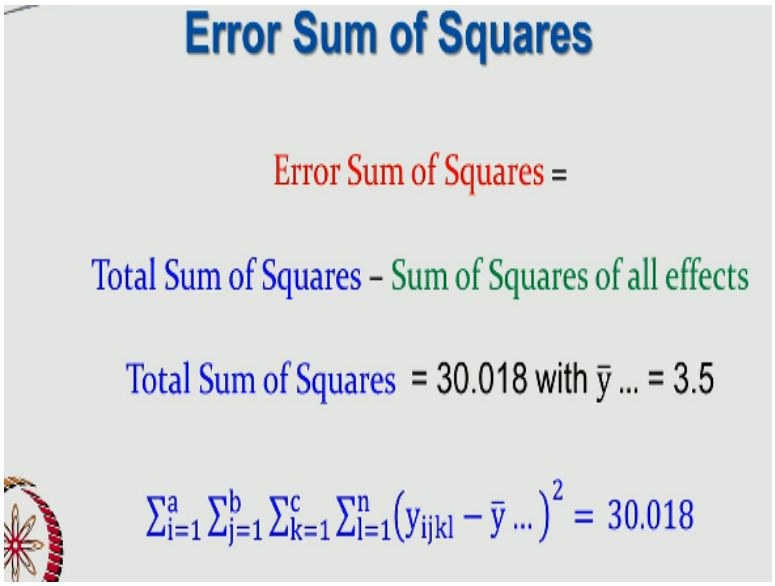
Statistics for Experimentalists
Prof. Kannan. A
Department of Chemical Engineering
Indian Institute of Technology – Madras

Lecture - 35
Factorial Design of Experiments Example Set (Part B)

Okay continuing with our example involving 3 factors in that 2 power 3 factorial design.

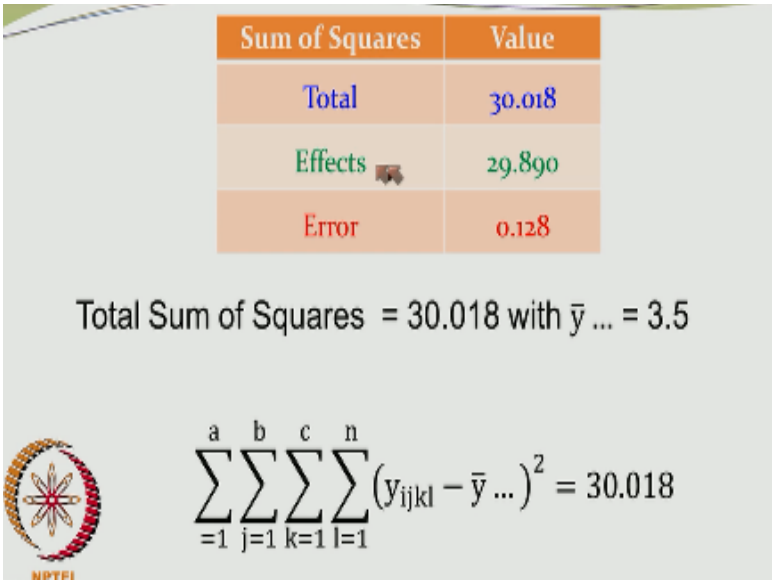
(Refer Slide Time: 00:20)



First we will calculate the error sum of squares. We have to calculate the total sum of squares first and then we subtract from the total sum of squares the sum of squares of all the effects. So the total sum of squares is 30.018 with the grand mean equals 3.5. i= 1 to a the index I running from the levels of a. Here we are having only 2 levels and j= 1 to b is representing the 2 levels for b and k equals 1 to c will represent the levels of c obviously a= b= c= 2.

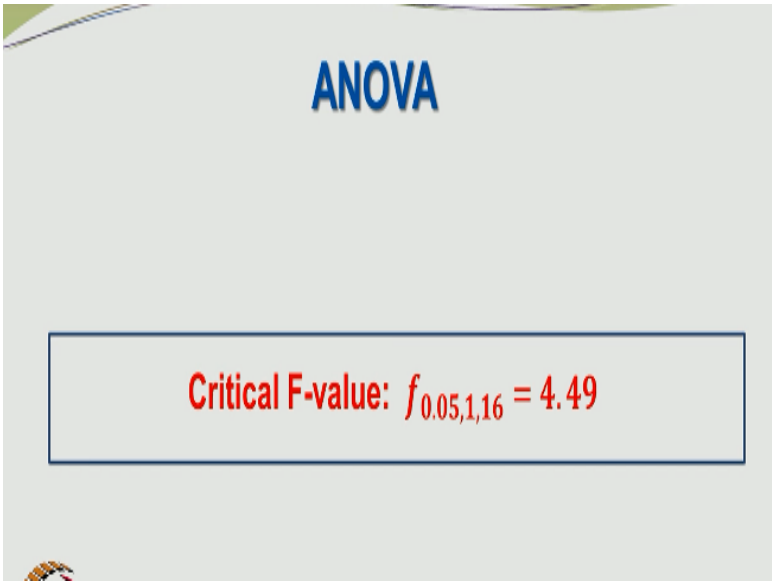
And n represents the number of repeats. So that is given by the index l, l running from 1 to n. So you have Yijkl i for factor a index j for factor b index k for factor c index and n corresponding to the repeat index. This is the overall mean and when we carry out these computations we get 30.018 and this represents the total sum of squares.

(Refer Slide Time: 01:40)



And the sum of squares due to the effects would be 29.890 from a b ab interaction, ac interaction bc interaction how did we calculate the effects sum of squares we have these formulae which I have shown you earlier. So using these formulae we can calculate the respective sum of squares for the different effects and with this we get the error sum of square which is rather small when you subtract 29.89 from 30.018 we get 0.128.

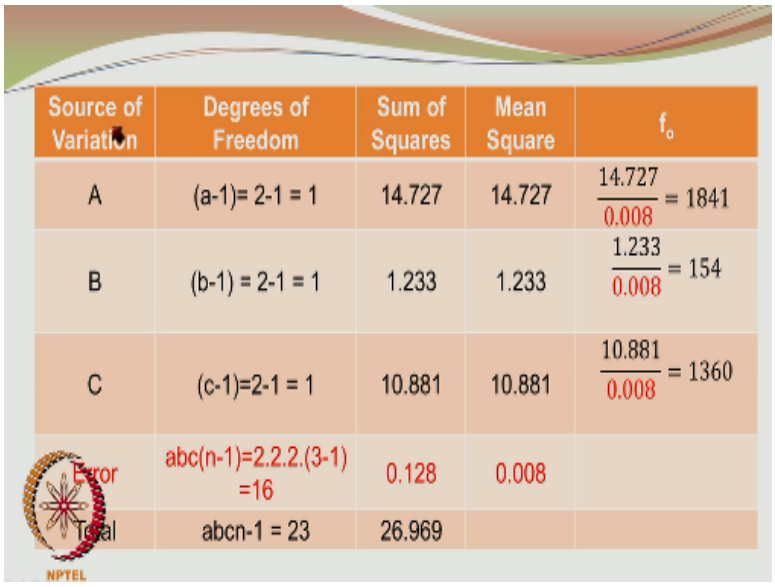
(Refer Slide Time: 02:26)



And for carrying out the analysis of variance we first of all calculate the critical F value which is f 0.05 where 0.05 represents alpha which is the level of significance 1 numerator degree of freedom because each effect has 1 degree of freedom and 16 represents the degrees of freedom in the denominator. It also represents the degrees of freedom for the error and the degrees of freedom for the error would be abc *n-1. ABC would be 2*2*2 which is 8. N-1 would be 3-1 which is 2 so 8*2 would be 16.

So that is why you have 16 degrees of freedom in the denominator.

(Refer Slide Time: 03:18)



And hence we can divide the ANOVA table into source of variation degrees of freedom sum of squares, mean squares f 0 as before and you can see that the degree of freedom are the same for all the effects and the interactions. Error I just told you as 16 degrees of freedom. The sum of squares due to A would be 14.727 B is 1.233 C is 10.881 error is 0.128 and the mean square is obtained by dividing the sum of squares by the degrees of freedom.

Here it is quite simple 14.727/1 is the same value 1.233 divided by 1 is 1.233 10.881/1 is 10.881 again, but here it is 0.128/16 which comes to value of 0.008. And this total is 26.969 and once you have the mean square you can divide by the mean square error to get the f values the f values are pretty large at 1841 154 and 1360. The f values I repeat are obtained by dividing the mean square of that particular effect or the interaction with the mean square error.

So we have the numerator degrees of freedom corresponding to degrees of freedom for the source of variation which we are considering and the denominator degrees of freedom corresponding to the degrees of freedom for the error. And similarly you can do for the binary interactions and the ternary interaction and we get these values. And interestingly the critical f value if you recollect was 4.49 these values considerably exceed this 4.49 and hence they lie in the rejection region.

The critical f value was 4.49 the actual f value is much higher than the critical f value. So these lie in the rejection region and hence these indicate that ABC are important factors and hence the null hypothesis which says that those factors are insignificant should be rejected.

(Refer Slide Time: 05:54)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	f_o
AB	(a-1)= 2-1 = 1	2.94	2.94	$\frac{2.94}{0.008} = 368$
BC	(b-1) = 2-1 = 1	0.0267	0.0267	$\frac{0.0267}{0.008} = 3.34$
AC	(c-1)=2-1 = 1	0.0817	0.0817	$\frac{0.0817}{0.008} = 10.21$
ABC	(a-1)(b-1)(c-1) = 1	0.000963	0.000963	$\frac{0.000963}{0.008} = 0.12$
Error	abc(n-1)=2.2.2.(3-1) =16	0.128	0.008	
Total	abcn-1 = 23	3.177		

Similarly, you have AB interaction to be quite significant it is much higher than 4.49. However interesting BC interaction is negligible because it is lying in the acceptance region. The f value is lying in the acceptance region the critical value was 4.49, but this f value is only 3.34 and hence you can reject the other null hypothesis, but for this case BC you accept the null hypothesis but the fact that BC or the interaction between factor b and c is negligible.

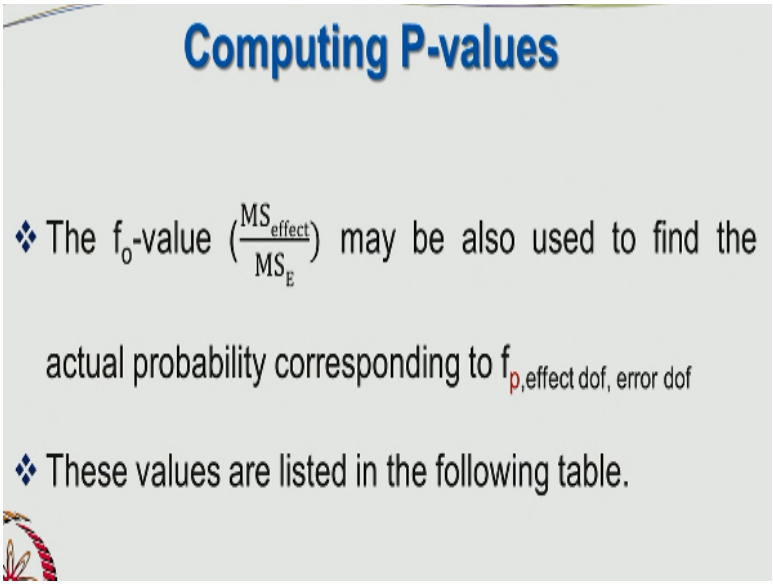
Now when you look at the ABC interaction again the f 0 value is 0.12 which is considerably is lower than 4.49 and hence we also accept the null hypothesis that the ABC interaction is not having any affect on the process. So this completes an ANOVA table, but we can also find the P values. P values are very useful. It tells us that by what margin the effect was considered to be significant or it was constructed to be insignificant.

What I really mean here is did effect A for example or factor A for example became significant narrowly it just quick through or it was a very dominating factor. If the P value is very small, then that particular factor or that particular interaction is strongly significant. It is lying well in the rejection region if the p value is let say 10 power -3 10 power -4 that indicates that the f statistics was lying firmly in the rejection region and you can reject the null hypothesis comfortably.

But if the p values is 0.049 and or 0.0501 it is a marginal case it is hovering between rejection and acceptance. So it tells us the P values are a better indicator of how much we are able to reject the null hypothesis or accept the null hypothesis. Usually p values of 0.049 or 0.0501 are unusual. The P values are usually the order of 10 power -3 10 power -4 for cases where you reject the null hypothesis or you get rather high p values like 0.46 or 0.35 or even 0.10 when you accept the null hypothesis.

