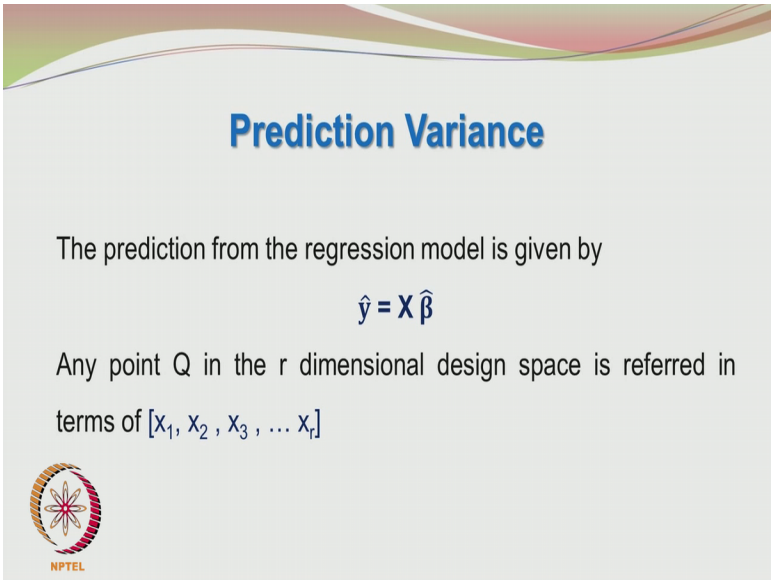
Statistics for Experimentalists
Prof. Kannan. A
Department of Chemical Engineering
Indian Institute of Technology - Madras

Lecture - 45
Orthogonal Model Fitting Concepts - Part B

Okay, after the break discussing about prediction variance, we will be the dabbling a bit with linear algebra, and the pictorial representation of a point in the design space, and the distance of a particular point from the origin and so on. It may look a bit difficult for some, but concepts are very straightforward, and brief recap on linear algebra would be very helpful at this stage. So I request you to take up any book on linear algebra.

And just look at the concept of vectors the distance of a point in the 3 dimensional co-ordinate system and multiplication of vectors, and the inversion of matrices.

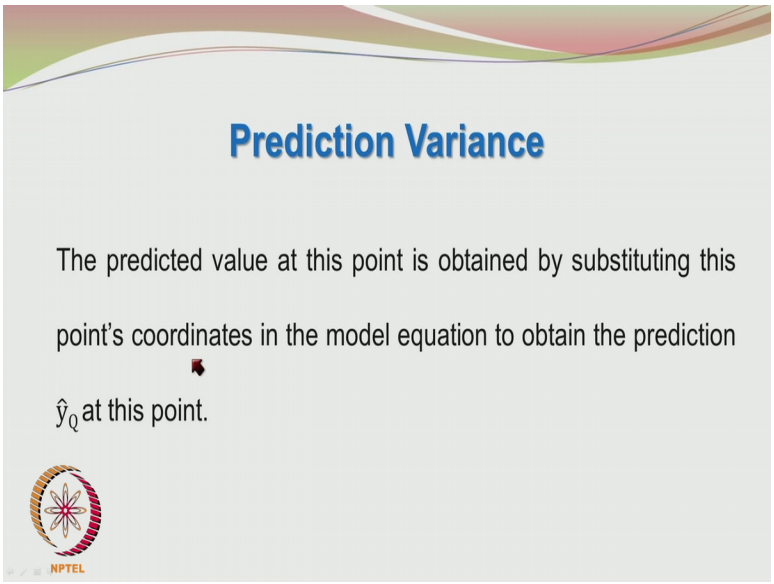
(Refer Slide Time: 01:14)



So once we develop a regression model, we are going to use it for predicting purposes. And how do we predict it? We multiply the vector of the estimated parameters with the X matrix, so again the X matrix is very important here, and that would give us the column vector of predictions. This is nothing new to us, especially after our review of regression concepts. Now let us say that we want to predict the performance of the experiment at a point Q in the experimental design space.

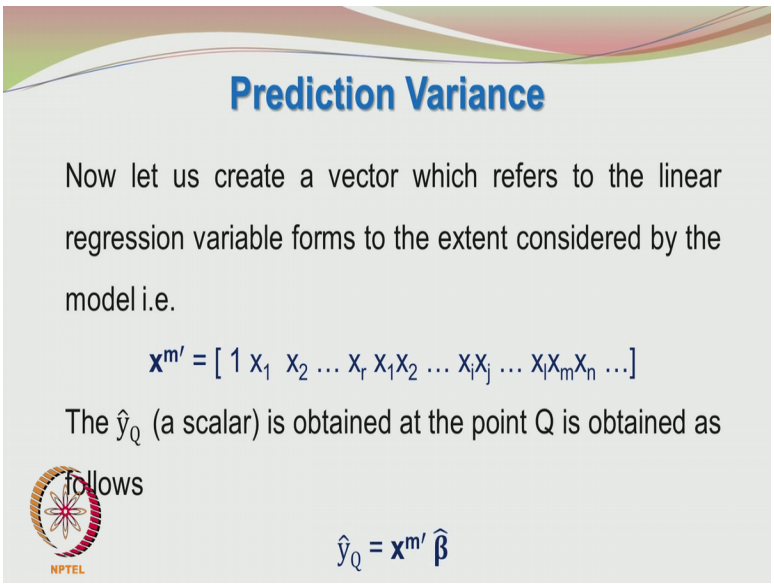
And this point Q in the r dimensional design space is given in terms of some co-ordinate values, so if you are having an r dimensional design space, it would be x1, x2, x3 so on to xr.

(Refer Slide Time: 02:26)



So if you want to predict the experimental response at this particular point Q, then what you have to do is substitute the coordinates of Q in the model equation, and that would give you the prediction y hat Q at this point, so there is a typo here I just make a correction to the typo. So the predicted value at this point is obtained by substituting this points coordinates in the model equation to obtain the prediction y hat Q at this particular point.

