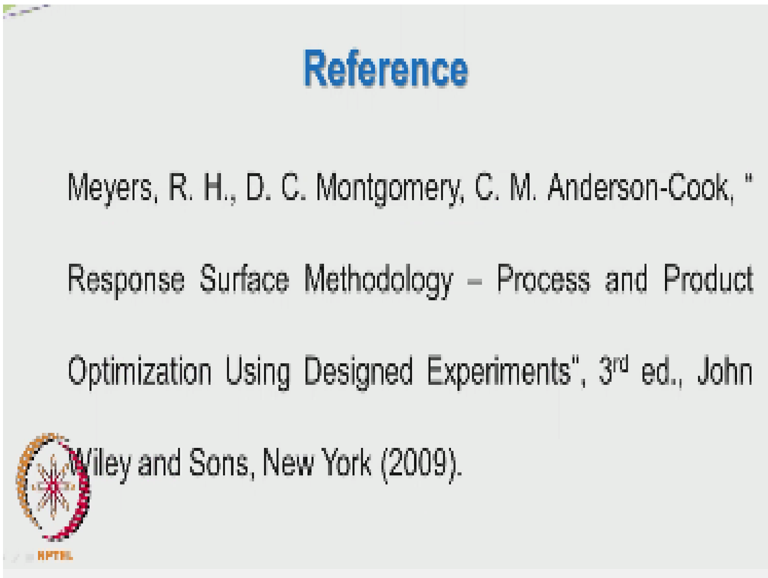
Statistics for Experimentalists
Prof. Kannan. A
Department of Chemical Engineering
Indian Institute of Technology – Madras

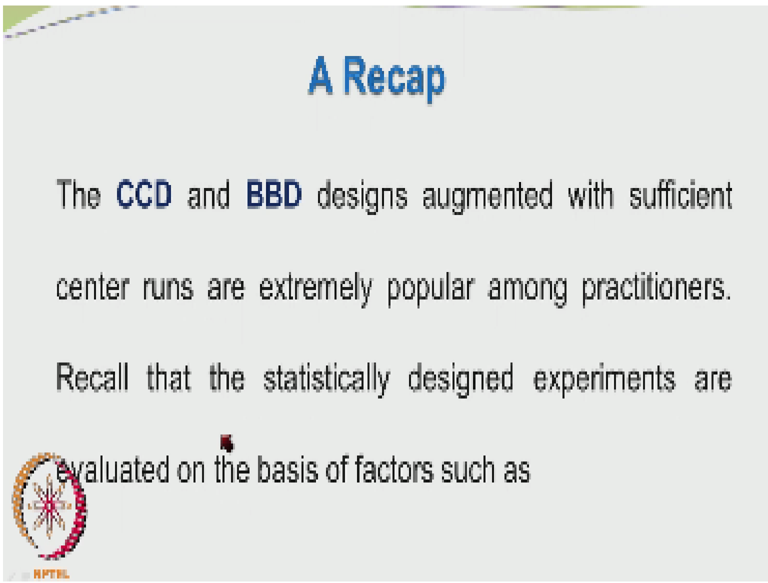
Lecture – 51
Optimal Designs - Part A

(Refer Slide Time: 00:23)



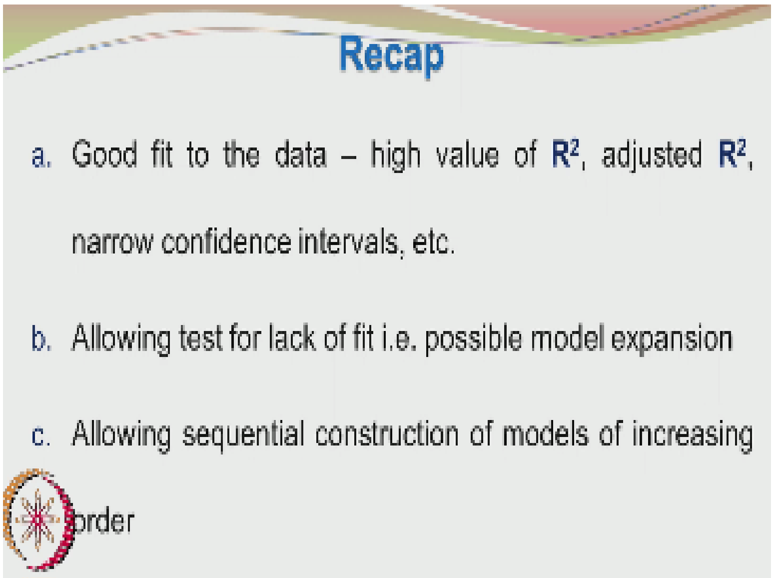
Hello welcome back, today we will be looking at the optimal designs, the reference for this lecture material is based on the book written by Meyers Montgomery Cook, the title of the book is Response Surface Methodology, Process and Product Optimization Using Designed Experiments, 3rd edition, John Wiley and Sons, New York, 2009.

(Refer Slide Time: 00:42)



So, we have looked at regular factorial designs and then we moved on to second order designs, where we talked about the central composite design and the Box Benkhen design. These are very popular design among the practitioners of this method, so the criteria for a good statistically designed experiment are listed in the following slides.

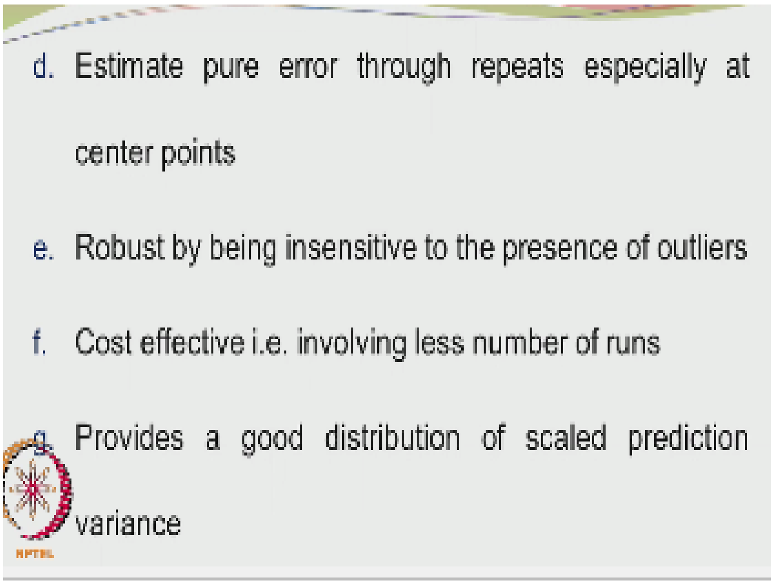
(Refer Slide Time: 01:12)



We use a statistical design strategy to develop a mathematical model eventually, based upon the significant factors in the process and that should have a good fit to the data; experimental data of course, high value of R square, adjusted R squares and narrow confidence intervals etc. The model should also have some degrees of freedom for lack of fit, so that it can be expanded to include higher order terms.

And it should also be amenable to building the model sequentially starting with the simplest model first and then gradually adding more factors or more interactions between factors or higher order terms involving the factors, so that we can see the benefit of increasing model complexity. At some point, we can say that okay; we are not getting any further improvement from the model.

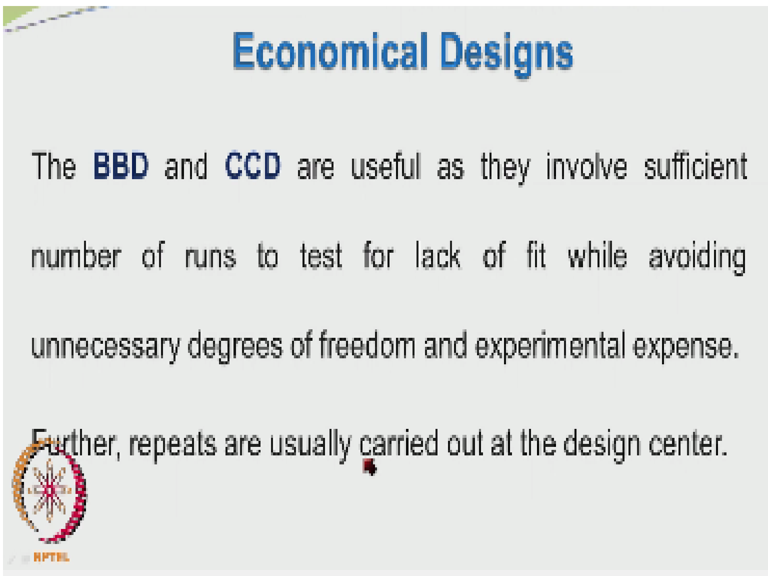
(Refer Slide Time: 02:40)



So, we will stop at this particular point for example, if the adjusted R squared is beginning to rise up to a certain model expansion and begins to fall, then there is no point in adding further parameters and also it should be having repeat points especially, at the center of the design, so that the pure error may be estimated and it should also be robust by being insensitive to the presence of outliers.

