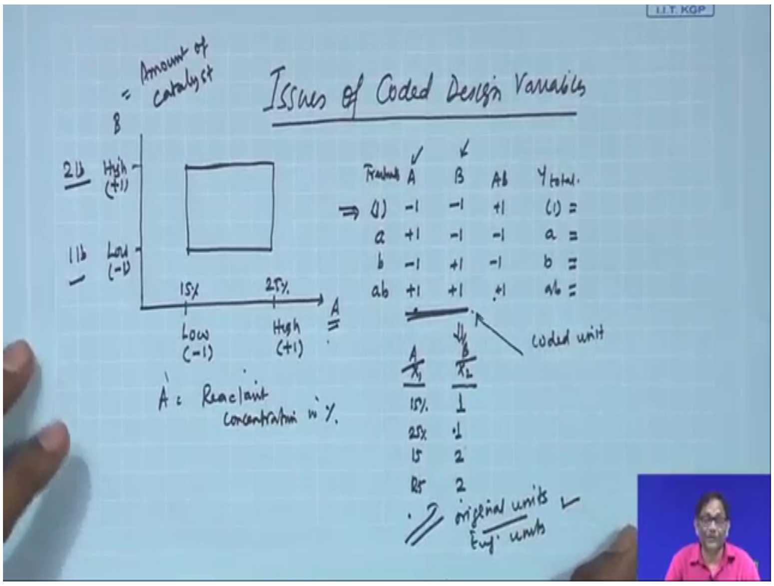
Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture – 41
2k Factorial Design – Issues with Coded Design Variables

Welcome. So, we will resume 2 to the power k factorial design. Today will discuss a special thing that is the issues related to coded design variable issues of coded design variables.

(Refer Slide Time: 00:37)



If you go back and see the subsequent lectures what I have delivered in last couple of classes. You have seen that I have consider 2 factors A and B, and I say that A has 2 levels; low, high. Also B has 2 levels low and high and this low we have denoted by minus 1, high we have denoted by plus 1. Here also minus 1, here also plus 1 and the individual, then these are. This is basically the settings and then we have created design matrix with reference to suppose A b, then we say this is minus 1 plus 1 minus 1 plus 1, then minus 1 minus 1 plus 1 plus 1 and these treatment combination we denoted like this treatments. This is your 1, this one is A, this one is B and this one is Ab. And; obviously, you have created Ab interaction by multiplying a column A and B minus 1 minus 1 plus 1, and then you find out the Y value.

Actually we have used Y total, and this is also Y total, we have denoted like this also 1 A B Ab here 1 A B Ab, their treatment combination here by 1, we said the what is the total here, what is the total value, what is the total value, what is the total value. So, now see what about this A and B, they are real physical variables. For example, if I say A is, its a chemical process, A is reactant concentration, reactant concentration in percentage and B is amount of catalyst, amount of catalyst. Then when you actually conduct experiment, there is no minus 1 plus 1 or low and high. What is there? There are certain percentage for A, it is the designer or the analyst of access operator which have been knowledge, he knows what is the range.

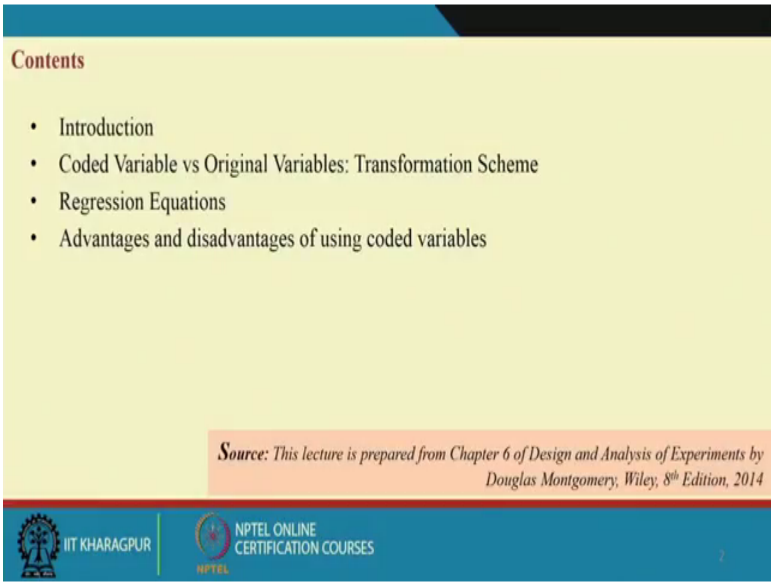
And in the beginning class I said that usually, may be minimum value will be low and maximum value will be high or some value in the operating zone will be low and high late this is. So, in reality, it may be low may be 15 percent for a high, may be 25 percent here, low may be 1 pound and high may be 2 pound that is the amount of catalyst. So, there nay essentially if I write down this one ; that means, in place of A also, if I write X 1 and in it is 1 write X 2, then or you can write A b no problem. So, minus 1 means A is 15 percent B is 1 plus 125 percent.