(Refer Slide Time: 03:07)



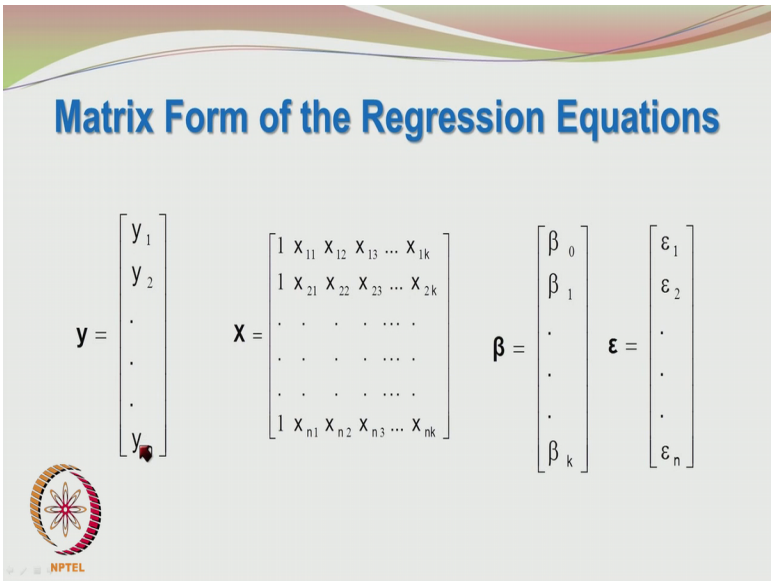
And what we are going to do next is to create a vector, this vector is different from the X values corresponding to the coordinates of Q, because the vector we are going to construct accounts for the model in consideration. Whereas the previous set of X values corresponding to the location of Q in the experimental design space, so please do not confuse between the coordinates of Q and the x m prime factor which we are going to construct shortly.

So the vector which we are going to construct now refers to the linear regression variable forms to the extent considered by the model. So we have estimated the parameters based upon a uncertain model considered by us, and then we look at the model form, and the model would have the main factors up to r factors, the binary interactions, ternary interactions and so on. Obviously, you may not have considered all the possible combinations of the regression variables.

You would have limited the model to a certain extent depending upon your requirement, so that is the model you are going to work with, and that is the model which is going to give you the x m prime vector. So let us say that you had considered only the r main factors in your model, then the x m prime vector would be 1 x1 x2 so on to xr that is it, we do not have the binary interaction terms. Suppose you are model had considered the main factors and also the binary interactions then you would have to go from 1 x1 x2 so on to xr.

And then you go and exhaust all the binary interactions, and that is where you stop the x m prime vector. And so you have the co-ordinates of the point Q, the y hat Q which is a scalar is obtained at the point Q by substituting the x m prime in this equation given here, y hat Q =x m prime beta hat.

(Refer Slide Time: 05:47)

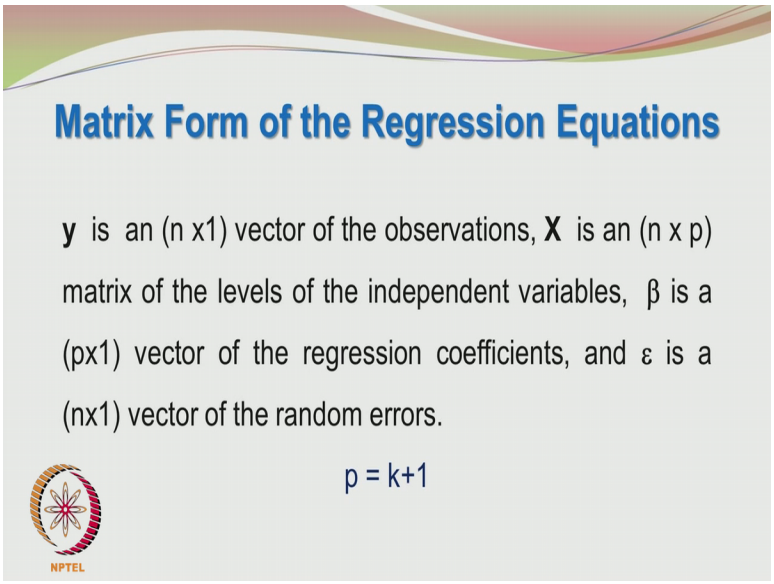


So we have the matrix form of the regression equations, where you have the y which is the vector of responses, and then you have the matrix x, and you can see that the model form is getting reflected in each of the rows. So obviously, you are going to have n rows, where n refers to the number of experimental settings. And then if you go horizontally along each row, then you are dealing with model.

So these are the regressor variables, and they are given in the matrix notation as row number and column number. So this can be x1, this maybe x2, this maybe x3, and then the last one maybe the last binary interaction term maybe x1 x3 or x2 x3, and this is the vector of the parameters which we want to estimate, and this is the error term. I am giving you the slide again from our regression lecture for you to remember, how the x matrix looks like.

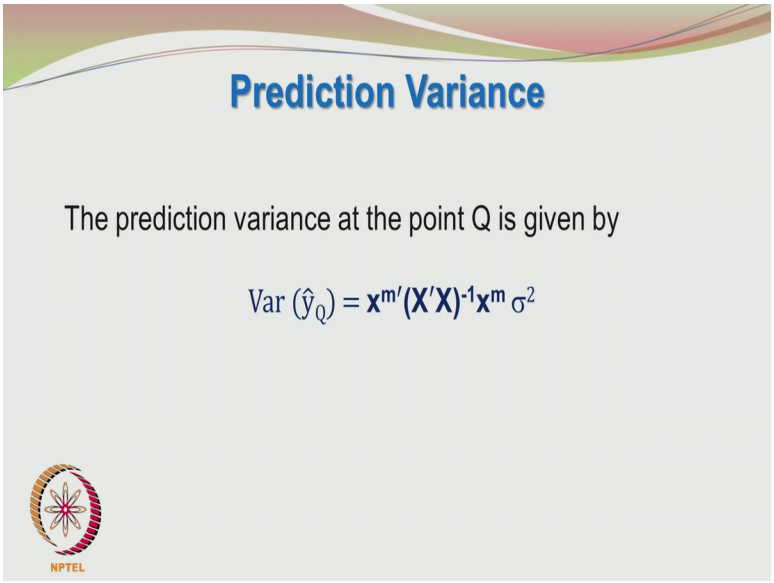
And most importantly if we look at the x m prime vector that refers to the form carried by the row of the x matrix here, so if we are considering only terms according to this model, then your x m prime will also be dictated by these entries in the row of the x matrix.

(Refer Slide Time: 07:35)



So y is an n by 1 vector of the experimental responses, and X is an n by p matrix of the levels of the independent variables, you have n rows and p columns p=k+1 where k is the regression coefficients, and the 1 refers to the intercept beta hat 0, and the beta is a p by 1 vector of the regression coefficients and epsilon is the n by 1 vector of the random errors, and you have p=k+1.