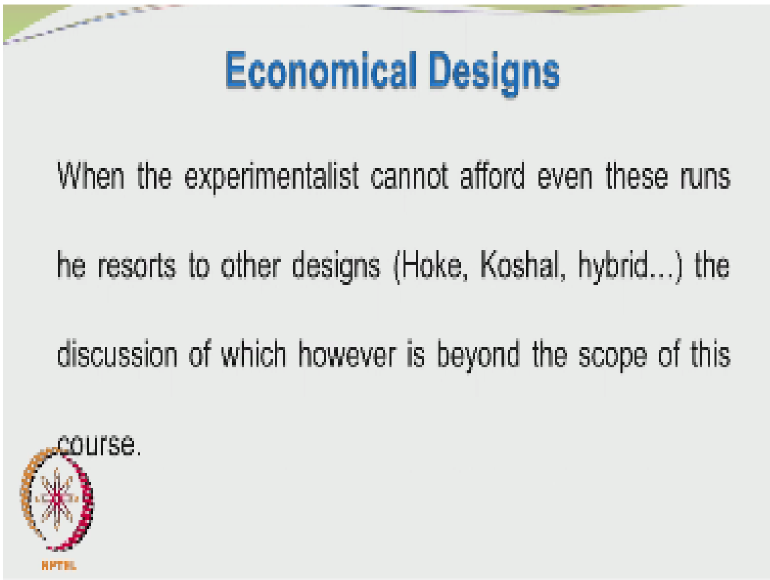
The presence of outliers should be clearly seen and their presence should not alter the model structure drastically, it should also be cost effective and design involving less number of runs would be more attractive in that sense and we have also seen this, it should provide a good distribution of scaled prediction variance. We saw that model should be able to predict well within the design space.

(Refer Slide Time: 04:02)



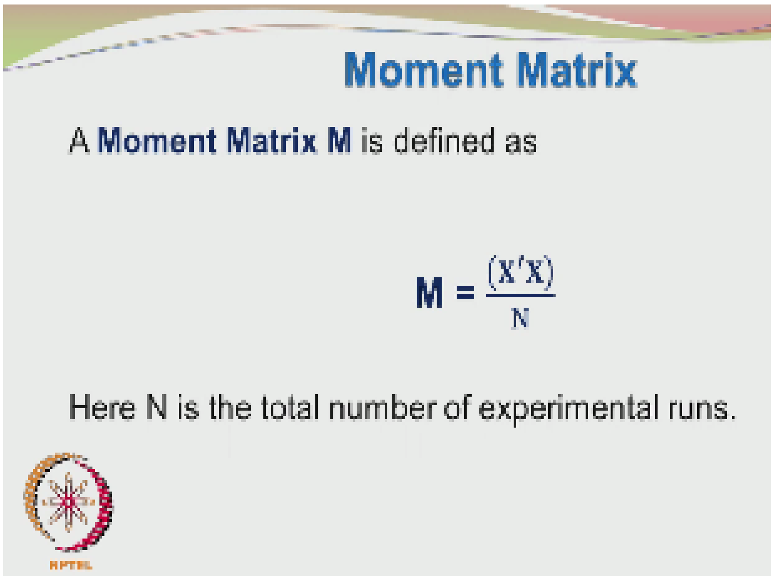
And since the model is built on experimental data and there is uncertainty associated with the experimental data because of random fluctuations and random errors and so on, there is also uncertainty associated with the model prediction and the variability in the model prediction should not be too much in the design space. So, the Box Benkhen design and the central composite design are useful.

(Refer Slide Time: 04:17)



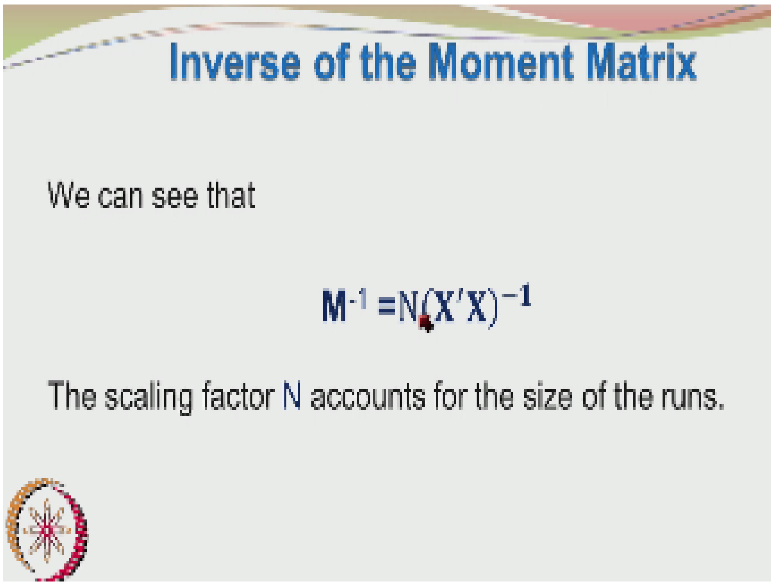
As they involve sufficient number of runs to test for lack of fit while avoiding unnecessary degrees of freedom and experimental expense and repeats are also carried out at the design center. When the experimentalist has to go for more economical runs, there are other versions available, which are discussed for example, the reference book by Montgomery; I just referred to; so those books are having further details on economical designs.

(Refer Slide Time: 04:49)



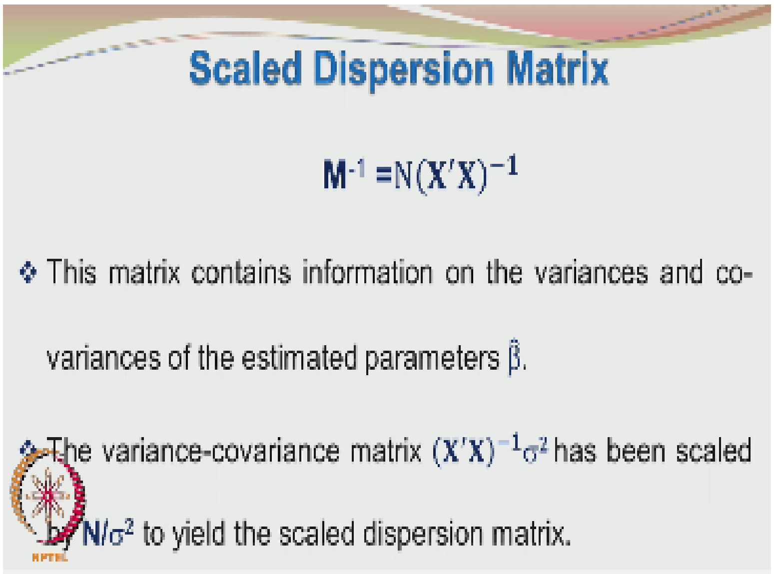
We will not be going into these economical designs for want of time and it is also not within the scope of the current subject, right. So, first we will define a moment matrix, we have already seen this earlier; moment matrix M is defined as the ratio of X prime X to N, where X matrix is the experimental design matrix containing the column of 1's, X1, X2 and so on and N is the size of the experimental run.

(Refer Slide Time: 05:19)



So, from the definition for M, which is X prime X/ N, we can easily show that the M inverse would be given by X prime X inverse * N, we are adding N to account for the size of the run okay.

(Refer Slide Time: 05:43)



Now, this matrix M inverse matrix is called as the scaled dispersion matrix, we are scaling it by N, we are scaling X prime X inverse by N, so that the size of the run gets cancelled out with the elements of X prime X inverse. Suppose, you have first order design; orthogonal design with 8 runs for a to2 power 3 factorial design, X prime X inverse would have 1/8 along the diagonal and then you are multiplying it by 8 and reducing it into an identity matrix.

