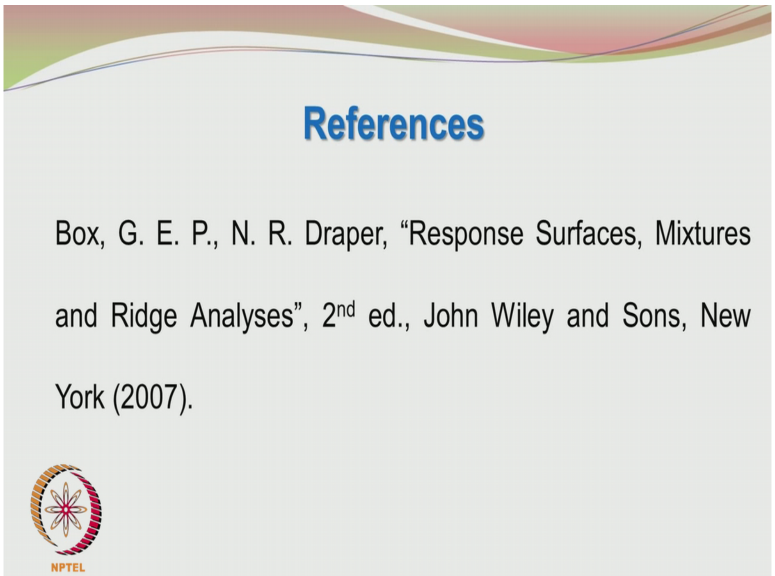
Statistics for Experimentalists
Prof. Kannan. A
Department of Chemical Engineering
Indian Institute of Technology - Madras

Lecture - 44
Orthogonal Model Fitting Concepts - Part A

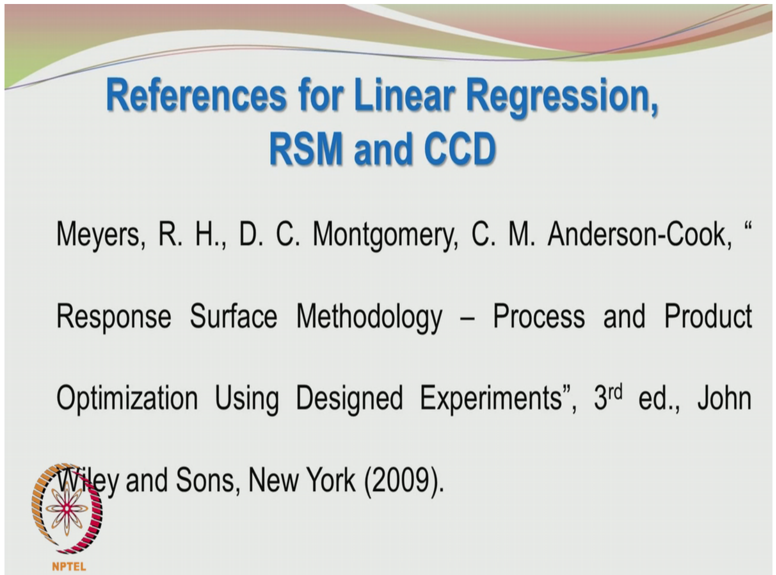
Hello, in this lecture 28, we will be looking at concepts pertaining to orthogonal models.

(Refer Slide Time: 00:23)



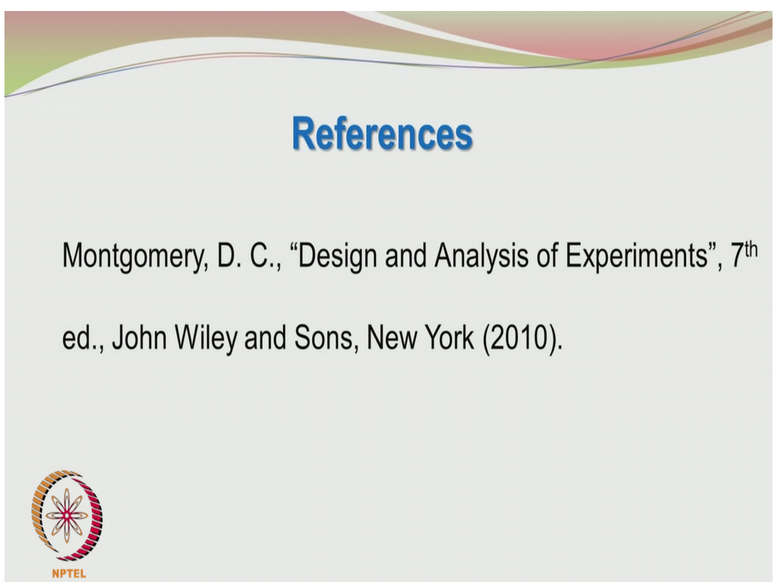
The references are given in the current slide and the slide to follow. The first book is by Box and Draper, Responses Surfaces, Mixtures and Ridge Analyses, it is slightly advanced book for people who want to deepen their understanding and knowledge on the subject may refer to this book.

(Refer Slide Time: 00:43)



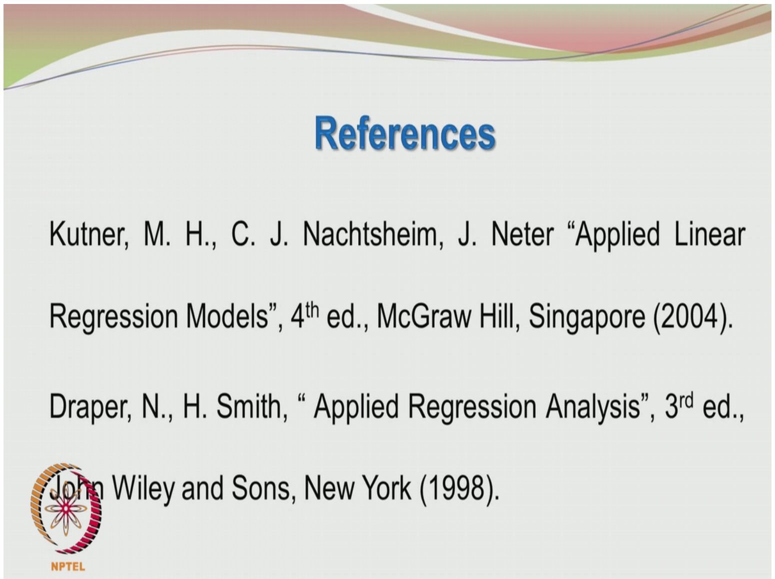
The other book where this slide material is mainly based upon is the one written by Meyers, Montgomery, Anderson-Cook, the title of the book is Response Surface Methodology Process and Product Optimization Using Designed Experiments, 3rd edition, John Wiley and Sons, published in 2009. This is an excellent book, where the concepts are explained in a very clear manner, it is not very mathematically regress, so people with the basic knowledge in linear algebra should be able to follow the material given in this book quite easily.

(Refer Slide Time: 01:28)



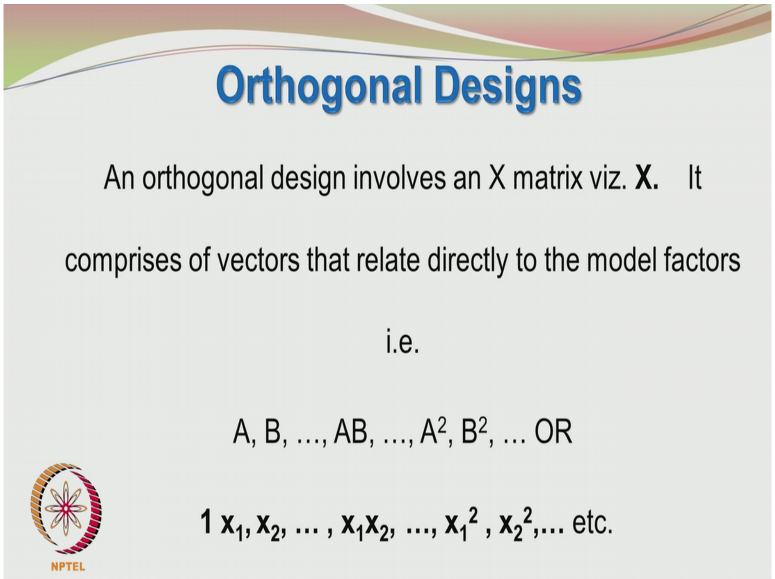
And of course the other book is the book written by Montgomery Design and Analysis of Experiments, 7th edition, John Wiley and Sons, New York 2010.

(Refer Slide Time: 01:38)



Other few references are by Kutner, Nachtsheim and Neter, Applied Linear Regression Models. And Draper and Smith Applied Regression Analysis. So there are quite a few interesting books written on this subject.

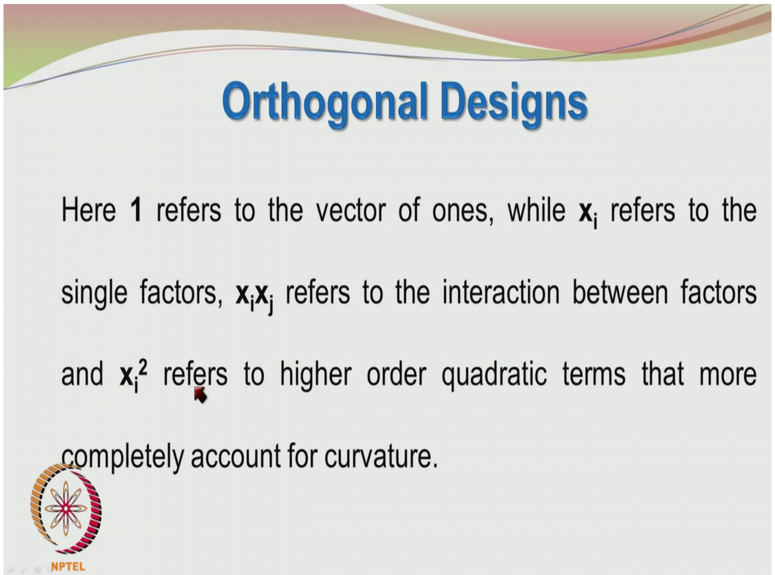
(Refer Slide Time: 01:48)



Now let us come to orthogonal designs. What is an orthogonal design? An orthogonal design involves a matrix X, and it comprises of vectors that relate directly to the model factors. So if your model is having the contribution from factor A, factor B, interaction AB, A squared, B squared, then the model will look something like this, it will have 1 x1, x2, x1 x2, x1 squared, x2 squared, these are all column vectors.

