Statistics for Experimentalists

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

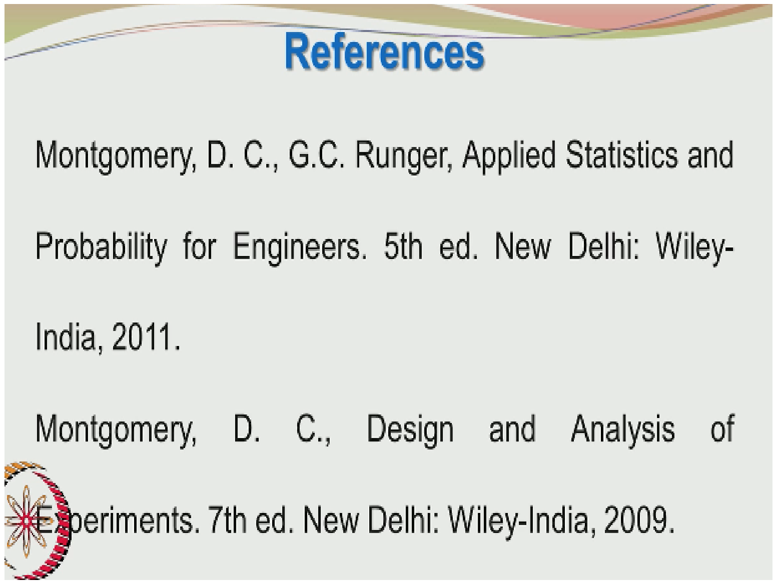
Lecture - 32
Fractional Factorial Design- Part A

Welcome back to the lectures on statistics for experimentalists we have completed the factorial design and also the general factorial design we learnt how to identify the main effects the interactions from the factorial designed experiments. We also carried out the analysis of variance to identify the important effects influencing the response. The factorial design is an economical and efficient way to carry out experiments.

Especially in industry when we want to do experiments it would mean a lot of investment in manpower time resources. So, it is important that we try to get the maximum information from I would not say minimum efficient economical number of experiments. So, in the industry and also when doing research, we really do not know what are all the factors that may influence the response.

We can look at literature and see what other people have done when there are a large number of factors especially in product design or development of a new process. We need to identify the appropriate factors, so we will be looking at fractional factorial design and as the name or the title suggests we will be looking at fraction of the factorial designs so how to identify the fractions we will form the outline for this course.

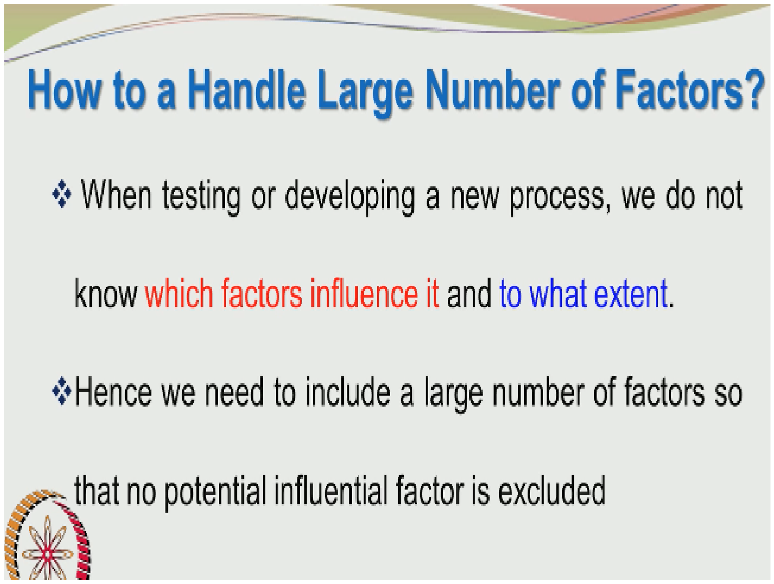
(Refer Slide Time: 02:36)



So, looking at the references we have the books written by Montgomery and Runger applied statistics and probability for engineers. This is a concise book starting with aspects of probability moving on to inferential statistics hypothesis testing parameter estimation method of maximum likelihood, probability distributions and then it also deals with factorial designs fractional factorial designs.

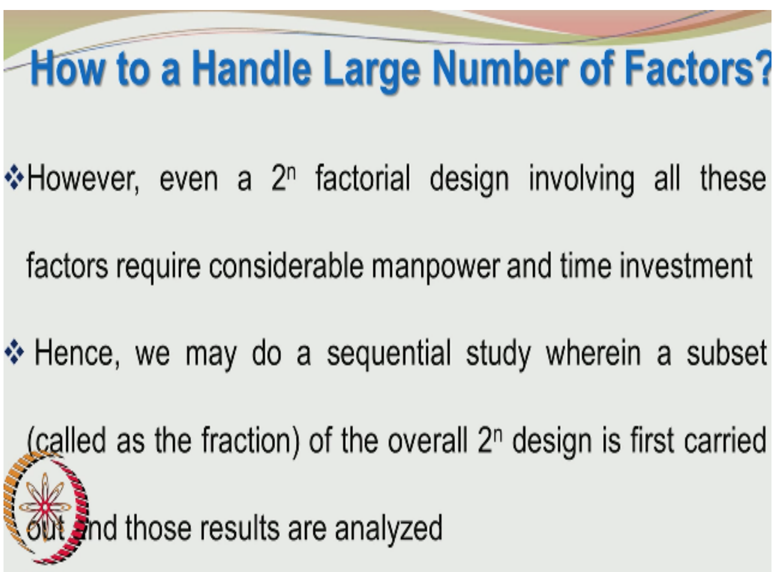
So, it is a single point reference for both students who want to learn the subject and also for practitioners. Then we have a more detailed book written by Montgomery and this gives more designs more advanced designs in addition to the basic ones. We will be referring to both these books during this lecture.