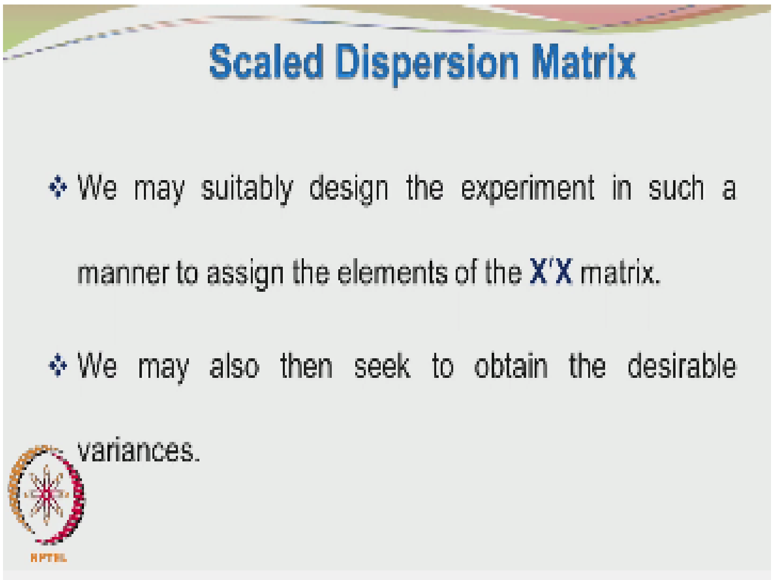
So, whether it is 16 runs or 8 runs that number of runs is removed from the analysis, this X prime X inverse matrix is a very important one because it contains the information on the variances and the covariances of the estimated parameters beta hat. We use the regression concept to find the regression coefficients beta hat0 beta hat 1 so on to beta k; beta hat k, so we found k + 1 parameters from the regression exercise.

And these regression parameters also have variances and covariances associated with them for example, for each regression parameter would have the variance associated with it and you will also have covariance between different parameters and to get these variances and covariances, we make use of the variance covariance matrix, which is nothing but X prime X inverse and we are multiplying by N and dividing it by sigma square to make the resulting structure or matrix independent of the size of the run.

And independent of the unknown variance also, so we are able to have a uniform; we have an uniform basis for comparison between different runs. The variance covariance matrix is actually given by X prime X inverse sigma squared and what we do is; we divide it by sigma square, so that the sigma square vanishes and you still have X prime X inverse. As I told you, the X prime X for an orthogonal design will take values along the main diagonal corresponding to the number of runs in the design.

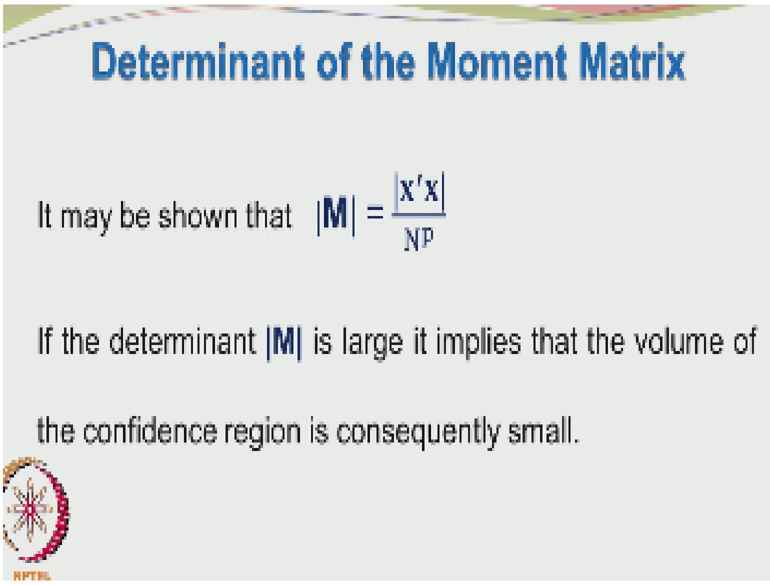
And when we take inverse of that we will take 1/N along the diagonal main diagonal terms, so when N increases, it will appear as if the variances of the parameters are less, so deliberately by choosing a large number of runs, I can claim that the variance of the parameters are reduced. There is a more economical efficient design involving less number of runs may seem to have a high variance of the regression parameters.

(Refer Slide Time: 09:36)



In such cases, it is important to put them on a uniform basis and so we multiply the X prime X inverse matrix with the size of the run N. So, now that we have the X prime X matrix and we know that it is playing the central role in the variance of the estimated coefficients, we can see how we can exploit the structure of X prime X such that we get low regression coefficient variances.

(Refer Slide Time: 10:29)



So, that we are estimating the regression coefficients more precisely or with less degree of uncertainty. So, we have to locate the factorial points or the experimental design points in the X matrix in such a way that the X prime X inverse is reduced. So, when we take the determinant of the moment matrix; it can be shown that the determinant of the moment matrix is nothing but the determinant of X prime X/ N power P, where P is the total number of parameters including the intercept beta hat 0.

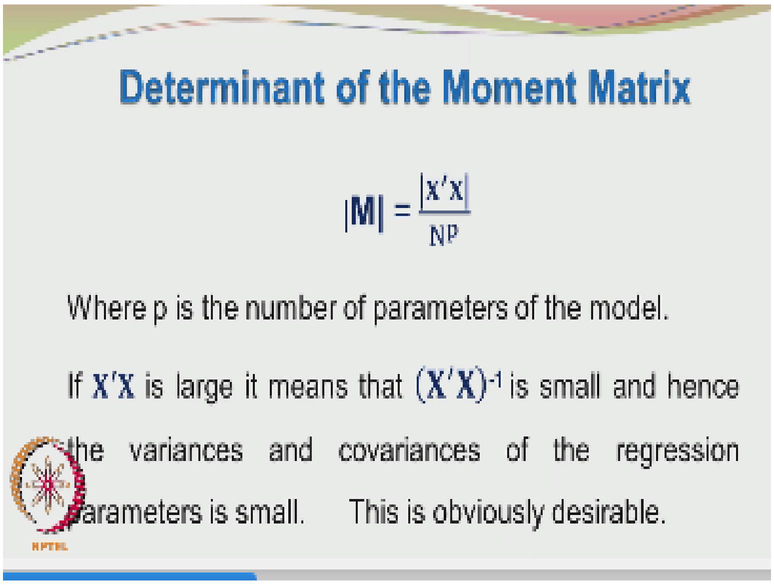
So, again I repeat, we are finding the determinant of the moment matrix, we defined the moment matrix as X prime X/ N and when we take the determinant of this particular M, we end up getting determinants of X prime X/ N power P because we are dividing it by N by property of determinants, we get a matrix divided by N becoming the determinant divided by N power P and also it can be seen that if the determinant M is large, it implies that the volume of the confidence region is small, okay.

The advantage of having a determinant is; you can get one single value and that can be used as a criterion for evaluating different designs. So, when the determinant M was large it also means that X prime X would be large; the determinant of X prime X would be large and but the same token you can also feel that determinant of X prime X inverse would be small. So, if we can imagine linking the confidence region, we are constructing in the parameter space.

The N dimensional space would comprise of N estimated parameters, suppose I am estimating 3 parameters, I would have a 3 dimensional space, if I am estimating 4 parameters, I would have a 4 dimensional parameter space and we can imagine a confidence volume in this multi-dimensional space and it is not good, if this confidence volume is large. It means that the models regression coefficients may take values between one number and another number.

And the 2 numbers are widely separated apart, that means the confidence levels for each parameter is 95% let us say, then the upper and lower limits of the conference intervals would be quite far apart from each other. So, this would make the volume of the confidence region pretty large and it will also indicate that the parameters are not estimated precisely. So, our aim is to make the volume of the confidence region quite small.

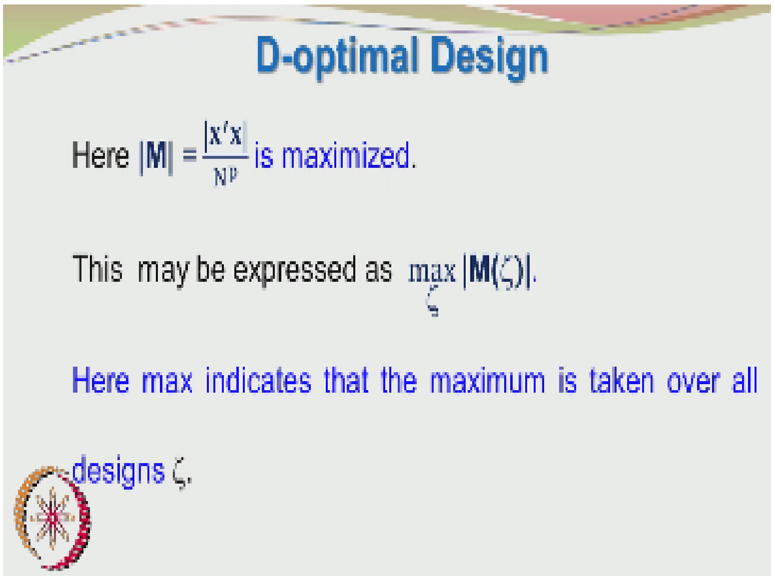
(Refer Slide Time: 14:24)



And in other words, we have to also then make the determinant of X prime X quite large. If the inverse of the confidence region is quite high, it means the confidence region is quite small and for having a small volume of the confidence region; it would be good, if we have a large determinant value for X prime X. So, if X prime X is large, it means that X prime X inverse is small and hence the variances and covariances of the regression parameters is small.