(Refer Slide Time: 08:19)



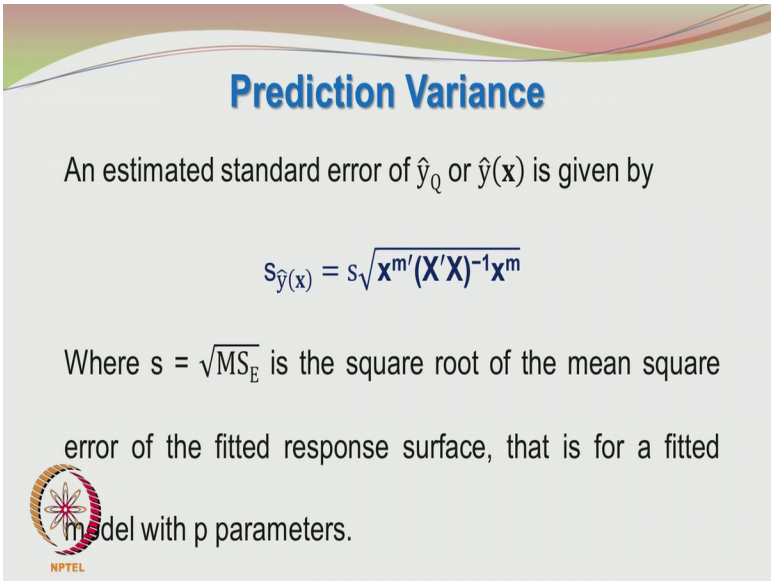
So the prediction variance at this point Q is given by variance of y hat Q =x m prime X prime X inverse x m sigma squared, this is a very very important equation. We want to see how good the prediction is at the point Q, so if the point is very far out into the design space, what is the measure of its prediction, is it a good prediction or is it a bad prediction. How does the

production capability of the model change, when you go further and further away from the center of the design space.

So you have variance of y hat Q=x m prime, we saw just now how x m prime was constructed, and then you have X prime X inverse x m*sigma squared, the sigma squared is the variance of the error. And we assume that the errors are normally and independently distributed with 0 mean and variance sigma squared, unfortunately we do not know the values of sigma squared, and so we use the residual mean square to get a estimate of sigma square which we call as sigma hat square.

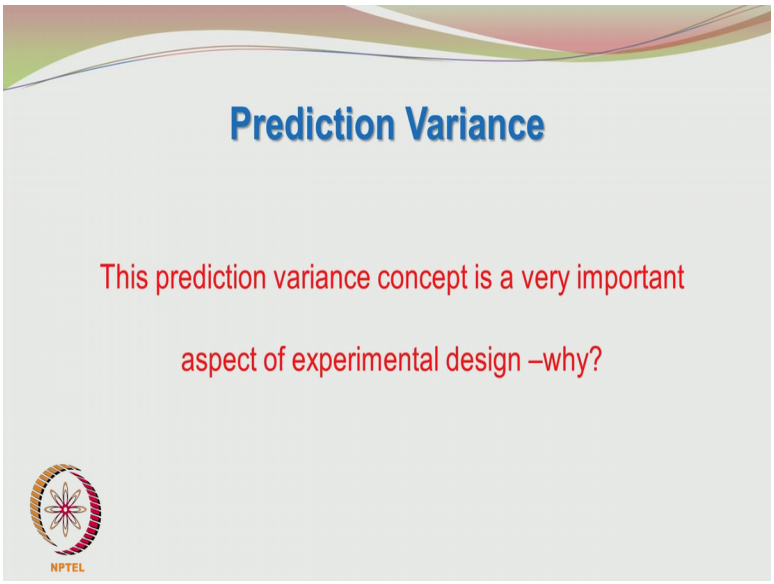
Again we have seen these things in the regression lecture. So variance of the prediction y hat at the point Q =x m prime X prime X inverse x m*sigma squared.

(Refer Slide Time: 10:05)



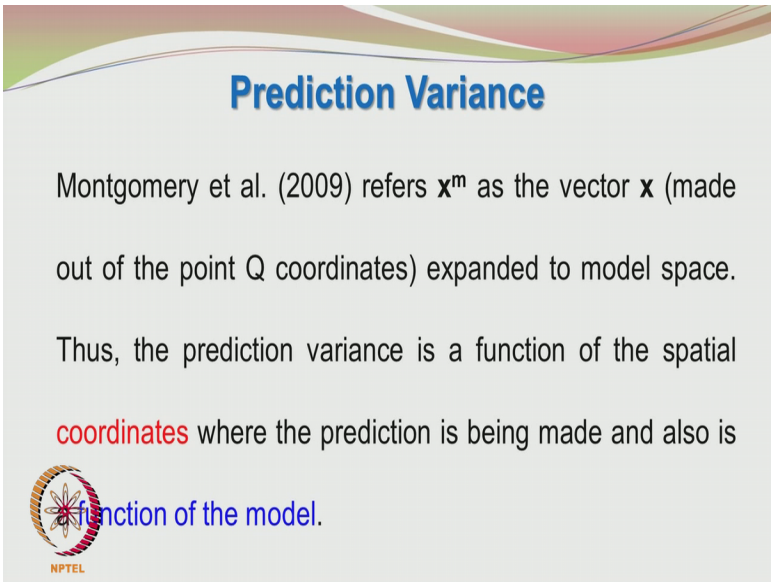
So how do you find the estimated standard error of y hat Q or y hat at x is given by square root of this particular equation, and we replace sigma squared with the mean squared error. How did we find the mean square error? It is nothing but the sum of square of the residuals divided by the degrees of freedom of the residuals that is n-p. So we have Sy predicted x=S* square root of x m prime X prime X inverse x m, where S=square root of the mean square error. And the residual mean square is obtained from the total sum of squares-the regression sum of squares.

(Refer Slide Time: 11:05)



So the prediction variance is a very important concept in experimental design, let us see why?