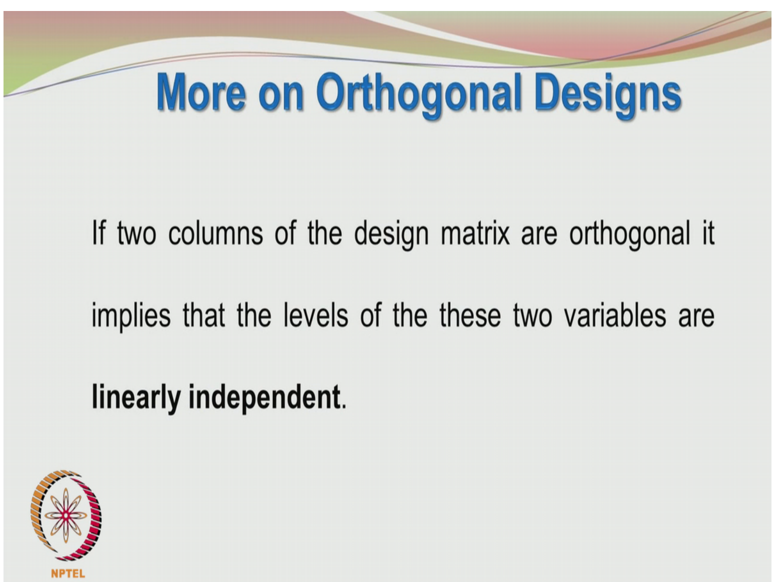
So when you enter the model in the matrix form, the first column could be the matrix of 1's, the second column would be the values corresponding to the settings of factor x1, similarly, the settings of factor x2, and then settings corresponding to x1 x2, then x1 squared, x2 square etc. depending upon the complexity and length of the model.

(Refer Slide Time: 03:12)



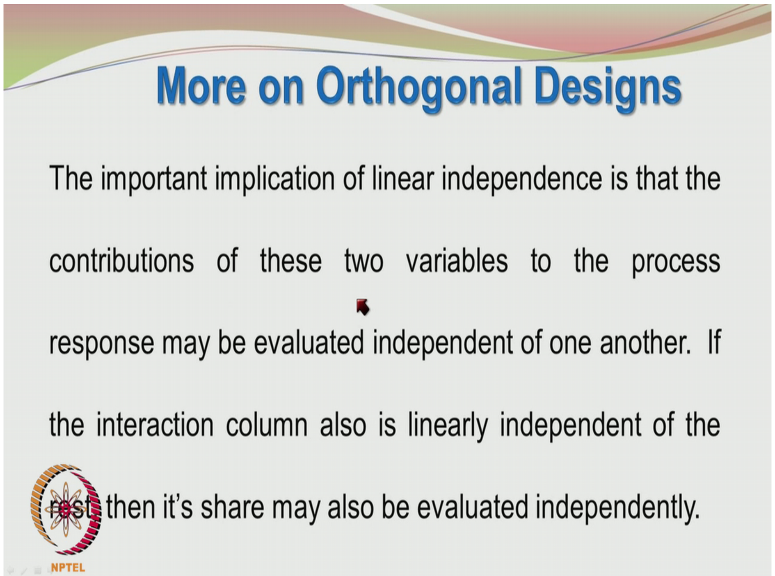
So here 1 refers to the vectors of 1's, and xi refers to the single factors, xi xj refers to the interaction between factors, and xi square refers to higher order quadratic terms that more completely account for the curvature.

(Refer Slide Time: 03:34)



Now if 2 columns of the design matrix are orthogonal, it implies that the levels of these 2 variables are linearly independent.

(Refer Slide Time: 03:40)

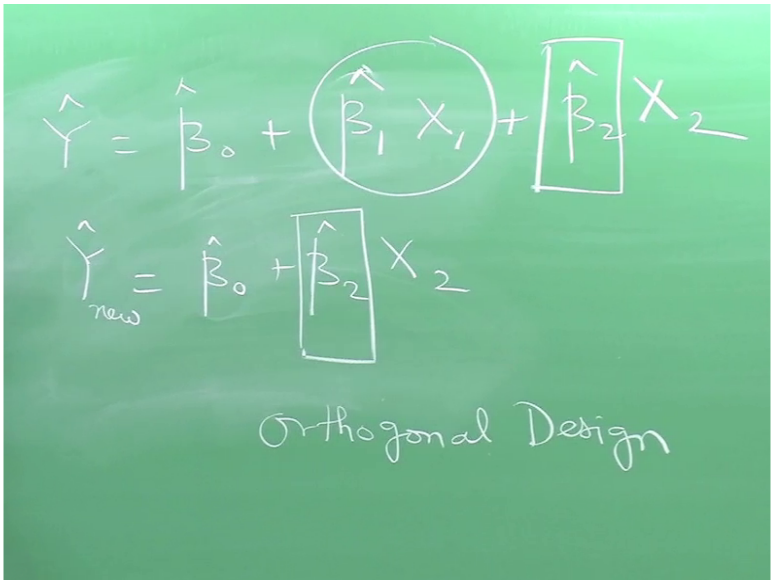


The important implication of linear independence is that the contributions of these 2 variables to the process response may be evaluated independent of one another, if the interaction column is also linearly independent of the rest then its share may also be evaluated independently. So this is a very useful concept in regression analysis and design of experiments, in planned design of experiments the way the experimental points are set, it makes the design orthogonal in nature.

So we can say that the factor A contributes to the model independent of how factor B is contributing to the model. Similarly, even if you have interaction AB in an orthogonal design, AB contributes to the model independent of how A and B contribute to the model, because the column vectors in the matrix are linearly independent of one another. So this is a very beautiful concept, what I am trying to say here is suppose you will have a model.

And in the model you say that the Y hat or Y predicted =beta 0 hat +beta 1 hat x1+beta 2 hat x2, then you will find the coefficients beta 1 hat and beta 2 hat. Suppose you for some reason decide to omit the factor A altogether in the orthogonal design, then your model will be beta hat 0+beta hat 2 x2, the value of beta hat 2 will be same as it was in the previous full model. So let me go to the board and explain what I mean by that.

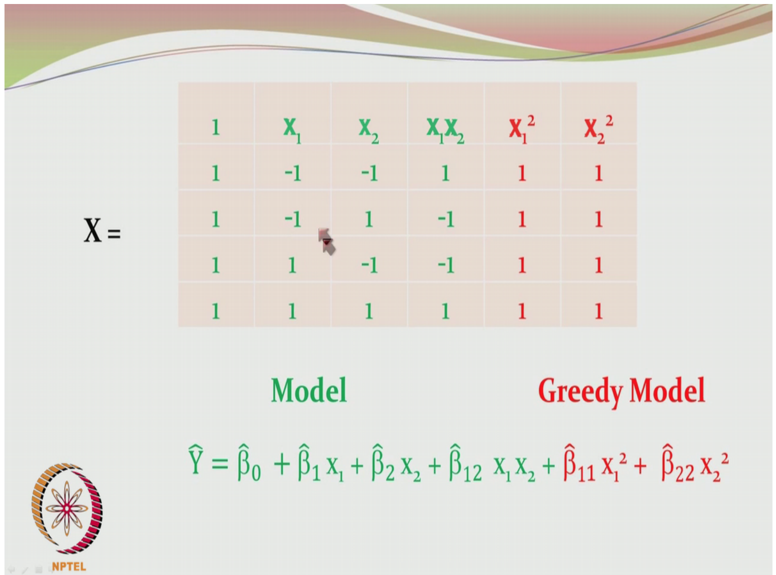
(Refer Slide Time: 05:52)



Suppose you have an existing model okay, so this is the existing model and you want to try a new model where you put only beta hat 0+beta hat 2 X2, you do not have the contribution from beta hat 1 X1, then you will find that the parameters beta hat 2 in this old model is =the parameter beta hat 2 in the new model, this is because of the orthogonal design. So the contribution of factor X1 and the contribution of factor X2 are evaluated independent of each other.

Their contribution is independent of each other to the observed response, this is only true in the case of orthogonal design, so you can see the advantages of this. We have already seen examples of the orthogonal design in our examples set on regression analysis.

(Refer Slide Time: 07:47)

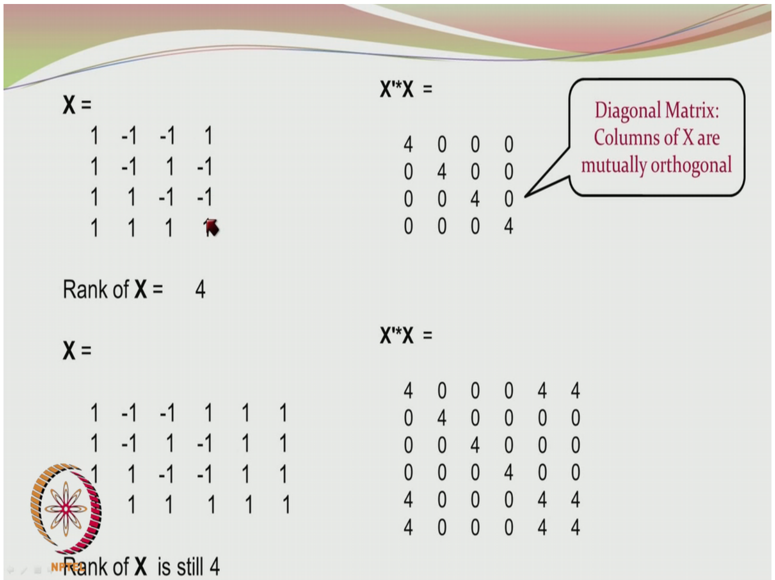


So let us look at the X matrix and you can see that this is 2 power 2 design, this is the column vector of 1's, this would be the settings corresponding to factor 1 -1, -1, 1 1. This would be the settings corresponding to factor 2 -1, +1, -1, +1. And you will have x1 x2 and that would be product of these 2 columns -1*-1 would be+1,-1*1 would be -1, 1*-1 is -1, 1*1 is 1, x1 squared is obtained by squaring these terms in this column, x2 squared is obtained by squaring these terms in this column okay.