But the p values here are very useful. It tells us whether the effects were rejected by a comfortable margin or they narrowly got rejected.

(Refer Slide Time: 09:00)



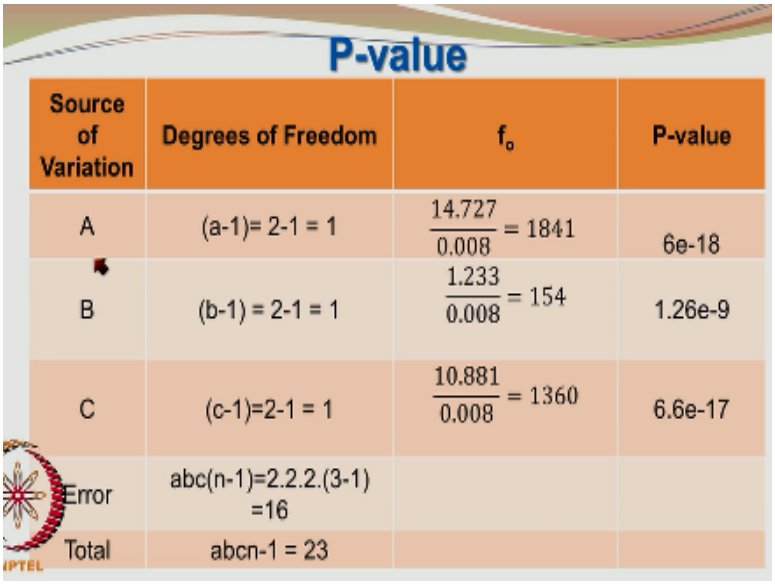
So instead of finding out F alpha effect degree of freedom and error degree freedom you find fp effect degree of freedom and error degree of freedom with the associated mean square error ratio you find the probability associated with that ratio. So mean square effect by mean square error gives you the value and you find the probability associated with that particular value.

So this is inverse of what we did earlier we just had the f value computed the critical value and compared the actual f value with the critical value and then said whether we accepted or rejected, but here we are given the f value and then you are also given the degrees of freedom you have to find the probability associated with that f value and this probability we compare it with 0.05.

If it is much lower than 0.05 we reject the null hypothesis if the computed p value is>0.05 we

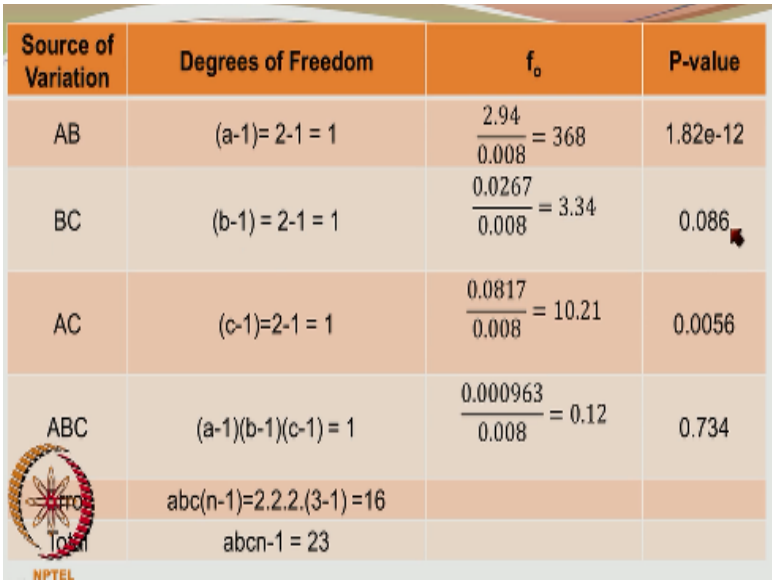
accept the null hypothesis.

(Refer Slide Time: 10:21)



So when you look at these p values corresponding to ABC you find these values are pretty small and so you can comfortably reject the null hypothesis saying that these effects are useless. So you have to say with lot of confidence that ABC effects are significant.

(Refer Slide Time: 10:48)



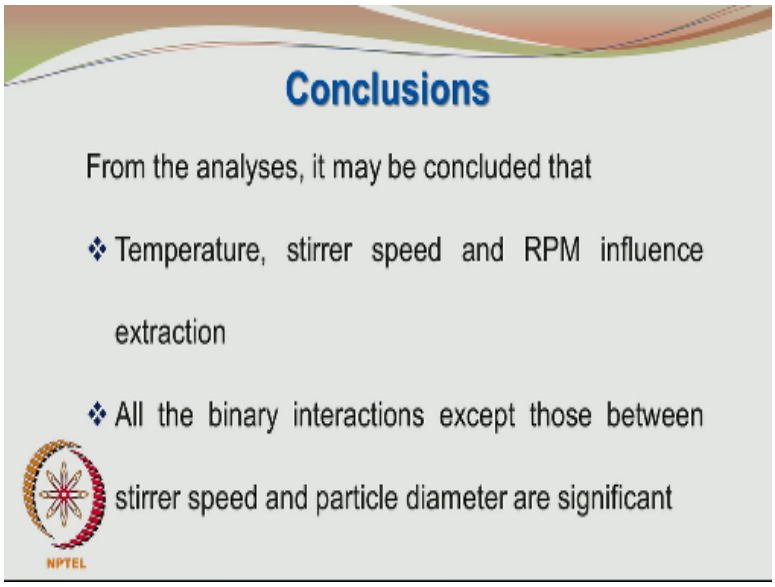
But when you look at BC the P value is 0.086 which is>0.5 and hence BC is a insignificant interaction in the process whereas ABC is having a p value of 0.734 which is quite huge and hence you can accept the null hypothesis that ABC effect is insignificant. The same conclusion we saw earlier also. We did not consider BC and ABC to be significant.

(Refer Slide Time: 11:28)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	f_o
AB	(a-1)= 2-1 = 1	2.94	2.94	$\frac{2.94}{0.008} = 368$
BC	(b-1) = 2-1 = 1	0.0267	0.0267	$\frac{0.0267}{0.008} = 3.34$
AC	(c-1)=2-1 = 1	0.0817	0.0817	$\frac{0.0817}{0.008} = 10.21$
ABC	(a-1)(b-1)(c-1) = 1	0.000963	0.000963	$\frac{0.000963}{0.008} = 0.12$
Error	abc(n-1)=2.2.2.(3-1) =16	0.128	0.008	
Total	abcn-1 = 23	3.177		

Let us go back to that particular table. BC 3.34 was smaller than 4.49 ABC is 0.12 was again smaller than 4.49 and hence we consider BC and ABC to be insignificant. Insignificant factors will have high P values and significant factors or significant interactions will have very low P values please remember this.

(Refer Slide Time: 11:58)

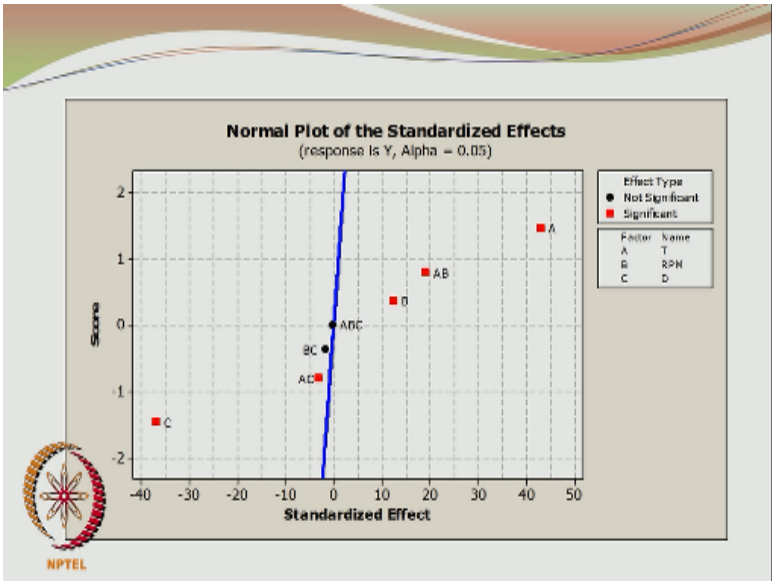


From the analysis it may be concluded that temperature stirrer speed and RPM influence the extraction. All the binary interaction except those between stirrer speed and particle diameter are significant. A ternary interaction between temperature stirrer speed and particle diameter is not significant. So we have the data most important analysis after the experiments are carried out and the model has been developed is to analyze the residuals.

We have seen the analysis of variance table from which we can detect the important effects.

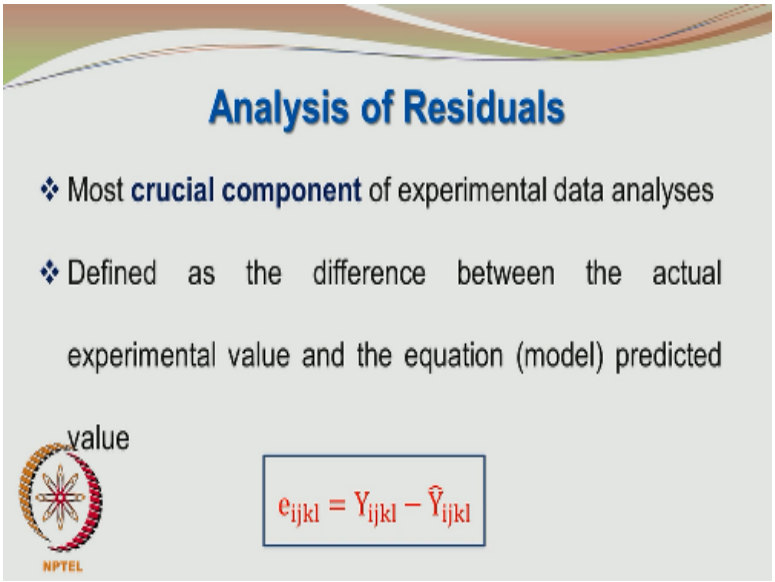
Another way to detect the important effects which I have not indicated so far is the use of the normal probability plots. The different factors and their interactions are plotted in the normal probability plot and we can find out which of the effects are significant.

(Refer Slide Time: 13:07)



So in this plot we can see that ABC and BC are lying close to the solid line whereas cac bab and a are lying quite further apart. This means that the effects corresponding to factor cac bab and a are significant whereas BC and ABC are insignificant.

(Refer Slide Time: 13:47)



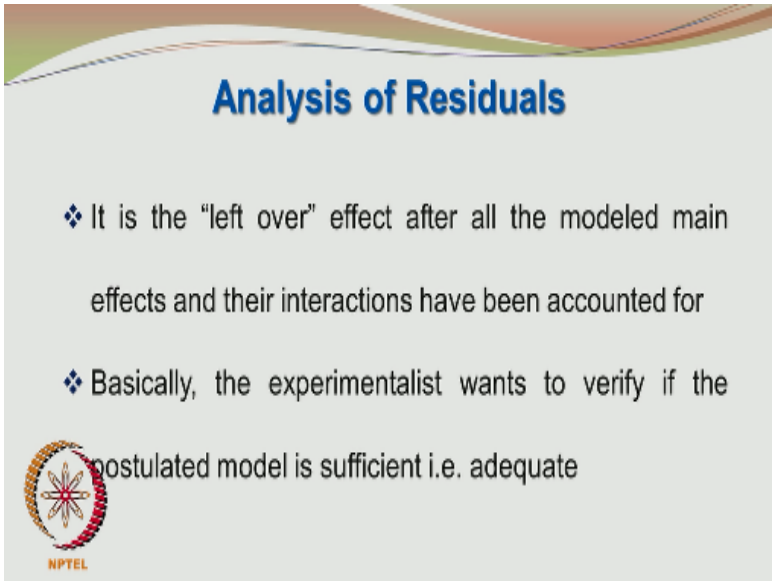
The next important concept in design of experiments is to analyze the residuals. The residuals are defined as the difference between the actual experimental value and the prediction from the model. So you can see that eijkl this is the residual corresponding to a single point. It is a different between 2 values. The first value is the response corresponding to the ith setting of

factor a jth setting of factor b and kth setting for factor c because now we are looking at 3 factors and l corresponds to one of the repeats corresponding to the setting of ij and k.