(Refer Slide Time: 03:44)



So, when we want to develop a new process or design a new product sometimes we may not know which factors influence it and to what extent and hence we do not want to leave out any factor and we want to include as many factors as possible. But when you add more and more factors to be on the safe side you are also adding on to the size of the experiment the number of experimental runs you have to make the investments you have to provide.

(Refer Slide Time: 04:27)

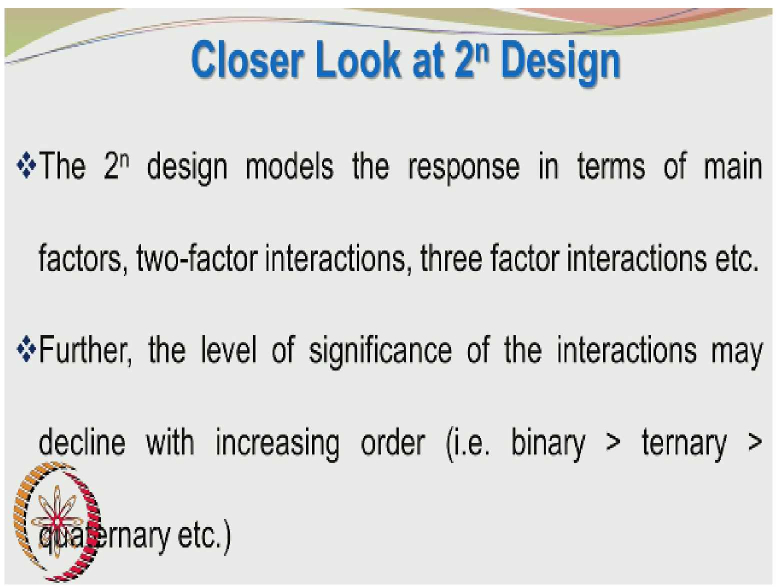


So, what would be a logical step is to do the experiments in a sequence we carry out a fraction of the factorial design even though the factorial design usually 2 level factorial design is efficient and reduces the number of runs. When the number of factors increase to a large value we need to even further economize for example if you are having 5 factors 2 power 5 is 32 and if you want to do at least 2 repeats.

It would mean 64 runs 64 runs may take a long time to complete so we need to do a fraction of the possible experiments given by the factorial design. See what information it provides even though it may not give complete information we use the information to get an idea it may even tell at least that 1 factor is not important okay that itself is a good outcome of doing only a fraction of the experiments.

So, what we may do is divide the experiment into equal fractions and do the experiments sequentially such that we keep building upon the information from the different fractions and we may stop at any time once we feel that okay we have enough information or we have enough handle on the process and we can take it up from there.

(Refer Slide Time: 06:16)



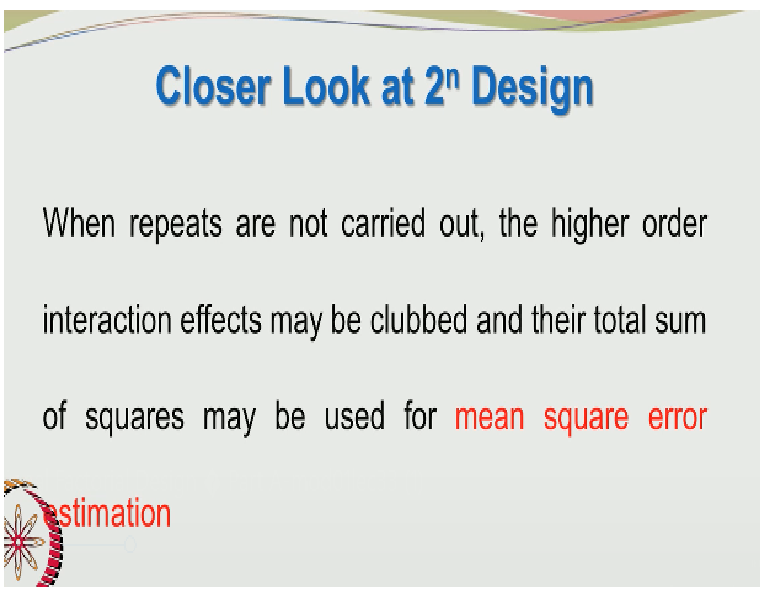
So, the characteristic feature of a 2-power n design is the response is given in terms of the main factors 2 factor interactions 3 factor interactions and so on. Actually, if you look at the information provided by the factorial design we sometimes or oftentimes may not be able to really make sense of the interactions. What is really a 2-factor interaction mean in a physical sense? for example if you are looking at temperature and pressure.

You may superficially believe that temperature and pressure for example T*P what is the meaning of T*P these having units of kelvin and pressure is having units of pascal or newton square meter per square. How can you combine T*P? Well normally we do coding, so all these factors become dimensionless entities when they take certain values so by coding we restrict them in the range of-1+1 and so on.

So, they are actually dimensionless, and temperature and pressure interaction simply mean that the response depends upon a particular factor, but that factor variation depends upon the level of the second factor and so on. So, we can have 3 factor interactions like this so when you have such a situation and you have a considerable number of interactions usually these interactions up to order 2 or 2 factor interactions are significant.

Very rarely 3 factor interactions become important and even more rarely 4 factor interactions would be important. But I have seen experiments general factorial designs where even 3 factor interactions were important. So, we cannot rule anything out and when you do a 2-power n design with n being the number of factors and you have a large number of factors then the number of higher order interactions start to increase in a rather alarming manner.