So we are having essentially 1, 2, 3, 4 columns, and we also have the additional column shown in red x1 squared and x2 square. Let us now develop the model Y hat =beta hat 0+beta hat 1 x1+beta hat 2 x2+beta hat 12 x1 x2, you can see that the green colored model is having 1, 2, 3, 4 parameters, and that would be the maximum number of parameters which can be estimated from this design, because you are having only 4 independent settings.

And so the maximum number of parameters would be 4, but if you have a greedy model and you also try to evaluate the additional terms like beta hat 11 and beta hat 22, then can you run into trouble because you do not have 6 independent settings to obtain these 6 parameters.

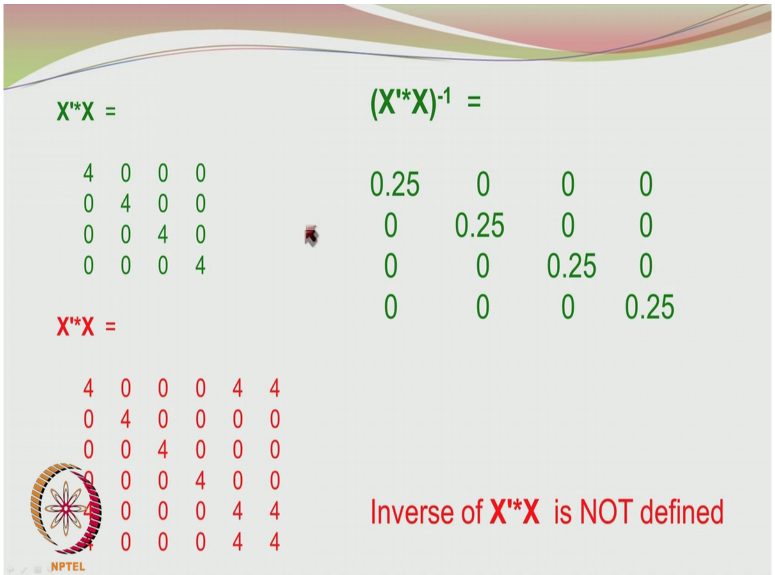
(Refer Slide Time: 09:48)



Now let us look at the X matrix, this is X matrix again exactly same as the X matrix corresponding to the green colored vectors. So the rank of this X matrix=4, and X prime X would be taking transpose of the X matrix and multiplying the transpose with X matrix again, you will get a diagonal matrix having 4 along the diagonals. And when you look at the matrix which is made up of the entries corresponding to x1 squared and x2 squared.

Then you have in addition to the old X matrix also these 2 additional columns of 1's, and of course you can see that this x1 squared is exactly same as the vectors of 1's, x2 squared also replica of the vector of 1's. Now when you look at this augmented matrix X, the rank of X is still 4, and when you look at the X prime X you get a matrix as shown here, and the problem is we run into trouble when we take the inverse for this particular matrix.

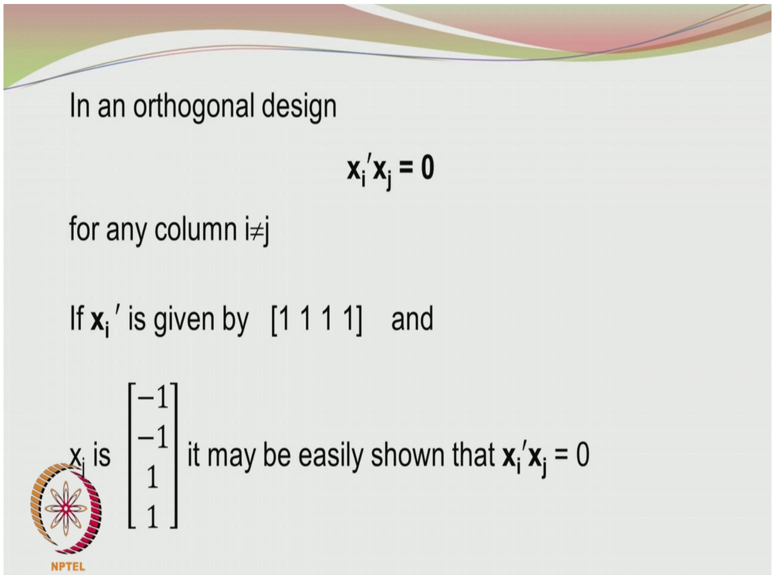
(Refer Slide Time: 11:22)



So you have X prime X inverse, for this case has 1/4, 1/4, 1/4, 1/4 along the diagonals, there is no problem in evaluating this particular matrix. But if you try to take the inverse of this X prime X matrix as given here then you will find that it is not defined, so what I am trying to say here is do not try to expand the scope of your model when the number of independent settings in the model is limited.

In the current case of a 2 power 2 design, we had only 4 independent settings and so we could estimate only 4 independent parameters, and the moment we try to increase the scope of the model by adding the quadratic terms, then we found that we cannot estimate the parameters because the X prime X inverse matrix was not defined in such a situation.

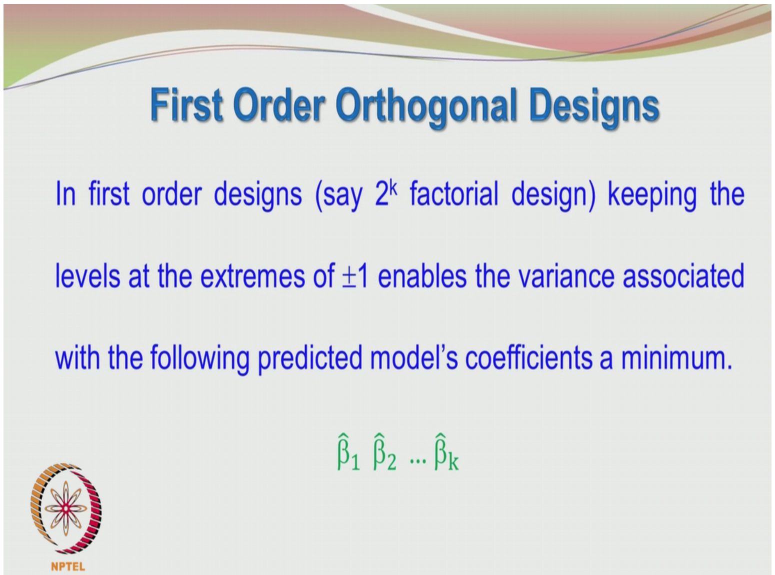
(Refer Slide Time: 12:25)



So what is an orthogonal design if you take the transpose of any vector and then multiply the transpose with the another vector a different one, suppose I am taking the ith column vector take a transpose of ith column vector, then I pre-multiply the transpose of the ith column vector against the column vector xj i !=j I will get 0. So if you take this particular case, if I take the transpose of this it will be a 1 row by 4 column vector a transpose of a vector, and that would be 1, 1, 1, 1 horizontally.

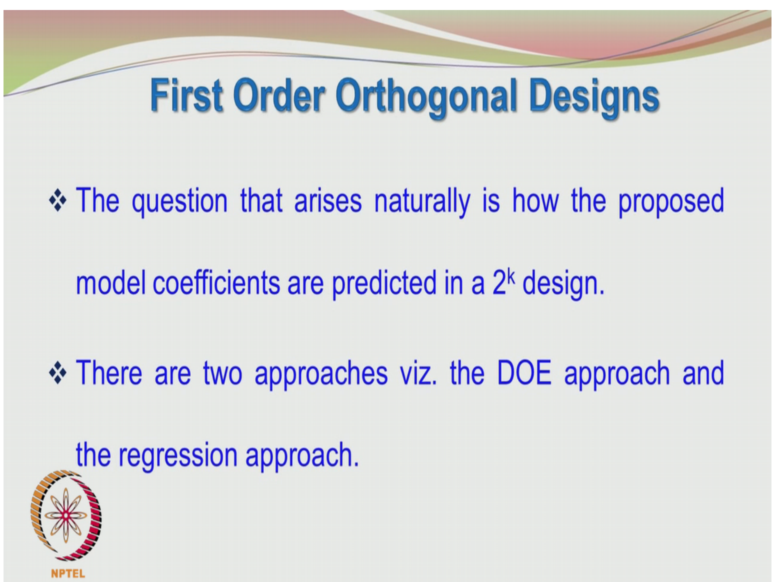
And then I multiply this with the column vector here -1, -1, 1, 1, I will find that it will be a -1, -1, -2, +1, +1,+2, so -2+2 will become 0. The same concept applies for any binary combinations of vectors 1 being a transpose and another being the regular column vector, if I multiply that too, I will get 0. So this is the property of the orthogonal design, the important precautions you should take is that you should not take i=j. Then the negative and positive elements would be squared, and all of them would be positive and the sum will !=0.