And this is the predicted value and the difference between the 2 would be defined as the residual. Now ideally we would like the residual to be 0, but in a real world there will always be random error component. We want to see whether the difference between the actual experimental data and the model predicted values corresponds to random behavior. The difference is the residual and we want to see how random the behavior of the residual is.

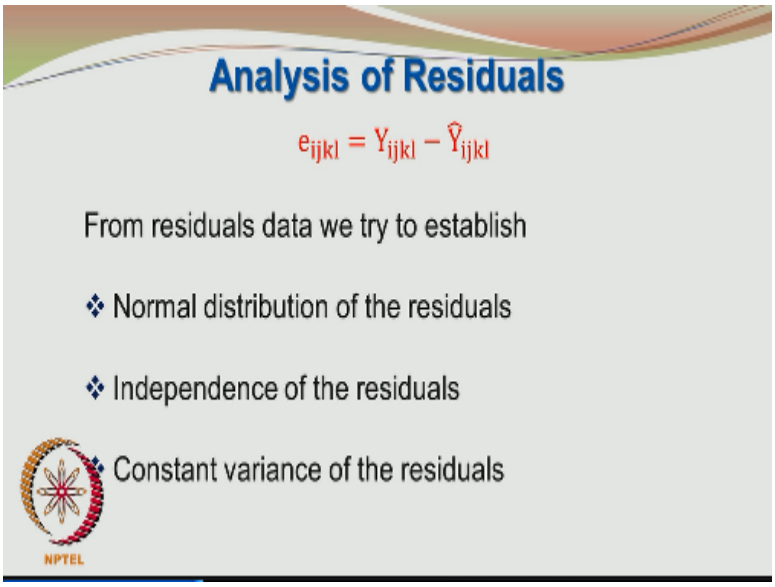
Similarly, we calculate the residuals for all other experimental data points.

(Refer Slide Time: 15:47)



So we may say residual has the left over effect after all the modeled main effects and their interaction have been accounted for. So the experimentalist wants to see whether the model he has developed is adequate or additional terms are necessary. So we will be looking at lack of its concept in regression where more would be said about the adequacy of the proposed model.

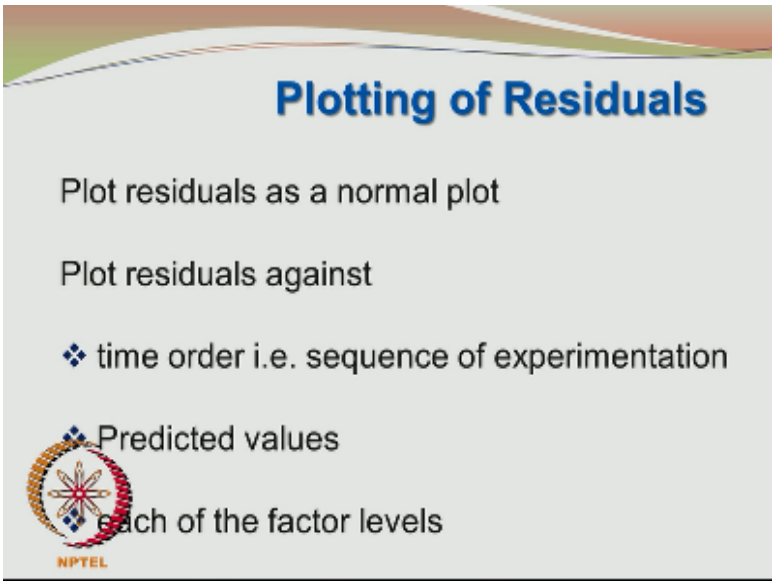
(Refer Slide Time: 16:18)



So what do we do with the residuals? We want to see whether the residual are normally distributed whether the residuals are independent and whether the residuals are having constant variance. We know that the basic assumption in the model development in the linear model development was the error terms where independent and identically distributed in a normal distribution with 0 mean and constant variance sigma square.

By looking at the residuals trend or they plan out we can check these assumptions whether the residuals are normally distributed whether they have constant variance whether they are independent from one another.

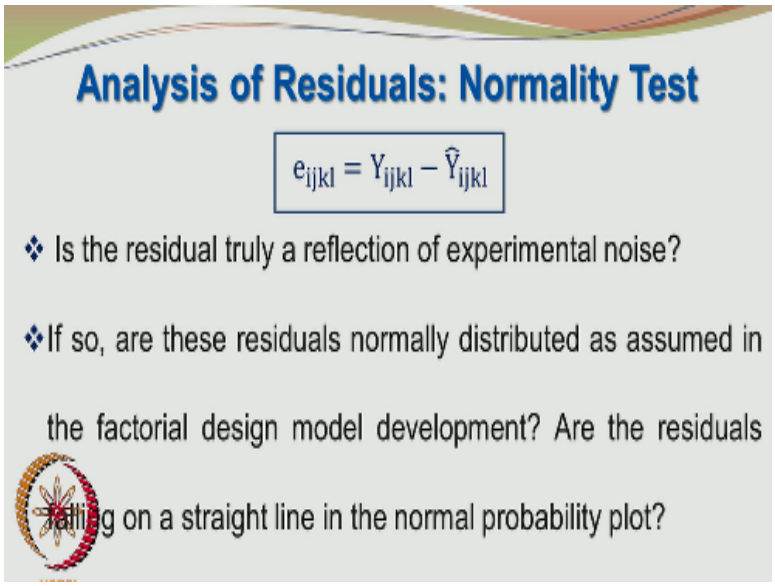
(Refer Slide Time: 17:17)



So the residuals may be plotted in a normal plot that is one option. The residuals may also be plotted in the time order the sequence in which the experiments were carried out. The

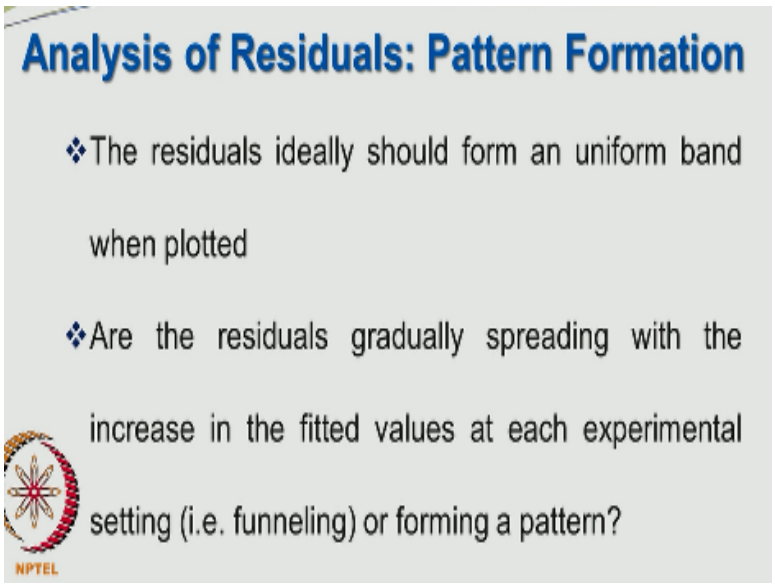
residuals may be plotted against predicted values and residuals may be plotted against each of the factor levels to see the spread of the residuals at each factor level.

(Refer Slide Time: 17:45)



So if the residuals are normally distributed they will fall on the straight line on the normal probability plot. So as far as this experiment is concerned where we considered the 3 factorial design involving the yield of the medicinal compound. We see that the residuals are behaving reasonably well they are falling n a straight line more or less. We do not see any gross deviation of the residuals from the straight line.

(Refer Slide Time: 18:28)

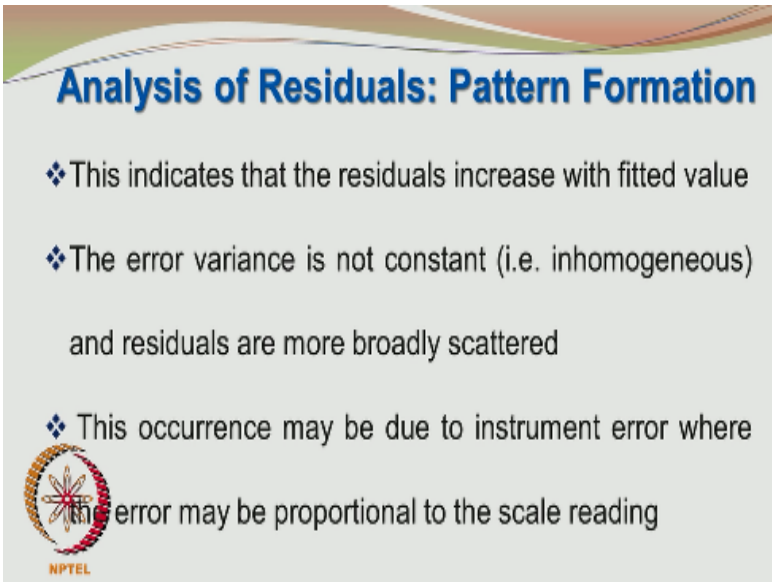


The next check would be to inspect the pattern formation of the residuals. The residuals are ideally speaking impartial quantities. They do not really depend upon the sequence of the experimentation; they do not depend upon the fitted value depending on the experimental

setting. So when you look at the spread or pattern of the residuals you should really see no distinct trend if the residuals are plotted against the fitted value.

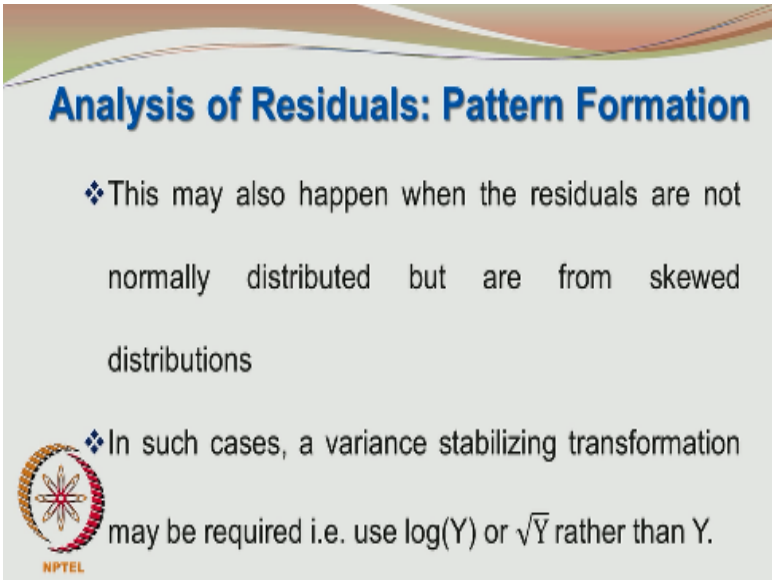
And they show up a definite trend then something is amiss or something is not correct. If the residuals when plotted against the fitted value show a funnel kind of pattern. So they are spread over a smaller distance at low value of the fitted value and if the residual spread out when the fitted value increases. Then something is not correct.

(Refer Slide Time: 19:53)



So this indicates that the residuals increase with the fitted value and the error variance is not constant that means it is not homogenous and residuals are more broadly scattered when the fitted value increases. For example, when the instrument error is proportional to the reading measured then the residuals are going to increase with the increasing value of the fitted value.

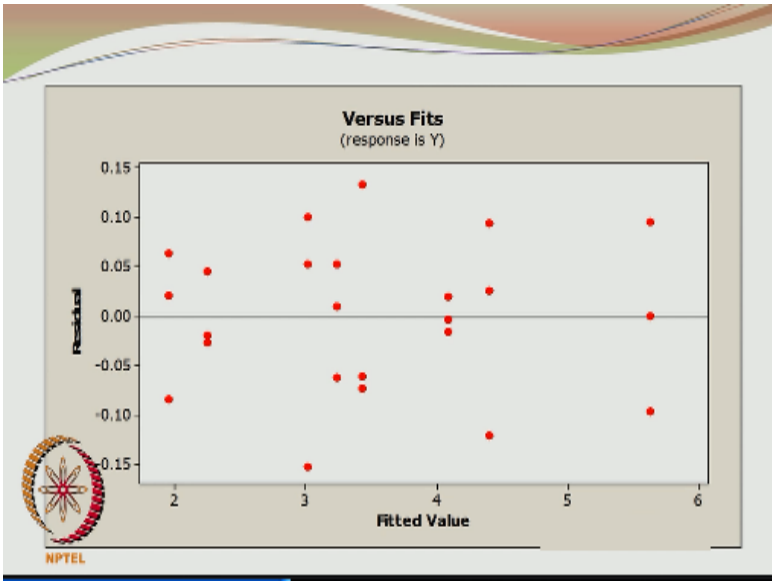
(Refer Slide Time: 20:26)



Sometimes the residuals may not be normally distributed and they come from skewed distributions. In such cases you may have to go in for a stabilizing transformation for the variance and you may instead of using y directly may want to consider log Y or root Y. So the residual analysis is subject (()) (20:50) and a fascinating one at that and because of lack of time we cannot go or dwell deeper into this subject.