(Refer Slide Time: 11:11)



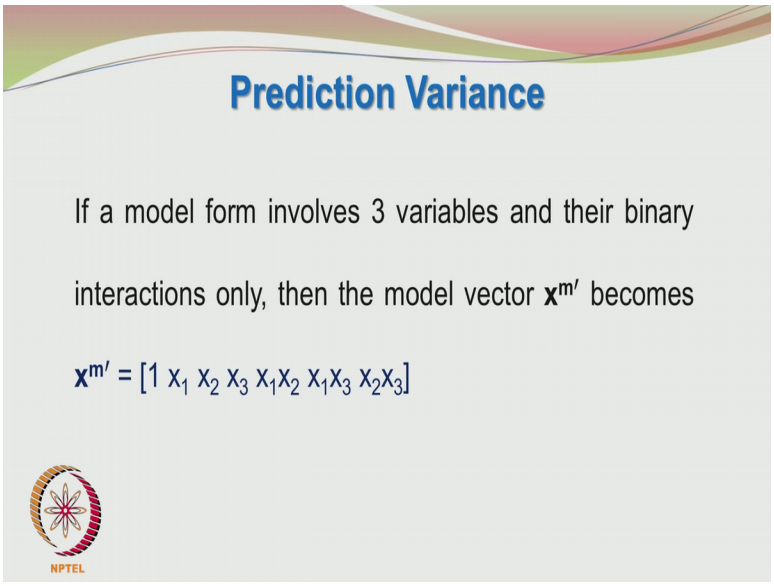
The book written by Montgomery et al 2009 refers x m as the vector x made out of the point Q coordinates expanded to model space, so x m is !=x, it is not the collection of coordinates corresponding to the location of Q, but it is comprised of elements which correspond to the model equation in consideration. We have already seen this a couple of slides back. Now the prediction variance is the function of the spatial coordinates where the prediction is being made and is also a function of the model.

Let us look at that so the x m prime depends upon the coordinates, because the coordinates of Q x1 x2 so on to xr, determine the value of x m prime. So what is going to happen to the variance of the prediction, when you move very far out into the experimental space, when you move very far out in the experimental space the x1 x2 so on to xr, co-ordinate values of the point Q will increase.

So we can intuitively expect that the variance is likely to increase, when we go further and further away into the model space into the extremes of the model. So that is not only aspect, the location of the point Q is not the only aspect, it also depends upon the X prime X inverse matrix, this X prime X inverse matrix is strongly determined by the design we have chosen okay. So that is also to be remembered.

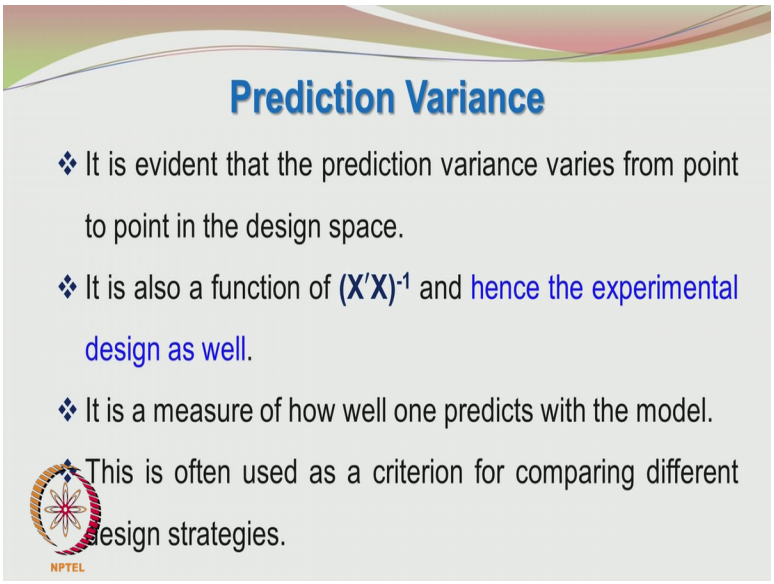
The experimental design we have taken into consideration also influences the variances of the predicted value at a point Q out in the design space. So there are 2 factors how far is the distance Q out in the model space, and what is that nature of the experimental design which dictates the X prime X inverse matrix.

(Refer Slide Time: 13:42)



So let us take a model which is involving 3 variables and their binary interactions only, then the model vector x m prime becomes x m prime=1 x1 x2 x3 x1 x2 x1 x3 x2 x3, so we have 4+2=6+1=7 terms in this particular x m prime.

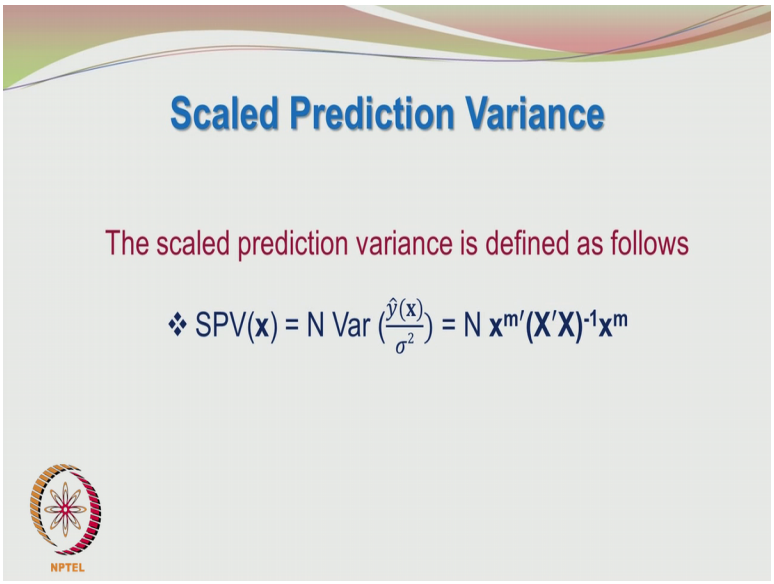
(Refer Slide Time: 14:08)



So the prediction variance varies from point to point in the design space, it is also a function of X prime X inverse, and hence the experimental design. And it is a measure of how well one predicts with the model, and this is often used as a criterion for comparing different design strategies. So when you are choosing a particular design strategy, the expected question from your management or your supervisor would be, why do not you choose this particular design? Why not some other design?

So you should be able to use this prediction variance as one of the different criteria for justifying the choice of your design.

