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

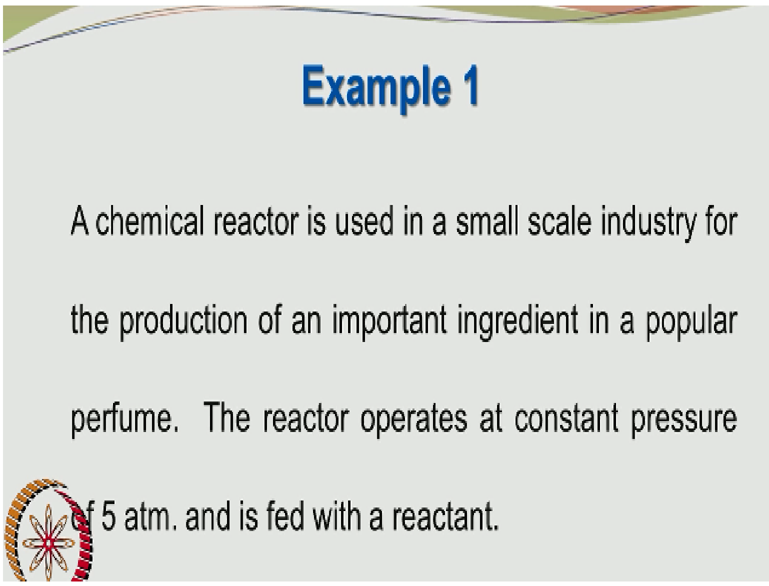
Lecture - 34
Factorial Design of Experiments Example Set (Part A)

Welcome back. In this lecture, we will be solving a few problems involving factorial design of experiments.

(Refer Slide Time: 00:27)

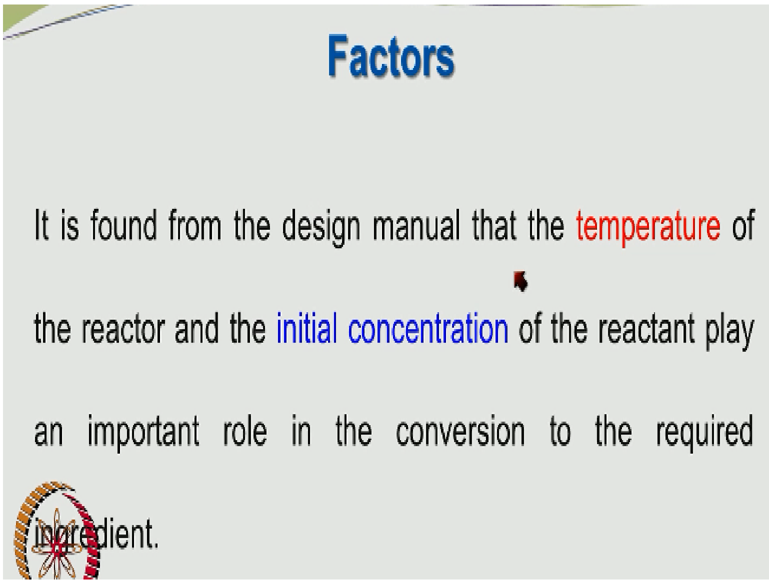
The MINITAB plots in this exercise set comprises of residuals, normal probability plots and the effects and their interactions.

(Refer Slide Time: 00:46)



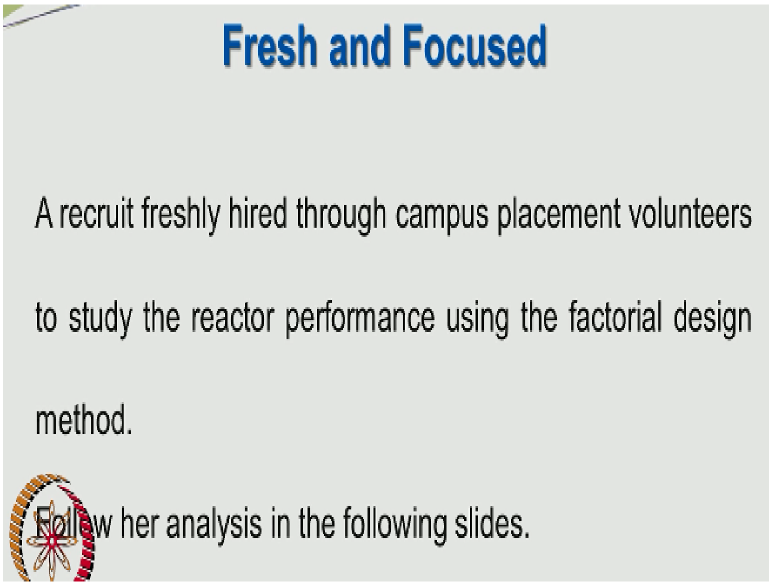
Example 1 a chemical reactor is used in a small scale industry for the production of an important ingredient in a popular perfume. The reactor operates at constant pressure of 5 atmospheres and is fed with the reactant.