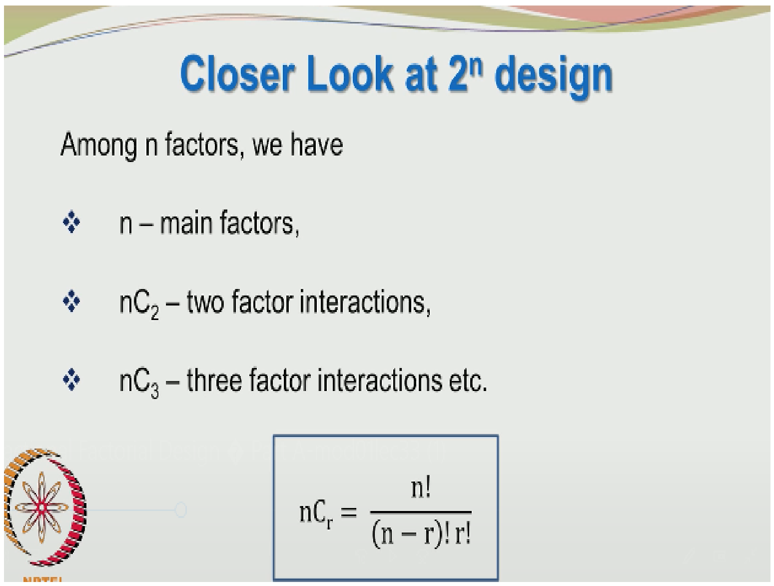
(Refer Slide Time: 09:11)



Okay so 1 interim possibility is to carry out the full factorial design and not do repeats, so you carry out the full factorial design and then you say that the higher order interactions are not important, and they may be clubbed or combined and may be used to have an idea of the random fluctuations. Okay so this is one way of doing that but as I said there may be instances where even 3 factor interactions maybe important.

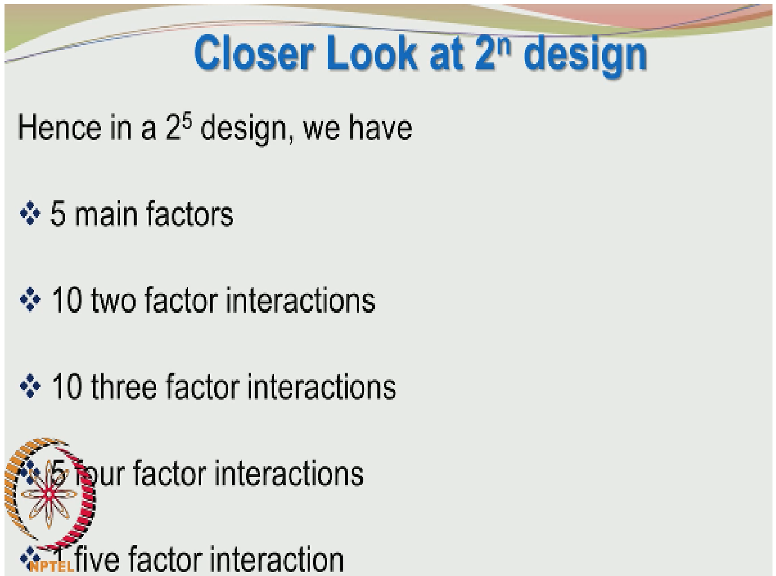
And clubbing them into other higher order interactions and calling them as a random contribution may lead to wrong conclusions. So, a better way is to carry out the fractional factorial design. So, how to carry out this fractional factorial design we would see without further ado.

(Refer Slide Time: 10:26)



In a 2-power n design you have in a 2-power n design n main factors and number of 2 factor interactions can be calculated by nC2 and higher order third order interactions may be calculated as nC3 and so on. And I hope you know that when nCr is being computed it is n factorial/n-r factorial*r factorial.

(Refer Slide Time: 11:10)



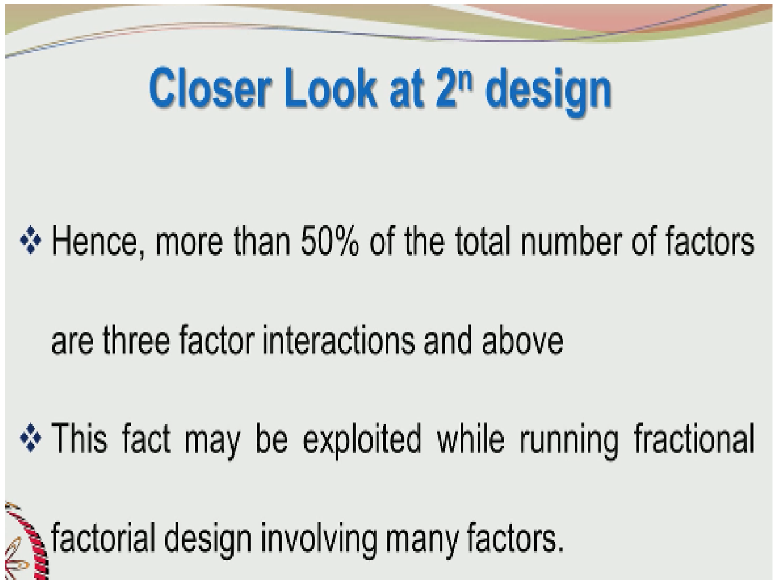
Applying this formula in a 2 power 5 design we have 5 main factors 10 2 factor interactions. How do you get 10 2 factor interactions? it is 5C2 and so when you have 5C2 it is 5 factorial/3 factorial *2 factorial, 5 factorial is 120 and n-r 3 factorial 6 6*2 12. So,120/10 is 120/12 is 10 5C3 is also 5C2 and that also leads to 10 3 factor interactions and when you have 5 4 factor interactions which is 5C1 and then you have 1 5 factor interaction itself.