This is obviously desirable. So, we are focusing a lot of attention on the variance covariance matrix, the variance of the regression parameters are given by the diagonal elements of the variance covariance matrix and the variance covariance matrix is nothing but X prime X inverse sigma squared. So, we are looking at ways and means through which the diagonal terms, the variances of the estimated parameters may be made as small as possible.

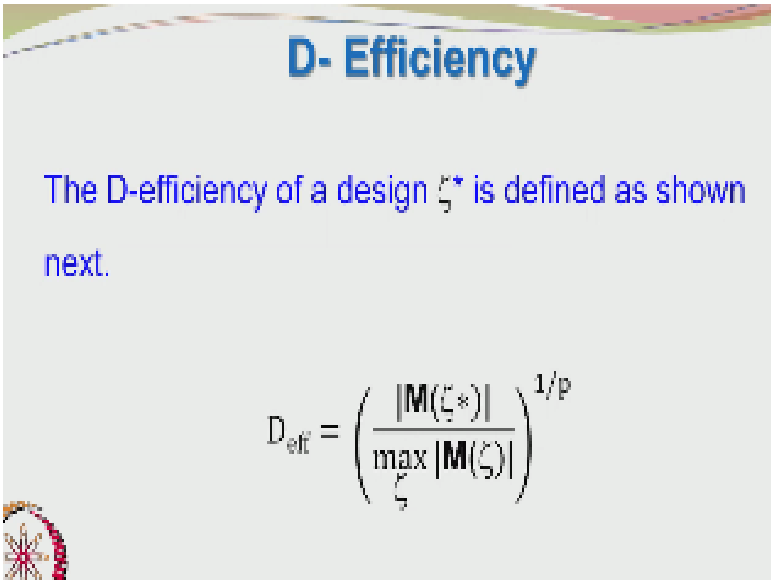
(Refer Slide Time: 15:25)



And that would be small, if the X prime X inverse is small. So, now we are looking at certain alphabetical criteria based designs; statistical designs, the first one is the D optimal design; it is alphabetical criteria based design because we are using the letter D, to denote it as the optimal condition. So, here you have determinant of M is = determinant of X prime X/ N power P and this may be maximized.

If this is maximized, then obviously the determinant of X prime X inverse would be minimized and that would also minimize the volume of the confidence region in the multi-dimensional space and our parameters would be estimated more precisely, so we can express this criterion as maximum of Zeta of determinant of M of Zeta. So, we are having several designs and we are choosing that design as D optimal, which will maximize the determinant of M for a particular design.

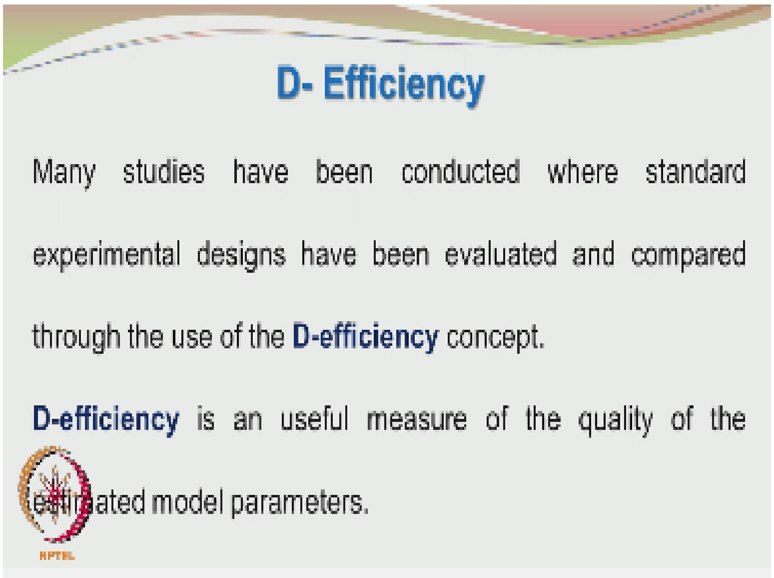
(Refer Slide Time: 16:44)



So, we can define a D efficiency of a design as the determinant of the moment matrix for a particular design under consideration divided by the maximum value that may be taken among all possible designs and this ratio, we are scaling by a power of 1/P, so this is called as D efficiency, it may look a bit complicated and highly mathematical, on the other hand it is very simple.

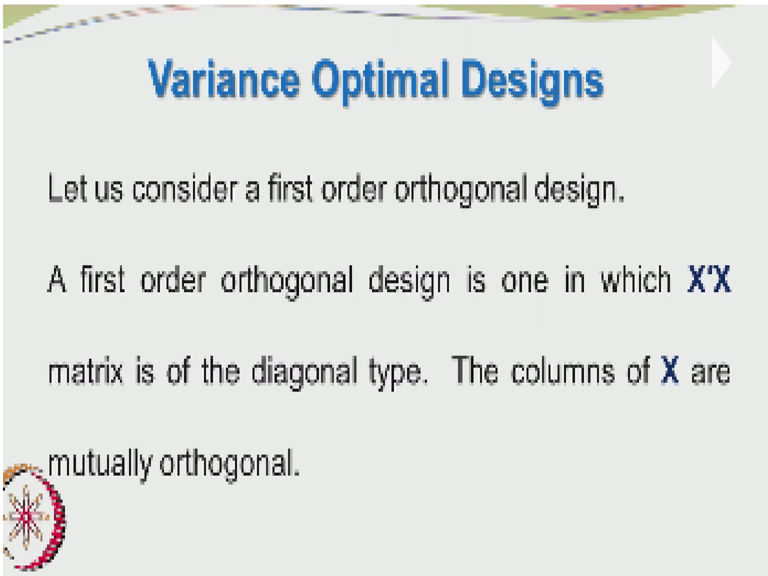
All you are finding is the determinant of the moment matrix and we also define the moment matrix as X prime X/ N. So, we have defined M earlier as X prime X/N and we are simply taking the determine of this M matrix for a particular design under consideration. Then among all possible designs, we are trying to find that design, which will maximize the determinant value of M.

(Refer Slide Time: 18:13)



And we are also scaling this entire ratio by 1/P, where P is the total number of parameters and so now that we have defined the D optimality, we may use this criterion to compare between different designs and it is also a useful measure of the quality of the estimated model parameters.

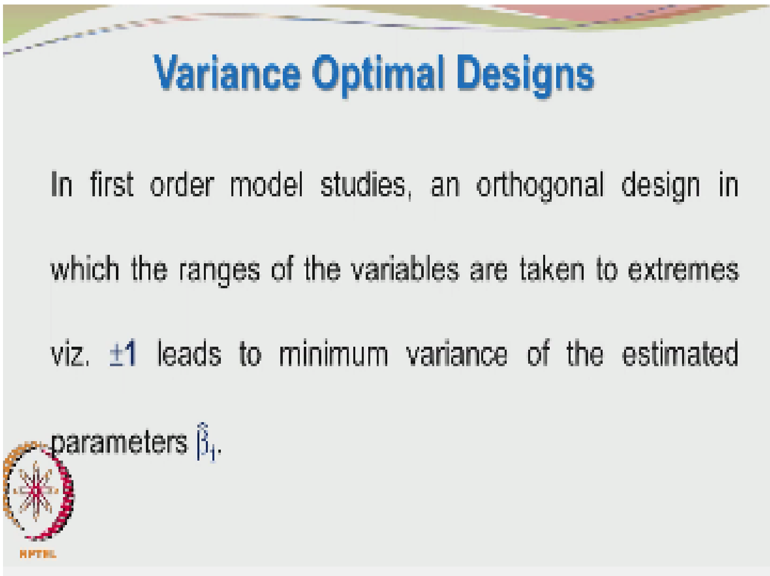
(Refer Slide Time: 18:35)



So, the next criterion is the variance optimal design and for illustration purposes, we will consider a simple first order orthogonal design and a first order orthogonal design is 1, in which the X prime X matrix is of the diagonals type and this I think, you should know by now, you take 2 power 2 regular factorial design and then you set up the X matrix and then when you compute X prime X, you will get a diagonal matrix.