(Refer Slide Time: 14:58)

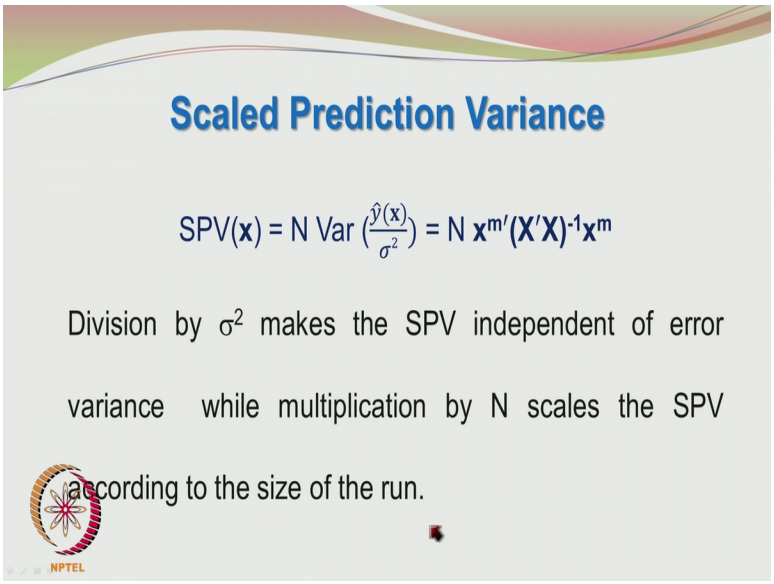


So let us now define the scaled prediction variance at location x as N variance of y hat x/sigma squared. So when we want to compare different designs, we really do not know sigma square, and we have not even conducted the experiment for us to have the value of the mean square error. So what is the point in having a sigma square, so we divide the prediction variance with sigma square, so that it becomes independent of the error variance.

And then we also multiply it by N, because when you have a large number of experiments conducted, then the unscaled or the prediction variance would decrease owing to the high value of this N which is the total number of runs performed. So to prevent artificial reduction in the prediction variance, we multiplied this variance with N, so we divide by sigma square and multiplied by N.

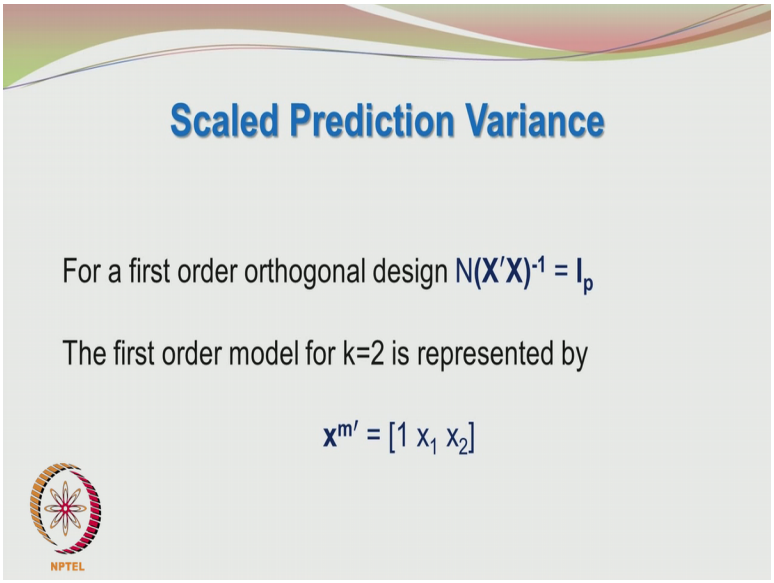
We divided by sigma square so that the prediction variance becomes independent of the error variance, and we also multiplied by the total number of experiments performed N in order to make the prediction variance independent of the size of the run. So we get the scaled prediction variance of x as x m prime X prime X inverse x m.

(Refer Slide Time: 16:45)



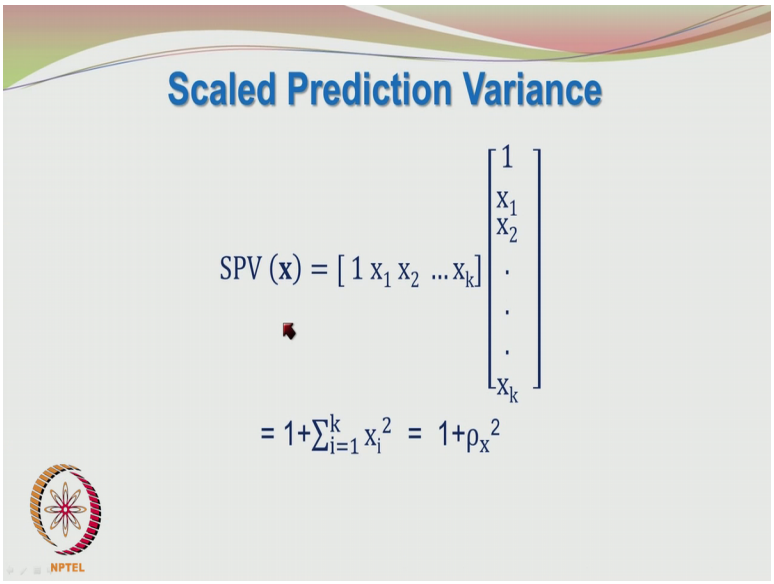
So division by sigma squared makes the scaled prediction variance independent of error variance, while multiplication by N scales the scaled prediction variance according to the size of the run.

(Refer Slide Time: 16:59)



For a first order orthogonal design N X prime X inverse =Ip, where I is the identity matrix, and the size of the identity matrix=p, where p=k+1 the total number of regression parameters. So you will have an identity matrix of with p rows and p columns. Let us take an example, the first order model for k=2 that means 2 regression parameters, p will be=3 is represented in terms of the model vector x m prime as 1 x1 x2, so we are going to find the beta hat 0, beta hat 1, beta hat 2. So we are having model vector having 3 terms.

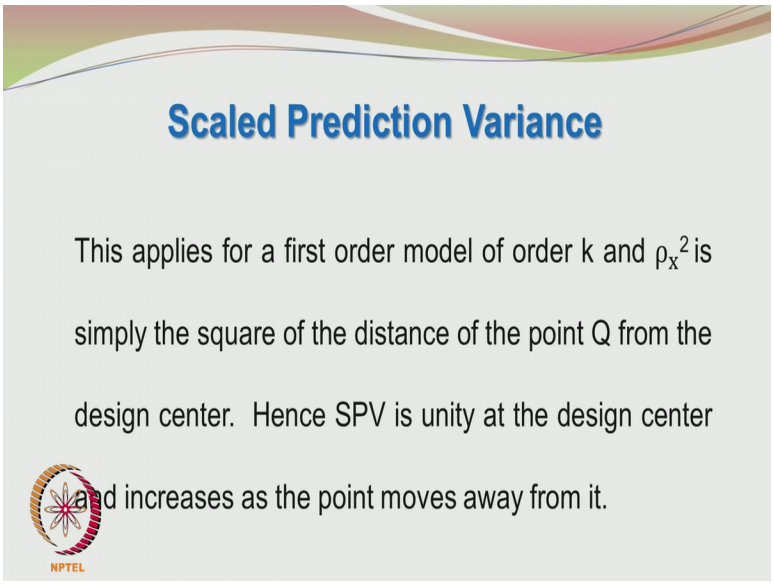
(Refer Slide Time: 17:54)



So when we want to look at the scaled prediction variance, I will correct a small typo at this point okay. When we are looking at the scaled prediction variance, we have x m prime X prime X inverse*x m*N, so ultimately this is the form we get because the X prime X inverse multiplied by N will become an identity matrix, and so we can directly have this product of the vectors as shown here.