(Refer Slide Time: 01:05)



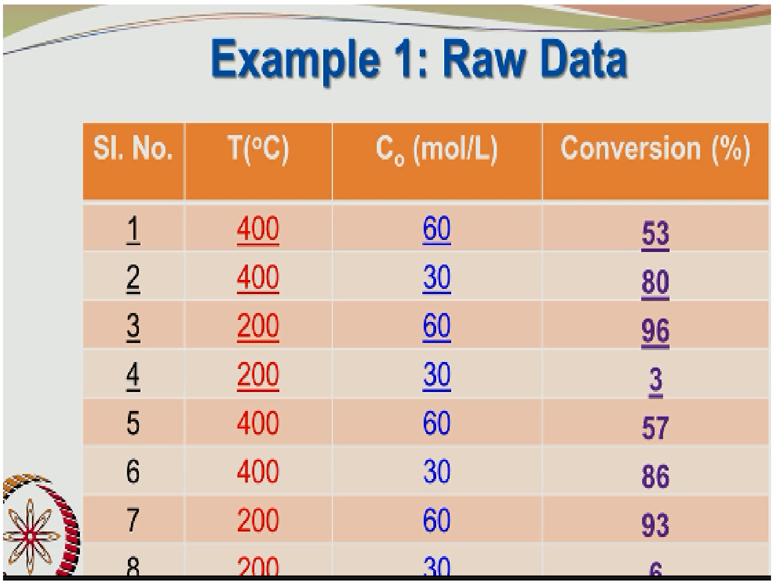
There is found from the design manual that the temperature of the reactor the initial concentration of the reactant plays important roles in the conversion to the required ingredient.

(Refer Slide Time: 01:17)



A recruit freshly hired through campus placements volunteers to study the reactor performance using the factorial design method. Follow her analysis in the following slides.

(Refer Slide Time: 01:28)

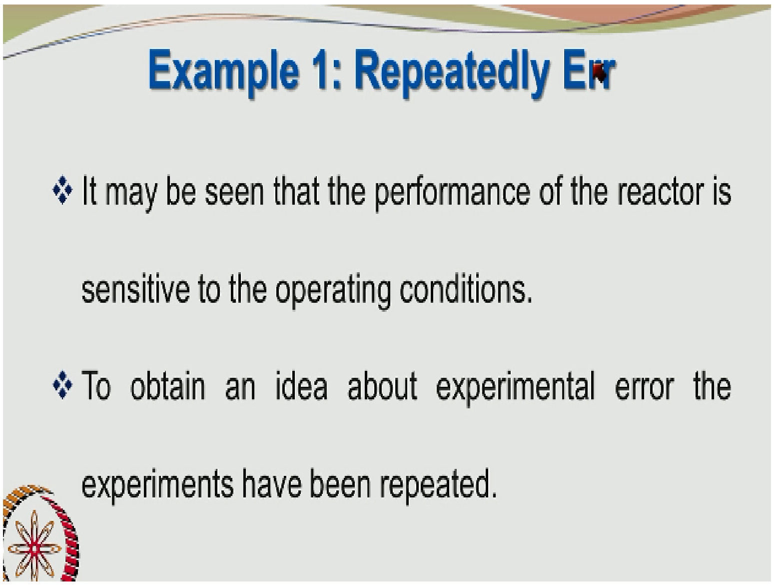


Example 1: Raw Data

Sl. No.	T(°C)	C_o (mol/L)	Conversion (%)
1	400	60	53
2	400	30	80
3	200	60	96
4	200	30	3
5	400	60	57
6	400	30	86
7	200	60	93
8	200	30	6

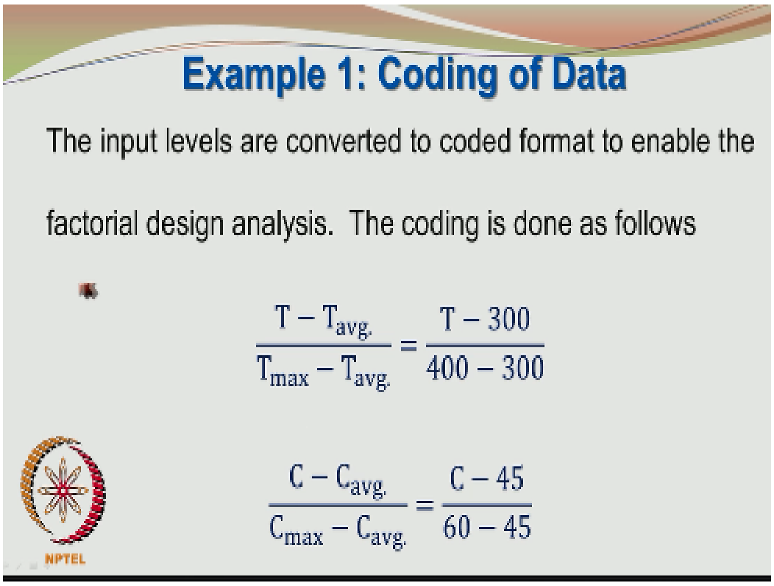
So this is the data you can see that it involves 2 factors temperature in degree centigrade and initial concentration in moles per liter. The 8 runs have been performed it is a 2 power 2 design repeated twice and the conversion values are given here and there are certain operating conditions where the conversion is pretty high at 96% and 93%. There are also certain conditions where the conversion is very low even going down to 3%.

(Refer Slide Time: 02:04)



So it may be said that the performance of the reactor is sensitive to the operating conditions. To obtain an idea about the experimental error the experiments have been repeated. So it is very important for us to first convert the given raw data into a coded format -+ or -1+1 format. So that it is confirming with the factorial design scheme.