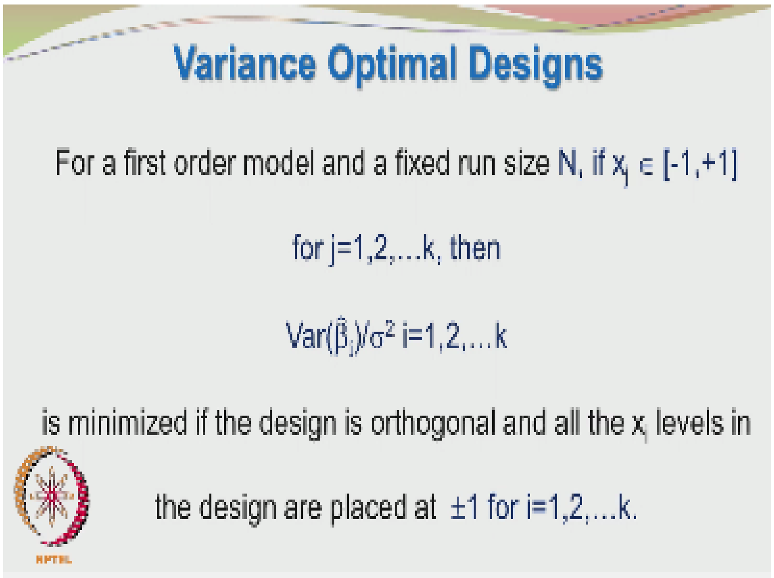
And we can say that the columns of X are mutually orthogonal for a first order factorial design and so when I take the transpose of one column vector and multiply that with a column of; with the another column vector, I should get 0 as the sum because in an orthogonal design, you have equal number of positives and negatives and when I multiply one column with respect to another column, the net answer should be 0.

(Refer Slide Time: 19:51)



This is assured in orthogonal designs. We also know that in first order model studies, an orthogonal design is such that the variables are located at extremes, we have coded the variable levels as lying between -1 and +1, the lowest limit of the variable is called as -1 and the upper limit of the variable is called as +1 and the experimental points are kept at the extremes, they are located at -1 and +1.

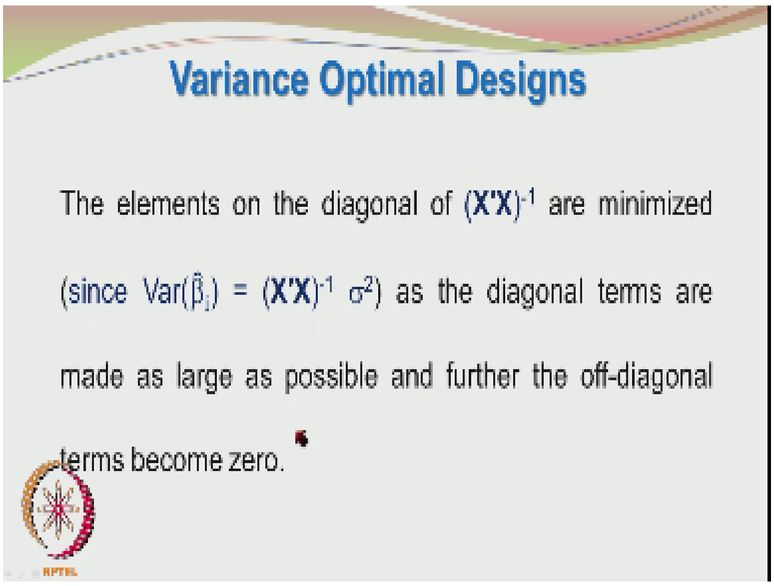
(Refer Slide Time: 20:42)



Since these values are located at extremes, the X prime X inverse matrix would be quite favourable to us, it will be quite small. Let us say that we are considering a first order model and run size of N, capital N, let xj values be defined such that they are falling between -1 and +1 and for parameters j is = 1, 2 so on to k, we can find the variance of beta hat i/ sigma square from the variance covariance matrix nothing but X prime X inverse matrix.

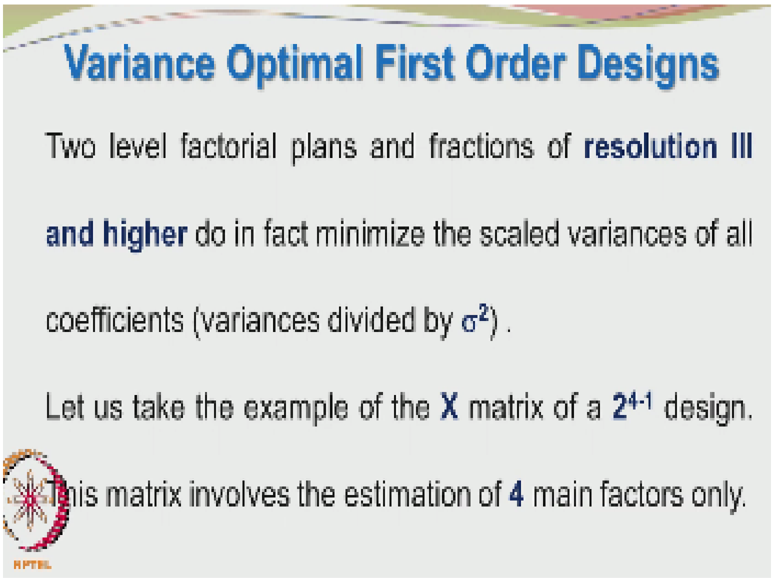
And in such cases, the variance beta hat/ sigma squared is minimized, if the design is orthogonal and all the xi values in the design are placed at plus or minus 1, for i is = 1, 2 so on to k. So, what I am saying is this variance of beta hat prime/ sigma square is lowered or minimized, if the experimental design is orthogonal in nature and all the xi levels are located either at -1 or at +1.

(Refer Slide Time: 22:02)



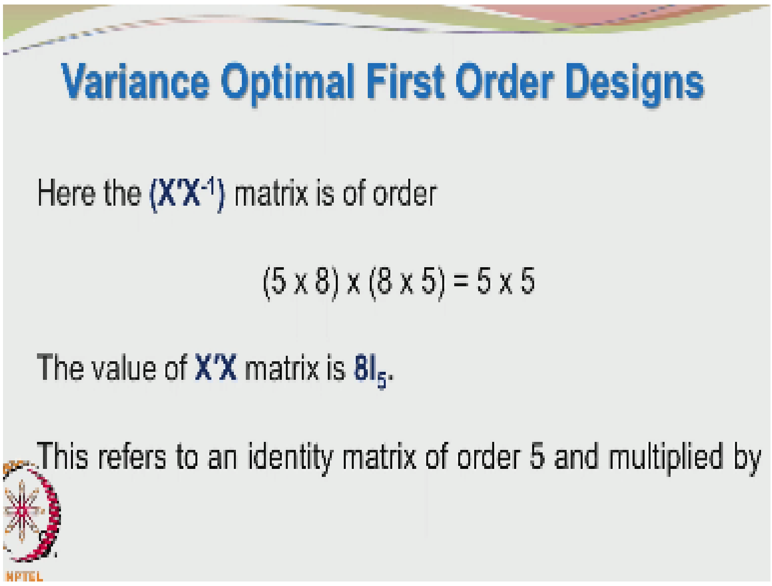
So, the elements of the diagonal of X prime X inverse are minimized as diagonal terms are made as large as possible and further the off diagonal terms become 0, in orthogonal design, the off diagonal terms are 0, so that seems to a lot of headache and at the variance of the estimated parameters decrease, when X prime X inverse is minimized.

(Refer Slide Time: 22:30)



So, now let us look at variance optimal first order designs, so we are now looking at 2 level factorial plans and fractions of resolution 3 and higher, do in fact minimize the scaled variances of all the coefficients; variances/ sigma square are minimized in designs of resolution 3 and higher. So, now let us take an X matrix coming from a 2 power 4 - 1 design, we are looking at a fractional factorial design, half fraction design, 2 power 4 – 1.