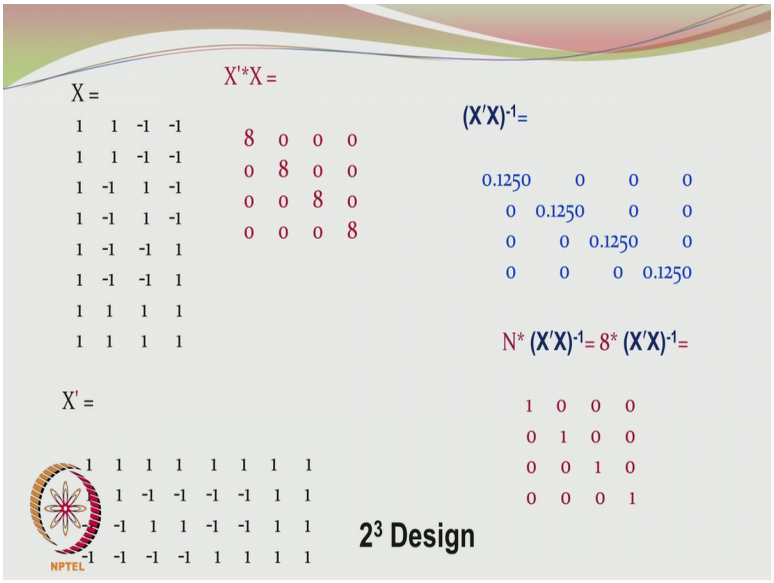
And that would be 1+x1 squared+x2 squared+ so on to xk squared, and that would be 1+rho x squared, because x1 squared+x2 squared+ so on to xk squared is nothing but the square of the distance of the particular point from the origin.

(Refer Slide Time: 18:59)



So this is applicable for a first order model of order k and rho x squared is the square of the distance of the point Q from the design center. Hence, the scaled prediction variance is unity when x1=x2=so on to xk=0, and it increases as the point moves away from the center.

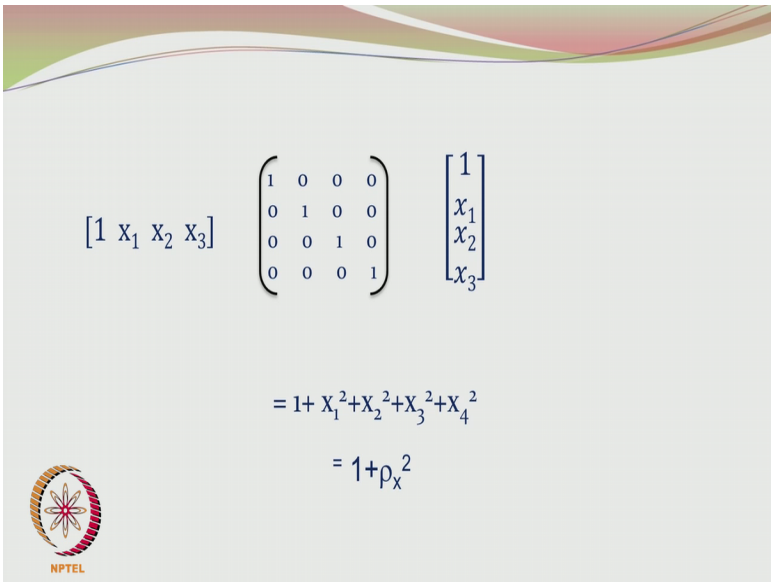
(Refer Slide Time: 19:32)



So let us look at the design here we are having a 2 power 3 design, so you have 8 rows and you have a full 2 power 3 design, and the X prime X matrix becomes a diagonal matrix, so the diagonal elements having the value of 8. And when you take the inverse of this, we will get 1/8, 1/8, 1/8, 1/8 or 0.125 throughout the diagonal. And when we do N*X prime X inverse, we are multiplying everything by 8, and we get the identity matrix of dimension p, p =k+1, and that would be 4.

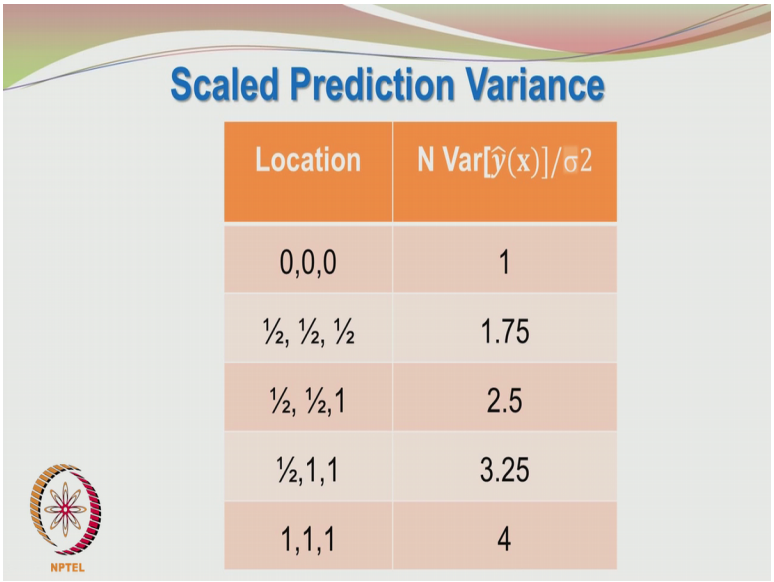
So we are having a 4 by 4 identity matrix. So if this is X prime the transpose of the matrix is obtained by changing rows into columns and columns into rows.