(Refer Slide Time: 14:03)



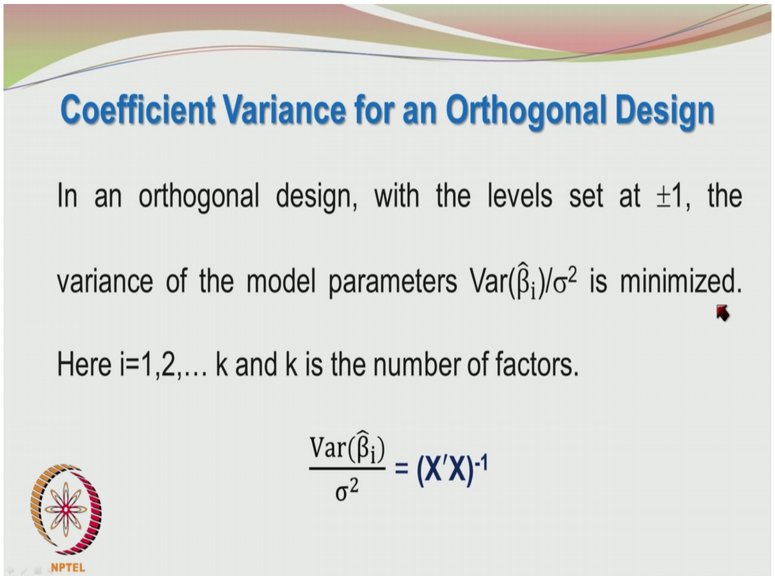
Now let us look at first order orthogonal designs, let us take care 2 power k factorial designs where k is the number of factors keeping the levels at the extremes of + or -1 enables the variance associated with the predicted model's coefficients at a minimum. So this is another advantage of factorial design, we are keeping the design settings at the extremes -1, +1, -1, -1, +1, -1, +1, +1. So we are having a design space the experimental settings are kept at the extreme ends, and this helps to minimize the variance of the estimated parameters.

(Refer Slide Time: 14:52)



So the question that arises naturally is suppose you have an orthogonal design, how do you go about estimating the parameters. Well you can do it in 2 ways, 1 you can use the regular design of experiments approach, and estimate the effects and then the model positions, or you can use the matrix approach in linear regression.

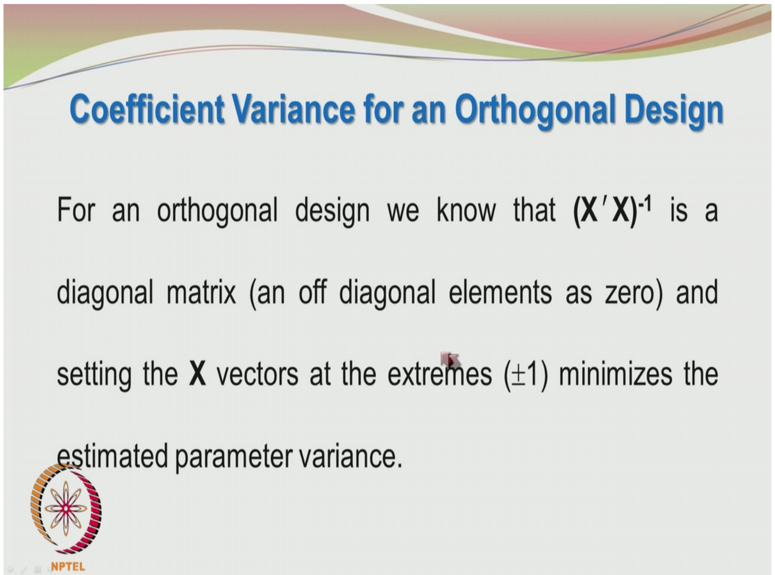
(Refer Slide Time: 15:15)



So as I said earlier in an orthogonal design with levels set at + or -1 the variance of the model parameters variance of beta hat i/sigma square is minimized. Here, i =1, 2, so on to k, and k is the number of factors, here obviously you are not including that intercept beta hat 0, and if you can include that also you will have k+1 which =P number of parameters.

The variance of the predicted coefficient is given by the variance of beta hat i and that you scale it by sigma squared you will get the variance covariance matrix X prime X inverse. So this X prime X and the X prime X inverse matrix are very, very important.

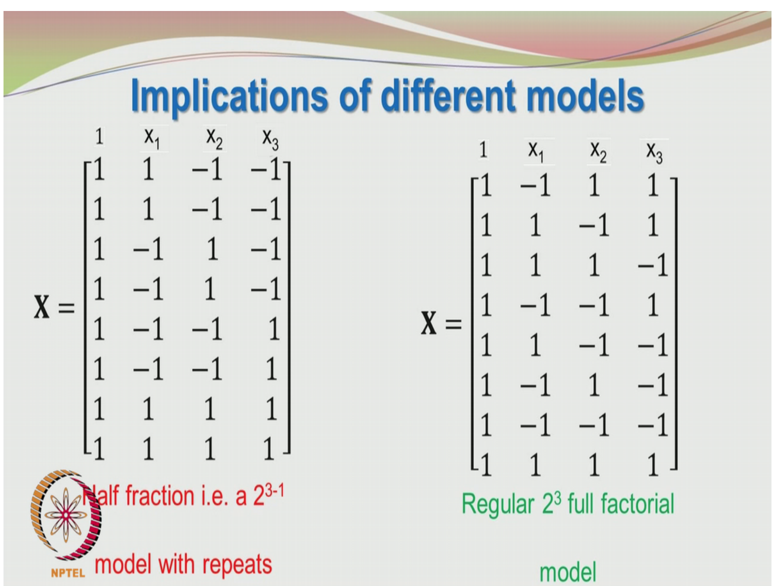
(Refer Slide Time: 16:17)



And for an orthogonal design, we know that the X prime X inverse is a diagonal matrix with the off diagonal elements are 0, and setting the X vector at the extremes minimize the estimated parameter variance. So we have already seen that for an orthogonal designed such as the one we considered previously, this is an orthogonal design that 2 power 2 design, and you can see that the X prime X matrix is a diagonal matrix.

And the X prime X inverse matrix is also diagonal matrix at 1/4, 1/4, 1/4, ¼, so this arrangement minimizes the variance of the estimated regression coefficients.

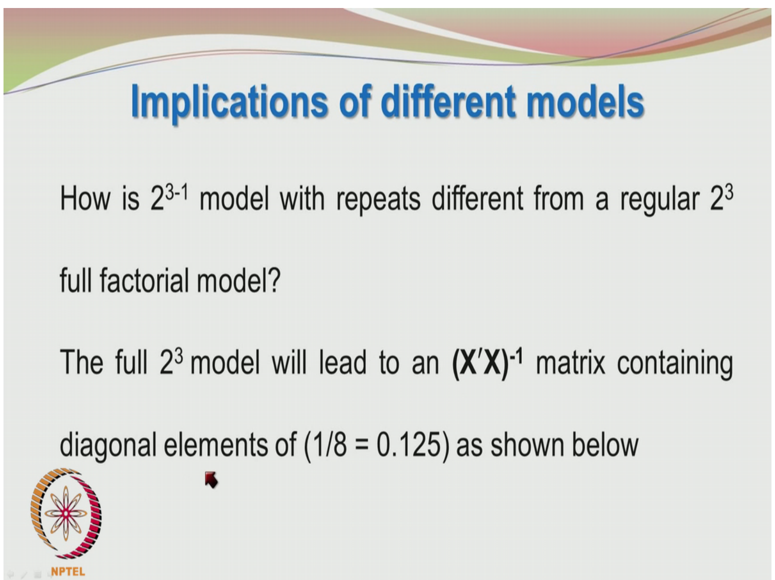
(Refer Slide Time: 17:07)



Now let us look at the implications of different models. So the first model we are considering on the left hand side is a half fraction of 2 power 3-1 design that means we should be having 4 runs only, but it is clear that we are having 8 runs here, the reason for that is quite simple we are repeating the 4 experiments of a 2 power 3-1 design. So you can see that the first 2 rows are identical, the 3rd and 4th rows are identical, the 5th and 6th rows are identical and the 7th and 8th cross are identical.

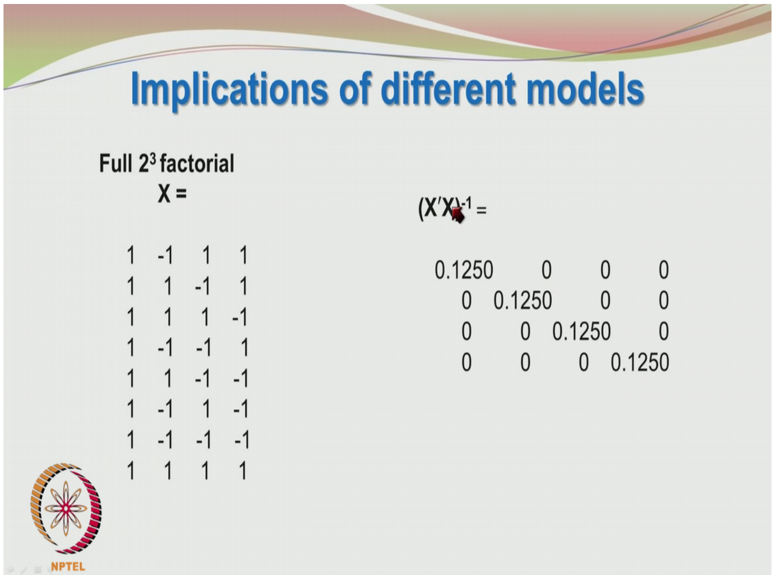
That means each experimental setting in the 2 power 3-1 design is repeated. We also look at the regular 2 power 3 full factorial model, here you have all the possible settings for a 2 power 3 factorial design, and the main thing to notice is there are no repeats. On the other hand, for 2 power 3-1 design you had repeats, but on the other hand you would not have the full set, you do not have the full set of independent settings possible.

(Refer Slide Time: 18:22)



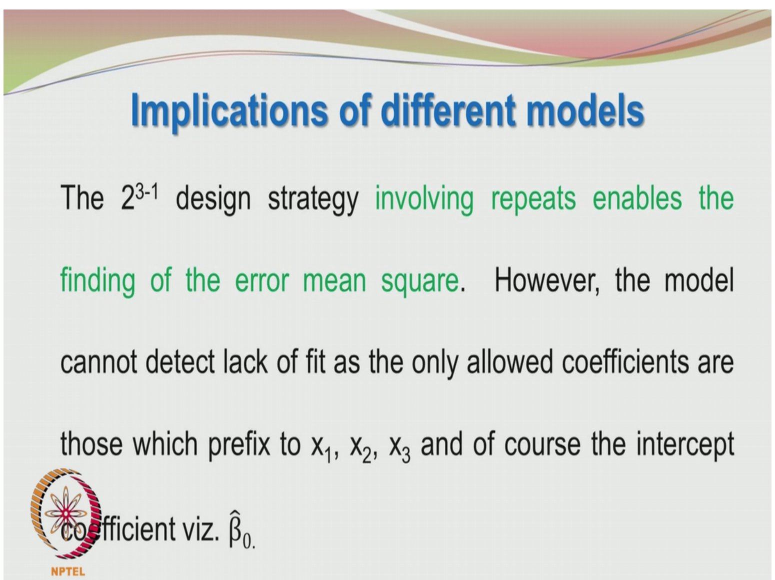
So when you look at the 2 power 3 full design it will lead to an X prime X inverse matrix containing diagonal elements of 1/8 which =0.125.

