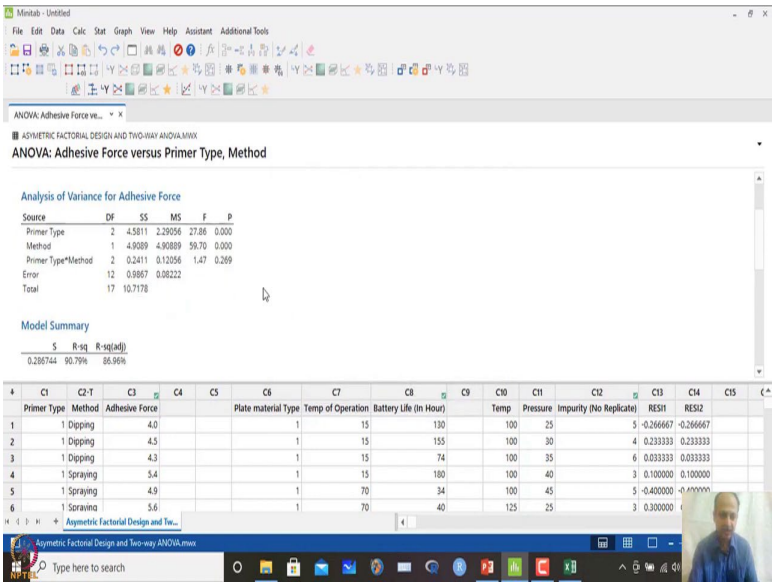
Ok, but and also if due to manufacturing capacity dipping and spraying has to be combined like that I do not have any control over here, then which type of primer type I should use I should use primer type 2, because it is always giving a higher mean as compared to any other primer types like that. So, that is also another interpretation we can make out of this. If I cannot control one what should be the setting of the other, which I can control basically. So, that we have to think and then adopt, which is the combination.

So, what we have told like that? So, we are doing model adequacy checks, we are seeing interaction plot, we are also seeing the comparison test, multiple comparison test, pairwise comparison test of two case tests we are adopting, which combination should be selected like that and interpretation is same like analysis of variance interpretation. So, what we can do is that ANOVA analysis.

(Refer Slide Time: 26:05)



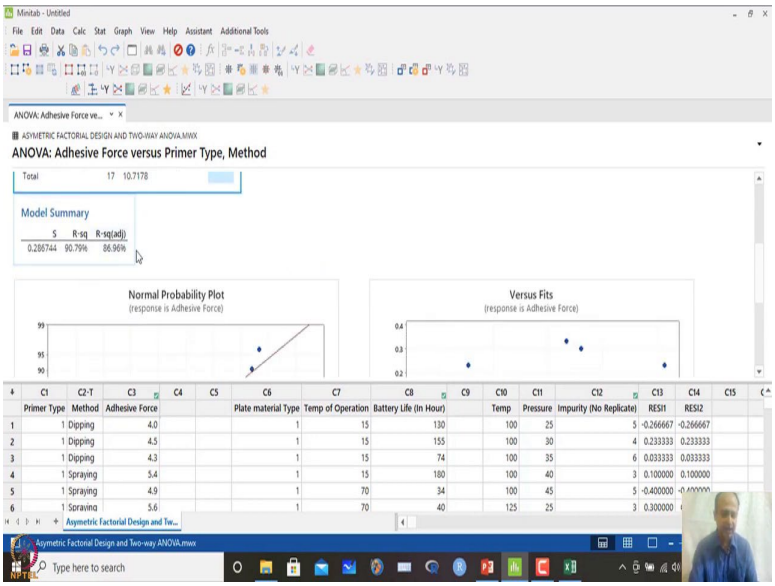
(Refer Slide Time: 26:08)