On the other hand, there are some excellent books on residual analysis one is the book by draper and smith and the next one is the one by Montgomery they have been cited in the references.

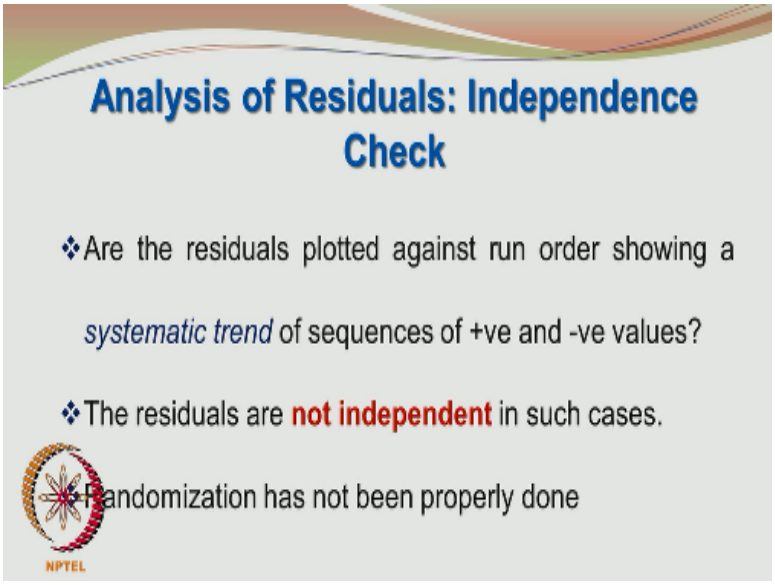
(Refer Slide Time: 21:20)



So when you look at the residuals as against the fitted value you can see that they are pretty much scattered uniformly. You do not see any kind of funneling arrangement. For example,

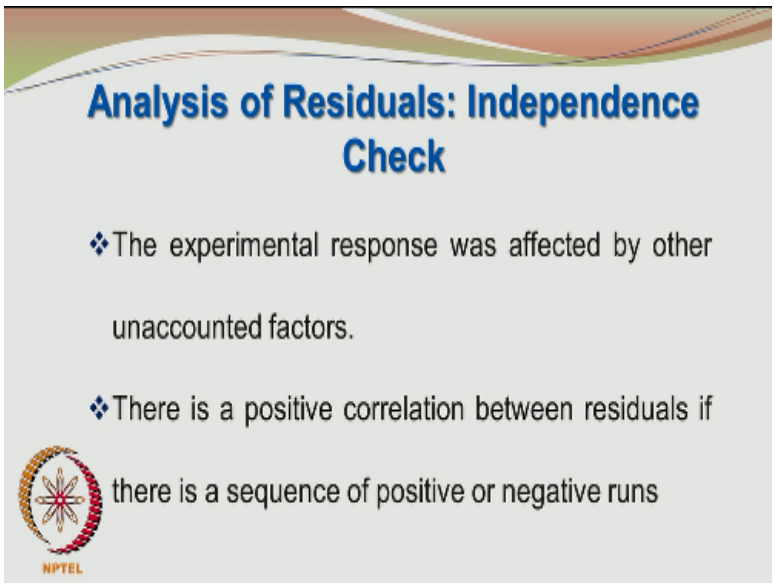
you do not get a kind of diverging residuals.

(Refer Slide Time: 21:40)



The next check is to see whether the residuals are independent of one another. So the residuals are plotted against run order and we have to see whether there is a systematic trend of positive residuals and negative residuals. What I am trying to say here is whether the residuals are over a certain time period increasing continuously and over the another time period they are decreasing continuously. This also means that the randomization is not properly done.

(Refer Slide Time: 22:25)

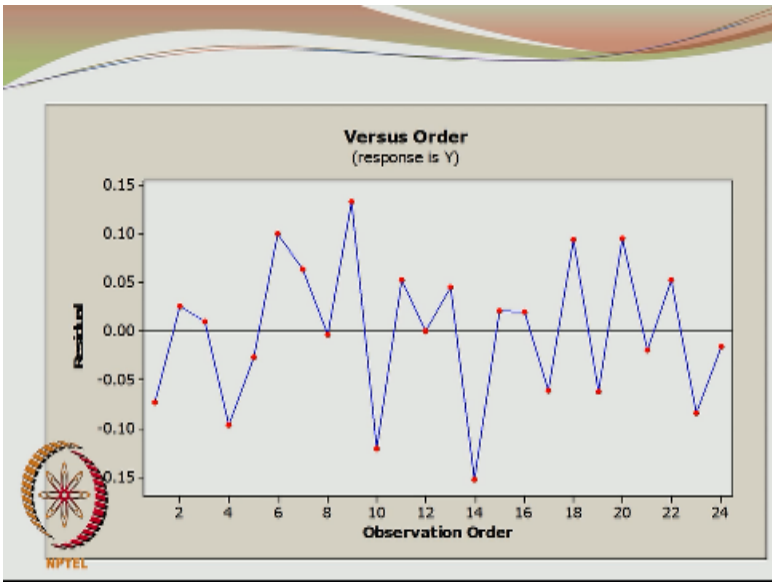


And the experimental response was affected by other unaccounted factors. For example, if the temperature was not controlled properly or there was a sudden shoot up in the temperature then the sequence of residuals may show a continuously increasing trend or there may be

some other unaccounted factor. So over a sequence of experimental runs the residual values may be continuously declining/

So this shows that the randomization was not proper and there were unaccounted factors which affected the experiment in a certain systematic manner.

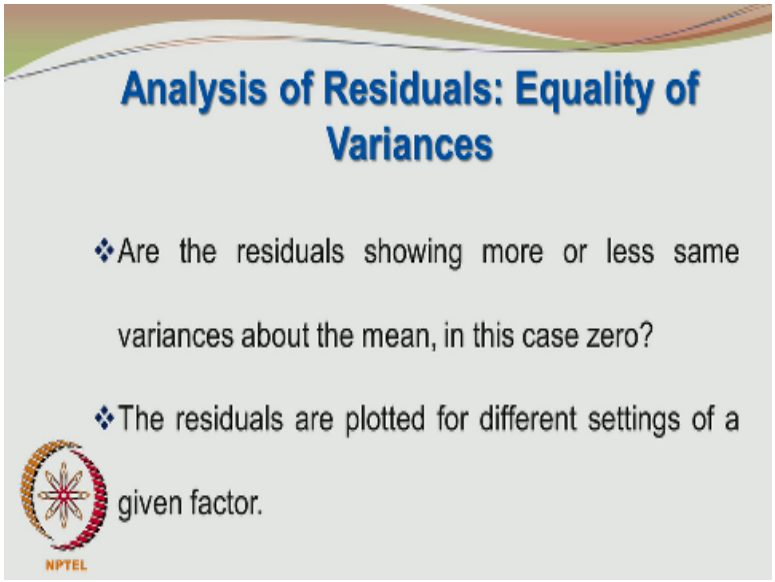
(Refer Slide Time: 23:14)



When you look at this particular plot you can see that you do not find any sort of residual showing a continuously declining a trend over a significantly long time interval or along the order of observations you can see that it is declining here, but then it is increasing. And here also it is increasing decreasing. So you do not really get 4 or 5 consecutive positive values or 4 or 5 consecutive negative values.

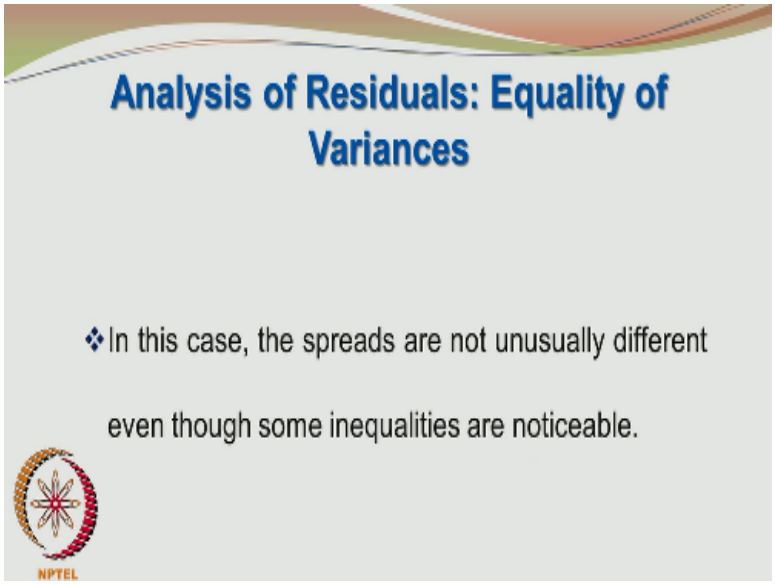
So you do not see a systematic trend of sequences of positive or negative values as given here.

(Refer Slide Time: 24:07)



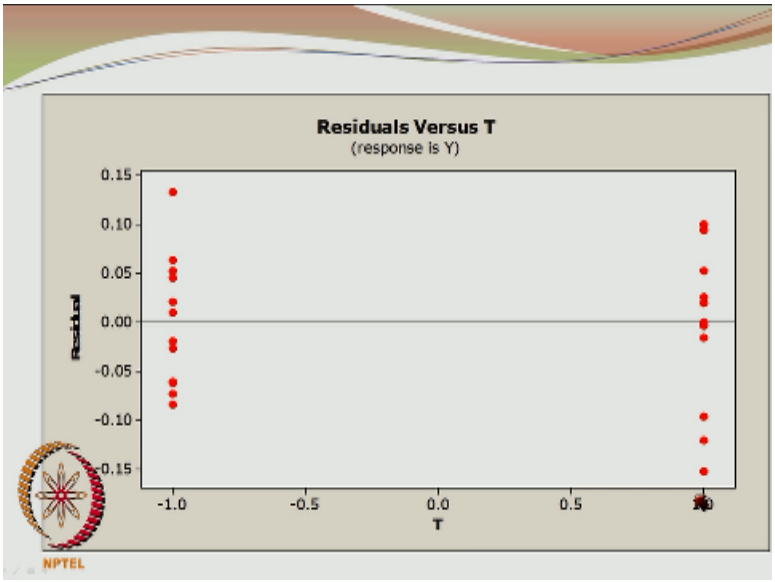
And another important thing is the errors are having constant variance sigma square. So are the residual showing more or less the same variances about the mean in this case 0 the residuals are plotted for different setting of the given factor.

(Refer Slide Time: 24:24)



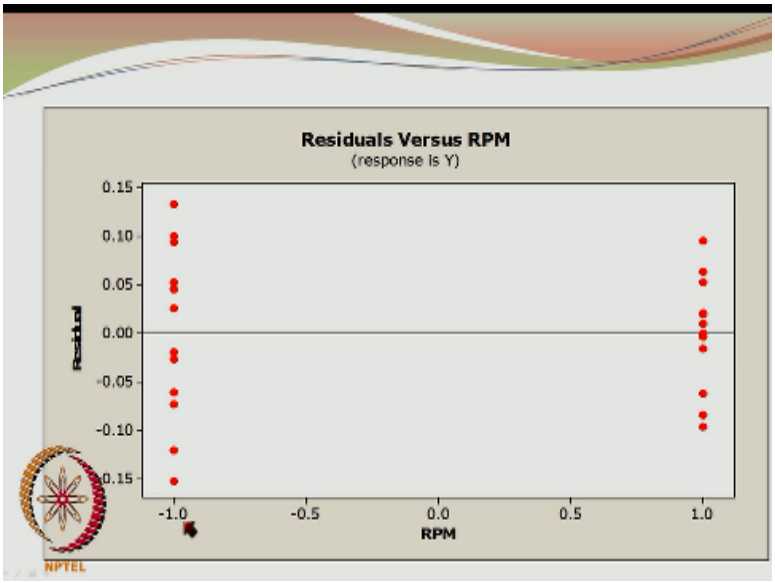
And if the spreads are not unusually different even though some inequalities may exist then we may say that the assumptions of constant variance is not bad.