This is your 1, then again 15 percent, 2 then 25 percent 2. So, this is the original value or ; that means, you original units or other way I can say these are engineering units, but when we are defining like this, we are converting this engineering unit to coded units minus 1 plus 1, irrespective of whether it is the ah, what is the quantity.

So, suppose for B it is 1 and 2 and A it is 15 and 25 percent, but here you see it has been, it has been transformed to minus 1 and plus 1, and it has become unit less kind of thing ok. So, when you do the experiment, you tune the process with reference to the factors in terms of its engineering unit, which is know original unit, but when we are basically giving you the treatment matrix, these are design matrix we. We are basically doing in terms of coded variables issues is that, how this transformation takes place number 1? Second issue is that, as we have designed what are the advantage of using coded variable, it definitely makes our mathematics calculation and calculation simpler.

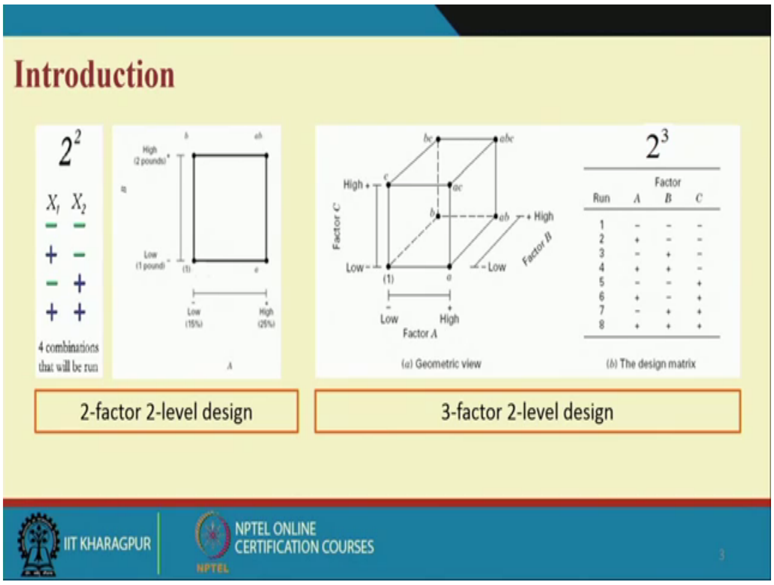
Now, third one is, if we do how ah, when we do regression kind of analysis, and develop the response surface or the effects, what way, the coded variable using coded variable, when the effects computed by regression and using your original units. The effect and their significance, will there be any difference? if there is a difference why these. And finally, if in, is there a way to even if your analysis is done using coded design variables will be, can, can be converted to the original units or engineering units and if there are such problems, some problems, then why should we use coded design variables, why should not be go for only engineering variables. So, these are the issues which will be discussing in this lecture today ok.

(Refer Slide Time: 07:58)



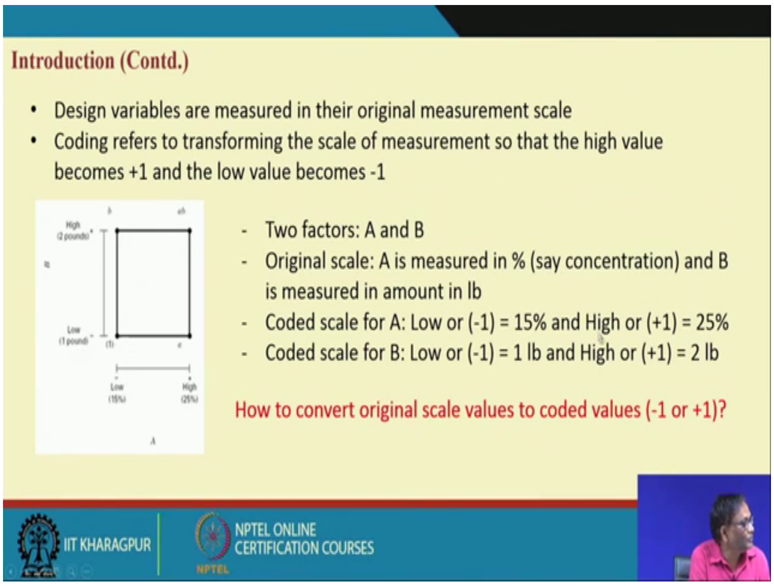
So, let us see the contents. I have already given you, we will give you the scheme for transformation from code; that is means original to coded and coded to original, then will see some regression equations and then we will discuss the advantage disadvantage of using coded design variables. And finally, I will show one example of the, for that advantage and disadvantage point of view.

(Refer Slide Time: 08:30)



So, you can recollect these kind of diagram. These are geometric view of the design matrix, this is for the 2 factor case, this is 3 factor case and 2 factor case, this is minus 1 plus 1 scheme 3 factor case and this is the 2, but 3 design. These are all known to you by this time.