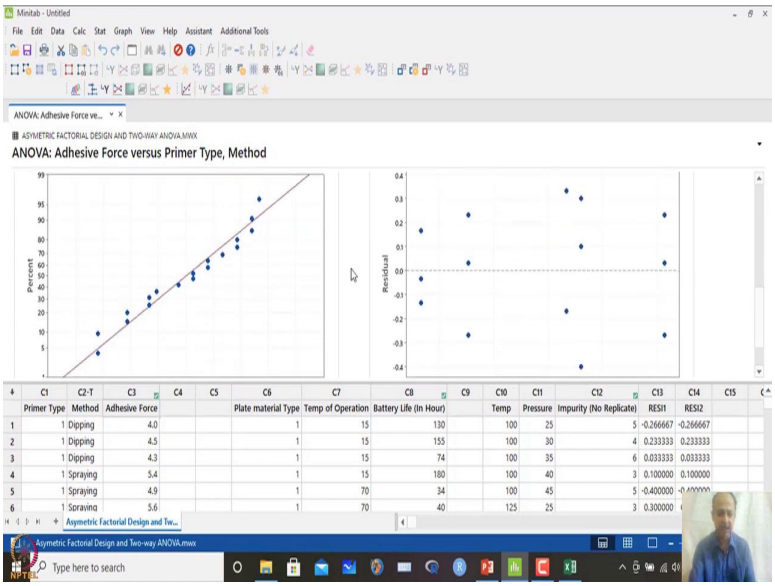
So, balanced ANOVA we have to go; adhesive primer type and the primer type and method that is option and when you click this one, you will get the ANOVA table which will indicate that which factor is significant. So, primer type is significant, what we are seeing method is significant over here, but interaction is not significant which is around 0.269 what we are getting ok.

So, when they are when they are together acting in that case, it is not impacting why expected value of y at least what we can interpret out of this.

(Refer Slide Time: 26:31)



(Refer Slide Time: 26:35)

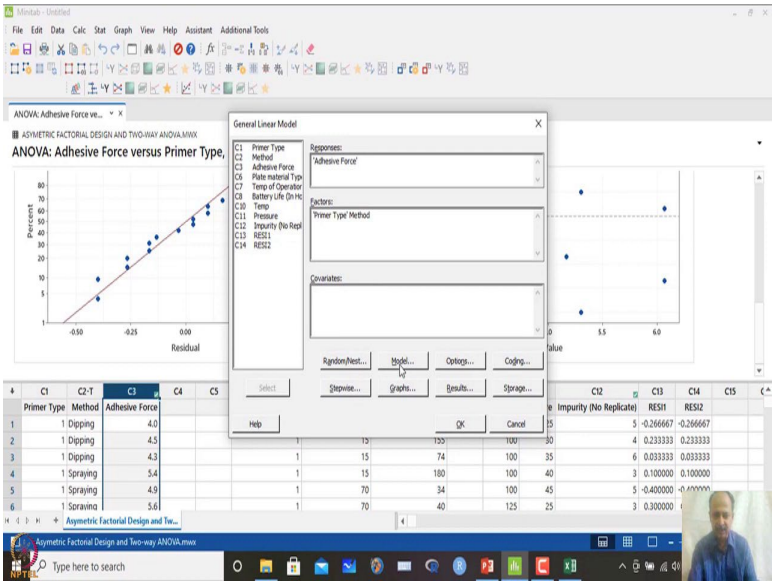


Model is summarized, R^2 is quite high; that means, the model the factor that we have selected is quite adequate factors and it is explaining about 90 percent of the variable, 86 percent of the variability over here.

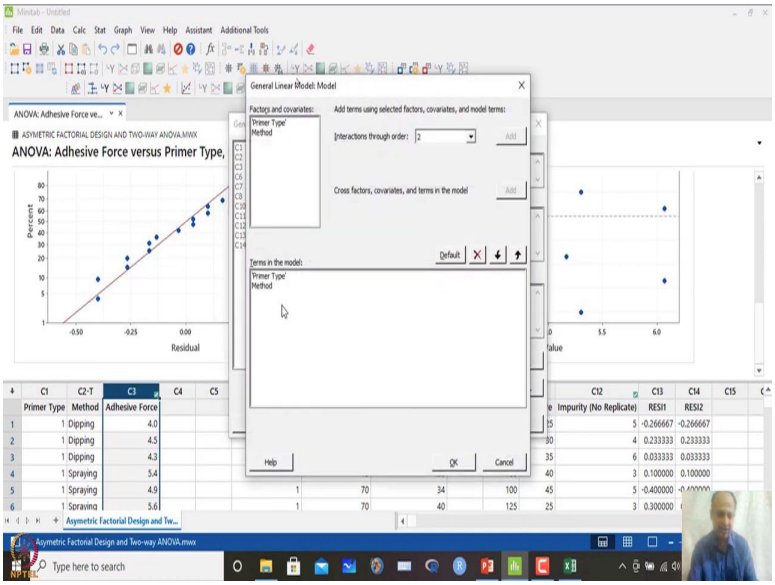
Normal probability and distribution assumptions are also quite ok and we can do all these tests like that so, over here. And there is another option; that means, we can also develop some regression equation based on this primer type although this categorical variable and adhesive force over here. We can also develop there is an option like

analysis of variance, what I have told is in one way analysis also we told we use generalized general linear model over here.