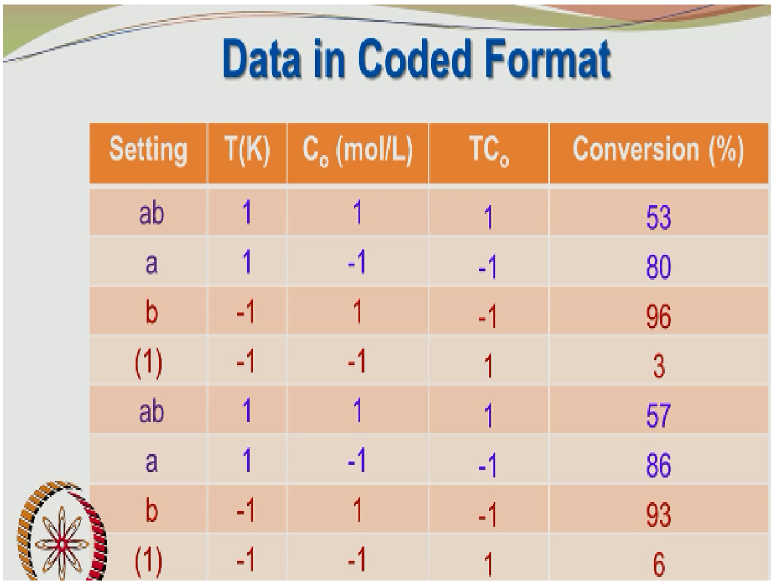
(Refer Slide Time: 02:30)



So how do you do the coding? There are 2 levels of temperature at 200 and 400 degree centigrade so the average would be 300 degree centigrade. So you have the coding formula T-T average/ T max-T average. Average is 300 degree centigrade so we have T-300/400-300. Similarly, the concentration values may also be coded and expressed as either -1 or +1. So we have C-C average/ C max-C average which is C-45/60-45.

The maximum concentration is 60 and the minimum concentration was 30 moles per liter 60 and 30. So when T is T max the value would be +1 and when T is T min this will be 200-300 which is -100., -100/+100 will be -1. Similarly, it would be C max would be 60 then this would become 1 and when C min is 30 that will be -15/+15 it will be -1.

(Refer Slide Time: 03:55)



Setting	T(K)	C_o (mol/L)	TC_o	Conversion (%)
ab	1	1	1	53
a	1	-1	-1	80
b	-1	1	-1	96
(1)	-1	-1	1	3
ab	1	1	1	57
a	1	-1	-1	86
b	-1	1	-1	93
(1)	-1	-1	1	6

So now we have coded the raw data the experimental settings and this is what we have and we can also create a column of TC0 which is the product of the elements in these 2 columns and we get 1-1-1 and so on. You can see that when both temperature and concentration were at the highest levels then that is called as setting A and B. For setting A temperature is at level 1 and initial concentration is at -1.

So obviously temperature is in factor A and C 0 is factor B. There should be a correction here now that we have expressed these in terms of dimension less variables we need to remove the units. Now they are in coded format. So we may as well call them as A, B and AB. So this is the corrected version of this slide. So the temperatures and concentration are no longer expressed in units since they have been converted into dimensionless form and they are expressed in coded form.

They simply take 1-1 values. For example, AB again this is repeat. So 1 B, A and AB are the 4 corners of the square and again these experiments have been repeated. Even though they are expressed in the same sequence the experiments have been done through proper randomization so that any systematic effect is not influencing one particular run or a couple of runs.

So experiments have been randomized and so here you have the conversions. Now how to calculate the effects we have already seen the table involving contrast.

(Refer Slide Time: 06:04)

Data in Coded Format

Setting	A(-)	B (-)	AB	Conversion (%)
ab	1	1	1	53
a	1	-1	-1	80
b	-1	1	-1	96
(1)	-1	-1	1	3
ab	1	1	1	57
a	1	-1	-1	86
b	-1	1	-1	93
(1)	-1	-1	1	6

So we contrast for A would be ab+a-b-1.

(Refer Slide Time: 06:18)

$$A = \frac{1}{2n}\{[ab - b] + [a - (1)]\} = \frac{1}{2n}[ab + a - b - (1)]$$

$$B = \frac{1}{2n}\{[ab - a] + [b - (1)]\} = \frac{1}{2n}[ab + b - a - (1)]$$

- interaction effect:

$$AB = \frac{1}{2n}\{[ab - b] - [a - (1)]\} = \frac{1}{2n}[ab + (1) - a - b]$$

$$\text{Effect} = \frac{\text{Contrast}}{n2^{k-1}}$$

Ab+a-b-1 and that we divide it by 1/2 because we are taking the average of ab and a and then subtracting it from the average b and 1. And then we are dividing it by n also to account for the repeats. Similarly, you can calculate the effect for b and the interaction effect ab also may be calculated. The effect is defined as contrast the contrast is what we have inside these brackets divided by n that n is the number of repeats and 2 power k-1.

K is the number of factors so this would be contrast divided by n * 2 to the power of 2-1 which should be 2n and that is what we have here.

(Refer Slide Time: 07:07)

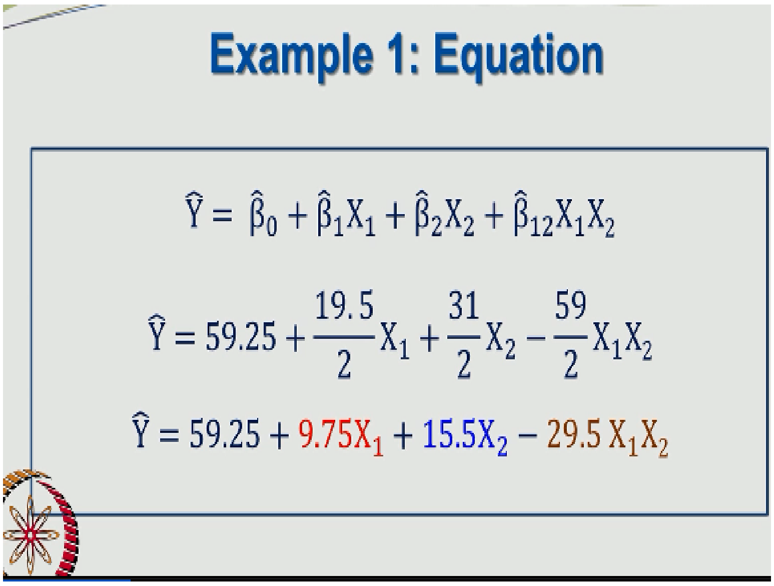
Coefficients : Effects Equation

Sl. No.	Effect	Value	Coefficient
1	Main Effect A	19.5	9.75
2	Main Effect B	31.0	15.5
3	Interaction Effect AB	-59.0	-29.5