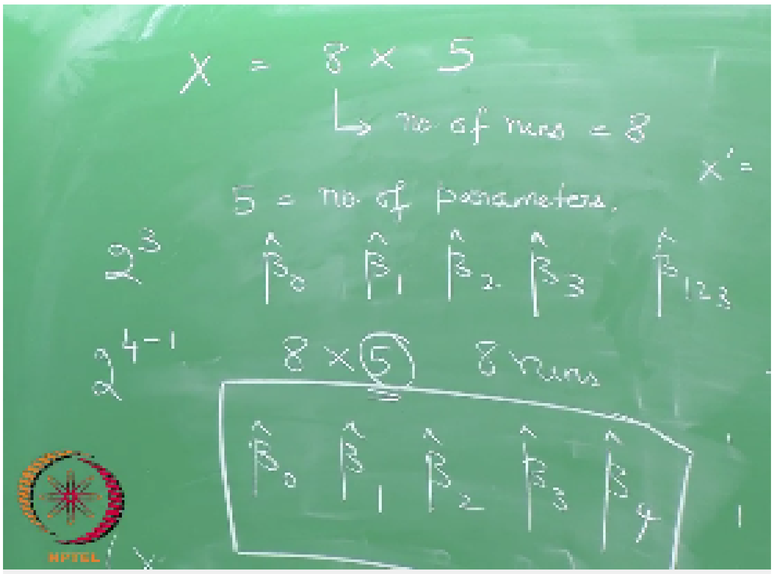
(Refer Slide Time: 23:29)



That means, instead of doing 2 power 4 experiments, we are doing only 2 power 3 experiments, so we are doing only one fraction; one 1/2 fraction of 8 runs, when there are totally 16 runs possible and in a 2 power 4 - 1 design, you can still estimate 4 main parameters; the X prime X inverse matrix in this case would be 5 by 8 * 8 by 5 or in other words, it would be a matrix of size 5 by 5 and the value of the X prime X matrix is 8i5.

And so, when you are looking at design of 5 by 8, so or 8 by 5 that means you are having the X matrix of dimensions 8 by 5, so you are having 8 runs, the 8 runs may be because of a 2 power 3 full design or a 2 power 4 - 1 fractional factorial design, so that is how you get 8 and how come there are 5 columns? The 5 columns maybe in the 2 power 3 factorial designs, the intercept factor 1, factor 2 and factor 3.

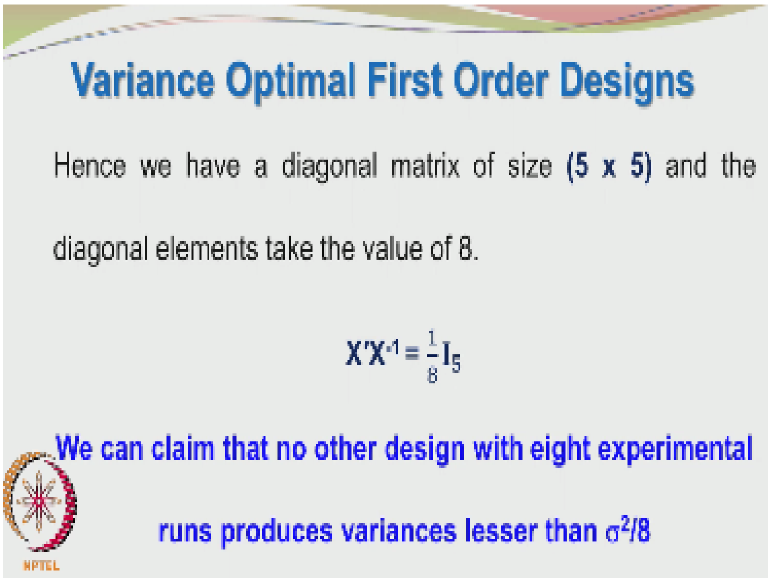
(Refer Slide Time: 25:09)



And then you also have the interaction between the factors, so let me just illustrate that also. When I am talking about X matrix; X matrix of 8 by 5; 8 rows and 5 columns, this may be number of runs and that is = 8 and 5 represents number of parameters. What could be the parameters? In the 2 power 3 designs, they can be beta hat 0, beta hat1, beta hat 2, beta hat 3, so this makes it 4 parameters.

And then, you can have one more parameter, let us say beta hat 123, this is just an example. If you are having a 2 power 4 - 1 design, this is what we are looking at. If you are having 8 by 5, it means we are doing only 8 runs and this 5 parameters will simply correspond to beta hat 0, beta hat 1, beta hat 2, beta hat 3, beta hat 4, so the 5 columns will comprise of the vectors corresponding to the estimation of the 4 parameters.

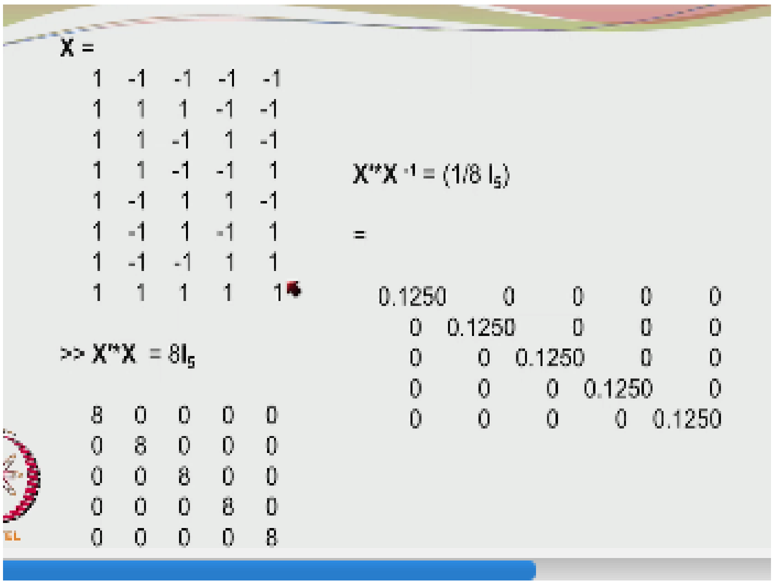
(Refer Slide Time: 27:25)



Here, beta hat 1, beta hat 2, beta hat 3, beta hat 4, in addition to the beta hat 0, so this is what we have in such kind of designs, so this 5 represents the number of parameters we are estimating. Now, that is clear, we can find the value of the X prime X matrix in such a situation involving 8 runs and 5 columns as 8i5, 8th multiplying an identity matrix of order 5 that is what I have said here.

Here, we have a diagonal matrix of size 5 by 5 and the diagonal elements take the value of 8 and what would then be X prime X inverse? That would be 1/ 8 * I5, the inverse of 8 is = 1/8 and the inverse of an identity matrix is the identity matrix itself of the same order. We can also even make a tall claim that there is no other design with the 8 experimental runs that can produce variances of the estimated parameters smaller than sigma squared by 8.

(Refer Slide Time: 28:18)



This is an important result. So, now I have a slide, which shows what I have been talking so far, you have the X matrix and let me just go back a few slides, we are talking about a 2 power 4 - 1 design, the slide I am going to show next is based on a 1/2 fraction of a 2 power 4 factorial design or a 1/2 of a 2 power 4 design or a 2 power 4 - 1 design, obviously this would require 8 runs and so these are the 8 rows in the X matrix.

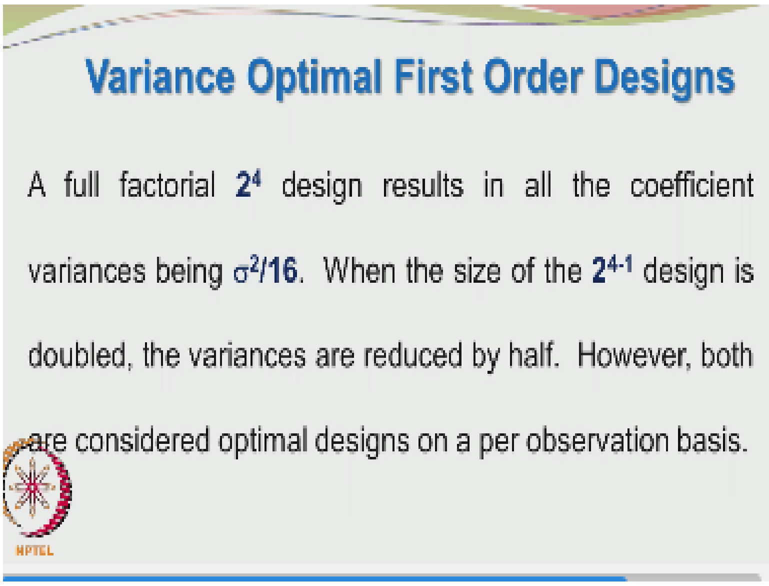
And you can see that we are having 2 power 3 factorial design here; -1, -1, -1, 1, 1, -1, 1, -1, 1, 1, -1, -1, and so on and the last column is obtained by multiplying the 3 columns that I think we know why and how, okay. Please see the lecture on fractional factorial designs, where we are talking about design generators and for this particular case, the design generator happens to be I is = A, B, C, D or D is = ABC.