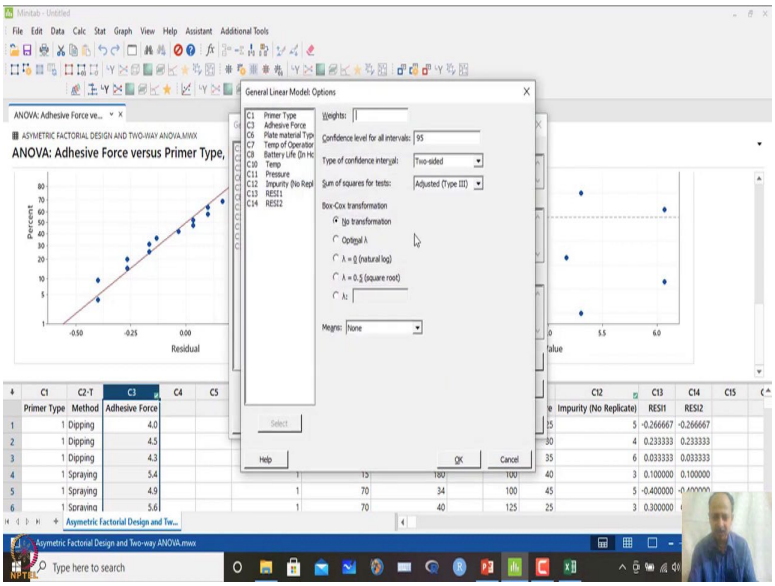
(Refer Slide Time: 27:13)



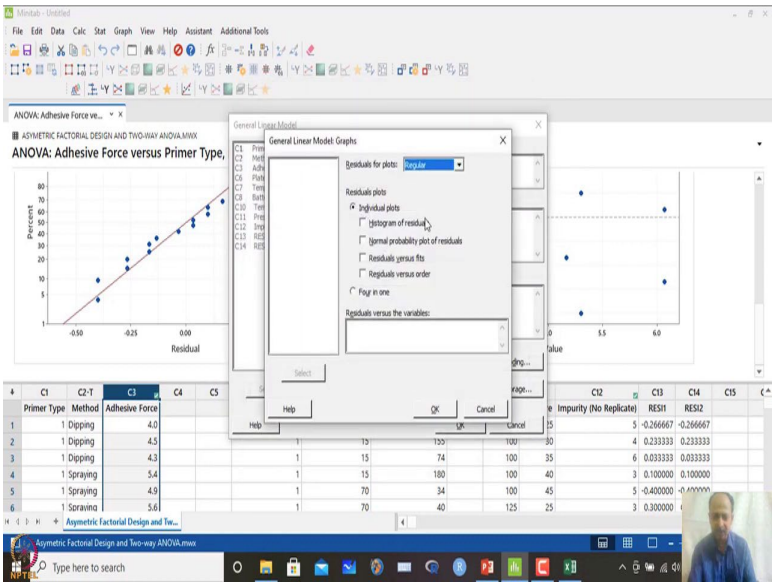
(Refer Slide Time: 27:24)



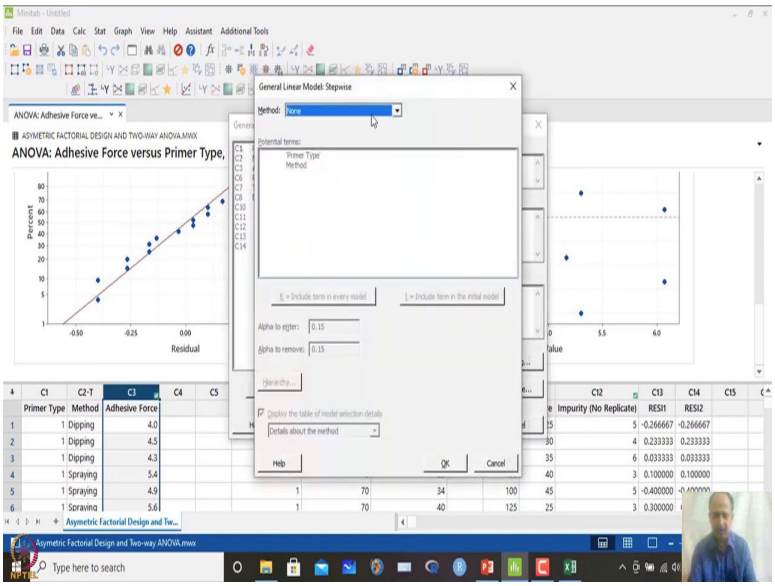
(Refer Slide Time: 27:34)