So, when you add up these 15 25 30 31, so you have 31 effects and then you also have the constant beta 0 in the model. So, when I look at a 2 power 5 design I can see that there are a huge lot of interactions out of 32 6 are pretty much nonexistent or ineffective. So, 6 out of 32 is about 20% percent of them are not really useful and if I even take the 3 factor interactions 16 out of 32 would be 50% 50% of the effects and determining are possibly not going to affect the response.

And then I have 5 main factors and 10 2 factorial interactions even among the 5 main factors there may be 1 or 2 main factors or 1 or 2 factors which may not influence the response of the experiment. So, let us assume that I am doing only a full 2 power 5 design and that would be 32 runs if there are 2 factors which are not really effective then it is actually a 2 power 3 design which is 8 runs only are required.

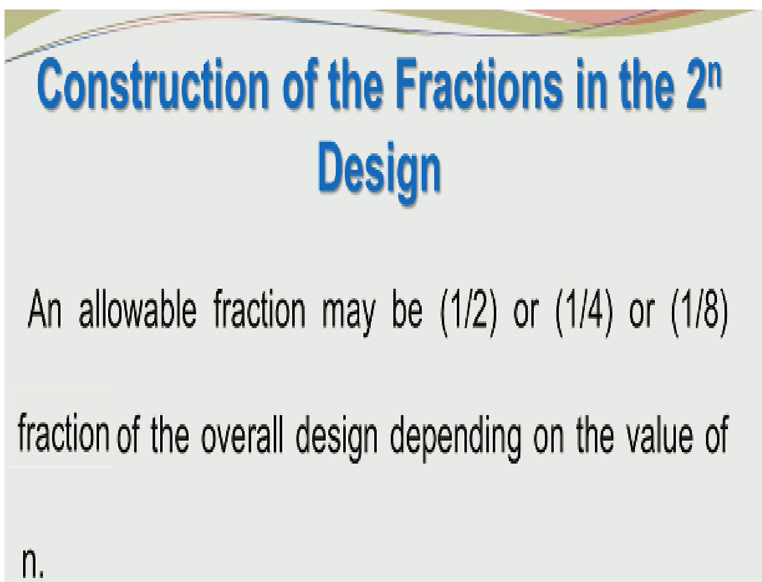
So, I am doing 32 runs so essentially, I am doing the experiment 4 times so these are some issues we will have to consider before we commit ourselves to investment of time manpower and money.

(Refer Slide Time: 14:26)



So, since there may be 50% wastage if you may want to put it like that in terms of the number of important factors then we may decide to go on to a fractional factorial design.

(Refer Slide Time: 14:47)

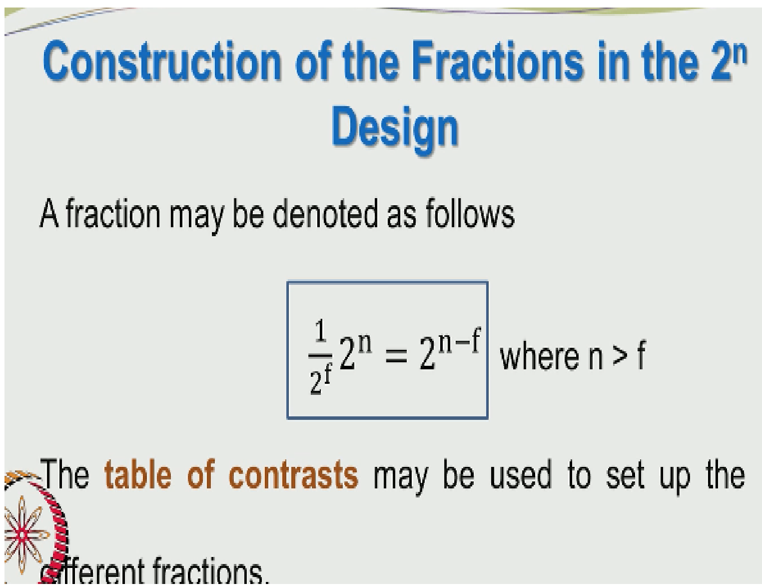


So, how to do it what we do is we take fraction of the total number of runs that fraction maybe 1/2 that is 1 half or 1/4 which is 1 quarter or 1/8 fraction of the overall design depending on the value of n. Obviously when you have 2 power 3 design you can go for a 1/2 fraction when you have a 2 power 4 design involving 16 runs then you may consider 1/4 fraction or 1/2 fraction would be more logical.

And 1/8 you may go for a when you have 6 factors that means 32 2 power 6 is 64. So, 1/8 would be a fraction involving eight experiments. So, you cannot reduce the number of experiments to an observed value for 2 power 3 that is 2 power 3 design you may want to do it fully but if you are in a hurry you may first do the 1/2 fraction for 2 power 3 design doing a 1/4 fraction is absurd you may go for 1/4 design for at least 2 power 4 factorial design.

So, that you have to do 4 experiments, 1/8 for a 2 power 4 is meaningless that would mean only 2 experiments that is not good you may want to go for 1/8 when you have a 2 power 5 or 2 power 6 design. So, after deciding the fractions let us see how to calculate the effects.

(Refer Slide Time: 16:59)



So, to put it formally a fractional factorial design maybe represented as 2 power n-f. We represented the 2-level factorial design as 2 power n and when you start looking at fractions we put it as 1/2 power f 1/2 power f *2 power n that would mean it is 2 power n-f where n is > f and f is of course and a whole number. Okay so you can have f as 1 2 so on to n-1 n also cannot take any real value it can take only whole numbers 1 normally 1 is not used n maybe 2 3 4 and so on.

Again, we resort to contrasts to set up the fractional factorial design when we set up the fractional factorial design we do it with the full knowledge that by analyzing only a fraction of the total number of runs we may be dealing with potential loss of information. What do you mean by that? we are taking all the variables in a fraction and to get all the variable effects uniquely we need a certain minimum number of runs.