So after we do the calculations correctly we get the main effect A as 19.5 the main effect B is 31 and the interaction effect AB is quite significant at -59. So in the model equation expressing the relationship between the conversions and the coded form of the variables A and B. The effect value is divided by 2. The model equation indicates the change in response to a unit change in the variables or the independent variables in the equation.

Now the effects have been calculated based on a change from -1 to +1 a jump of 2 units and hence we have to account for that change and so we divide the effect value by 2 to get the value of the coefficient. That is what the slide says and hence we have the coefficient as 19.5/2 that would be 9.75 31/2 is 15.5-59/2-29.5.

(Refer Slide Time: 08:35)

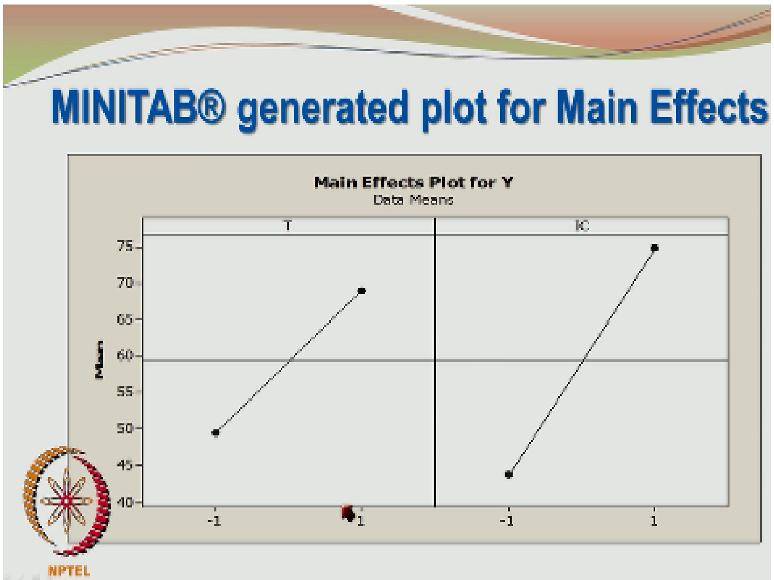


Okay and hence we have the equation Y hat= beta 0+beta 1 X1+beta 2 X2+beta 12 X1 X2. We have the hat over these entities y and beta to indicate that they have been predicted and hence we have Y hat is 59.25+9.75 X1+15.5 X2 and -29.5 X1, X2. A word of caution about using these equations you cannot plug in the value of X1 at let us say 1.5 and X2 as -2 and hope these equations would give you the correct conversion.

It may even give absurd conversions of exceeding 100% or below 0%. This equation is only meant for the range of variables studied. This is also true for any correlation developed from experimental data. The correlation is only valid in the range of experimental data is considered and another thing about this correlation is we really do not know how well the correlation would predict for an X1 value within the range.

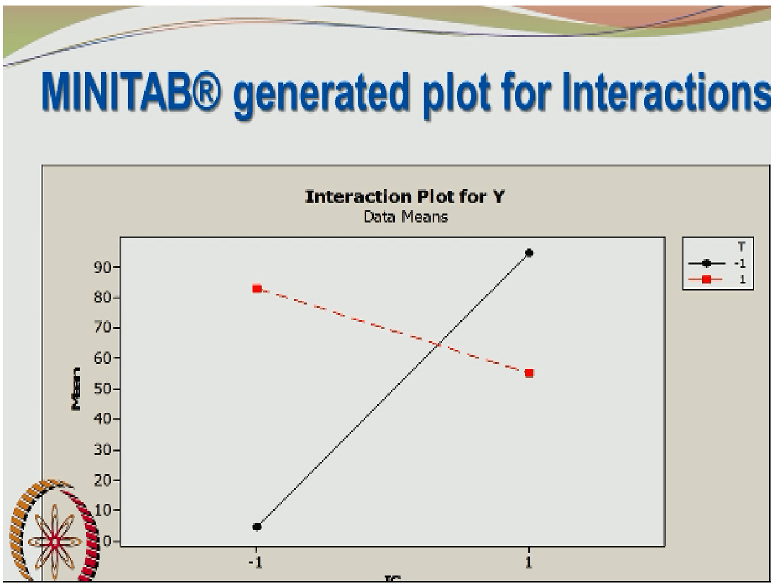
For example, even though you have done the experiments at -1 and +1 what would happen if I substitute the value of 0 and 0 is lying within the range. It is not outside the range of -1 and +1. So if I plug in the value of 0 will I get a proper conversion. You may or you may not because this is a kind of a fixed effects model and only applicable at the experimental conditions at which the runs were carried.

(Refer Slide Time: 10:29)



This is where we have the main effects for factors A and B. It shows that when the temperature rises from a lower level to a higher level the mean conversion increases and similarly the initial concentration or factor B in the coded form changes from -1 to +1 the mean conversion also increases. So it looks like you perform experiments at the maximum value of temperature and the maximum value of initial concentration to get the maximum conversion, but this is not the complete story.