(Refer Slide Time: 20:34)



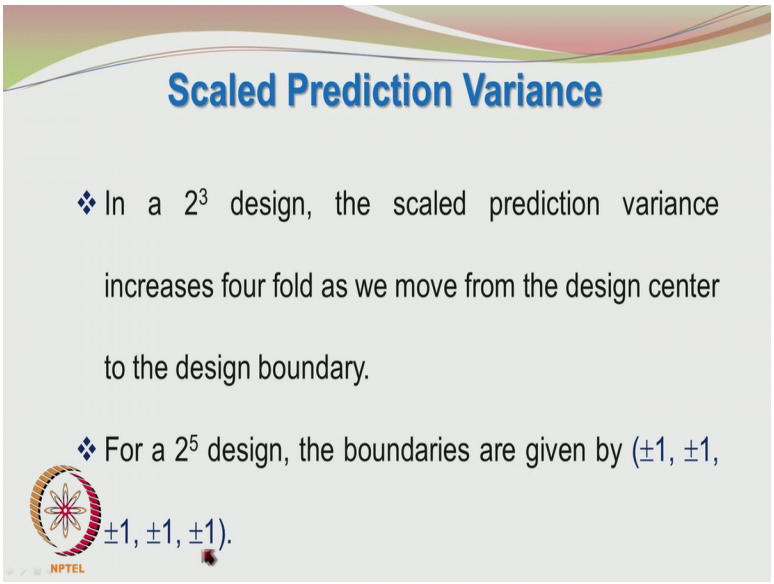
So when we try to calculate the scaled predictions of variance for the 2 power 3 design, so we again get 1 x1 x2 x3 multiplying 1 x1 x2 x3, and that would be 1+x1 squared+x2 squared+x3 squared+ that is x3 squared, I will correct a typo here. So I again we get 1+square of the distance from the design center, so for this 2 power 3 design the scaled prediction of variance is unity at the design center, and then increases as we move away from it.

(Refer Slide Time: 21:11)



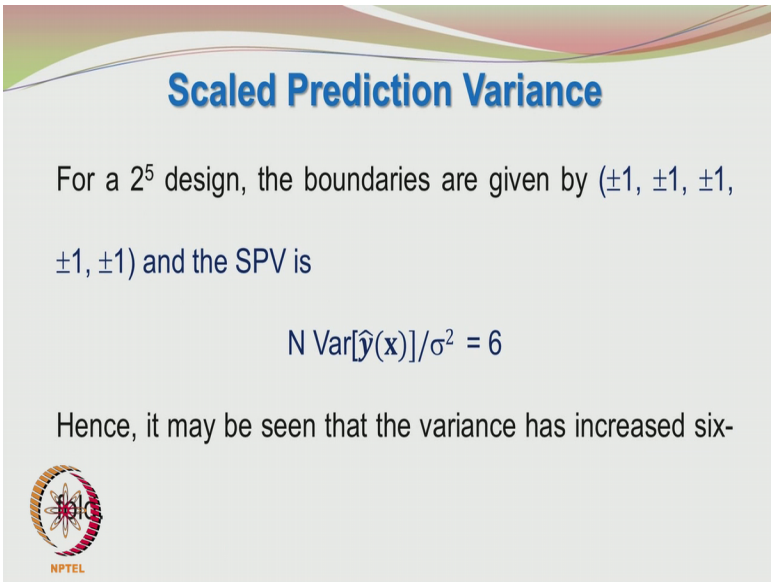
So this is for an orthogonal design, so we can see that the scaled prediction of variance at different locations in the design space, and it is unity at the center, and then as we go to the extremes of the design space, we get the scaled prediction variance as 4.

(Refer Slide Time: 21:29)



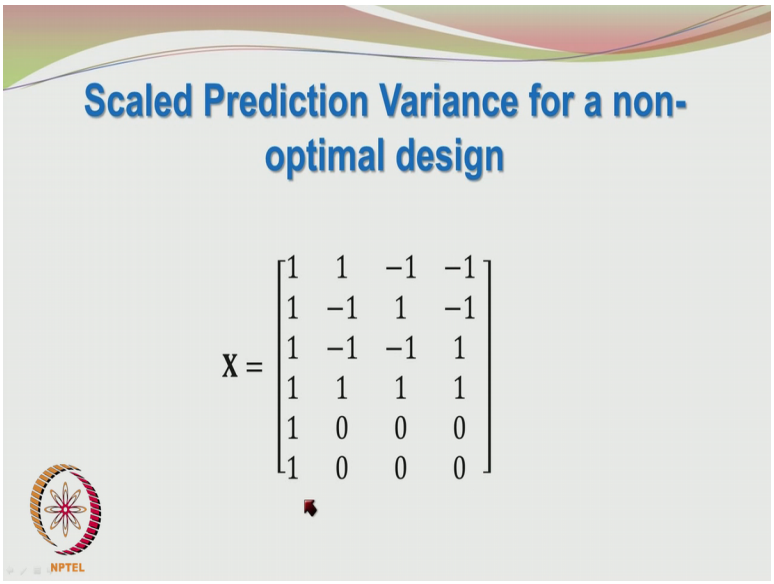
And if you have a 2 power 5 design, the boundaries are given by +or-1, +or-1, +or-1, +or-1, +or-1 that is 5 times.

(Refer Slide Time: 21:42)



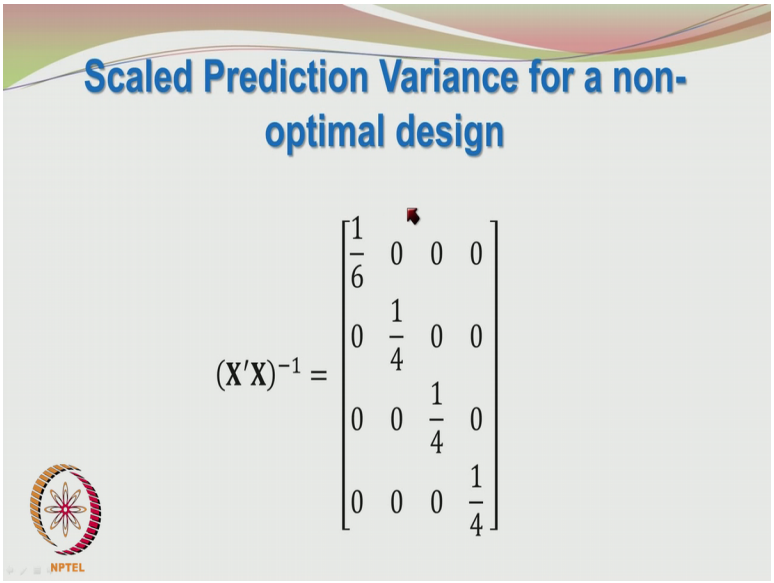
So the scaled prediction variance, we can easily show to be=6, so when compared to the design center where the scaled prediction variance is 1, when you go to the design boundaries or design extremes the scale production variance increased to 6, so it has increased 6 times.