(Refer Slide Time: 24:39)



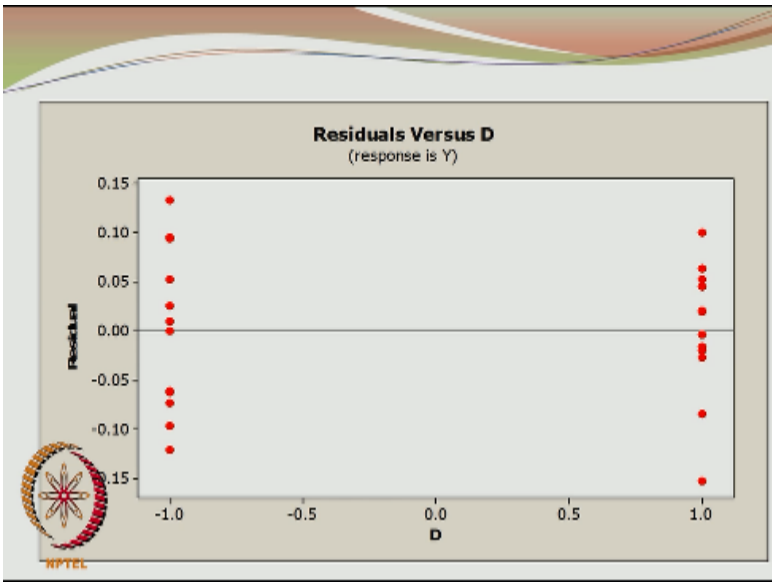
So here we have plotted the residuals as function of the 2 different settings of the temperature in the coded format and you can see that the spread is reasonably uniform around the origin. So the variance does not really change drastically from one setting to another setting.

(Refer Slide Time: 25:04)



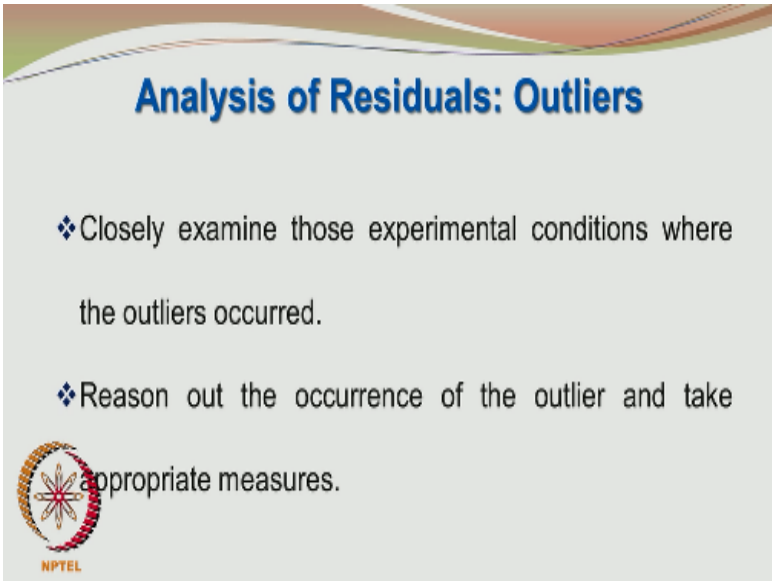
You have another case. Well you can say that this variance is smaller than this variance, but in real data we cannot exactly get a uniform spread and so on.

(Refer Slide Time: 25:20)



So unless there is a gross deviation we can assume the residuals are behaving in a reasonable fashion. For example, if the residuals was clumped over a very narrow region here and the residuals are occurring over a broad region here then the assumption of constant variance may not be true, but here the spread is comparable. I would not say exactly identical, but it is comparable and hence we can live with the assumption of constant variance.

(Refer Slide Time: 25:56)

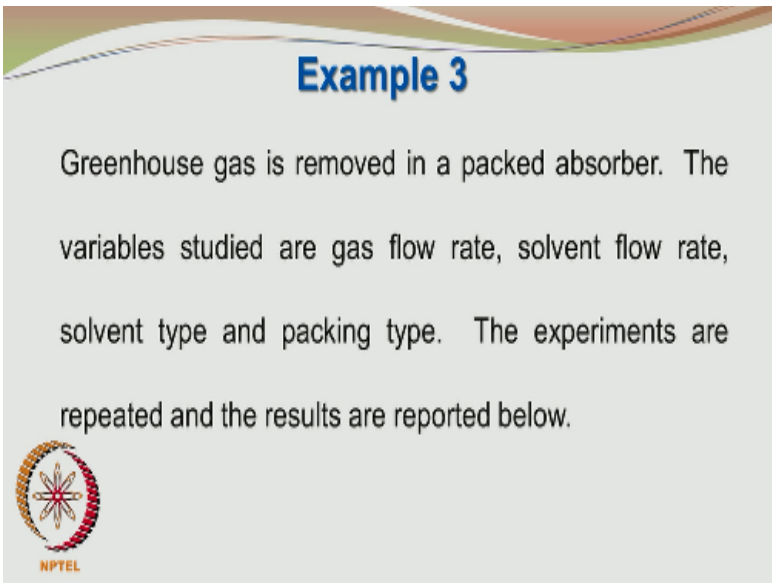


Another important thing is the presence of outliers. Outliers are unusual data points they may be taught of by the experiment as a rogue data point or rebellious data points something which is not following the general trend. It is lying way out of the general trend. The reasons for outliers may be pretty simple measurement error or a calibration error of the instrument. Some careless mistake okay or these settings were not noted properly.

Usually the experimenter gets into a routine and he does not make these kind of elementary mistakes even though they are possible and if you find any outliers it is very important that you do not brushed them under the carpet, but pay closer attention to the reasons for the occurrence of such outlier or outlying data. They may tell you something different something unusual is going on which is not accounted for in your approach to the experimentation.

For example, if there is a condition at which the data point shows an unusually high yield obviously that is going to be profitable to the company if it is true and the experimenter may like to inspect the data point more closely.

(Refer Slide Time: 27:36)

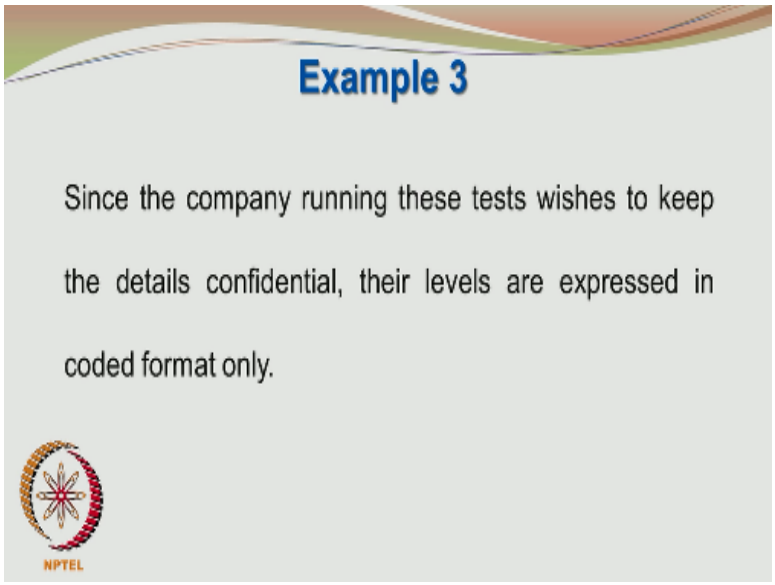


Okay now coming to example 3 the problem statements goes like this. Green house gas is removed in a packed absorber. The variable studied are gas flow rate, solvent flow rate, solvent type and packing type. The experiments are repeated and the results are shown in the table below. Looking at the background for this example we know that carbon dioxide is a green house gas and its emissions are very harmful to the environment leading to global warming and so on.

So one way to reduce the emission is to observe the carbon dioxide in using a certain solvent. The solvent may be monoethanolamin or diethanolamine and the equipment in which the gases are removed by absorbing or putting it (()) (28:29) dissolving them in the liquid solvent is called as the absorber and if you put packing in them it become a packed bed absorber and we want to make the process quite effective.

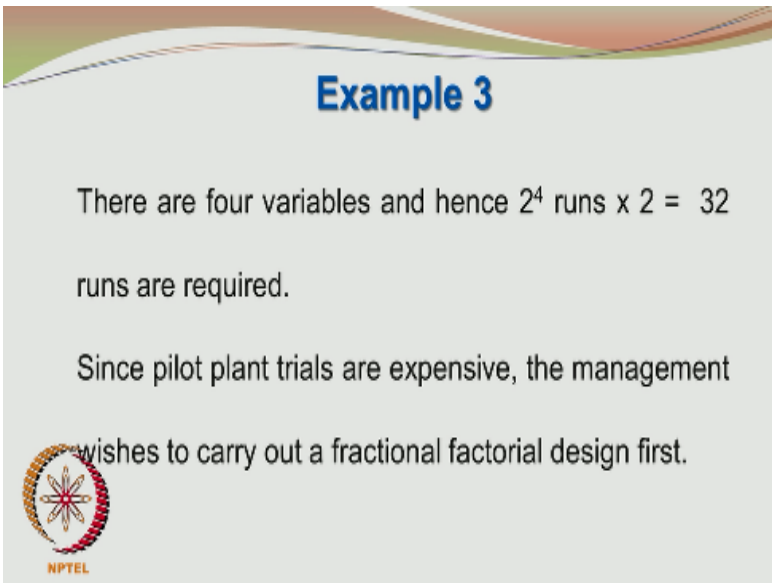
And we want to see what variables are significant. So the variable studied are the gas flow rate, the solvent flow rate, solvent type and packing type. The experiments are repeated in the results are given below.

(Refer Slide Time: 29:14)



Thus company apparently running these test wants to keep the details confidential so the levels are expressed in the coded format.

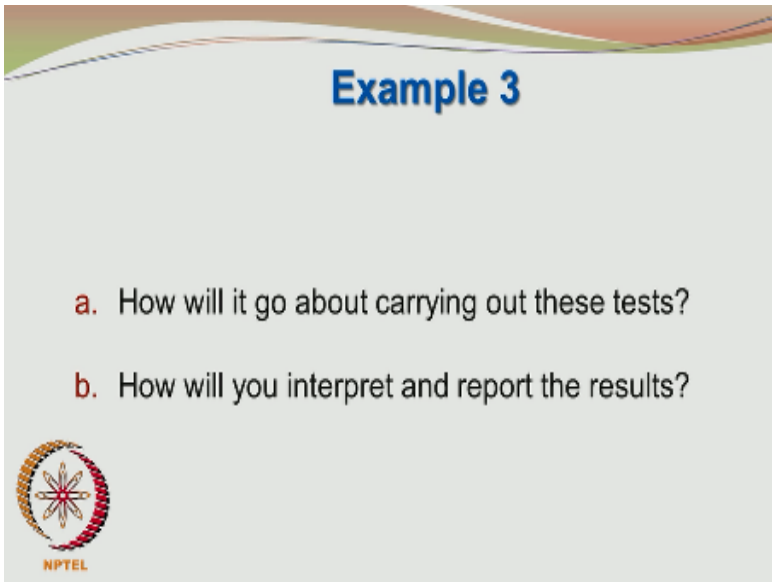
(Refer Slide Time: 29:25)



Since there are 4 variables you have 2 power 4 that is 16 runs and if you want to do at least 2 repeats per setting that would mean 16*2 32 runs. Pilot scale studies are quite expensive and hence the management may feel those 32 runs are quite a lot in terms of investments time manpower and so on. So it may even consider telling to you what would happen if you do a fraction of those runs and see what results we get and then we decide to move on from there.