For example, when you are doing a 2 power 2 design we do a minimum of 4 runs, the 4 runs are required to estimate the intercept or beta 0 and also to find the 2 main effects that makes it 3 and then we also need to find the interaction between the 2 factors and that makes it 4. So, we need at least 4 runs to get all the 4 coefficients like beta 0 coefficient to factor A then the coefficient to factor B and then also the coefficient to factor AB.

What are these coefficients these coefficients are present in the model developed to relate the response to the different factors? So, when you are doing only a fraction of the total number of runs involving a certain number of factors then you have to understand that there are more variables than the number of experiments you are doing, and this may lead to potential problem called aliasing.

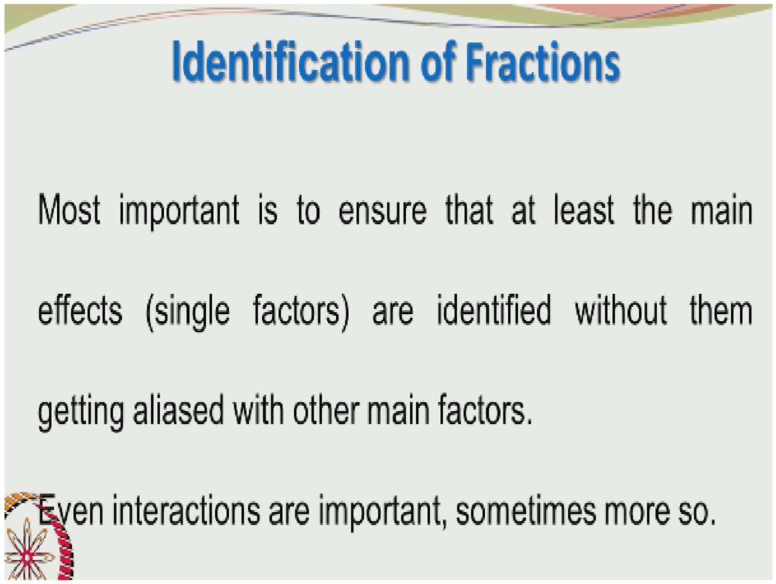
Sometimes when people mention in the news about criminals they sometimes tell so and so alias some other name and so on. So, the same person is having different names and that is called as alias. What does this got to do with our design? in our design whatever we are finding the effect the effect may in fact represent more than 2 more than 1 factor. Earlier in our discussion we found the effect and that effect was directed to a particular factor.

So, we said the main factor A is having this much value main factor B is having this much value interaction between A and B is having this much value. But when you do fractional factorial design once you calculate the effects you do not assign the effects uniquely to a given factor to 1 factor. You may assign the effect to more than 1 factor so what it means is the effect you have obtained may be the contribution from more than 1 factor.

And we are restrained from calculating the factors uniquely or we are restrained from calculating the effects of factors uniquely because we are doing only less number of runs. So, this we have to keep in mind to summarize when we do fractional factorial designs we do calculate the effects in the same way as we did earlier. But the effects may no longer be uniquely representing a particular factor it may represent a combination of more than 1 factor.

So, in our fractional factorial design what we are going to do is to see which are all the factors which are aliased to a particular effect.

(Refer Slide Time: 22:55)

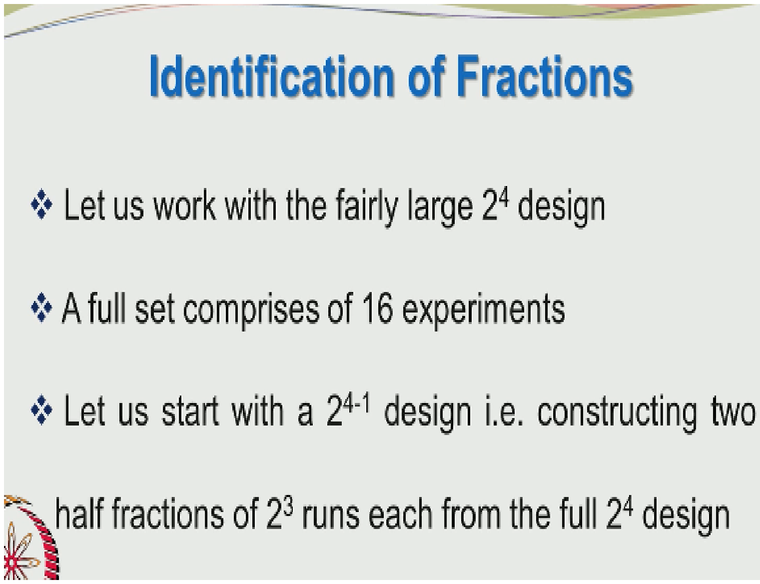


So, it is considered to be important that when the effects are combined or aliased we have to make sure that at least the main factors do not get aliased with each other. So, when an effect is getting calculated we do not calculate the contributions of A and B. The combined contributions of A and B, A is important B is important so when you have an effect based on both A and B then the information is not completely coming out from the design.

But if you find that the A factor or any main factor is aliased with 3 factor interactions or higher order interactions then most likely that the effect you have calculated is only because of the main factor. Even though other factors are present the higher order interactions are present they may be making a very small contribution to the effect and so the effect may be attributed to only the main factor.

Sometimes in certain designs the 2 factors may be aliased with one another okay then we may have to do the next fraction. It is important not to ignore second order interactions because the second order interactions are very essential.