(Refer Slide Time: 27:38)



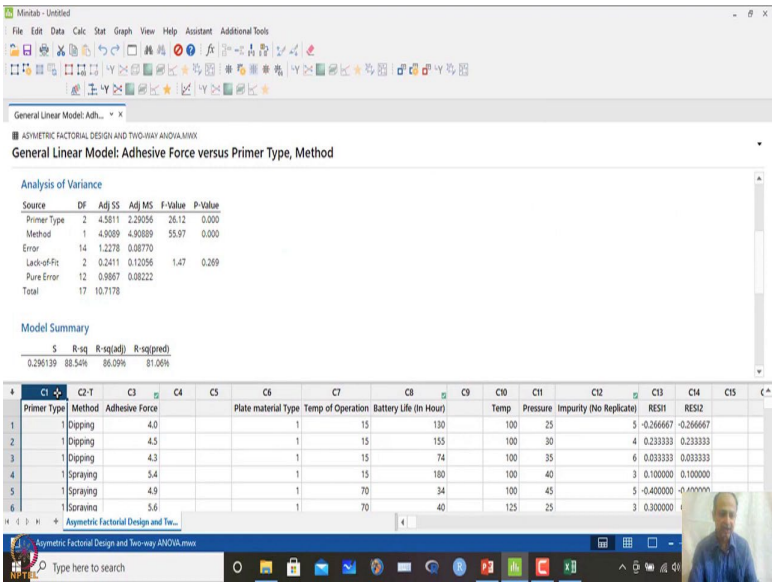
(Refer Slide Time: 27:43)



So, the assumption is normality assumption is taken is considered over here, but we can fit a general linear model also. And we can say that we want to see adhesive force over here and this is the primer type and method that we are adopting over here. So, then in this case models, because there is no interaction, I am not considering that interaction effects over here.

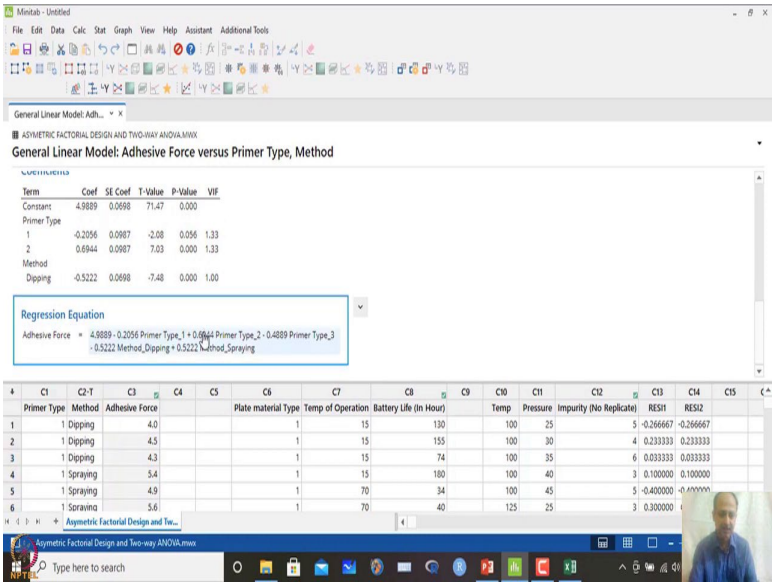
So, primer type and this one and if you click ok over here. So, in this case options we do not want to change anything and in graphs what we can do is that residual plots we can see that is also possible, but we are not using any stepwise regression. So, we are not using, because we have finalized that main effect is prominent over here.