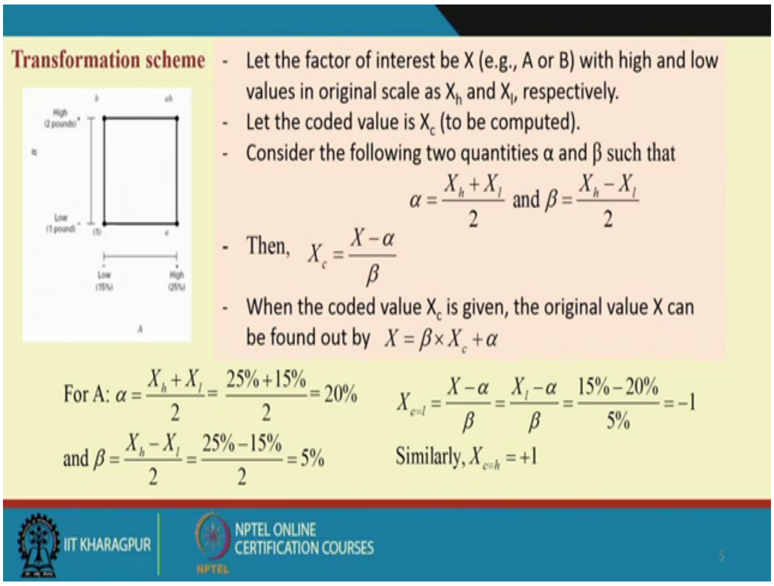
(Refer Slide Time: 08:57)



When coded to original to coded variable and reverse one. We will discuss with reference to 2 factors and please read this slide carefully, see that design variables are measured in their original measurement scale, what I have already told you. If these first factor is reactant concentration, it is measured in terms of percentage, and if it is the second factor is our amount of catalyst, it is measured in terms of units like pounds. So, now, coding refers to transferring the scale of measurement so that the high value become plus 1 and low value become minus 1. I have shown you that high or plus 1 low at minus 1.

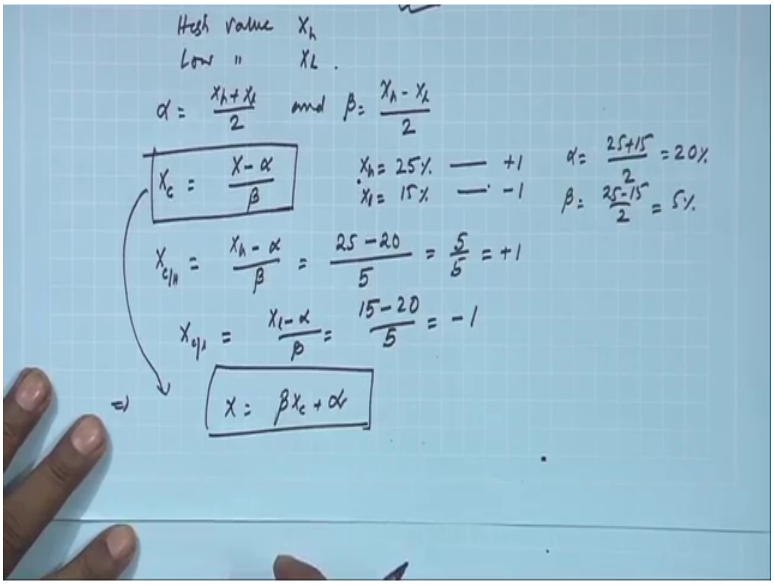
So, the transformation is, basically how the 25 percent will be transform to plus 1 15 percent will be transformed to minus 1. And similarly here 1 pound will be transformed to minus 1 and 2 pound will be transform to plus 1. So, in this example 2 factors A and B, and our job is that minus 1 correspond to 15 percent plus 1 correspond to 25 percent minus 1 correspond to 1 pound plus 1 correspond to 2 pound for factor A and B respectively.

(Refer Slide Time: 10:26)



So, what do we will do here? So, we will start with one factor only. For example, all k we are explaining 2 factor case, but we will do it separately for each factor at a time. Let the original value.

(Refer Slide Time: 10:46)



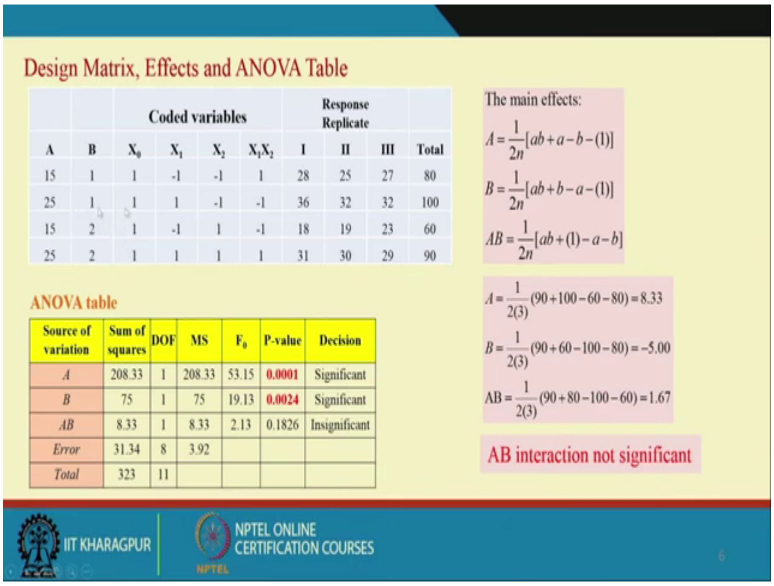
Let the original value be X. Other way I can say X is the variable. Now suppose the high value is, let it be, we are writing X h and low value we are writing as X low. So, create two quantity alpha which is X h plus X low by 2 and beta which is X h minus X low by 2 ok. Now, if the, if X is measured in original unit, then the coded unit X coded can be X minus alpha by beta ok. For example, for example, if we want that 25 percent to plus 1 and 15 percent to minus 1, because our X h is 25 percent and X low is 15 percent.