(Refer Slide Time: 24:51)

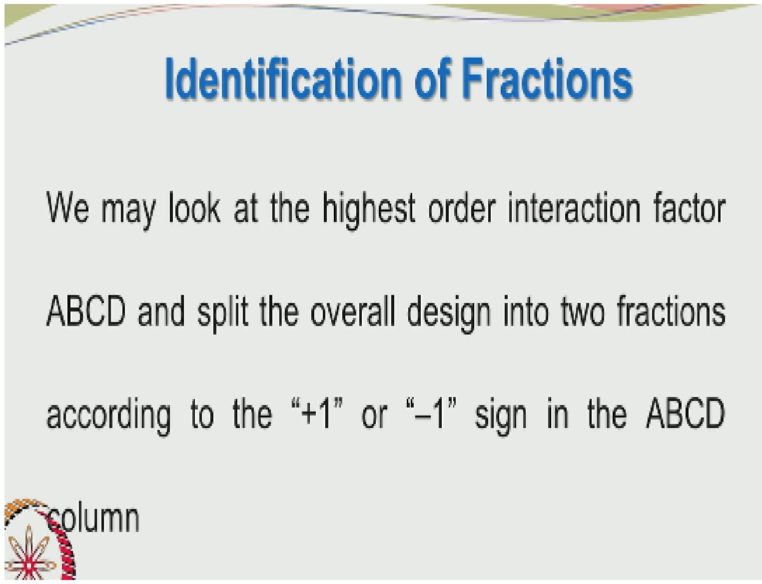


So, let us take a small example we will work with the 2 power 4 design 2 power 4 would lead to 16 experiments and multiplied by 2 to account for repeats and you have 32 runs and your boss may not really like the idea of doing 32 runs in the pilot plant. Okay the logical question he will ask is can you do a lesser number of runs let us take it from there. Obviously intuitively he is suggesting to you do a subset of the complete number of runs and see what results you get.

And this is precisely what the fraction factorial designs are about. So, we will talk about this in more detail with reference to a 2 power 4-1 design that means you are looking at 1/2 of 2/4. We are looking at 1/2 fraction of a 2 power 4 design and that means we are talking about a fraction involving 16/2 runs which is 8 runs okay and the first fraction will involve 8 runs and the second fraction will also involve 8 runs.

Of course, we may do repeats in the first set of 8 runs to get an idea about the random error. So, we want to construct a 1/2 fraction of the 2 power 4 factorial design. So, how to identify the elements of the fraction we cannot arbitrarily choose certain combinations and take it as a first fraction there is a systematic way of doing it. Let us see what this method is.

(Refer Slide Time: 26:59)



What you do is you write the full 2 power 4 design on paper there is no problem with this and look at the highest order interaction. When you are having a full 2 power 4 design the highest interaction is ABCD for 2 power 4 design ABCD the quaternary term is the highest order interaction and when you look at the design matrix you will find it having a combination of minus and pluses and in a 2 power 4 design.

You will have in each column corresponding to column A or column B or column AB BC CD AD ABC so up to ABCD you will find that the columns are containing minuses and pluses and the total number of minuses will=the total number of pluses. So, when you are having 16 possible runs you will have 8 pluses and 8 minuses in each column. So, please look at the column corresponding to ABCD that column will also have 8 minuses and 8 pluses.

You choose your first fraction according to ABCD the first 8 runs for the first fraction will correspond to the pluses in ABCD the ABCD column is having 16 entries 8 of them are positive or+1 and 8 of them are negative or negative 1 use the 8 pluses in the ABCD column to create your first fraction. So, I do not know how many of you followed this.

(Refer Slide Time: 29:14)

	A	B	AB	C	AC	BC	ABC
(1)	-1	-1	1	-1	1	1	-1
a	1	-1	-1	-1	-1	1	1
b	-1	1	-1	-1	1	-1	1
ab	1	1	1	-1	-1	-1	-1
c	-1	-1	1	1	-1	-1	1
ac	1	-1	-1	1	1	-1	-1
bc	-1	1	-1	1	-1	1	-1
abc	1	1	1	1	1	1	1
d	-1	-1	1	-1	1	1	-1
ad	1	-1	-1	-1	-1	1	1
bd	-1	1	-1	-1	1	-1	1
abd	1	1	1	-1	-1	-1	-1
cd	-1	-1	1	1	-1	-1	1
acd	1	-1	-1	1	1	-1	-1
bcd	-1	1	-1	1	-1	1	-1
abcd	1	1	1	1	1	1	1
	0	0	0	0	0	0	0

I will just demonstrate it, so we are having this design matrix as before these are the main factors binary interaction AB main factor C binary interaction AC BC binary interaction ABC ternary interaction.

(Refer Slide Time: 29:42)

	D	AD	BD	ABD	CD	ACD	BCD	ABCD
(1)	-1	1	1	-1	1	-1	-1	1
a	-1	-1	1	1	1	1	-1	-1
b	-1	1	-1	1	1	-1	1	-1
ab	-1	-1	-1	-1	1	1	1	1
c	-1	1	1	-1	-1	1	1	-1
ac	-1	-1	1	1	-1	-1	1	1
bc	-1	1	-1	1	-1	1	-1	1
abc	-1	-1	-1	-1	-1	-1	-1	-1
d	1	-1	-1	1	-1	1	1	-1
ad	1	1	-1	-1	-1	-1	1	1
bd	1	-1	1	-1	-1	1	-1	1
abd	1	1	1	1	-1	-1	-1	-1
cd	1	-1	-1	1	1	-1	-1	1
acd	1	1	-1	-1	1	1	-1	-1
bcd	1	-1	1	-1	1	-1	1	-1
abcd	1	1	1	1	1	1	1	1

And this also continues you have D AD BD ABD BCD and ABCD so that is done then if you look at the first column here these are the 16 possible settings of your 2 power for full factorial design. So, you have 1 a b ab c ac bc abc d and so on. And you should not have any difficulty in recognizing these for example 1 corresponds to the lowest setting of all the variables A means only factor A is at a+ level.