(Refer Slide Time: 18:38)



So you are having an X matrix corresponding to a full 2 power 3 design, if I take X prime X inverse matrix which is the straight forward think to do by now, I will get I think 1/8, 1/8, 1/8, 1/8 along the diagonals, and off diagonal terms would not be present.

(Refer Slide Time: 19:01)

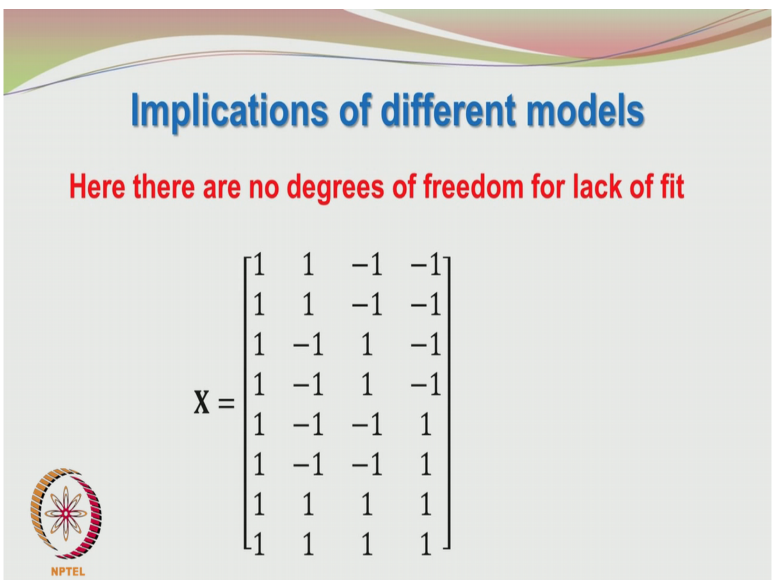


Right, but what is the advantage of here 2 power 3-1 design? Why do we have to go for 2 power 3-1 design? May be performing 2 power 3 full design was expensive, so the experimenter decided to go for 2 power 3-1 design. On the other hand, he still conducted the 8 experiments the same number as a full 2 power 3 design, so the experimenter was more focusing on the estimation of the pure error.

He probably had some insight into the model based on previous experience, and so instead of wasting time and resources and doing the complete set, he probably wanted to get more information on the pure error term. So a 2 power 3-1 design strategy will unfortunately not enable you to find the full set of parameters that are possible from a full 2 power 3 design, a 2 power 3 design has 8 independent settings and so you should be theoretically able to estimate 8 parameters of the model, so you can build your model up to 8 parameters.

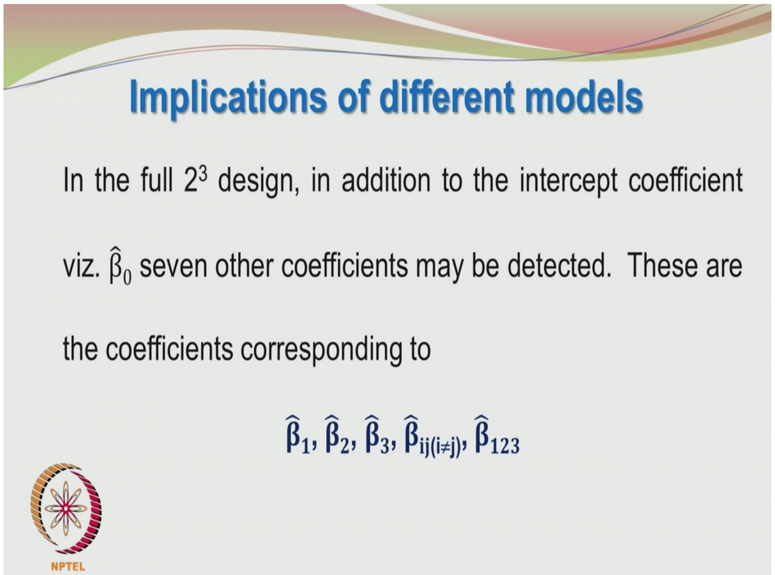
However, for 2 power 3-1 design you are having only 4 independent settings, and so you will be able to estimate only 4 parameters. So you are losing the ability to estimate 4 parameters, and so your 2 power 3-1 design strategy will not have any degree of freedom for lack of fit, if you estimate the 4 parameters. But on the other hand, if you look at the 2 power 3 design and you are estimating only 4 or 5 parameters, then you have sufficient degrees of freedom to test your model for lack of fit.

(Refer Slide Time: 21:02)



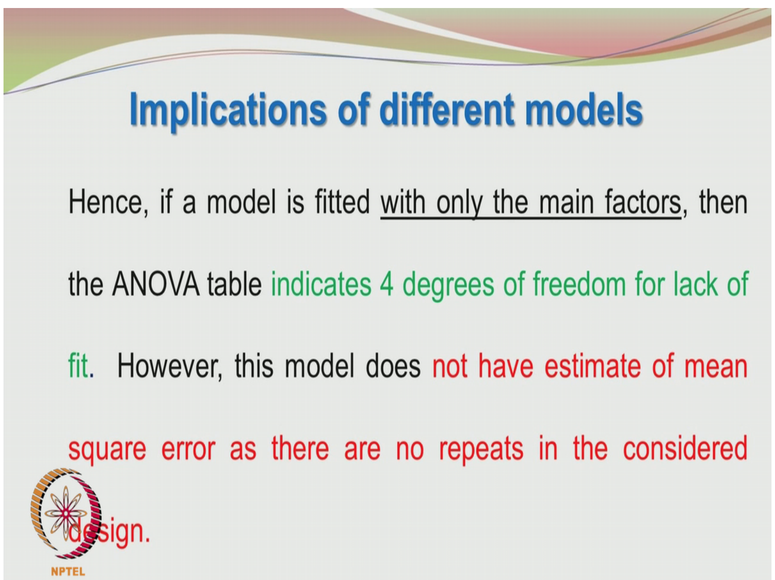
So when you look at the 2 power 3-1 design where each design setting is repeated you do not do the full design you do only half the complete design but you do repeats. So in this if you are estimating 4 parameters since there are only 4 independent settings, you do not have any degrees of freedom for testing the lack of fit.

(Refer Slide Time: 21:27)



So in the full 2 power 3 design, in addition to the intercept coefficient namely beta hat 0, 7 other coefficients may be detected these are the coefficients corresponding to the main factors beta hat 1 beta hat 2 beta hat 3, and then you can also look at the interactions between the 2 factors beta hat ij i !=j, and then ternary interaction term beta hat 1, 2, 3. So you can estimate the main effects, and you can estimate the binary interactions and you can also estimate the ternary interaction between the factors. So these are the 8 coefficients including beta hat 0.

(Refer Slide Time: 22:06)



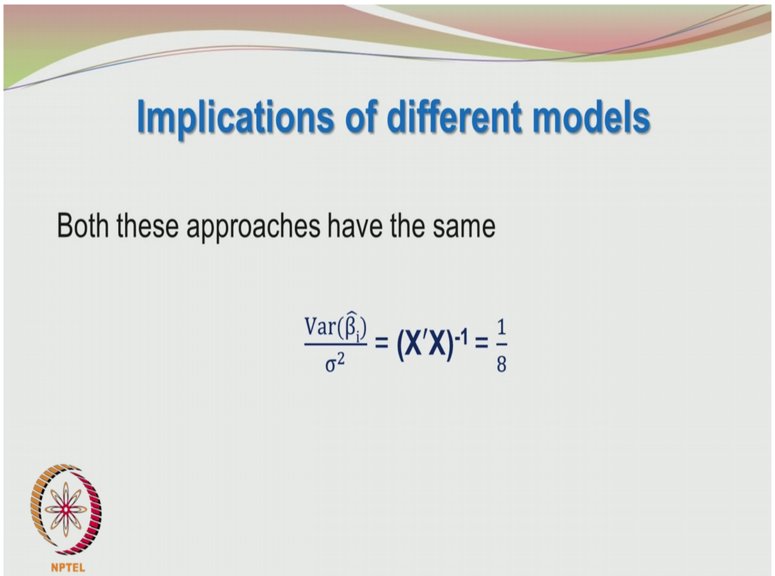
And if you are fitting a model with only the main factors that out means out of the 8 possible parameters you are estimating only 4 beta hat 0, and then the regression coefficient corresponding to factor A, regression coefficient corresponding to factor B, regression coefficient corresponding to factor C that would make it as only 4 parameters. And in such a case that our table indicates 4 degrees of freedom for lack of fit.

So 8-4=4, 8 parameters are maximum possible but you have estimated only 4, so you have 8-4=4 degrees of freedom for testing the lack of fit. But what is the drawback in the full 2 power 3 model you exhausted all the settings all independent settings have been exhausted, and that consumed all the full 8 runs, so you are not in a position to repeat your experimental settings. Suppose your management says that fine you can do a maximum of 8 experiments.

So one group proposes a 2 power 3-1 designs with repeats, on the other hand there is another group which goes in for a full 2 power 3 model, on one hand the group which proposed 2 power 3-1 design will have an estimate of the pure error but it not be able to estimate all the interactions including the ternary interaction. Whereas the group which went in for a 2 power 3 design will not be able to get an idea on the pure error.