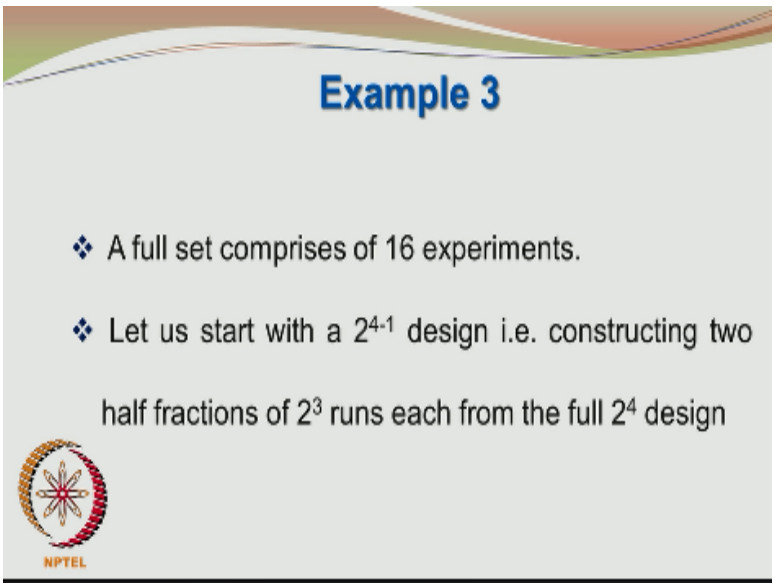
Factorial designs of experiments are so structured that it is indeed possible to construct a fraction design and implemented.

(Refer Slide Time: 30:17)



So the main thing is how will it go about carrying out these test and how major results be interpreted and reported.

(Refer Slide Time: 30:28)

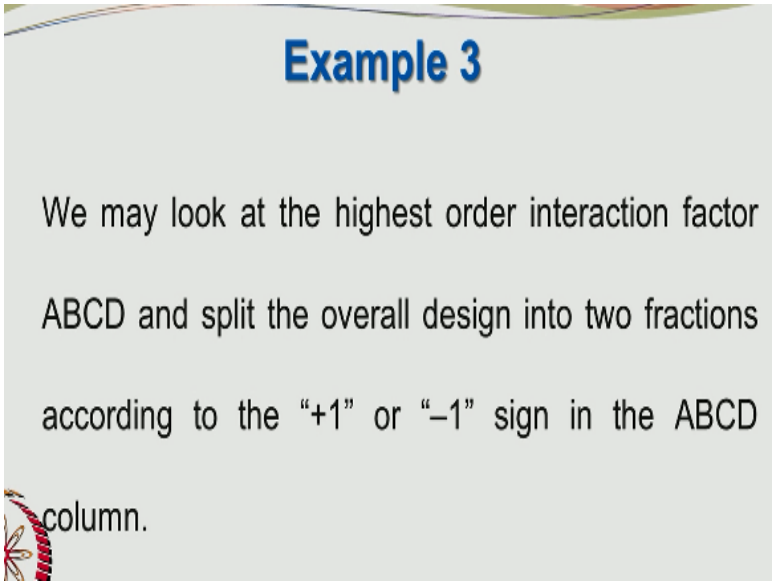


So a full set comprises of 16 experiments so we can start with one-half fraction of the full 2 power 4 design and we have 2 half fraction since there are totally 16 experiments one half fraction would involve 8 runs which would corresponds to a 2 power 3 design, but you may ask wait a minute you are having 4 variables you are having a 2 power 3 design by making a 2 power 3 design I will be able to handle only 3 variables.

What about the fourth variable it does not seem to be there. So the answer to the question is even the 2 power 3 design we are considering all the 4 variables. The only thing is we are not doing the full set of possible experiments as envisaged in the 2 power 4 design. We are only having a 2 power 3 design and we are investigating all the 4 variables with the truncated design.

So this means that while we are gaining on the effort in experimentation we are losing out on some information whether the losing out of the information is serious or not the results will tell us. So we are losing on information in trying to save on experimental effort.

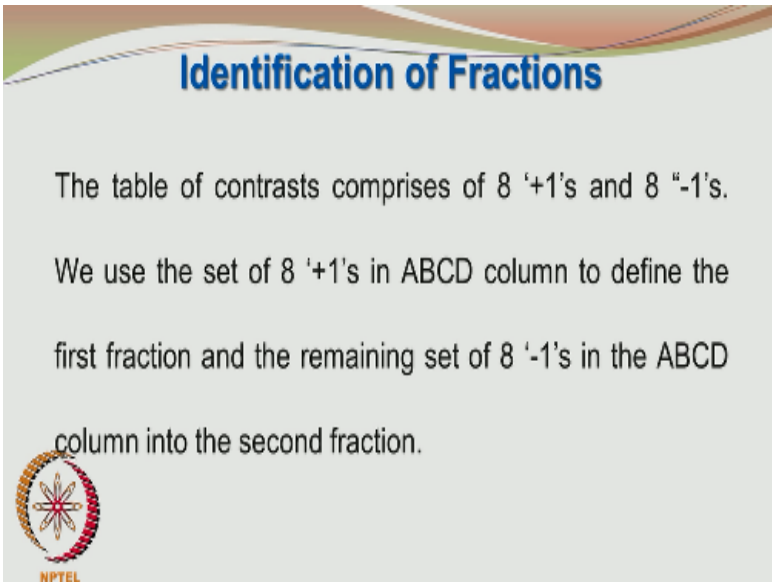
(Refer Slide Time: 31:56)



So how do you go about doing the design. You are having 4 factors we are considering all the 4 factors so now let us look at the highest possible interaction among the 4 factors. It is very simple it is ABCD. Now when you look at the design matrix of an experimental design you will find that it has a set of columns each column containing some pluses and some minuses the number of pluses would equal to the number of minuses and so in a 2 power 4 full design you will have 8 pluses and 8 minuses in every column.

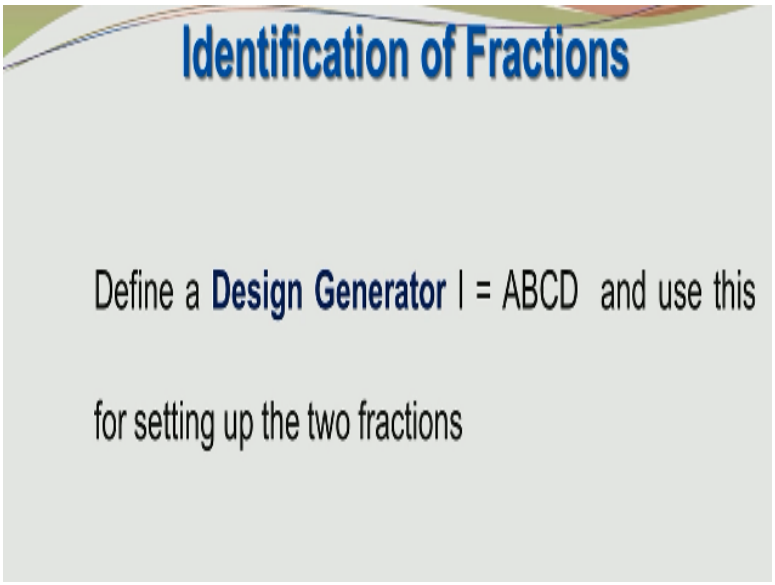
The column may be ABCD ABAC BCBD and so on. So it will have 8 pluses and 8 minuses. So look at the column corresponding to ABCD all the pluses will constitute one set all the minuses will constitute another set. Please consider the experimental settings corresponding to the plusses that will be the first fraction.

(Refer Slide Time: 33:10)



So the table of contrast comprises of 8+1s and 8-1s. We use the set of 8+1s and ABCD column to define the first fraction and the remaining set of 8-1s in the ABCD column into the second fraction. So all the pluses will be corresponding to one fraction all the minuses will be corresponding to the next fraction.

(Refer Slide Time: 33:32)



So we defined a design generator I=ABCD and use this to set up the two fractions. First we define a design generator I=ABCD and use this for setting up the 2 fractions.

(Refer Slide Time: 33:56)

	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

So you can see that the design matrix is set up you see some of them are blue and some of them are red what is the reasons for this it is quite simple.

(Refer Slide Time: 34:10)

	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

We take the ABCD column whatever is 1 we label it as color coded as blue and whatever is -1 we color code as -1. The experimental data were coded into +1 and -1 for the sake of convenience. Now I am color coding the +1 with blue and the – 1 with red. So we collect all the +1s and you can see that the experimental settings are also color coded according to whether ABCD is 1 or -1.

So this one represents a case where all the factors are at their low level ABCD are at their lower level. A is the setting corresponding to A at a higher level b is setting corresponding to B at a higher level. In such cases all other factors would be at their low levels. If you have AB

then a and b are their high levels and all other remaining factors c and b would be at the low levels.

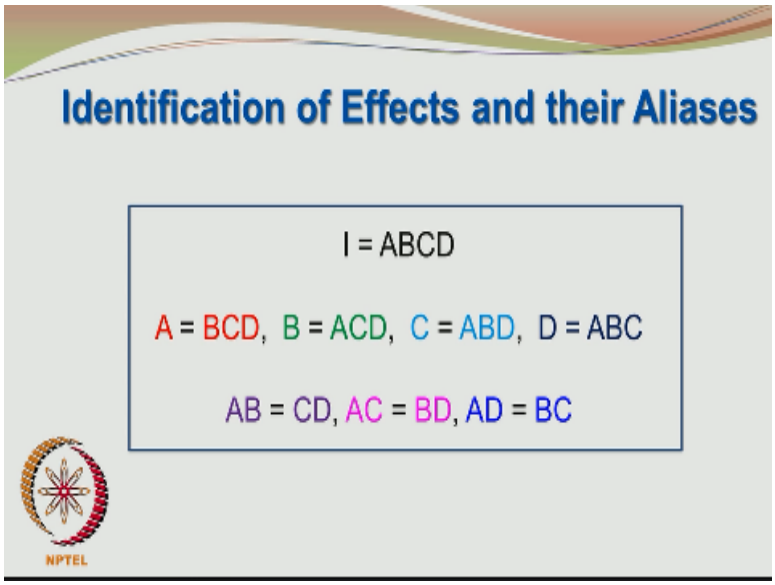
Anyone this we have already seen I am just bringing it to your attention.

(Refer Slide Time: 35:27)

	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

In case you have a doubt you just go back and then see these table things will become clear. It is +1 here and b a is -1 but b is +1. So that is quite simple, but we do not look at these. We look at abcd and we look at the color codes and we collect all the 1s together then we collect all the –1s for the second fraction and once we do that we will get the 2 fractions.

(Refer Slide Time: 36:02)



And we can now look at the design generator I=ABCD to find the aliases. You can see that A=BCD here B= ACD C= ABD and D=ABC. So main factors are aliased with only 3 factor

interactions. So even though the A and BCD are coming together BCD interactions are usually negligible. So the full effect of A is felt in even in the partial factorial design that is because the higher order interactions like third order interactions are usually not important that is very good.

But sometimes in addition to the main factors the interactions are also quite significant or quite important, but unfortunately in this design the 2 factor interactions are aliased with one another as you can see here AB is aliased with CD AC is aliased with BD and AD is aliased with BC. That means when the information is presented we are unable to uniquely determine the interaction due to AB and interaction due to CD both of them are felt together.

Similarly, AC and BD are felt together and AD and BC are also felt together, but whether that cause a serious issues depends upon the responses and depends upon the particular experiment we are considering and let us look at the results and see what really happened.