So, the columns corresponding to factor D is obtained by multiplying columns A, B and C, so this is what you get as the X matrix and then when you do X prime X, you get 8i5 that means you get a diagonal matrix having 8 along the main diagonal and if I take 8 outside, I will get 8 into an identity matrix of order 5 that is called as i5 and then, you are also having X prime X inverse as 1/8 i5 and that is = 0.125, 0, 0, 0, 0, 0.125, 0.125, 0.125 is nothing but by 1/8.

And that is what is given here, if I take 0.125 outside, I will get 1/8, which is = 0.125 and then multiplied by the identity matrix of order 5; 1, 2,3, 4, 5; 1, 2, 3, 4, 5 that is right, okay. Next, we will go on to the variance optimal first order designs, you can see that the variance is showing up so frequently, we have variance in the experimental measurements, we are having analysis of variance ANOVA.

Then, we have variances in the regression in several forms; standard error and then we also have variances in the regression coefficients, we have covariances between the regression coefficient and not only those, we also have variances in the model predictions and now we are talking about variance optimal designs. So, the design of experiments course is basically an attempt to understand the phenomenon of variance in experiments.

(Refer Slide Time: 32:26)



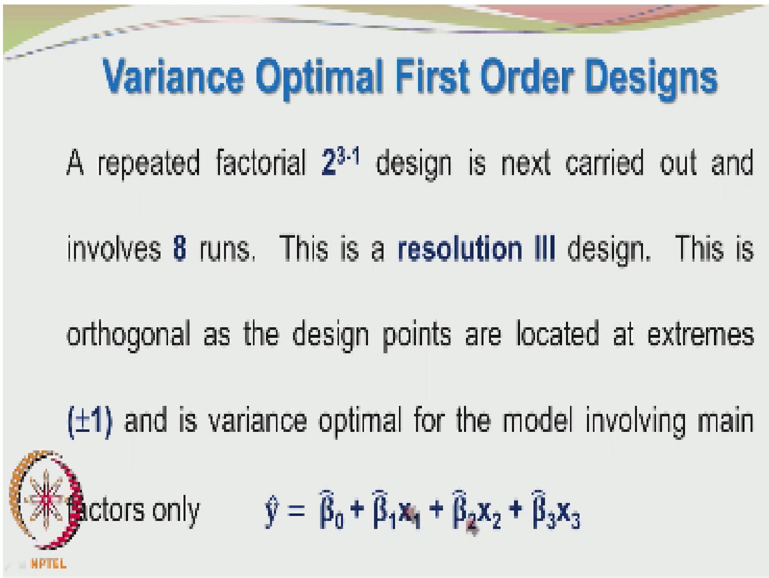
So, you have a full factorial 2 power 4 design and that would lead to the regression coefficients variances to be sigma squared/16. We know that the variance covariance matrix is given by X prime X inverse Sigma square and X prime X inverse happens to be diagonally matrix for orthogonal designs and the variances of the coefficients would then be sigma squared/ 16. So, here an important thing is the presence of 16, okay.

So, the diagonal terms is sigma squared/ 16, here we saw that the diagonal terms were having 1/8, so this was having 8 runs and then when you have a full factorial 2 power4 factorial design, you have 16 runs and the X prime X inverse sigma square becomes the diagonal terms take the value sigma squared/ 16, so it appears that for 2 power 4 design, the variance is lower; than variance of the coefficients are lower than the variance of the coefficients for a 2 power 3 design.

A 2 power 3 design had only 8 runs, a 2 power 4 design has 16 runs and we know that the variance covariance matrix X prime X inverse becomes 1/N along the diagonals into sigma square and that sigma square/N becomes the variance of the regression parameter. So, in such a case, you look for large set of experimental data thinking that it would reduce the variance of the estimated parameters that is not a correct way of looking at it.

Because this is an artificial way of reducing the variance, so the best way is multiply the variance covariance matrix by N. So, if you have N * X prime X inverse sigma squared that N will automatically cancel out the N coming inherently in the X prime X inverse matrix and then the variance coefficients would be independent of the size of D. So, then when you are scaling for the size of the run, then both the 2 power 3 design and the 2 power 4 design are may be considered as optimal designs on a per observation basis.

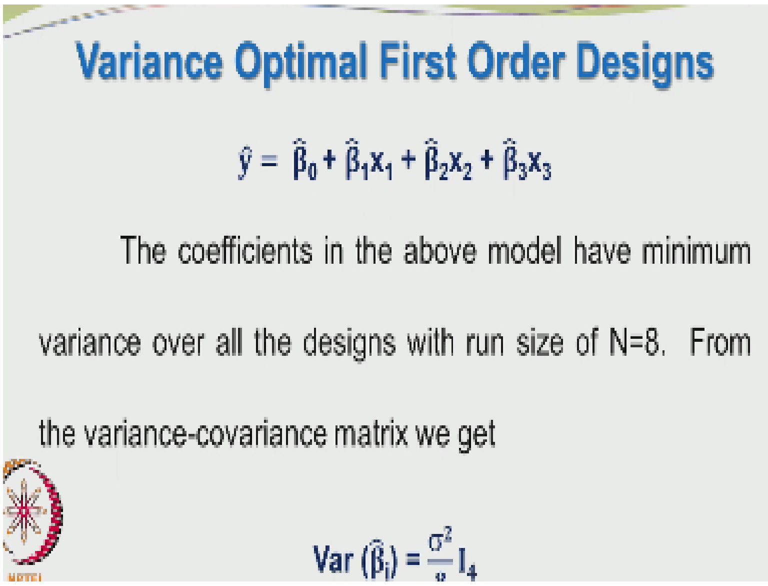
(Refer Slide Time: 35:47)



So, now let us look for a repeated 2 power 3- 1 design and this involves 8 runs, you are looking for a 2 power 3 -1 design and that would be only 4 runs but we are talking about 8 runs, so is there a mistake was it a 2 power 4 – 1 design or was it a 2 power 3 -1 design only, it is actually a 2 power 3 -1 design and you had 8 runs because you repeated the experiments. Each setting was repeated twice and so you have a 2 power 3 – 1 design having 8 runs.

Even though, there are only 4 independent settings, so this repeated the 2 power 3 - 1 design is orthogonal and the design points are located at extremes, we are located at +1 or -1, they are located at the boundaries and this variance optimal for the model involving main factors only. So, Y hat is = beta hat 0 + beta hat1 x1 + beta hat2 x2 + beta hat3 x3, so this is a saturated design, you cannot go beyond this.

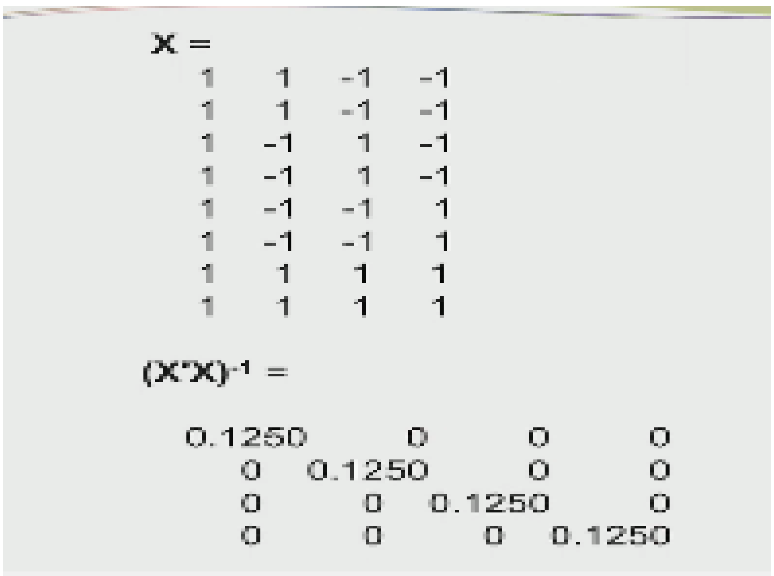
(Refer Slide Time: 37:22)



Because you are having only 4 independent settings and that means you can maximum estimate only 4 parameters even though, you have 8 runs, you are having only 4 independent settings and you can hence estimate only 4 independent parameters and now the regression coefficients in the above model have minimum variance over all the designs with run size of N is = 8, so from the variance covariance matrix, we can easily find out that the variance of the beta hat i.