(Refer Slide Time: 11:05)

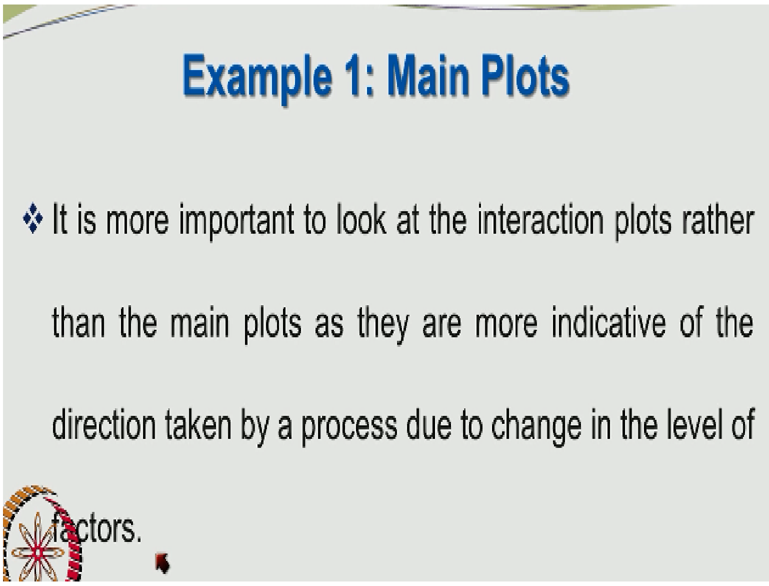


Here we see the interaction a strong one at that between the 2 factors. If the 2 lines had been parallel to each other then interaction effects could have been ignored, but the very fact that they are crossing one another indicates a very strong interaction between the 2 factors a and b. So this tells that at the lower level of temperature when the initial concentration rises from -1 to +1.

So at the lower level of temperature when the initial concentration goes from a lower value to higher value the percentage yield increases, but on the other hand when you performed the experiments at the highest temperature when the initial concentration rises from a lower level to a higher level the mean yield actually decreases. So the yield or the percentage conversion let me just check what percentage conversion. The percentage conversion depends upon the setting of the variables.

If the temperature is at the lower setting, then increasing concentration leads to higher conversion. On the other hand, if the temperature is at a higher value increasing the initial concentration actually reduces the percentage conversion. So just do not go with the main effects plots because they seem to indicate that when you increase the temperature and the initial concentration the percentage conversion increases.

(Refer Slide Time: 12:52)



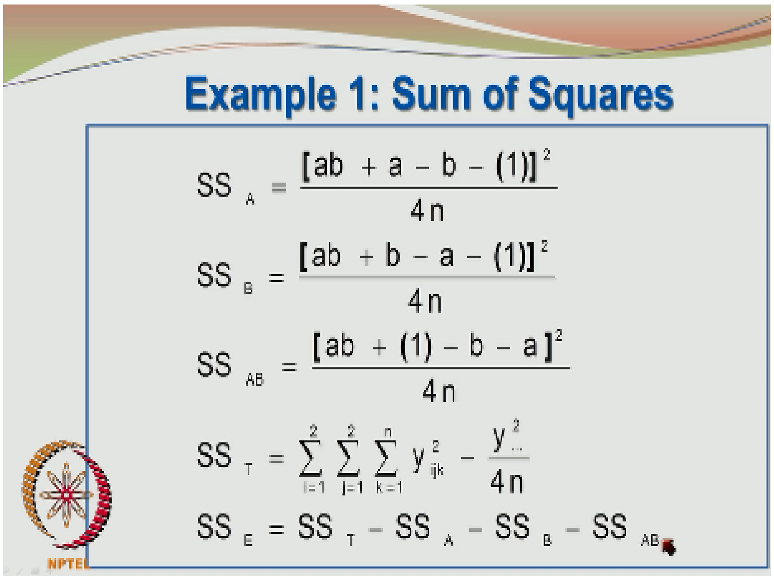
It is more important to look at the interaction plots rather than the main plots as they are more indicative of the directions taken by a process due to change in the level of the factors. Now we have to construct the ANNOVA table or the analysis of the variance table in order to see which of the effects are significant. By looking at the model equations coefficient we get the rough idea.

We do not have to worry about the actual value taken by the factors because these values are coded. So the A is -1 and +1 and B is also at -1 and +1. It is not as if one factor is of the order of 1000s so its coefficient would be in the order of 10 power -3 whereas another factor is in the order of 1. So the coefficient would be in the order of again 10 power 0 or 1. So you cannot claim that factor B coefficient is higher than factor A coefficient.

The coefficient sometimes has to also reflect on the value of the independent variable it is attached with, but when you do coding both the factors are varying from -1 to +1. They have been scaled nicely so that you do not have to really worry about the magnitude of the independent variables or the magnitude of the factors. So by looking at the coefficient themselves you can very well say whether a factor is likely to be significant or not.

But in order to confirm this. We have to do the analysis of variance or the ANNOVA. Let us see how to calculate the ANNOVA table for that we need the sum of squares. Again calculating the sum of squares is pretty easy it involves the contrast. We use the contrast for finding the effects. We will also use the same contrast to find the sum of squares.