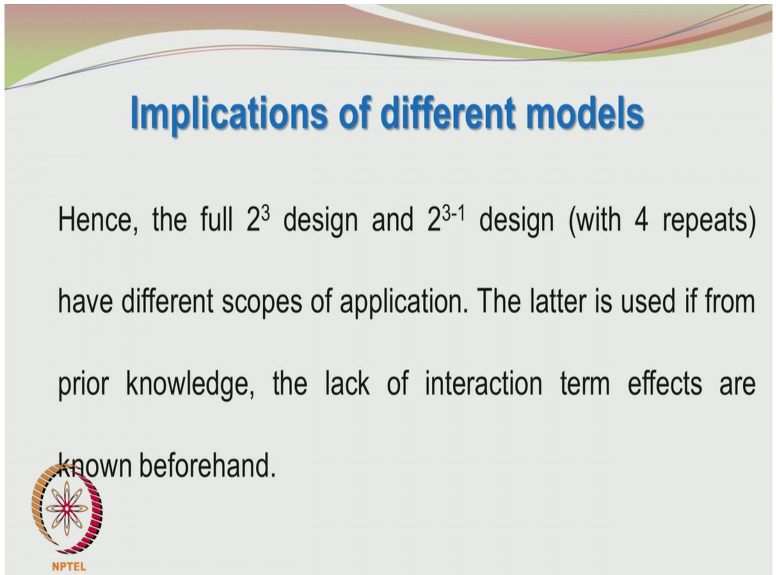
And on the other hand, it would be able to estimate as many as 8 parameters. So which model is good that depends upon the process and the prior knowledge on the process you have.

(Refer Slide Time: 24:10)



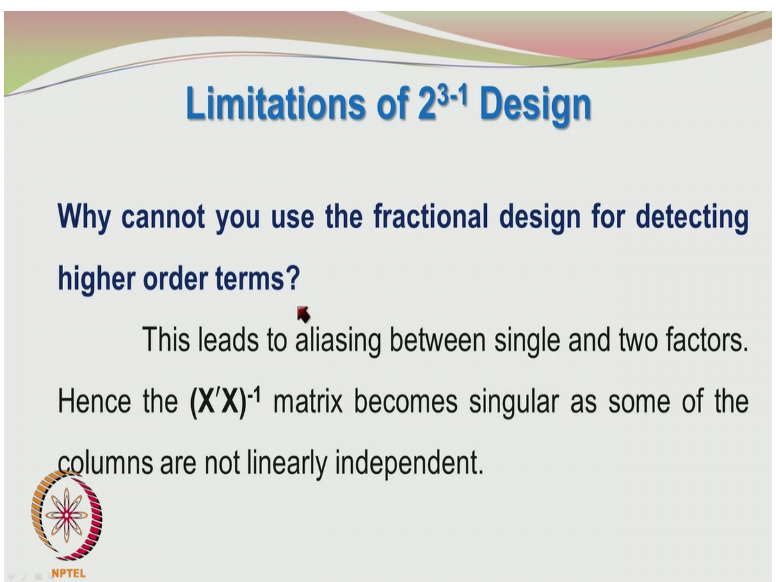
So when you look at the 2 approaches, they have the same variance of beta hat i/sigma square, because the X prime X inverse matrix is same in both the cases, and that is actually 1/8.

(Refer Slide Time: 24:26)



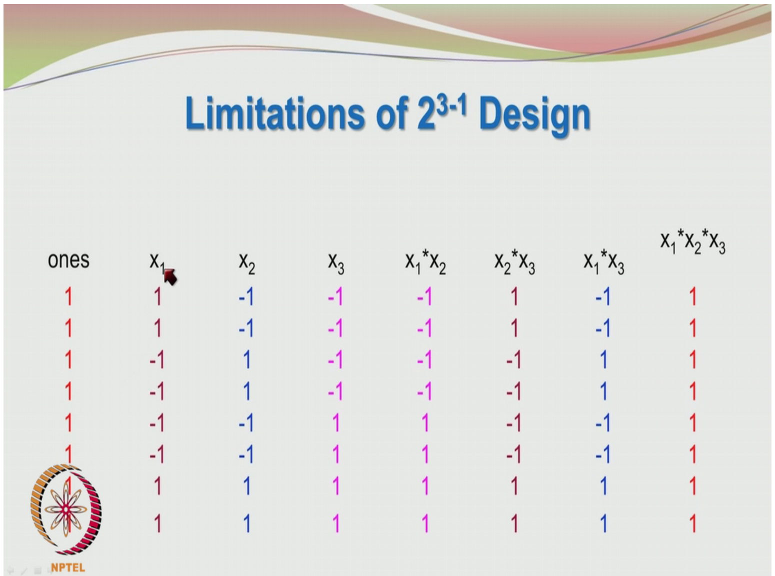
So the 2 power 3 and 2 power 3-1 design with 4 repeats have different scopes of application, the latter is used the 2 power 3-1 design is used if you have prior knowledge or experience that the lack of interaction term effects are known beforehand, so you know previously that there is no interaction between the factors of your model, so the interaction terms are neglected. And if the binary interaction terms are neglected, then there is even a very little chance the ternary interactions would kick in.

(Refer Slide Time: 25:09)



So let us expand on this topic a bit more the query is, why cannot you use the fractional design for detecting higher order terms? So the main problem is there is aliasing between single and 2 factors, and the X prime X inverse matrix becomes singular as some of the columns are not linearly independent.

(Refer Slide Time: 25:35)

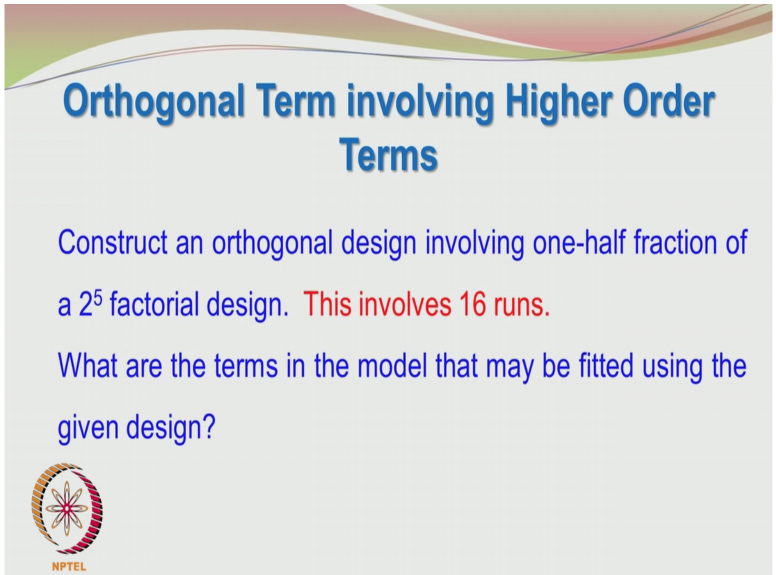


So let us look at 2 power 3-1 design, so you have the 1's, you have the settings corresponding to factor 1, settings corresponding to factor 2 and settings corresponding to factor 3. So obviously the runs are repeated so you are doing 8 runs, but you have only 4 independent settings. Now I can estimate only the intercept beta hat 0, beta hat 1, beta hat 2 and beta hat 3, suppose I am trying to estimate the interaction also.

If I do X1 X2 to bring in that interaction effect into that model unfortunately, this X1 X2 column vector will be exactly identical to the X3 column vector. Similarly, X2 X3 column vector which is the second binary interaction would be identical with the X1 column vector, and you can figure out that X1 X3 column vector would be identical with X2 column vector, and X1 X2 X3 is aliasing with the column vector of 1's.

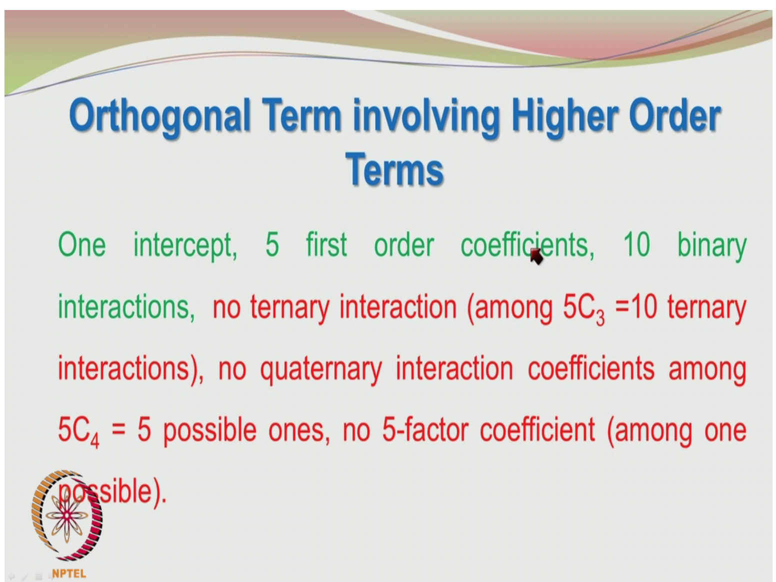
So this X matrix made up of all these elements is definitely going to have a lower rank, and the inverse of the X prime X matrix will lead to difficulties, it cannot be estimated, because the columns in the design are not linearly independent of each other.

(Refer Slide Time: 27:11)



Now let us look at the orthogonal design involving an experimental scheme with as many as 5 factors. So let us construct an orthogonal design involving one-half fraction of 2 power 5 factorial design, the full factorial design will involve 2 power 5 which =32 runs, obviously this is too many, so we want to restrict the number of runs. So we going for a 2 power 5-1 factorial design which is one-half fraction of a full 2 power 5 design and so we have 16 runs. So what are the terms in the model that may be fitted using the given design?