(Refer Slide Time: 27:51)



So, in this case if you click ok what will happen is that, you this on coded variables, this will be coded C1 and C2 will be coded variables will be used and this will be modelled with adhesive and the equations will be given. So, regression equation you can see over here.

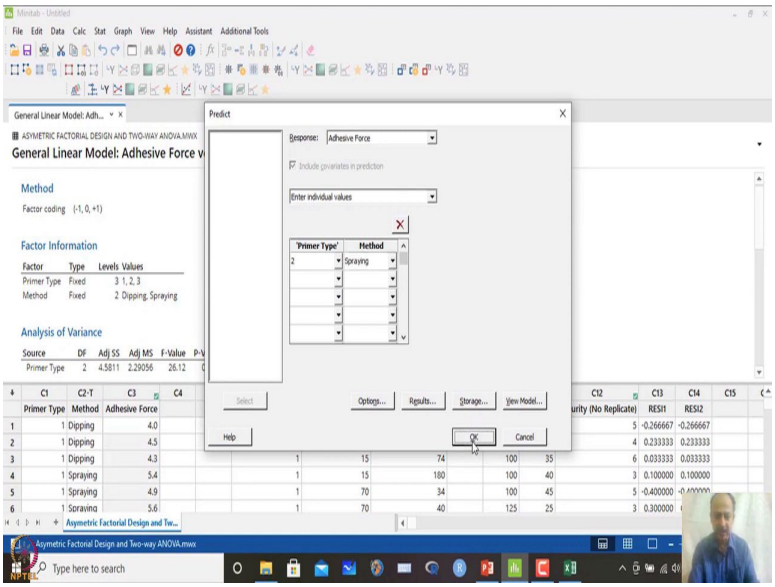
(Refer Slide Time: 27:59)



So, this is the regression equation that is developed and variation inflation factor is not there and these 2 are the prominent factors that is what we can see. And based on this basic regression model, what we can do is that we can also predict what value of

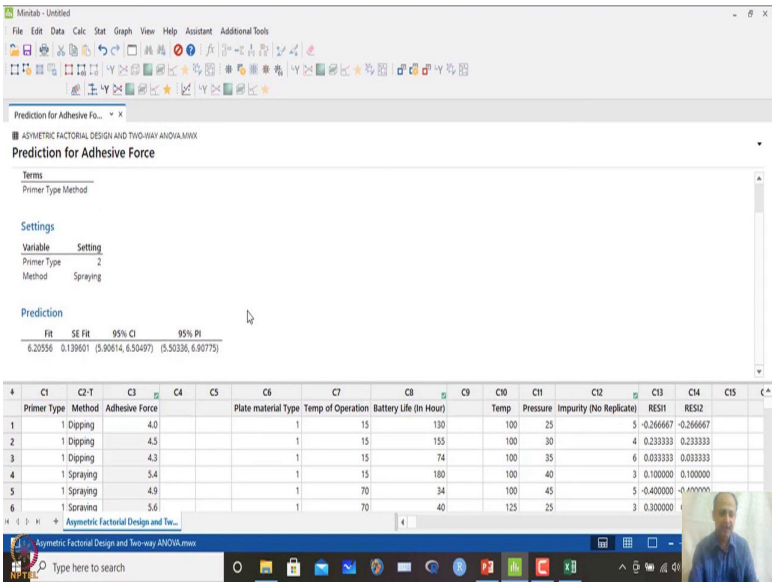
adhesive force is expected like that. So, if you go to analysis of variance general linear model using this model that was fitted by general linear model, then in that case we can also predict this one.

(Refer Slide Time: 28:28)