So, you can see that only a is at a+ level whereas b is at-1. So, ab would be-1 c is at-1 ac is at-1 bc would be-1*-1 abc is 1*bc which is+1.So you can easily create all the entries after you do for A B C and D that is the standard order. So, you are constructing the columns corresponding ABCD in a systematic standard manner and once you have ABC and D columns you can easily create the other columns AB BC CD and so on.

Even you can go up to the ABC BCD and also up to ABCD. So, the procedure is very simple, and that procedure is also outlined here and then these are also the different corners of the 2 power 4 design okay for example BD means that only factors B and D are at the higher settings factors A and C are at the lower setting why are these colored in blue and red I will explain this very shortly.

So, understand how the entire table was created first you do for A B C and then D which is going to be shown in the next slide. Once you have A B C and D then you can calculate everything else by simple multiplication. For this you can use the spread sheet to carry out the different calculations and produce the final table quickly. Right let us now look at ABCD ABCD is the highest interaction in this design mind you we are doing the full factorial design.

So, now you have 1 1 so all the blues I have colored as 1. How many+ 1 are there it should be 8 let us confirm 1 2 3 4 5 6 7 8. So, you have 8+1 and then you also you should have 8-1 so whatever+1 are there I am coloring it by blue and whatever-1 are there I am coloring it by red and my first fraction would be the entries corresponding to+1 in ABCD what it really means. We know that there are 8 pluses and 8 minuses.

So, I am going to only look at all the +1 to define my first fraction. How do I do it.?, So, corresponding to ABCD of+1 a is at-1 and here you can see ab is also at+1 and ac is 1 so these need not be all+1 they can be any value negative or positive, but they are corresponding to ABCD of+1. So, this is important, so I will collect all the +1 and put them together that means what I will be looking at all the settings corresponding to the blue color.

Remember the blue color corresponds to +1 in ABCD. I will be doing my first fraction at 1 ab ac bc ad bd cd abcd that means I will be doing my first fraction with lowest setting of all the factors and ab both a and b only at their high values ac both factors a and c at their high values and the remaining 2 factors b and d at low values word of caution here. Even though I am doing a 1/2 I am not excluding factor d factor d is also present here okay in the experimentation.

Its values are also changed levels are also changed during the experiments but instead of doing 16 I am doing only 8 of the total number of experiments. So, I am doing only 8 out of the 16 experiments. So, I will be doing at lowest level of ABCD.a and b at high levels c and d at low levels a and c at high levels and b and d at low levels bc b and c factors at high levels and a and d factors at low levels a and d factors at high levels.

b and c factors at low levels a and c factors at high levels and b and d factors at high levels. I hope I said it correctly a and c factors at low levels and b and d factors at high levels cd means both c and d are at high levels and a and b are at low levels abcd means all of them are at high levels. So, I am not doing the complete set I am first doing only 8 experiments 1 2 3 4 5 6 7 and 8.

So, this way I have done my first fraction the second fraction which I may decide to either do or omit are corresponding to the red entries a b c abc d abd acd bcd so those would be the settings corresponding to my second fraction. So, you can see that abc all the items highlighted in red color correspond to-1 entries in the abcd column. So-1 entries in the abcd column and they correspond to the second fraction.

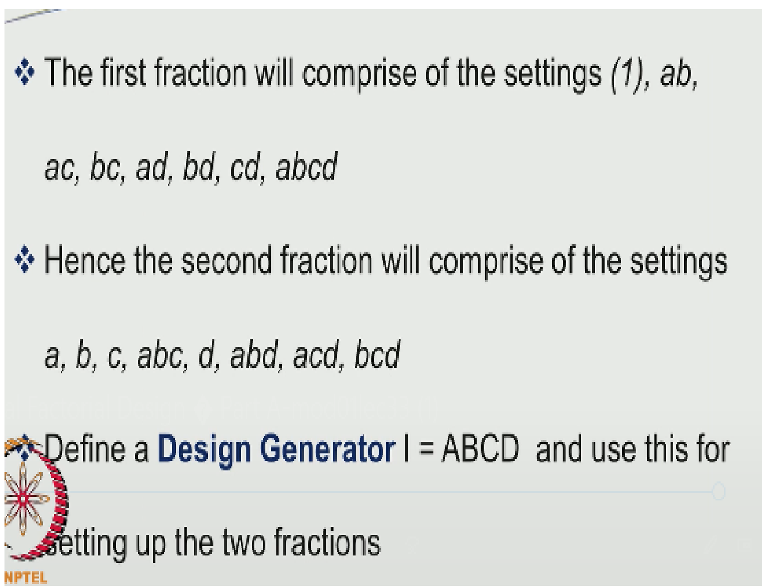
The first fraction we have taken based on+1 or is termed as the principal fraction. So, then you may ask what are all these terms doing. I have entry for A I have entry for D I am just multiplying the entry in A with entry in D to get the AD interaction these are used to define the table of contrasts and from the table of contrasts I can calculate the effects. Ideally speaking for a 2 power 4 design I would for a 2 power 4 design.

I would need to calculate the 16 entries this we have seen previously in the discussion on factorial designs 2 level factorial designs. I had been looking at the entries in each of the column and from this contrast I would be finding out the effect of A so my effect of A you would have 16 entries-1+a-b+ab and so on. But you would be correctly asking look how can you use all the 16 entries to find the effect of factor A.

Because you have done only 8 experiments and so in order to calculate factor A from the table of contrasts I will be constrained with the only 8 available entries. That means I am not able to find A uniquely, so I am not able to find A fully this is a problem. Similarly, same applies for any other interaction or main factor for example look at ABC in the usual case I would have had to go for the entire set of 16 entries to calculate the effect of ABC.