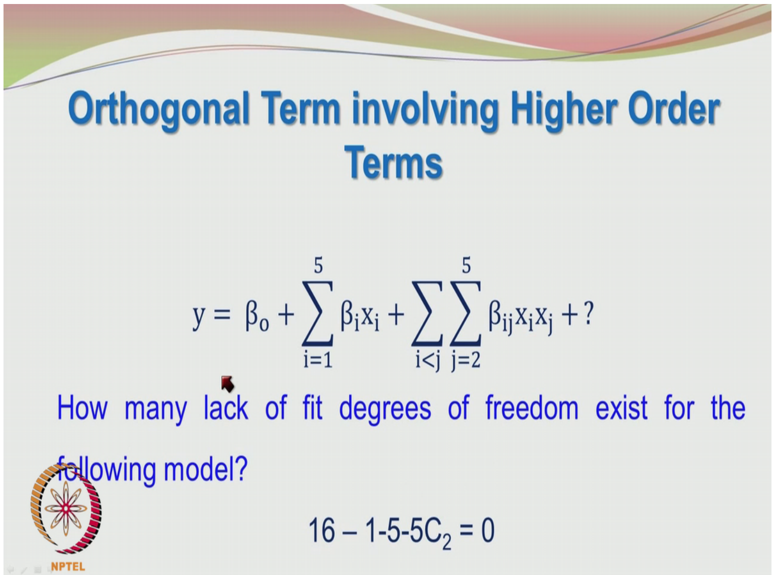
(Refer Slide Time: 27:59)



When we consider the 2 power 5 design we have 1 intercept, 5 first order coefficients, 5 main factors, 10 binary interactions, and then you have again 10 ternary interactions, and 5 quaternary interactions and 1 5 factor interaction. So this is the complete set in a model, where we consider only the main effects and the interaction between the factors. So that would be a total of 1+5=6, 6+10=16, 16+10=26, 26+5=31+1=32.

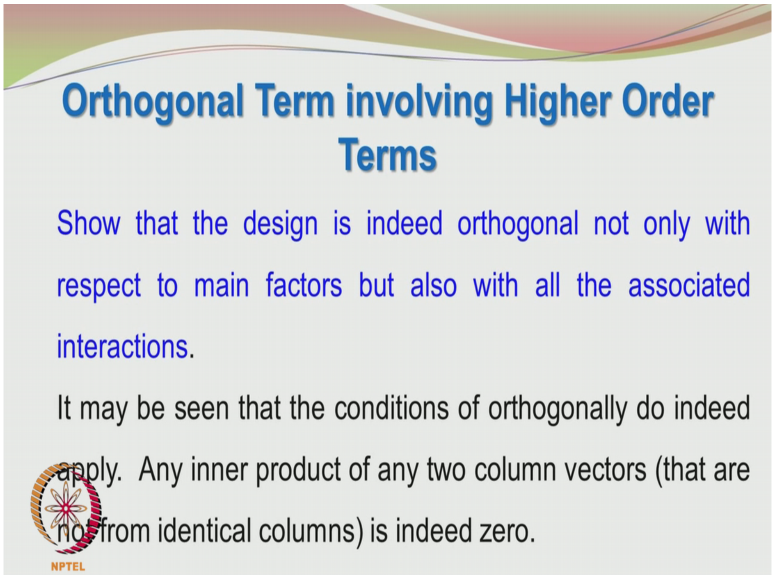
So in a full 2 power 5 design involving 32 independence settings, you would have been in a position to estimate the constant beta hat 0, the 5 main factors, 10 binary interactions, 10 ternary interactions, 5 quaternary interactions and one 5 factor interaction. But we are having only 16 independent settings and so. If we are going sequentially in the model development, we have possible the estimation of 1 intercept, 5 first order coefficient that makes it 6, and 10 binary interactions that makes it 16, beyond it we cannot estimate any more parameters.

(Refer Slide Time: 29:55)



So when you have such a case 2 power 5-1 design, and you have estimated the intercept, the main factors and the binary interaction, what would be the lack of fit degrees of freedom? Since you have exhausted all the 16 independent settings to estimate the parameters associated with these terms, the lack of fit degrees of freedom would be 0.

(Refer Slide Time: 30:30)



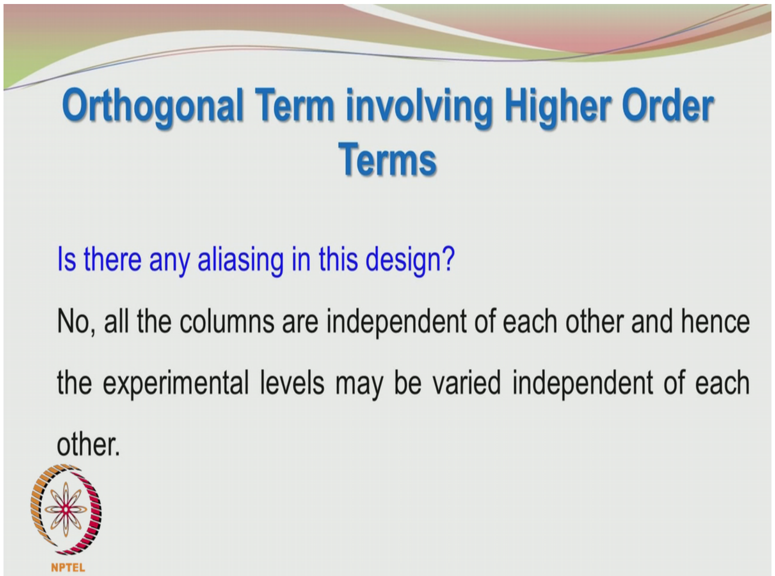
And the next query is quite simple, show that the design is indeed orthogonal not only with respect to main factors but also with the associated interactions. This is a very straightforward thing to do, you can write down the X matrix, you can first write the vector of 1's corresponding to the 16 experiments, and then you can write the half fraction design, so you write the full factorial for a 2 power 4 design.

And then we know from our concepts of fractional factorial design, how to accommodate for 5th factor, let us say the 5th factor is factor E, then to set up the column of factor E, we have to use the design generator I=ABCDE or E=ABCD. I request you to refresh these concepts, and you can write down the X matrix corresponding to a 2 power 5-1 fractional factorial design. And once you have written down this matrix.

And you have also included the possible binary interactions in the X matrix, you can see that they are comprising of number of columns, these columns are having a set of -1 and +1 values, except the first column which is the matrix of 1's which will have all elements as 1, the other columns would be having a mixture of +1's and -1's, and if you take any 2 columns you take the transpose of the first column and then you multiply that with the other column vector, you will get it as 0.

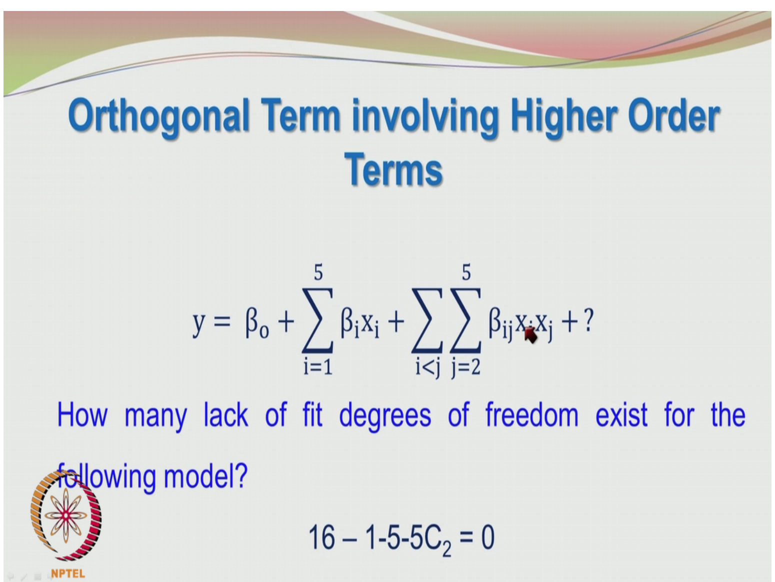
So any inner product of any 2 column vector that are not from identical columns is indeed 0 that can be easily shown.

(Refer Slide Time: 32:42)



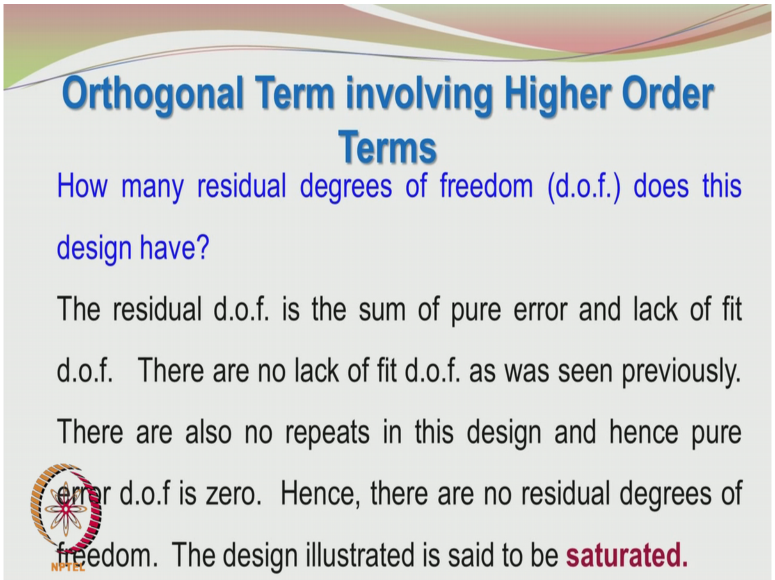
And is there any aliasing in this design? No, all the columns are independent of each other and hence the experimental levels may be varied independent of each other, there is no aliasing in this 2 power 5-1 design. As long as you restrict yourself to the constant, the main factors and the binary interactions. The moment you go for ternary interactions or quaternary interactions then there would be aliasing.

(Refer Slide Time: 33:14)



So up to this model you are not going to have the dangerous aliasing, because you can estimate the parameters from different experimental settings.