Then alpha equal to alpha equal to 25 plus 15 by 2; that is X h plus X L by 2. So, this is 20 or 20 percent and beta will be 25 minus 15 by 2. So, this is 5 percent. So, if I my, if observed X is X h, then what is the coded value, it will be X h minus alpha by beta, which is basically X h is 25 percent minus alpha is 20 20 percent, beta is 5 percent. So, this is basically 5 by 5 equal to plus 1 in a same manner, this basically X coded when high then the same manner for X coded for low will be that this will be X low minus alpha by beta.

So, X low is 15 percent minus 25 percent by 5 percent which is basically minus 1. So, if you if you do in the same manner for the other factors, like your B, factor B that is also in the same manner, you will be able to do. So, let us see what we have done here ok. I have noted yeah I have given here. So, for a alpha beta and for X low, this is minus 1. Similarly X I it is plus 1. Now it may so happen that the coded value is given. So, how do get back the original value, you will be getting back original value from this equation.

From this equation you can write X equal to beta X c plus alpha. So, let us see the example, this example we have seen earlier also.

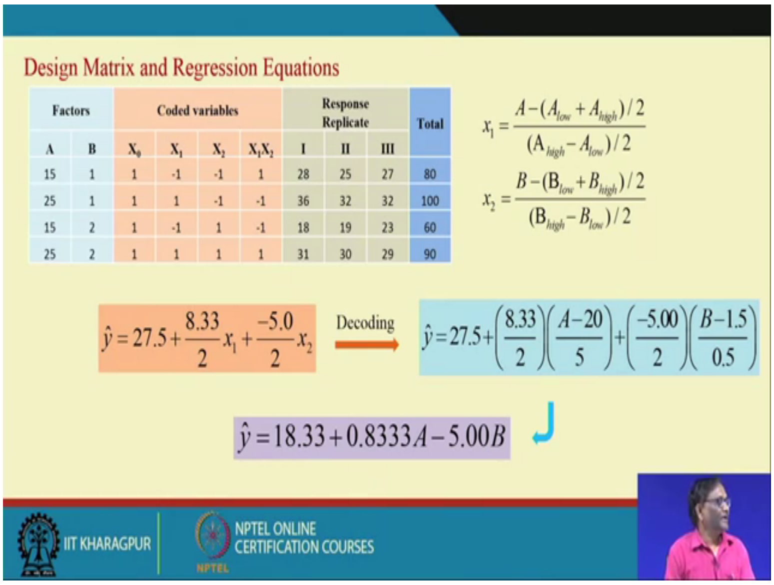
(Refer Slide Time: 14:25)



So, the advantage of using minus 1 plus 1 we are getting here, because of this minus 1 and plus 1, I have shown you using A by least square regression methods, that ultimately the effects are effects you have seen that this contrast used to compute the effect. So, we have shown you two is of 2 is of finding out the effects 1, is basically the from, the physical many point of view; that means, the effect of A, when A at low and A at high, the average difference that that way these are the things, and this is what is the data, and you are getting like this, and what we have done.

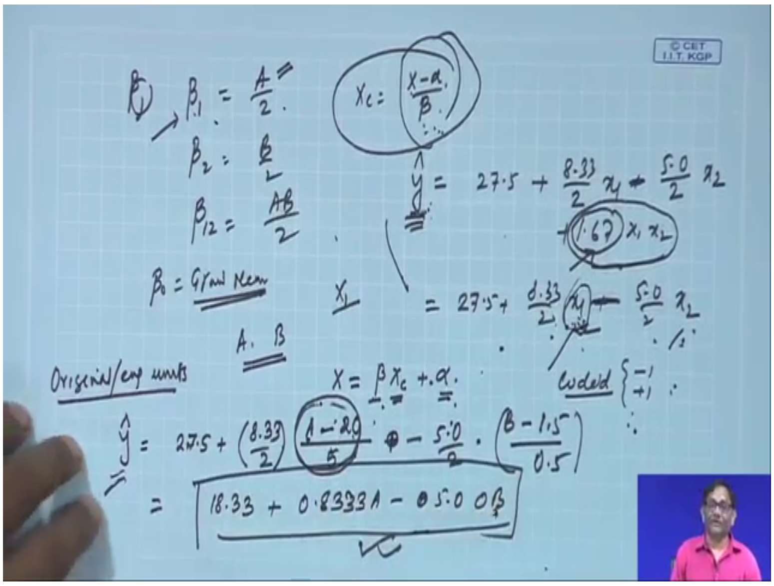
We have we have basically the shows the ANOVA table. And this table is known to you. I do not want to discuss further and you found out that that P value is A, not significant for Ab interactions, and this what we have done. This is just to remind you that this is the way things are done.