(Refer Slide Time: 38:04)

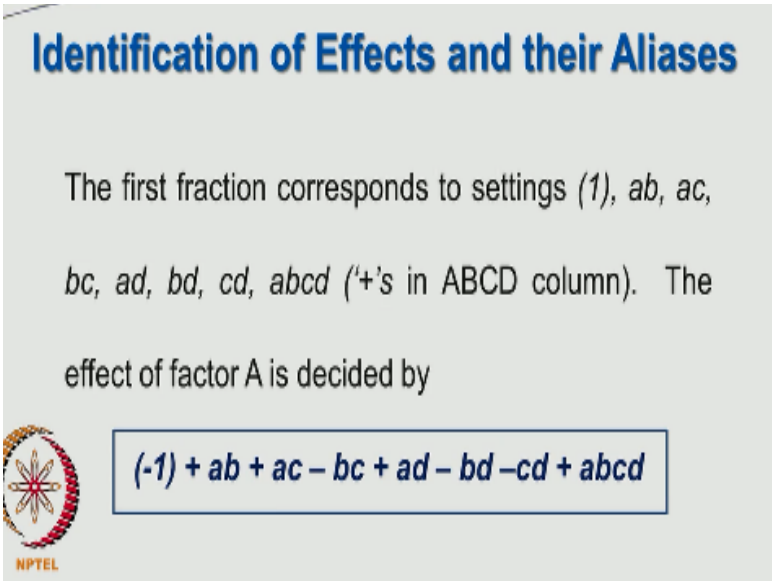
Setting	Organic	Packing	Aqueous	Solvent	% Extraction
(1)	-1	-1	-1	-1	19.22
ad	1	-1	-1	1	19.53
bd	-1	1	-1	1	22.58
ab	1	1	-1	-1	38.28
cd	-1	-1	1	1	21.78
ac	1	-1	1	-1	27.48
bc	-1	1	1	-1	26.43
abcd	1	1	1	1	24.72
(1)	-1	-1	-1	-1	17.52
ad	1	-1	-1	1	19.23
bd	-1	1	-1	1	23.08
ab	1	1	-1	-1	39.59
cd	-1	-1	1	1	22.17
[illegible]	1	-1	1	-1	26.86
[illegible]	-1	1	1	-1	26.25
abcd	1	1	1	1	25.34

First Fraction

NPTEL

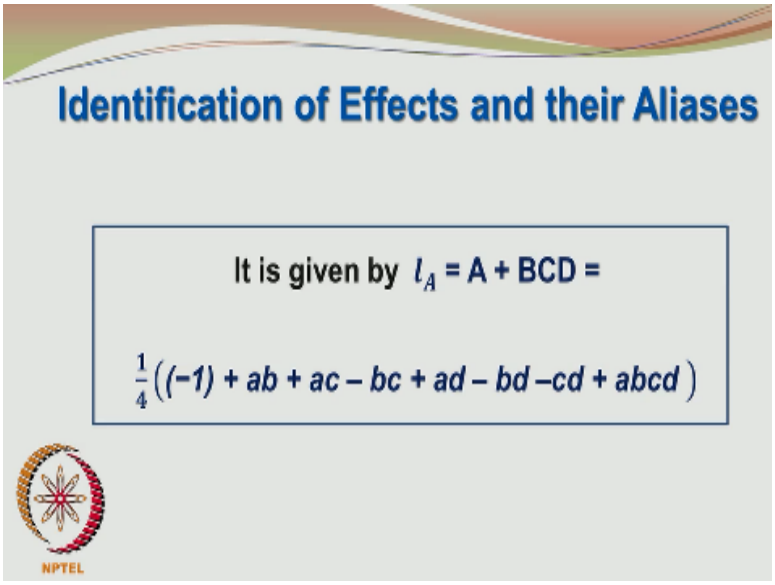
So the first fraction you have 1 ad, ab, ac, ad okay you can go in the sequence ab, ac, ad bc bd and then cd abcd so that is what you have as the first fraction and if you look at it is 1 ab, ac, ad, bc, bd cd and abcd so we are having the first fraction. Since we have done 2 repeats you have the same settings repeated twice and you have the experimental observations presented up to 2 decimals under the column percentage extraction.

(Refer Slide Time: 39:00)



So I have given you the design table once again for the purpose of calculating the effects. The effect of factor A is decided by just go to A so it is -1+ab+ac –bc+ad –bd-cd+abcd. So we are using the contrast in exactly the same way we did for the full factorial design.

(Refer Slide Time: 39:51)



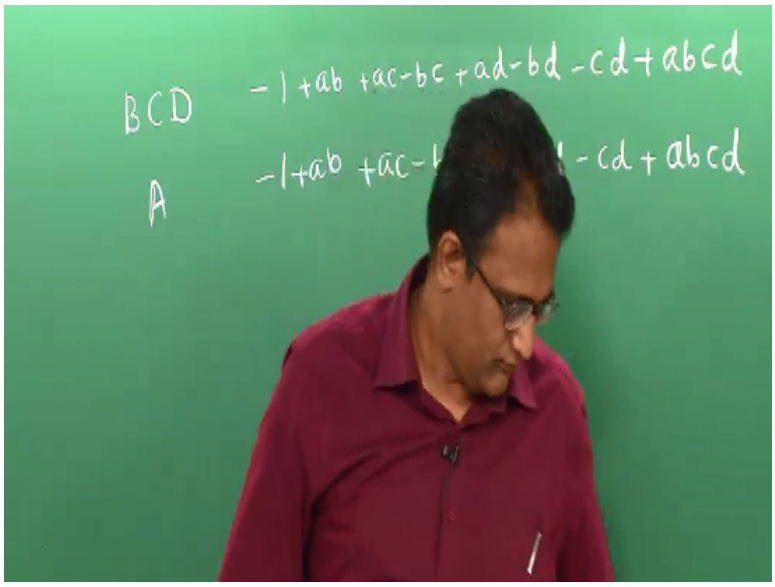
However, even though you are doing under A it is also having the aliasing with the BCD. So will BCD have the same entries in its column when compared to the (()) (40:06) A. The answer is yes because A is aliased with BCD you cannot really distinguish between A and BCD because their column entries are identical. So let us verify this.

(Refer Slide Time: 40:29)

	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

If you look at this bcd it also has -1+ab+ac-bc. So let me write it down so that there is no ambiguity.

(Refer Slide Time: 40:45)

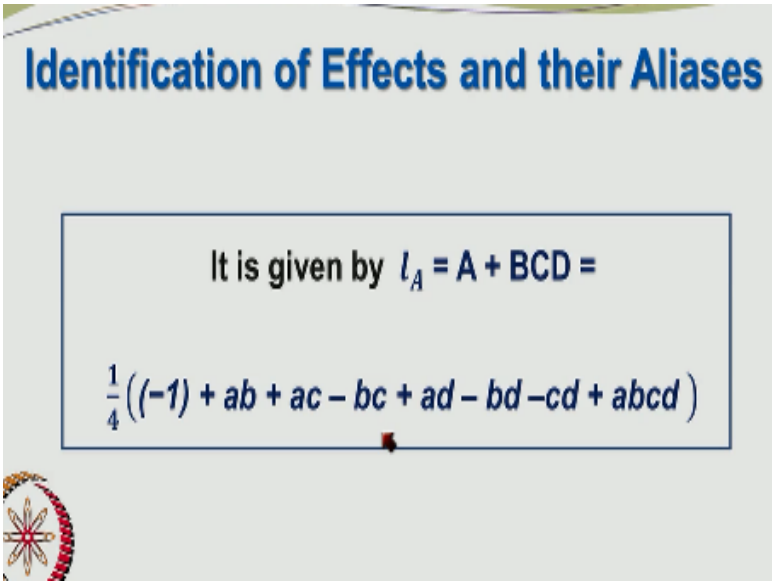


So we are looking at BCD-1 +ab+ ac- bc+ ad-bd- cd+ abcd. And earlier for A also it was -1+ab let me write it down so -1 +ab +ac-bc +ad-bd-cd+ abcd. So when you look at the entries for A and also BCD they are identical and when you are calculating the effect of A you are also finding the effect of BCD in addition. So the same idea will also apply for factor B which is aliased with ACD and for factor C which is aliased with ABD and factor d which is aliased with ABC.

If you want, you may take any main factor and its ternary alias and see whether the column entries are matching. It is a good idea to do this to verify so that we have set up the design

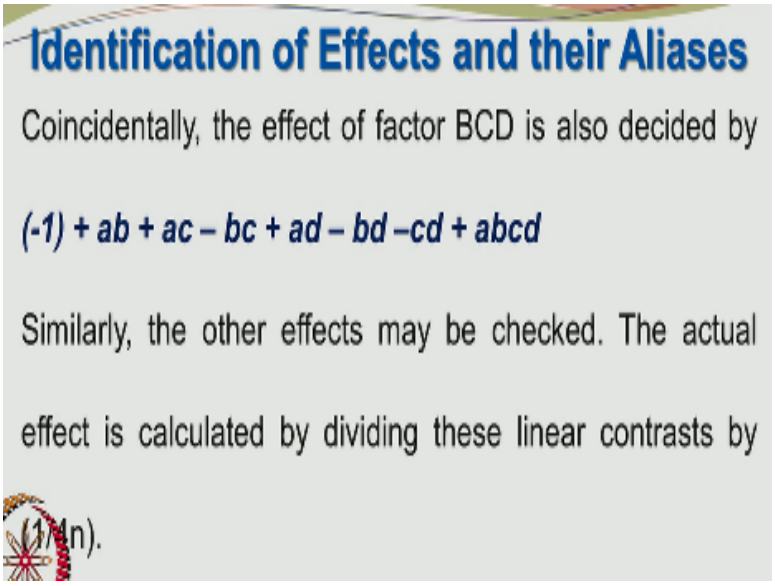
matrix correctly. Similarly, the 2 factor interactions are also aliased with the one another so you can easily find that for example AB would have the same column entries like with the CD and BC would have the same column entries as AD.

(Refer Slide Time: 43:27)



So when you are finding the effect of A even though you write the linear contrast for A as lA it is actually having both effects of A and BCD in it and you have +4 entries and -4 entries so you divide it by ¼ for averaging purposes.

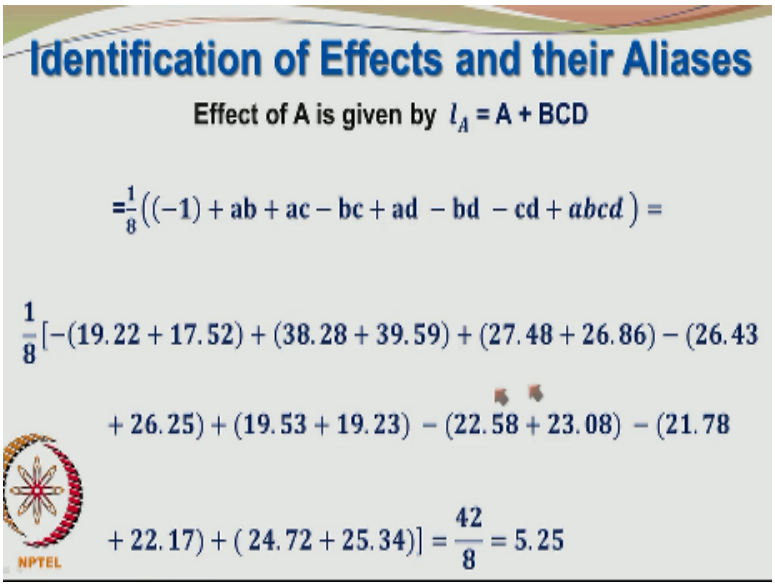
(Refer Slide Time: 43:56)



So as I showed on the board the effect of factor BCD is identical to the (()) (44:01) factor A. Since there are 4 positive and 4 negative entries we take the average with respect to 4. That means we divide by 4 and you are also having 2 repeats so we effectively divide the contrast by 8 4 for the averaging and n for the number of repeats. So using this formula we can

calculate numerically the effect of A and we find that A value is 5.25. So the entries are given here

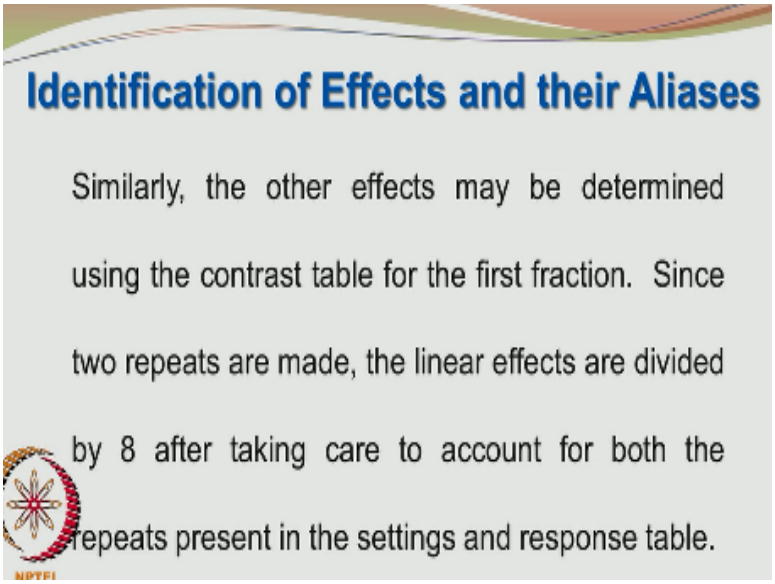
(Refer Slide Time: 44:44)



So this is -1 and that value is 19.22+17.52 it is – of 1. So 1 value is 19.92 and 17.52 so -1 would correspond to –of 19.22 and -of 17.52. Similarly, you can find the values corresponding to the other settings. Please note that we are doing repeats and so for each setting there will be 2 values or 2 responses. So effect of B is effectively given by 6.56 that includes both B as well as ACD.