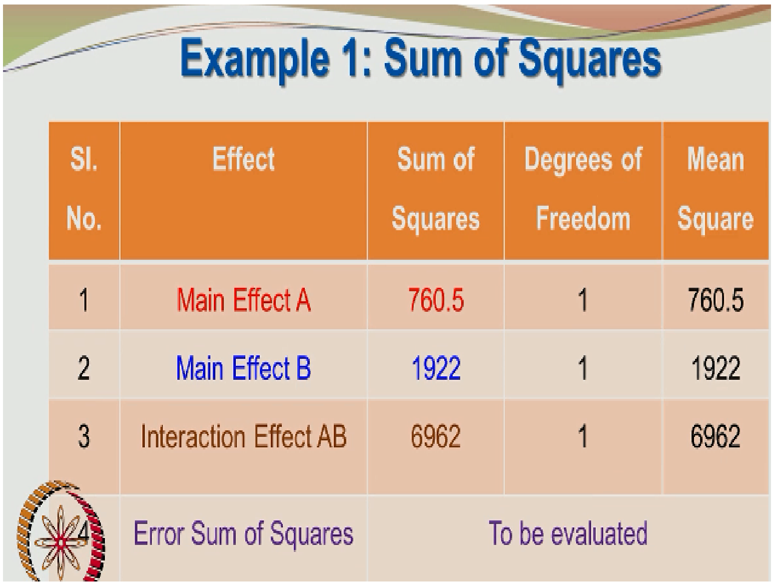
(Refer Slide Time: 14:51)



So you have SSA= ab+a-b-1 whole square/4n. Sum of squares due to factor b would be ab + b-a-1 whole square/4n. Please look at the table of contrast of the design matrix and indeed confirm that the sum of squares are given by these formula. The total sum of squares may be calculated by using this shortcut formula yijk square each entity in the table square the sum of squares of errors may be easily computed by taking the difference of the total sum of squares or taking the difference between total sum of squares.

And the combined sum of squares due to A B and AB. So this is very interesting and very straight forward. Please do the calculating using a spread sheet or even a hand calculator and please confirm whether the results which I am going to show next are correct.

(Refer Slide Time: 15:58)

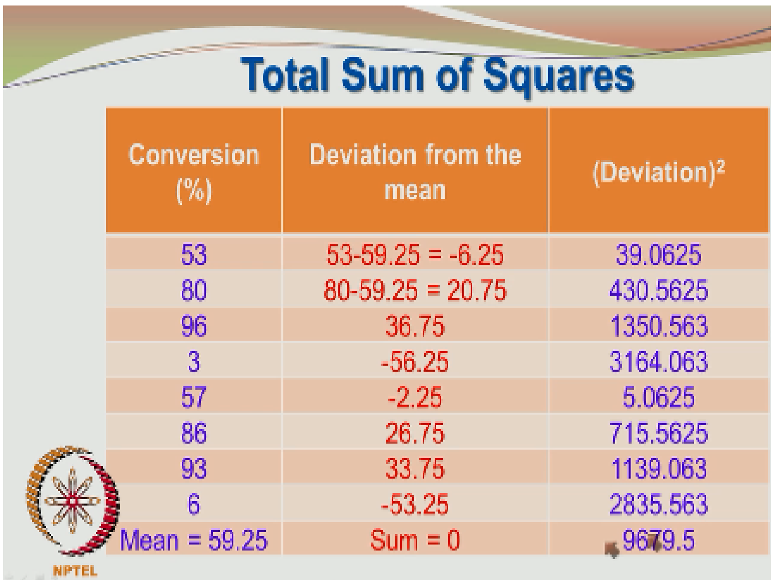


Example 1: Sum of Squares

Sl. No.	Effect	Sum of Squares	Degrees of Freedom	Mean Square
1	Main Effect A	760.5	1	760.5
2	Main Effect B	1922	1	1922
3	Interaction Effect AB	6962	1	6962
4	Error Sum of Squares	To be evaluated		

So you have example 1 the sum of squares the main effect A is having a sum of squares of 750.5 main effect B is having a sum of squares of 1922 and interaction effect AB is having a sum of squares of 6962. The degrees of freedom associated with each of these would be 1 and the mean square would be you just simply divide the sum of squares by the degrees of freedom and you will get these values.

(Refer Slide Time: 16:28)



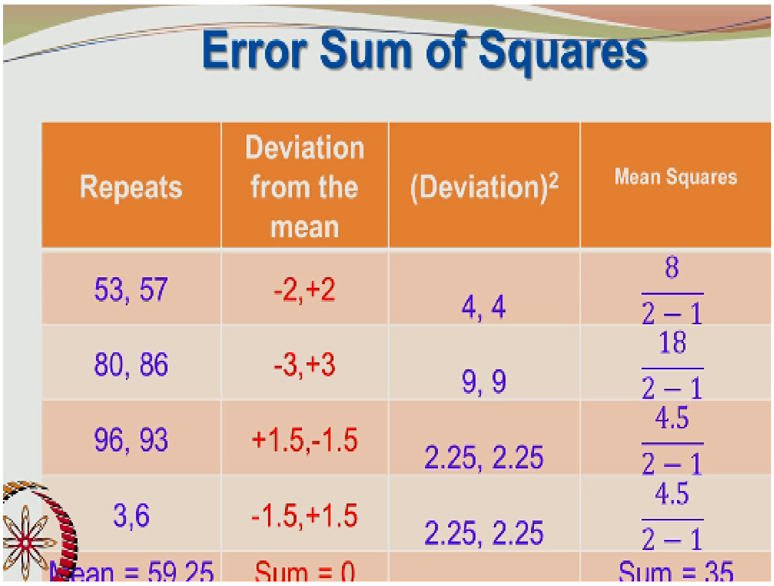
Total Sum of Squares

Conversion (%)	Deviation from the mean	$(\text{Deviation})^2$
53	53-59.25 = -6.25	39.0625
80	80-59.25 = 20.75	430.5625
96	36.75	1350.563
3	-56.25	3164.063
57	-2.25	5.0625
86	26.75	715.5625
93	33.75	1139.063
6	-53.25	2835.563
Mean = 59.25	Sum = 0	9679.5

Total sum of squares you can also use the shortcut formula I told you earlier. So you first find the deviation of each and every individual value from the mean and then that difference you square. And when you sum the square of such deviations you get 9679.5. So the mean value is 59.25 sum of deviations is 0 and sum of squares is 9679.5. Please note down the values of the other sum of squares A is 760.5 B is 922.