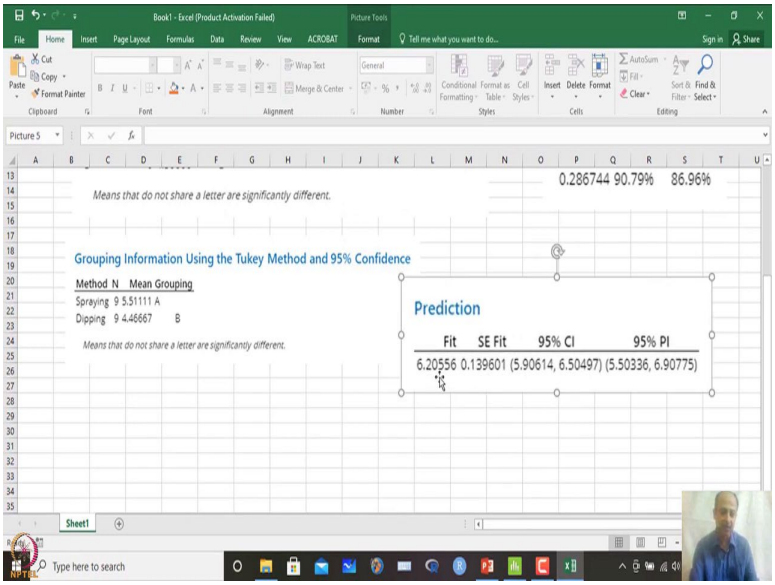
So, if it is primer type 2, in this case and the method is spraying over here, I want to predict what should be the expected value of adhesive strength like that; adhesive strength.

(Refer Slide Time: 28:40)



If you click ok over here and what you get is that around 6 is the fit that we are expecting over here.

(Refer Slide Time: 28:50)

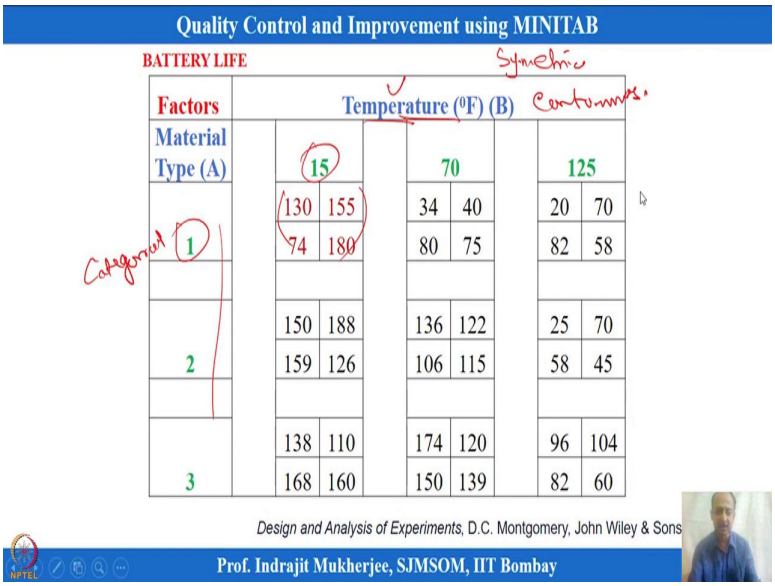


So, if I copy this one as picture over here and I paste it over here so, this what we will see is that the predicted value or expected value is around 6.2 and this is with some prediction interval and confidence interval. Based on the regression equation, general linear model fitting and this was done. We can think of general linear model as a as a generalized view of linear regression model, that we are discussing earlier like that. So, it is at a broader umbrella, you can think of ok.

So, this way also we can have a prediction model like that and this is categorical variable, they will be coded and based on that regression will be developed. You can see more on general linear model on MINITAB websites also. And you can see in any other books also you will find; general linear model how the models are developed, how the beta are estimated, how variables are coded, you can see.

So, this is one options that we have and these are all this both the factors are categorical over here. You see both the factors are categorical over here and, but scenarios can be that one is categorical, one is continuous like that; one is categorical, one is continuous that will be our next example. That is second example that we will discuss battery life design experimentation; battery life design experimentation where factors that is selected over here one is material type and one is temperature over here.

(Refer Slide Time: 30:00)



One is material type and one is temperature and the battery life is measured over here and these are the values of and 4 replicates are done at each combination of this material type and temperature. So, this is a balanced design, what we what you can see is that in every trial, 4 replicates are collected over here and also the number of levels of temperature and number of levels of this is same. So, we can think of as a symmetric design; symmetric design over here.

Earlier one was asymmetric this is a symmetric design and one of the factor is categorical over here categorical over here, but one of the factor over here is continuous that is temperature is a continuous variable. So, this is continuous.

So, earlier both the factors are categorical. So, how to analyse that, one we have seen. Now we are trying to see this experimentation, which was this data was reproduced from this design of experiment by Montgomery and how to analyse this data how to make interpretation out of this data. When one is categorical and one is continuous variable, but at discrete levels we have just experimented 15, 70 and 125 that we will see in our next session.

Thank you for listening.