(Refer Slide Time: 22:05)



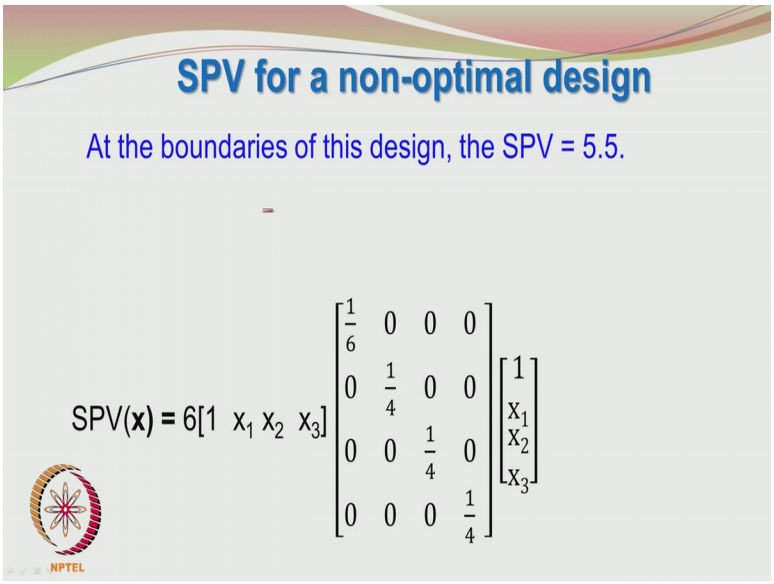
So when we have a non-optimal design, let us see what happens to the scaled prediction variance? So in this non-optimal design the X matrix is given as shown here, you also have repeats and that is why you have these 2 rows with 0's in them.

(Refer Slide Time: 22:30)



And when we look at the X prime X inverse matrix we get 1/6, 1/4, 1/4, 1/4, so this is the matrix we get in such a situation for a non-optimal design. Why is the non-optimal design? The design points are not only located at the extremes, but we have the design points located at the center also.

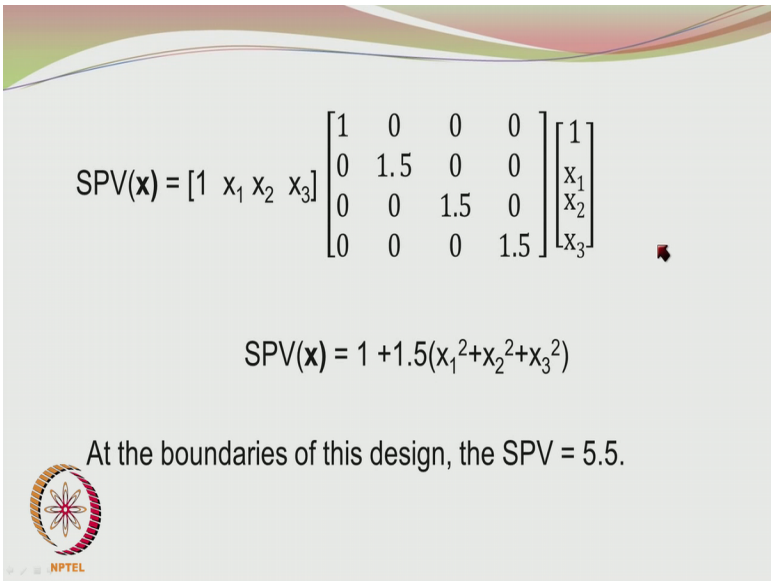
(Refer Slide Time: 22:54)



In such a case at the boundaries of this design the scaled prediction variance turns out to be 5.5. Now we are going to look at the scaled prediction variance for non-optimal design, it is a non-optimal design because the design points are not located at the extremes of the design space alone, in addition we are having design points located at the design center. In such a situation the scaled prediction variance is higher and we can easily calculate it.

The size of the run is as you can see it is 6, and when you have a size of the run to be 6 N=6, and so we multiply it by 6, and then you have 1 x1 x2 x3 we are considering a model involving the main factors alone. And the X prime X inverse matrix turns out to be 1/6, and then you have 1/4, 1/4 and 1/4 as the remaining terms in the main diagonal. And so when we do the multiplication we get 6/6 will become 1, 6/4 will become 1.5.

(Refer Slide Time: 24:21)



And hence you will get this particular expression, when you simplify this we will get scaled prediction variance of x=1+1.5*x1 squared+x2 squared+x3 squared, and at the boundaries if you put x1=x2=x3 as 1, 1 and 1, we will get 1+1.5*3 as 5.5 totally. So the scaled predictions variance at the boundaries of the design is 5.5, you may want to see what would be the scaled prediction variance for the optimal design, if the center points had not been there.

This concludes our discussion rather brief one at that on the orthogonal concepts, usually this topic is not covered in factorial design of experiments, it is supposed to be implicitly understood. But I thought having a separate lecture on this concept would put things into perspective. It will also explain, why in the regression analysis the adjusted sum of squares and the sequential sum of squares are identical.

For an orthogonal design it does not really matter in what sequence a particular factor enters the model whether it is coming in the beginning or in the end or coming as a sequence. But when you have a non-orthogonal design the adjusted sum of squares and the sequential sum of squares are different, and the order in which the parameter are the factor enters the design experiment assumes importance.

The order in which the regression parameters is introduced into the model assumes importance in non-orthogonal designs. In orthogonal designs it does not really matter, and therein lies the

advantage. So it is always better to go for planned experiments such that your design space is comprising of orthogonal vectors, and you can also code them uniformly, so you will have a column vectors of -1, +1, 0 and so on, the design looks neat.

In some cases it may not be possible, you may have to work with the available data to develop your regression. And then in such cases you can adopt the general approach. A very important advantage of factorial design of experiment is, it leads to an orthogonal design, and the parameters are estimated quite easily, the X prime X matrix is a diagonal matrix in such cases, and the estimation of the parameters is becoming very straightforward.

And the role played by a particular factor in orthogonal design experiments may be assessed independent of the other factors. So this concludes our discussion on orthogonal concepts involved in experimental design or simply put orthogonal designs. We will next to move on to second order designs such as the central composite design, which will lay the groundwork for response surface methodology, thank you for your attention.