And the interaction effects AB is 6962 and the total value is 9679.5.

(Refer Slide Time: 17:25)



To calculate the error sum of squares you can reduce or deduct the sum of squares of A B and AB from the total sum of squares that is one way of doing it. Another way of doing it also I am showing here. Here the repeats are having percentage conversions of 53 and 57. Please remember that repeats also indicate the same setting for both these runs. So 53 and 57 let us go back to the table 53 and 57 correspond to 460. 460 would be +1+1 again this is +1 +1 the coded format you get 53 and 57% conversion.

And so the deviation from the mean would be -2 and +2 the mean of course is 55. So the deviation from the mean is -2 and +2 the deviation square is 4 and 4 and then the mean square would be 8/2-1 that is only 8 rises from 4+4 is 8 and then you are having n-1. So it is like taking the variance so it is 2-1 and similarly 1886 would have an average value of 83 and so you have -3 and +3 which makes it 9 and 9 18/1. 96 and 93 would be 94.5 and so you have +1.5-1.5 2.25, 2.25 4.5/2-1 where this 4.5 came from adding up 2.25 with 2.25.

3 and 6 will have an average of 4.5 again you get -1.5 and +1.5, 3-4.5 is 1.5, 6-4.5 is +1.5 and so the square of these 2 terms would be 2.25+2.25 which is 4.5 divided by 1. So you add up

all these things 26 30.5 31+4 is 35. So the sum of deviation square of the deviation would be 35.

(Refer Slide Time: 19:49)

Repeats	Deviation from the mean	$(\text{Deviation})^2$	Mean Squares
53, 57	-2,+2	4, 4	$\frac{8}{2-1}$
80, 86	-3,+3	9, 9	$\frac{18}{2-1}$
96, 93	+1.5,-1.5	2.25, 2.25	$\frac{4.5}{2-1}$
3,6	-1.5,+1.5	2.25, 2.25	$\frac{4.5}{2-1}$
Mean = 59.25	Sum = 0		Sum = 35

$$S^2 = \frac{\sum_{i=1}^{n} \nu_i s_i^2}{\sum_{i=1}^{n} \nu_i} = \frac{35}{4} = 8.75$$

And once you have this you can find the error variance as nu 1 si square divided by nu 1 sum. Nu I is the degrees of freedom associated with each of the individual s1 square. So you are just calculating the variance of 53 and 57 and that you get as 8. So nu I would be 1 the degrees of freedom would be 1 so 1*8. Since all the degrees of freedom are 1 you get 35 1*18+1* 4.5+1*4.5 would be 35 as shown here divided by sum of the degrees of freedom would be 4. So the result would be 8.75.

(Refer Slide Time: 20:41)

Formulae for Sum of Squares

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{...})^2 = bn\sum_{i=1}^{a}(\bar{y}_{i..}-\bar{y}_{...})^2 + an\sum_{j=1}^{b}(\bar{y}_{.j.}-\bar{y}_{...})^2$$

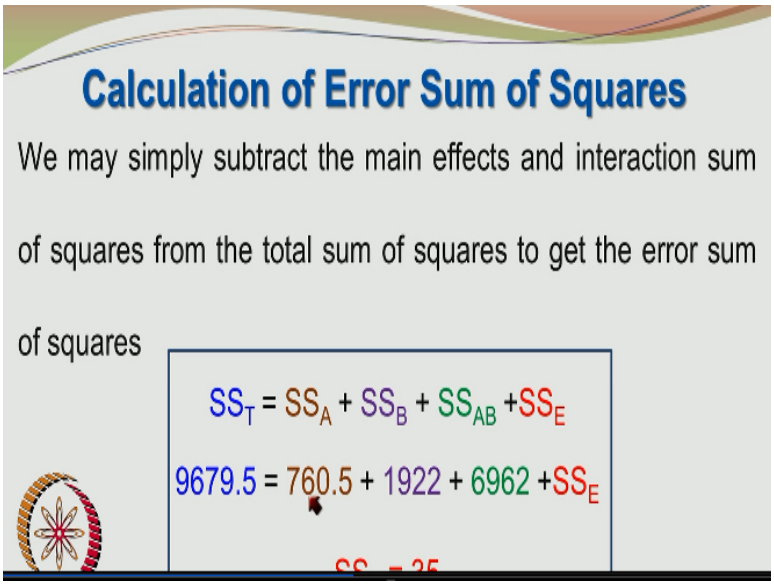
$$+ n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{ij.})^2$$

There is another way of doing it. So this is the overall sum of squares relation. This is the total sum of squares and that is resolved or divided * sum of square contribution from factor

A and sum of squares contribution from factor B and this represents the sum of square contribution from AB and this is the error sum of squares.