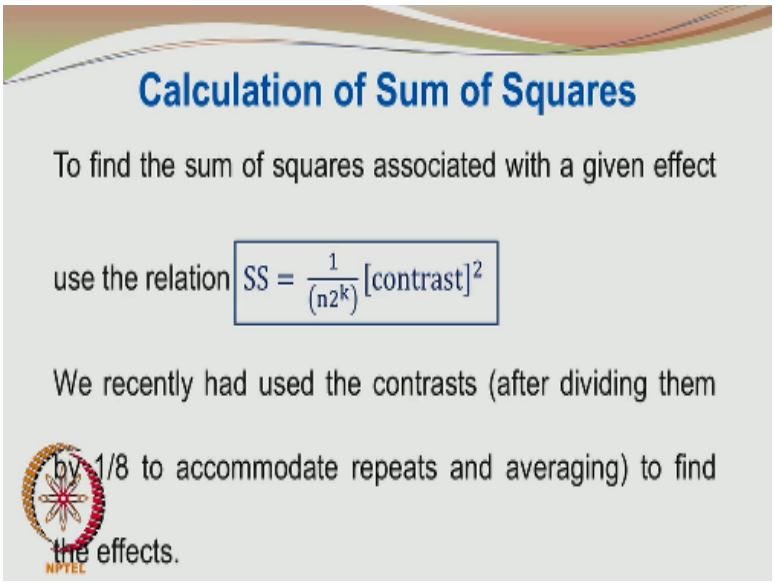
So let me make that correction. So again you have -1 so again you have –of 19.22+17.52 and then +ab 38.28 and 39.59 let us just check it just once. AB 38.28 and 39.59. So we are on the correct track so that others can also be written down in the similar fashion and we calculate the effect of B as 52.48/8 is 6.56.

(Refer Slide Time: 46:50)



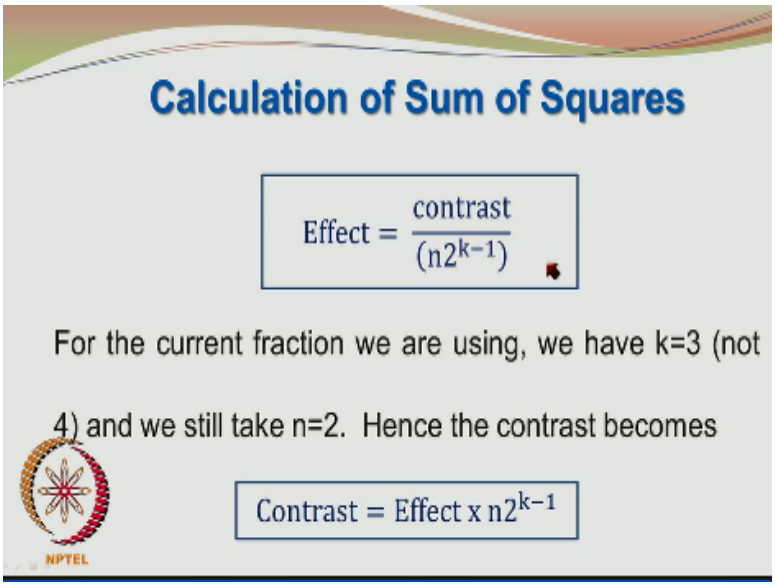
Similarly, you can find the other effects. So take care to divide the linear contrast by 8 to account for both the averaging and also for the repeating.

(Refer Slide Time: 47:05)



The next important step is to calculate the sum of squares. Once you have the contrast calculating the sum of squares is also a piece of cake. And the sum of squares is given by 1/n * 2 power k * contrasts squared and let us see how to calculate the sum of squares.

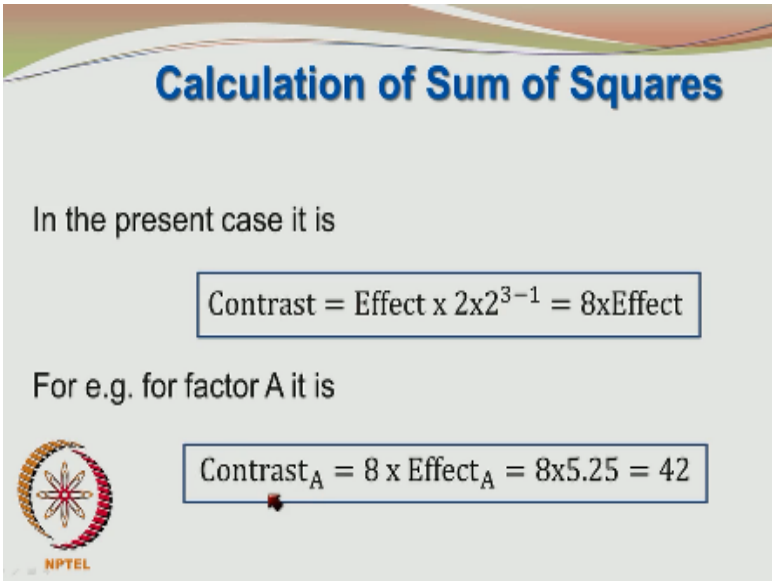
(Refer Slide Time: 47:37)



The effect was given by contrast/n * 2 power k-1. Here k=3 because we are only looking at a 2 power 3 design. One half fraction of a 2 power 4 design so we put k=3. So 3-1 is 2 2 power 2 is 4 2 repeats 4*2 8 and that is what we did when we divided the effects. We divided the contrast by 8. So that is an order and hence we can calculate the contrast as effect* n power 2 power k-1 or effect*8 will give you the contrast.

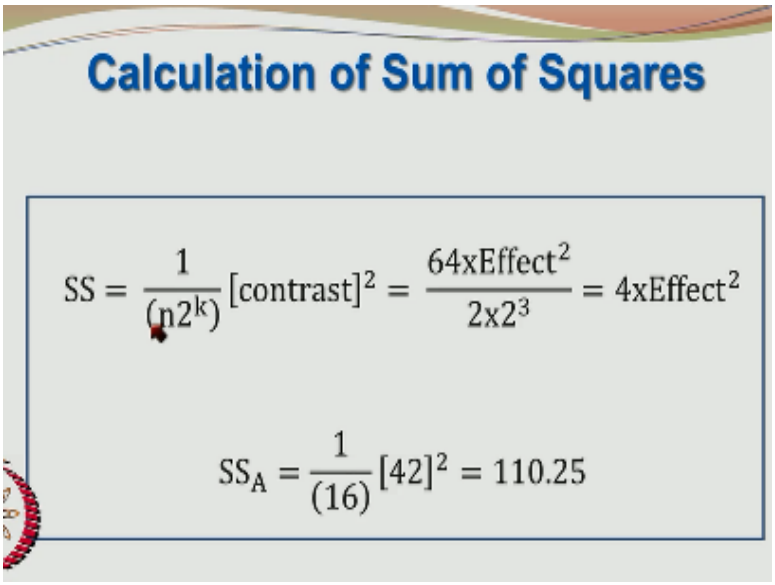
Once you get the contrast you can square the contrast and divide by a suitable number to get the sum of squares.

(Refer Slide Time: 48:35)



So contrast is 8 times the effect for contrast A you first find the effect of factor A and effect of A and then multiply by 8 and then you get back 42. This 42 corresponded 42 you had calculated earlier when finding the effect of A..

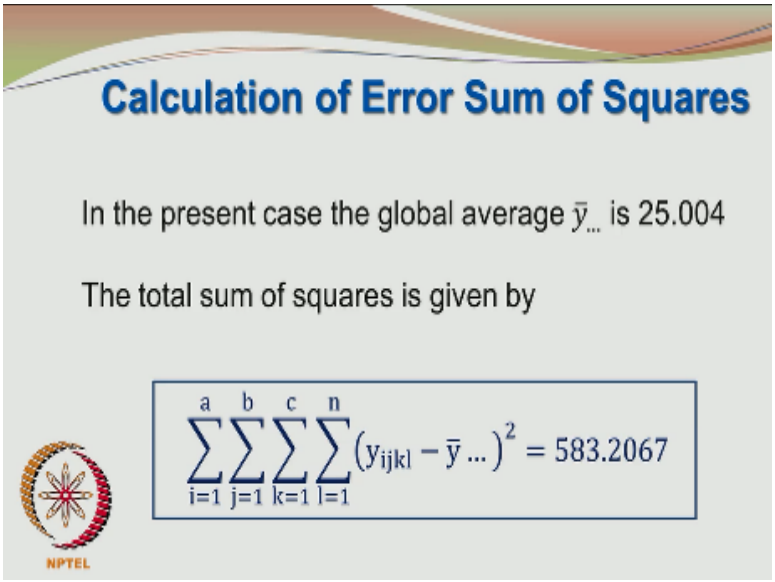
(Refer Slide Time: 49:01)



And once you have the contrast then you divide it by n*2 power k. So contrast square is 8 times the effect and that becomes 64 times the effect square/n*2 power k or 2*2 power 3. So you get 64 effect squares by 16 or 4 effect squares. Sum of squares of A is 1/16*42 square ns =2 2*8 is 16. Contrast we just now saw for A as 42 so you get 42 squared/16 which is 110.25. Likewise, we can find the contrast for B based on the effect of b and then square the contrast divided by 16 n*2 power k and get the value for B.

Similarly, you can get for C, D, AB, AC and AD. So using all this information we can construct the ANOVA table.

(Refer Slide Time: 50:22)



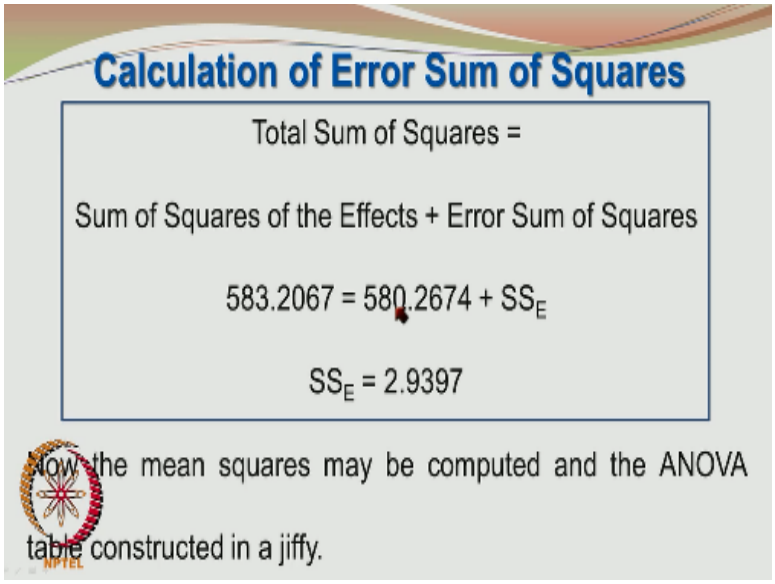
The error sum of squares for doing so first we have to calculate the total sum of squares and

the total sum of squares is given by this huge relationship yijkl-y bar triple dot. So we are finding the total sum of squares now we are subtracting each experimental data observation with the grand total not the grand total the grand average adds up all the numbers in the matrix divided by the total number of entries will get the grand average and that grand average is subtracted from each and every individual observation and that will give you the difference which is squared.

And when combined or added completely gives you the total sum of squares. To make sure that you have understood this correctly please do the calculation on your own and see whether you get the correct number because more than the correct calculation it is the understanding of the procedure which is more important. You have to implement the procedure in the correct manner.

We have found that the sum of squares due to the various effects by using the contrast square and dividing it by the n*2 power k I seem to forget it all the times. So that is 16 in our case. So we calculated the sum of squares for all these effects and once that is done.

(Refer Slide Time: 52:23)



You also have the total sum of squares with you the sum of squares from all the effects are calculated and so you can find out the error sum of squares. So total sum of squares= sum of squares of effects+ error sum of squares which comes to 583.2067= 580.2674 which is sum of squares of all the effects + sum of square of error. The sum of square of error comes to only 2.9397.

The next step would be to construct the mean squares and find the f values and see whether the f values are lying in the rejection region or not based on the critical f value. So we will continue with this example in the next lecture. Please revise the portions discussed in this slide, discussed in this lecture rather and make sure that you have understood the important concepts.

Make sure that you have understood how to form the different fractions because the understanding here will help you when you are actually carrying out the experiments please pay particular attention to the calculations of the errors how to account for errors in your experimental data and analysis. The residual analysis is also very important. We will continue with this example in the coming lecture. Thank you.