(Refer Slide Time: 15:37)



Now, from there straight way I am going to the regression equation. So, what happened, as you know that the regression equation beta j or other way I can.

(Refer Slide Time: 15:48)



In this case suppose beta 1 which will be A by 2 and beta 2 will be B by 2 and beta 1 2 will be Ab by 2. So, that you have seen earlier, because regression effect talks about unit change in X, what is the change in y a is talks about when we are changing minus 1 to plus 1, what is the difference, what is the change in y. So, as a result these will be A is divided by 2. Similarly other effects. So, as A and B are known and also the beta zero will be the grand mean, beta zero will be the grand mean.

So, that mean from the data that the simple average. So, you will you get if using this data, you will get this equation what about X 1 and X 2, because X 1 X 2 A B interaction is not significant. So, as a result this equation we got from the data given in this table. So, that mean the responds are phase, what you got this is 27.5 plus 8.33 by 2 into X 1 minus 5.0 by 2 into X 2 and we have not considered the Ab of it, which was basically effect wise, it was it was 1.67. So, that mean plus 1.67 ah. Then X 1 into X 2 and as you know we are using two symbols either A B C D or X 1 X 2 X 3 like this to represent the factors. So, they are interchange writing interchangeably.

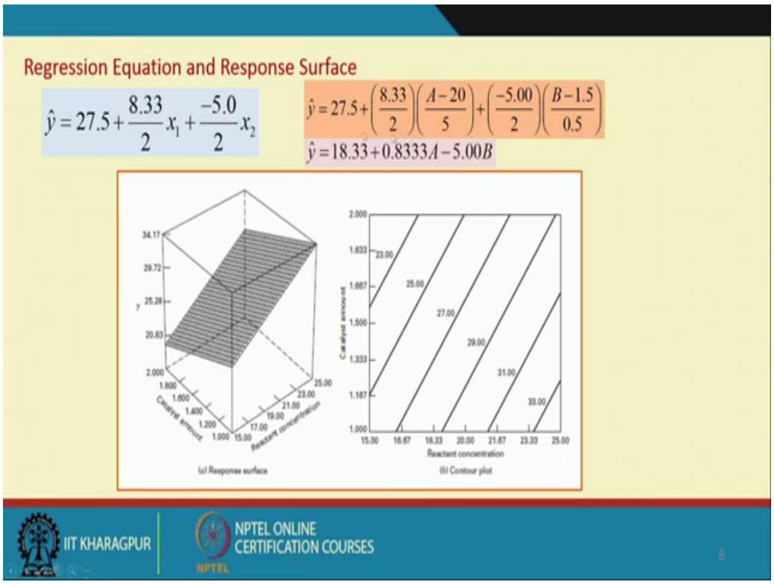
So, you must keep in mind this thing. Now this particular estimate value is not significantly from zero. So, thats why we have not included in this, in the final equation. So, the equation will be 27.5 plus 8.33 by 2 X 1 minus 5.0 by 2 X 2. Here these X 1 and X 2, they are coded. coded means minus 1 and plus 1, this is basically the. Now you the and the and the regression equation here it is with reference to the coded X 1 and X 2, you may be interested to go back 2 factor A and B in the original scale, instead of writing the coded scale for minus 1 and plus 1. So, I told you that if you want to go back, what you require, you have to use for that particular variable original unit beta X c coded unit plus alpha where alpha is Xh plus XL by 2 beta is Xh minus XL by 2 ok.

So, in this case, then we will use this. So, y we want in the original unit, original engineering units. What do you the regression equation? So, regression equation y cap will be 27.5 plus 8.33 by 2 into X 1 X 1. So, what you will write in place of X 1. So, I am writing in place of X 1 A, what is basically the, how do get what is X c X c. We are writing X minus alpha by beta and what is your beta is X h minus X L by 2. So, if we write these A, this is the original value minus alpha is 20 by beta is 5 ok.

So, this is my X c plus ah, no minus 5.0 by 2 into this will be B minus what was the alpha value 1.5 divided by 0.5. So, what is happening here from coded value to original values, you are getting. And if you simplify this one, you will get 18.33 plus 0.8333 A minus 5.00 B ok. So, you have coded values. Now in the ah, you just replace these by this ; that is what I have done. So, that is the beta value, and this X 1 is replaced by the original counterpart; that is X minus alpha by beta X. Here I am writing A or you can also write X 1 as logo that capital X 1, what we have write A and here we are writing B.