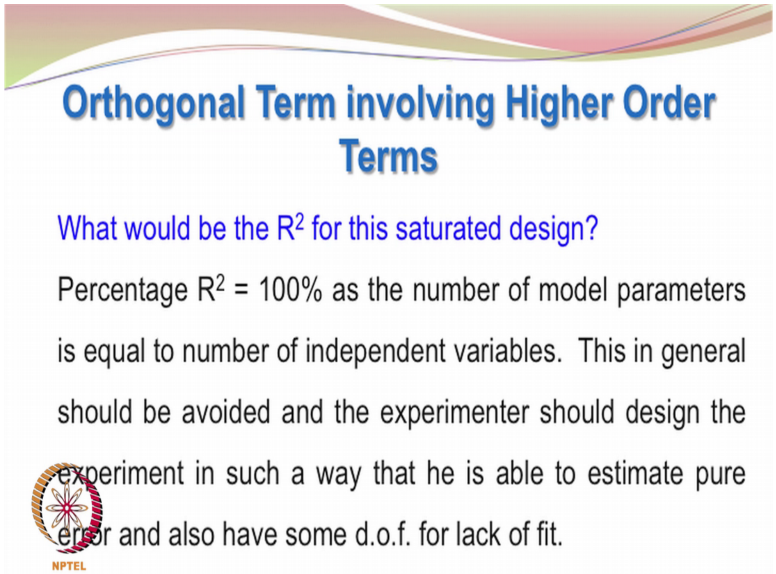
(Refer Slide Time: 33:25)



And how many residual degrees of freedom does this design have? If you have utilized all the 16 independent settings to find 16 model parameters, the residual degrees of freedom would be 0, the lack of repeats also ensure that the degrees of freedom for pure error is also=0, and you also unable to test for lack of fit, so the lack of fit degrees of freedom is also 0. The residual degrees of freedom is the sum of the pure error and the lack of fit degrees of freedom.

There are no lack of fit degrees of freedom as was seen previously, there are also no repeats in this design and hence pure error degrees of freedom is 0. Hence, there are no residual degrees of freedom, and the design illustrated is said to be saturated.

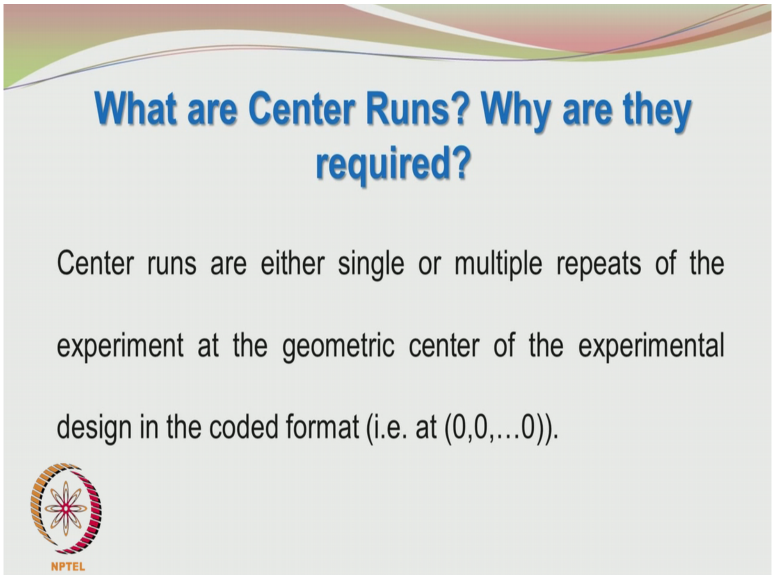
(Refer Slide Time: 34:21)



What would be the R square for this saturated design? R squared value would be =100%, because you are using all the independent settings to estimate a corresponding number of the model parameters. So we will be able to achieve the perfect fit to your model, and that R square value would be=1 or expressed in percentage would be=100%. Well this is misleading because your model has now got as many as 16 parameters.

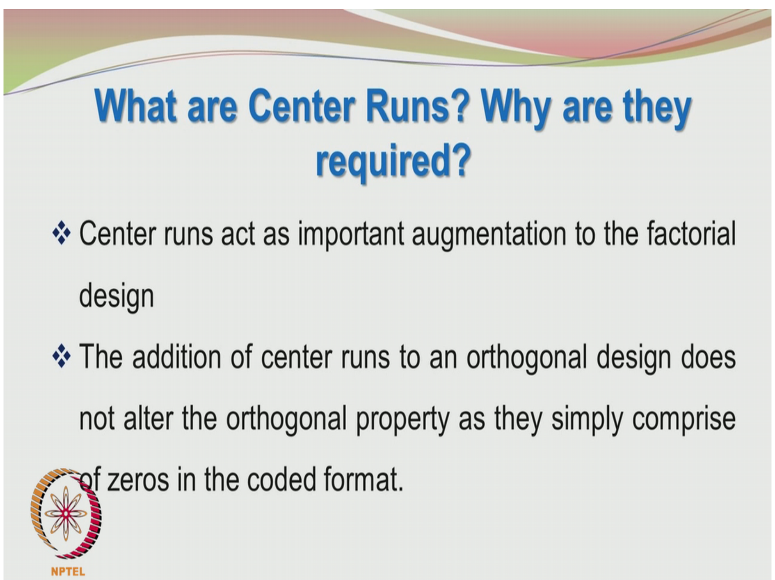
These 16 parameters are difficult to work with unwieldy, and when you tried for a different setting there maybe give completely different predictions. So there are a lot of issues with 16 parameters, normally your model even an empirical one or a linear regression model would be having and not more than 4 or 5 parameters at the most. Now let us talk about Center Runs.

(Refer Slide Time: 35:25)



So what are center runs? And why are they required? Center runs are either single or multiple repeats of the experiment at the geometric center of the experimental design in the coded format. Suppose we have an experimental design and it is coded in terms of -1, +1 and so on, we look at the geometric center of such a design and that would be at 0, 0, 0 corresponding to the midpoints of the independent factors.

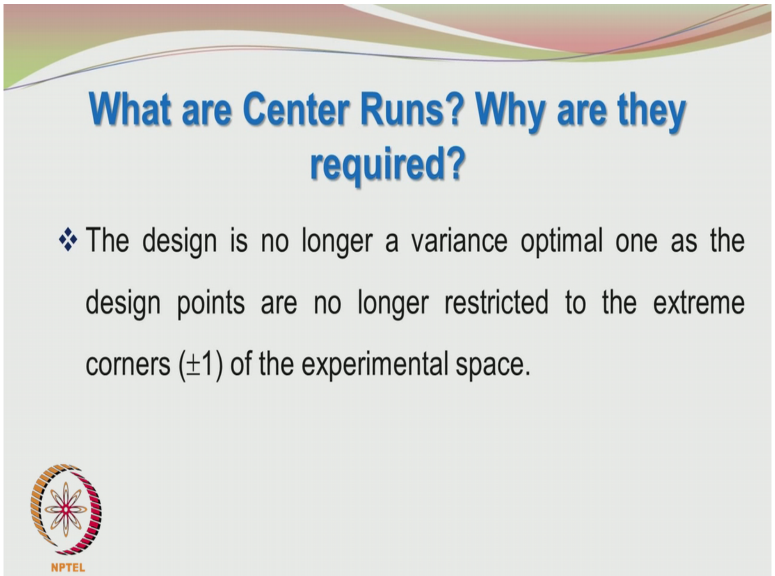
(Refer Slide Time: 35:56)



So center runs are very important to experimental design, you have a factorial design and you add center points you are enhancing or augmenting the factorial design. The addition of center runs to an orthogonal design does not alter the orthogonal property as they simply comprised of

0's in the coded format. So if you have an orthogonal design and you add center points to it, it does not disturb the orthogonality.

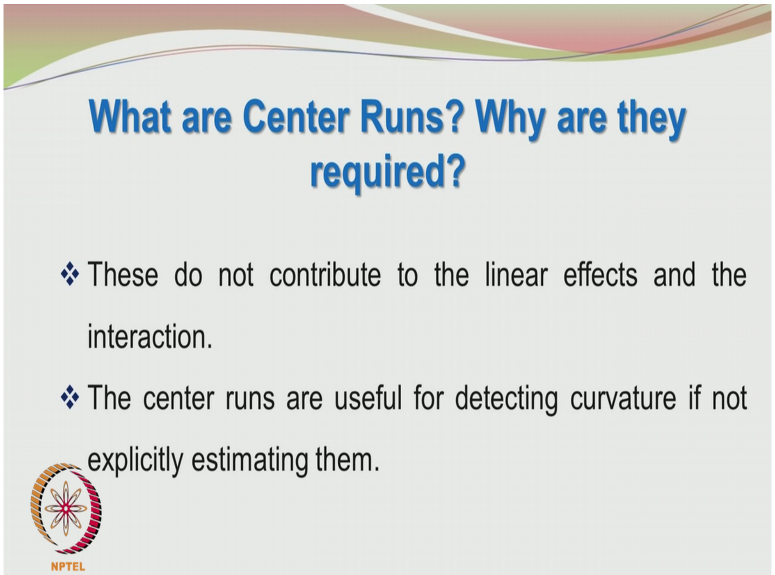
(Refer Slide Time: 36:41)



And the design including the center points will unfortunately make the variance of the parameters slightly higher. So it is no longer a variant optimal design, as the design points are no longer restricted to the extreme corners of the experimental space. In the previous factorial designs without center points all the experimental settings where at +1, -1 combinations, and since these are located at the very edge of the boundaries, the variance of the estimated parameters was minimum.

And so it was a variance optimal design, but the moment you have center points you are having some design conditions or experimental conditions at the center also, they are no longer at the extreme ends. So this makes the design a non-variance optimal one. So if that is so why do we need center runs? What help or what additional benefit do they bring about?

(Refer Slide Time: 37:43)



Before we get into that let us understand about the center runs a bit more, the center runs do not contribute to the linear effects and the interaction, they do not contribute to the main effects and the interaction. Suppose you have a factorial design without center runs and you estimate the main effects and the interaction, then you include the center runs in this orthogonal design the main effects and the interactions values and the corresponding coefficients will not be altered. Suppose you have a design with the center runs, you have another design without center runs.

Both of them will predict the same value of the main factor affects and the interaction effects. So it does not really matter whether you have center runs or not, as far as the estimation of these coefficients are concerned. So center runs are repeated runs, so they help you to get an idea about the experimental error which is very important. Only when you know the extent of experimental error would you be able to comment upon the significance and relevance of the factors in the experiment.

And depending upon the significance or relevance of the factors in the experiment that corresponding coefficients will appear in the model developed. And so in addition to the estimation of the pure experimental error, the center runs are also helpful for detecting whether curvature effects are important are not. The center runs cannot explicitly bring in the contribution due to curvature, but it can only indicate whether the curvature effects are important or not important. So we will take a small break here.