But now I can do only with 8 entries corresponding to the blue color. So, I would have to use this-1 then I would have to go for this-1 then these 2-1. Okay so only 8 values may be used so I am not able to get the complete information.

(Refer Slide Time: 41:05)

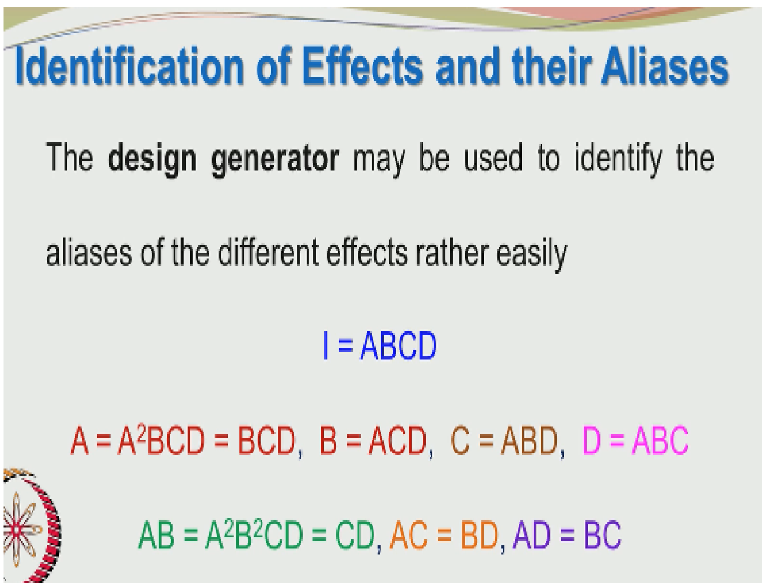


So, I hope you have followed the discussion so far it is pretty simple actually it is just that you are having a lot of entries and you have to keep track properly and not make any mistake by putting a+1 instead of-1 and so on and that would lead to complete confusion. One quick way to check is to see whether the number of pluses are equal to the number of minuses so what are the blue entries?

So, they are 1 ab ac,bc,ad,bd,cd and abcd 1 ab, ac,bc,ad,bd,cd and abcd bd and cd are there yeah bd and cd are there ad bc ac ad bc ac so that is correct, and the remaining settings will constitute the second fraction. So, these are the 1 in the blue color and these are the 1 in red color to make it clear. Let me put the color right now so these are the 1 in blue color and these are the 1 in red color so then full screen so here you go so this is the first fraction and this is the second fraction.

Now we introduce an important concept called as the design generator and the design generator I=ABCD and we use this to set up the 2 fractions remember that we used ABCD we used ABCD to set up the 2 fractions and this is ABCD is called as the design generator.

(Refer Slide Time: 43:23)

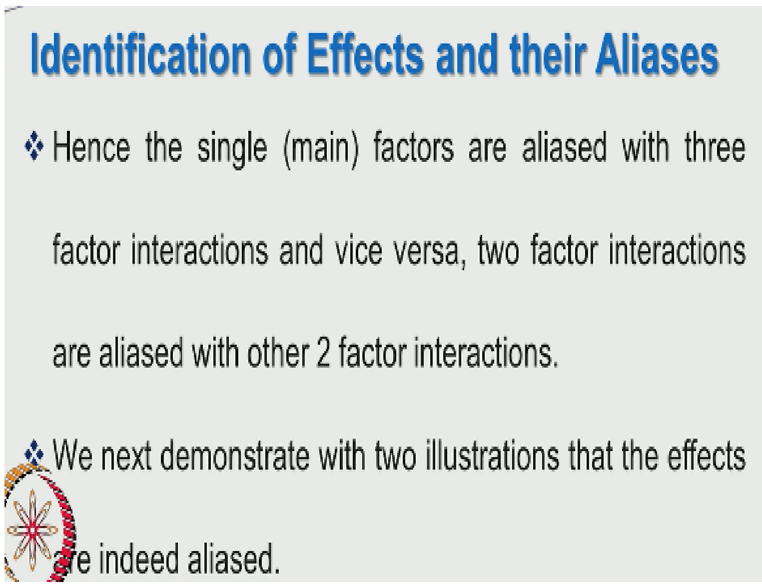


Now using this we can identify the aliases very easily the designs generator is given by I=ABCD so A may be written as A squared BCD so A*A will lead to all + 1 so that vanishes and so easily a is Aliased with BCD and B is Aliased with ACD because you put B here it becomes A B squared CD that becomes ACD because B will then become all B squared will then become all+1 and so B aliased only with ACD.

Similarly, C is aliased with the non-C terms ABD and D is Aliased with ABC. So, you can say that the main factor is aliased with 3 factor interactions what happens to the 2 factor interactions we see that the 2 factor interactions are Aliased with the other 2 factor interactions. AB, I put AB here A Squared B squared that becomes 1 because A squared is a column of all+ values and B squared is a column of all+ values.

So, they really do not contribute to the sign taken by AB, so you have AB you have A Squared B Squared A Squared B squared will be all+1 so that really does not affect the sign. It is only determined by CD. So, AB=CD that means AB is aliased with CD on the same lines we can easily show that AC is aliased with BD and AD is aliased with BC.

(Refer Slide Time: 45:22)



So, we can say that the single factors are aliased with 3 factor interactions and 2 factor interactions are aliased with other 2 factor interactions. So, what is this aliasing what do you really mean by A=BCD and B=ACD what does it mean? So, in the next part we will be seeing that A is indeed aliased with BCD in other words you cannot distinguish between A and BCD and again the 2 factor interactions are aliased with other 2 factor interactions.

So, refer that is what we will be seeing and then we will be seeing how to calculate the effects. So, we will take a small break now and continue shortly.