So, you are writing this, what is the difference between this two equations. If you want to use this equation you have to use ah. Suppose you want to get the predict the value of y given, given the factors values, then the original values you have to convert to coded value and use the coded value here for X 1 X 2 and get this value, but here what will happen, here what happen, you use the original values and then get the (Refer Time: 21:58) value ok. So, this is what is our; that is what the way coded transformation and original transformation from coded variable.

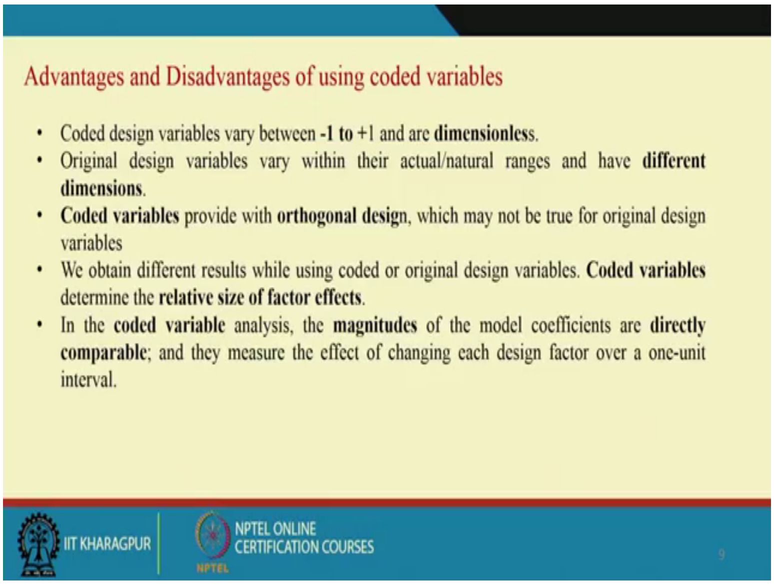
(Refer Slide Time: 22:05)



And now you know that using either the coded or the original one. Either these or these you can find out the response surface. Here what happen response surface is a flatter one the, the reason is basically there is no interactions here, and the controls are parallel to each other, and the how to interpret the control, I told you several times earlier ok.

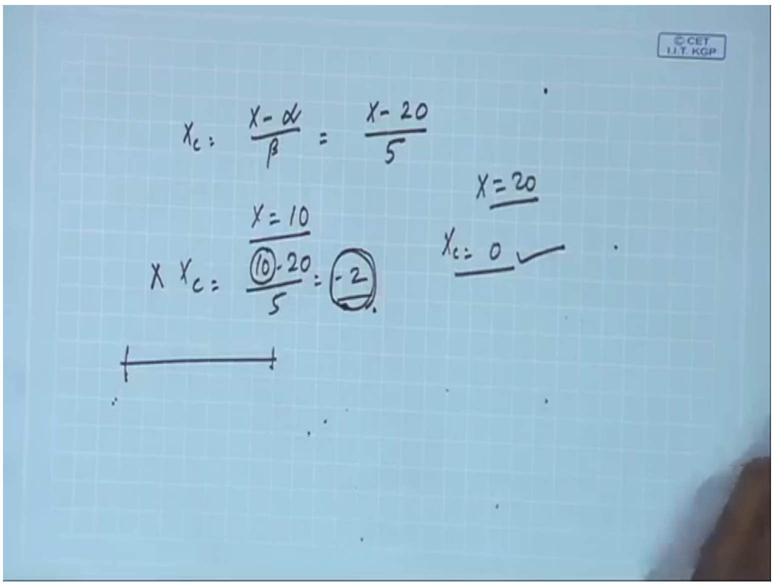
So, now, you know that the transformation from original to coded and again when you have the regression equation in coded variable, how it is to be converted to original variables that also you know. Now also know, you know that the original regression equation with original variables can be used to develop the response surface and the control plots and accordingly you can interpret and use the results for the purpose, it is, it was designed that sense.

(Refer Slide Time: 23:08)



Now, I will show you some of the advantage and disadvantage of using coded variable. You see coded design variables vary from minus 1 to plus 1 and they are dimensionless. Original design variable vary within their actual and natural ranges and different dimensions. So, this, so; that means, the variability for coded variables is from minus 1 to plus 1 in between anything is possible.