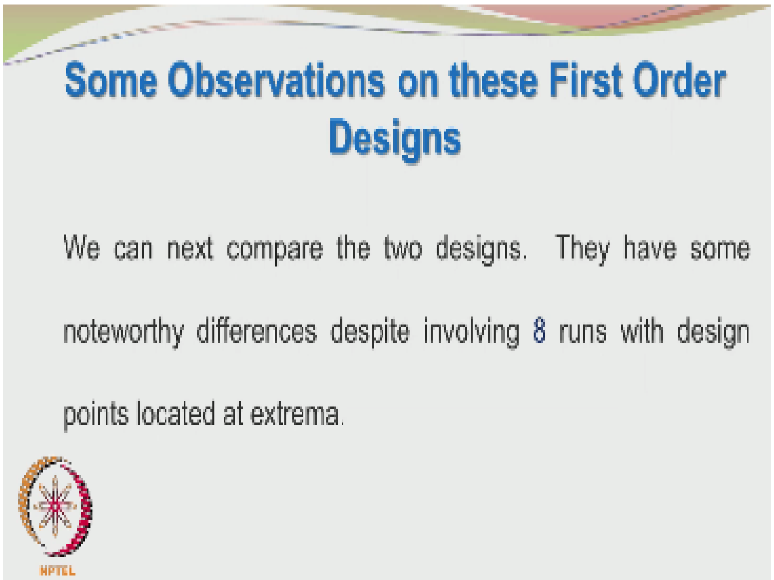
The regression parameter is sigma square / 8 * i4; sigma square is the unknown error variance, 8 is the size of the run and I4 is a diagonal matrix of order 4. So, very interesting and since we do not know the sigma square, the experimental error variance, we use the mean square error, we find the residual sum of squares divided by the degrees of freedom for the residual sum of squares and then we get the mean square error.

(Refer Slide Time: 38:32)



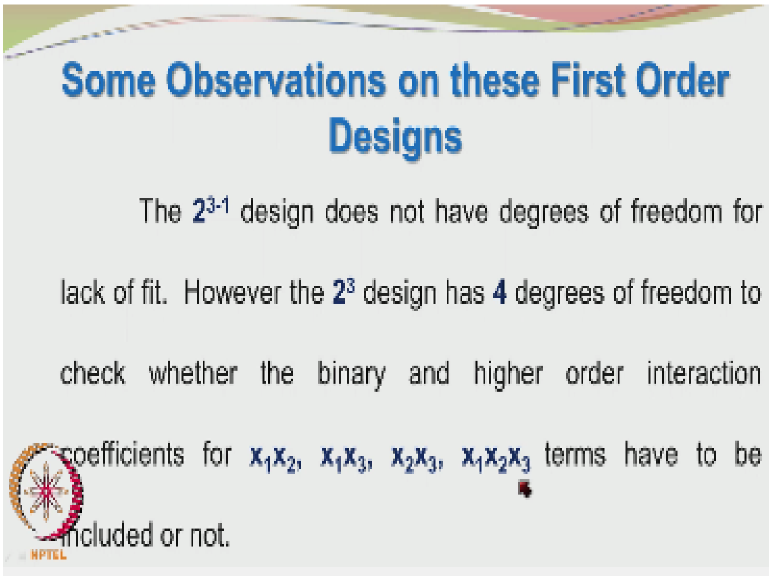
We use mean square error instead of sigma square. So, how did we get that for a 2 power 3 - 1 design, you can see that each setting is repeated twice 1, 1, -1, -1, 1, 1, -1, -1, so they are repeated twice and hence we have 8 runs, so X prime X inverse in such a case would be 1 by 8, 1 by 8, 1 by 8 along the main diagonal and 0 along the off diagonals.

(Refer Slide Time: 38:59)



So, now we can compare the 2 designs and so, what are we actually comparing? We can compare a 2 power 3 full factorial design involving 8 experiments, 8 independent settings but no repeats, then we are also taking into consideration a 2 power 3 – 1 design involving again 8 runs but only 4 independent settings and each independent setting has been repeated twice. So, even though both the designs are having only 8 runs, they have some important differences.

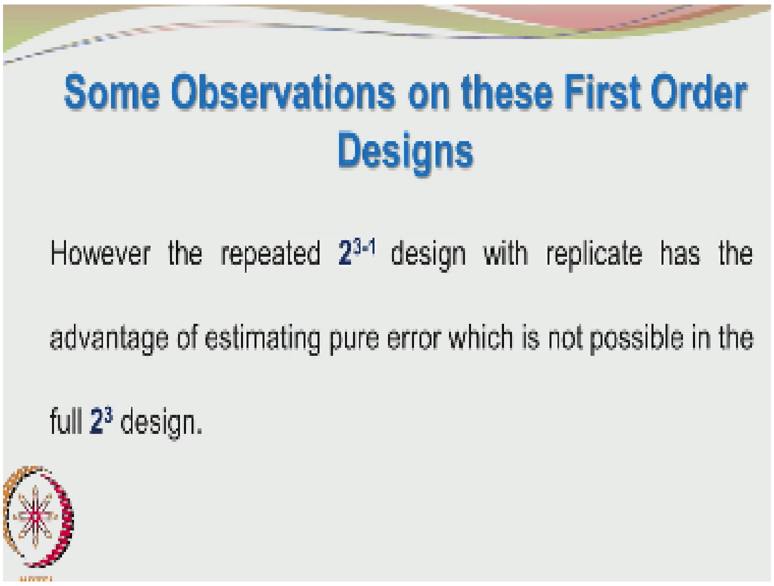
(Refer Slide Time: 39:55)



Both are variance optimal designs but what are the differences? The 2 power 3 – 1 design does not have degrees of freedom for lack of fit however, why does not the 2 power 3 – 1 design does not have degrees of freedom for lack of fit? That is because you are having only 4 independent settings, you have estimated all the 4 parameters; beta hat 0, beta hat 1, beta hat 2, beta hat 3 and you are left with no degrees of freedom thereafter for identifying more parameters.

However, in a 2 power 3 design, you have 8 independent settings and you have found only 4 parameters, the same once I listed just a bit earlier; beta hat 0, beta hat 1, beta hat 2, beta hat 3, so you have estimated 4 parameters but there are 4 more degrees of freedom for expanding the model. So, you can use those degrees of freedom additional or extra available degrees of freedom to estimate the 3 binary interactions and 1 ternary interaction.

(Refer Slide Time: 41:06)

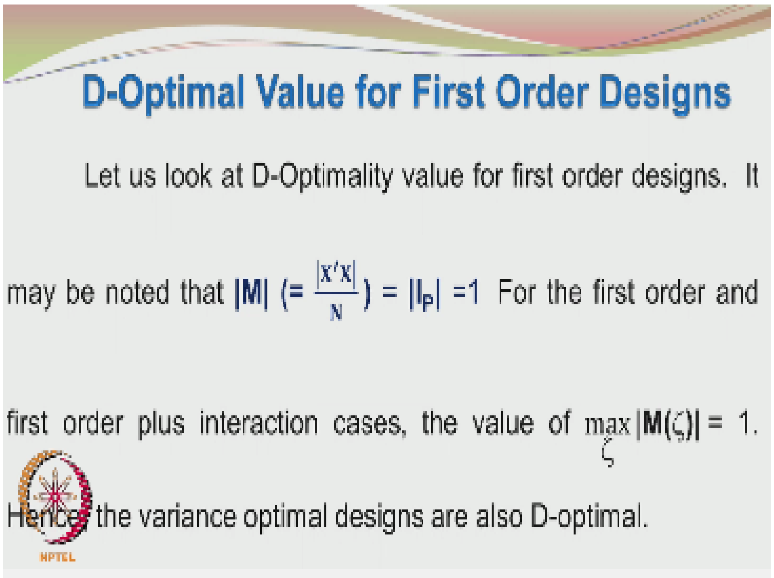


However, even though the 2 power 3 - 1 design does not have degrees of freedom for lack of fit and hence you cannot expand upon the basic model, you have center points repeated sorry; not center points repeated, the factorial points are repeated and hence you can have a good idea about the experimental error. So, on one hand, you can expand upon the model and go for a more sophisticated model.

But since there were no repeats in such a design, you cannot have an idea about experimental error, on the other hand in the second design, which again involved 8 experiments but with the repeats, you cannot expand upon the basic model but you can have idea about the experimental error. So, which design you will go for depends upon what information you already have with you.

If you know that there are no interactions in your model based on prior process experience, then you can work with 2 power 3 – 1 design with the repeats, so that you have an idea about the experimental error. On the other hand, if you suspect that interactions are there and you already have an idea about the experimental error based on previous knowledge, then you can go for a full 2 power 3 design and try to estimate all the interactions; higher order interactions in addition to the main effects.

(Refer Slide Time: 42:41)



So, what is the D optimality value for first order designs and so the determinant of the moment matrix is equal to determinant of X prime X/ N and that is = 1.