(Refer Slide Time: 23:45)

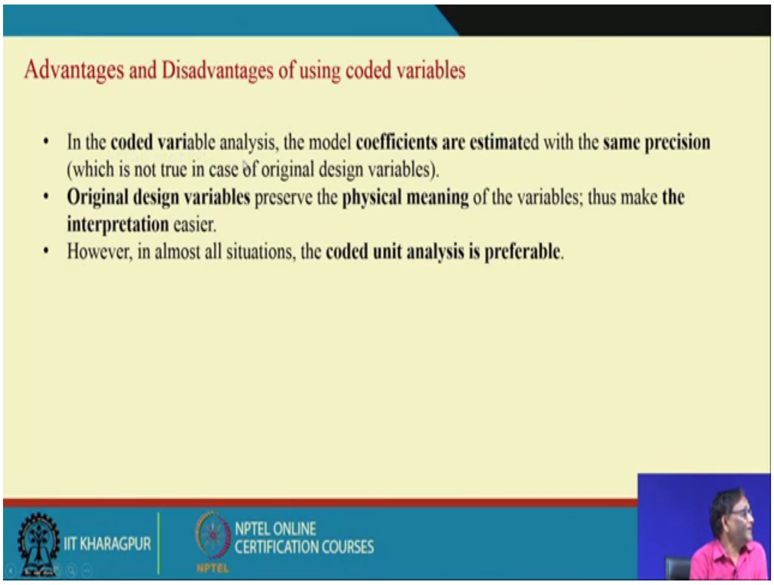


For example for example, what I have given you that X c equal to X minus alpha by beta and in our example for reactant concentration. These minus 20 by X minus 20 by 5, when X equal to X high you have got a 1, when X equal to X low you got minus 1, but X can be in between minus 1 to plus anything. Suppose it is not the minimum 1 like this one, 15 or 25 percent. Let X equal to 10 what will happen to coded value; that is 10 minus 20 by 5, it will become minus 2, suppose, no no X is 20, then coded value become zero ok. So,, but its range is always what I want to say that ah, as this 10 is not within the A range.

So, this is, this 10 is not within the range, this is not considered. You are basically considering low to high, this is the range of experiment is A, so minus so on. So, this we do not consider. This one will consider second. So, as they dimensionless comparison is easy, second one is that that original variable have different dimensions. I want to compare that 1 unit changes in reactant concentrate with 1 unit change in your other one, that amount of catalyst. So, the, as the ah, as the units are different. So, the meaning is different ok. Now coded variables provide with orthogonal design, this is very very important. This provides you orthogonal design, but which is, may not be true for original design variables.

The orthogonality means, what happened that the output out of to the coded variable, the values if you consider they will become zero. So, the independent party will be important here, we obtain different results while using coded or original design. Variables, coded variable determine the relative size of factor effects, relative size of factor effects in the coded variable analysis, the magnitude of the model coefficient the directly comparable. And they measure the effect of changing each design factor over 1 unit interval, which is not possible, which is not possible in original one.

(Refer Slide Time: 26:45)

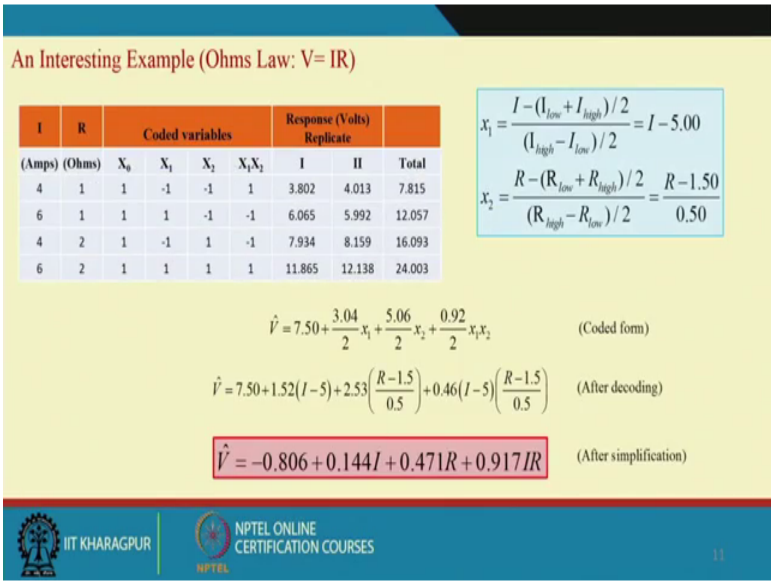


In the coded design variable, the model coefficient are estimated with the same precision standard error of the model, estimate will be same irrespective of what is the, what way it is measured in the original units, if you code it, and then each of the that regression estimates, there will be there will be estimated with same precision standard error, same standard error.

Original variables preserve the physical meaning of the variable; that is the, that is the main advantage of using original or engineering units, because they preserve the actual meaning of the variable. Thus make the interpretation easier, but when you transform to coded variable; that is last, it become dimensionless.

So, you have to be careful while interpreting the results, whatever may be the case, whatever may be the thing, the coded variable is advantages and it is preferable in analyzing the experimental data. Particularly in the 2 to the power k design, k factorial situation, because you have already saying that even if you have the equation, regression equation or response surface in coded variable terms, it can be converted to original variables quite easily.

(Refer Slide Time: 28:27)

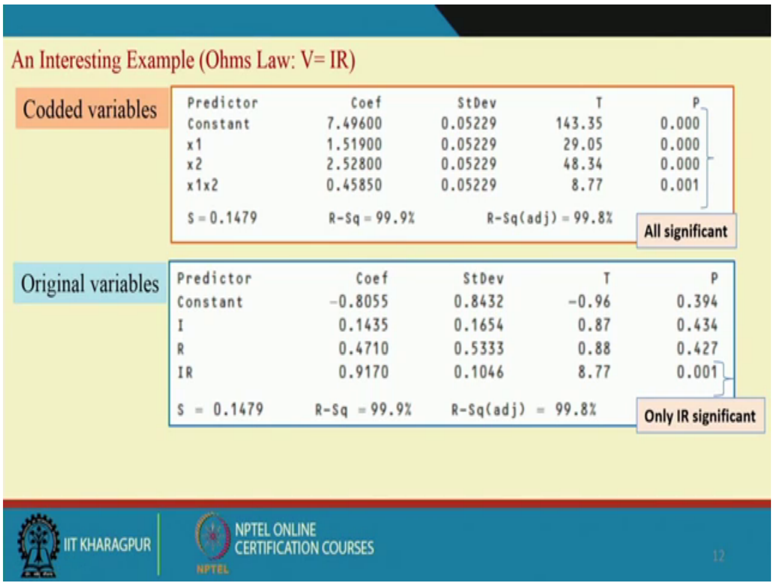


So, now I will show you another example that the, if you go for coded variable and keep original variable the, how the regression estimates changes ok. Suppose I, here we all know V equal to I R is the ohms law, V is the voltage, I is current and R is resistance. And suppose a laboratory based experiments was conducted and then keeping that I in at 2 labels 4 amperes 6 amperes and resistance at 2 labels 1 and 2 ohms. So, then this is basically 2 to the power 2 design, and this is this is basically design may takes for regression. And suppose that the response, the response volts is replicated twice, then these are the responses obtained and this is the total.

So, what happened, basically from here, I have shown you how this codification is done. This is now known to you, and now with coded variable and ultimately the regression variable equation is found out, and you see that the regression equation with coded variable is voltage equal to 7.5 plus this by 23.04. This is the equation, this is the equation. Now you can convert it to original variables, also find and you will get this. This is equation fantastic. Now actually what you do now.

So, what you are now getting here, if you use coded variable what you are getting here, not V equal to IR, what is the ohms law, what you are getting V equal to this plus this plus this plus this. So, here there is a interaction term is there, but now if you ; that means, what will happen if this in order to preserve this law; that means, it is constant and other two are the terms, this should be, this should be insignificant statistically, then will not consider them, will consider only I interesting terms and will get the results.

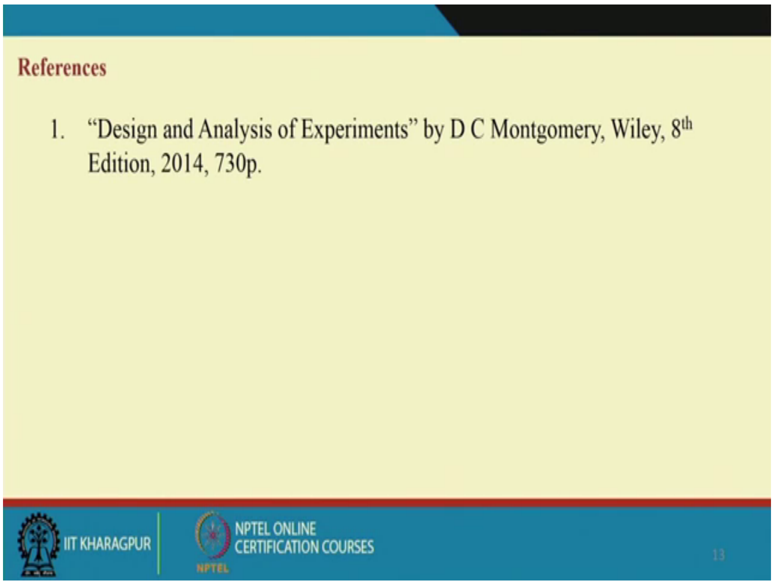
(Refer Slide Time: 30:36)



Now, see, you will see the output of the regression using coded variables and you see this side; that means, the constant is 7.5 X 1 is 1. 52 X 2 is 2.52, everything alright R square is very high and T values are very high. So, all the parameters are significant, that mean the original, the main effects significant inter resistance is significant ok. Now if you do the same thing in the original units, what is happening here? You see only interaction effect is significant, but the other things including constant they are insignificant. So, that mean now what will happen after that, you do not consider the intercept and the main effect only consider the interaction term and do the regression once more, you will find out that I R will be this value, will be almost 1. So, that mean that is satisfying the physical law here that physics law, but which is not done, satisfying here.

So, these type of issues will be there. So, when you basically analyze the experimental data is in coded variable, you have to be very careful about the about the physical meaning of it. And another issue is that basically when there are such physics behind this one, something is known. So, this also that physics is to be taken into consideration ok, while explaining the outputs and this is true, not for every models, not only related to the regression here.

(Refer Slide Time: 32:22)



So, lecture is prepared based on the book by design and analysis of experiment by Montgomery, and in fact, we have prepared most of the lectures, almost all from this book.

Thank you very much.