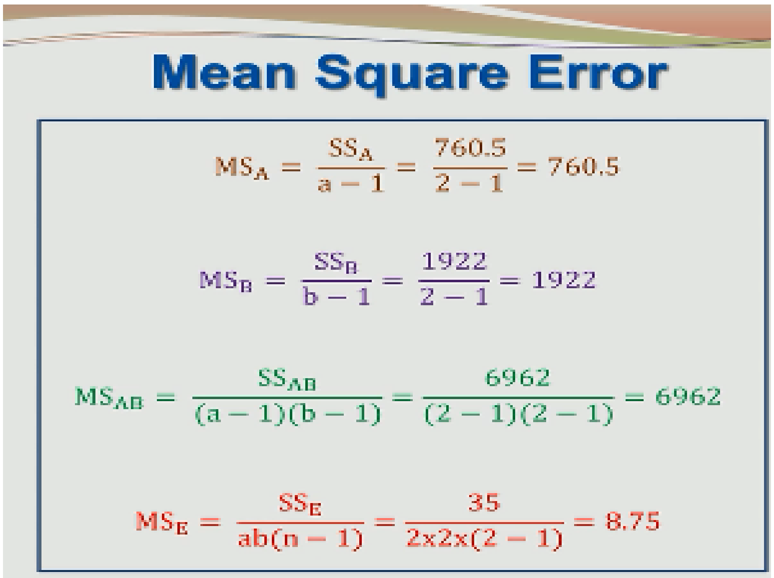
So you have this and you can calculate the total sum of squares first calculate all these sum of squares and then calculate sum of squares of error or you can also calculate it independently as I have shown in the previous 2 slides.

(Refer Slide Time: 21:23)



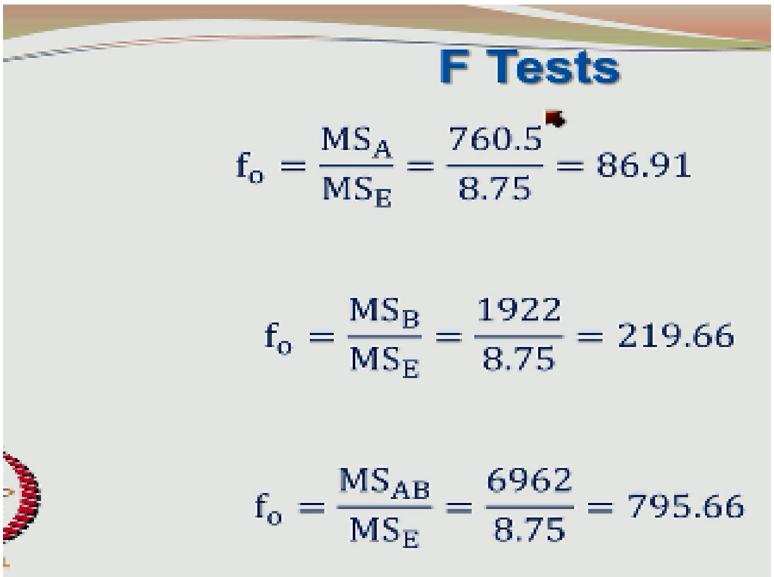
So we have 9679.5=760.5+1922+6962+ sum of square of the error. And you can also see coincidentally that the sum of squares of the error based on this equation is 35 and this is exactly what we obtained by our independent calculations. So 2+2 is 4, 4+5 is 9 that matches here and then 6+2 8+6+3 17 so that 7 matches here 8+9 17+9 26. So 6 matches here and 3+6 9. So 9679 and then this extra 0.5 is there so the totaling is correct.

(Refer Slide Time: 22:18)



And hence now that we have calculated all the sum of squares for different factor we can divide it by the number of degrees of freedom and get the mean square for AB, A, B and also for the error. So you can find it as 760.5 1922, 6962 and 8.75 here.

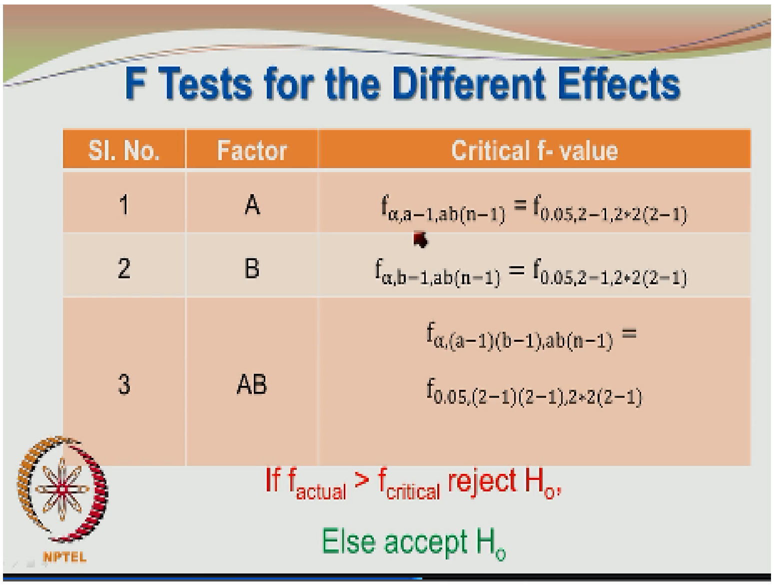
(Refer Slide Time: 22:41)



Now we can carry out the F tests the F statistics is given by mean square of the effect divided by mean square error and for A it is 86.91 for b it is 219.66 and for AB it is 795.66. These values are pretty high and so it is likely that these values are lying in the rejection region please remember that the null hypothesis we have taken is all the effects are negligible and the alternative hypothesis is at least one of the effect is important.

And it contributes to the variability in the process. Now it can be seen from these F values which are pretty high that all these effects A B as well as the interaction they are likely to be important.

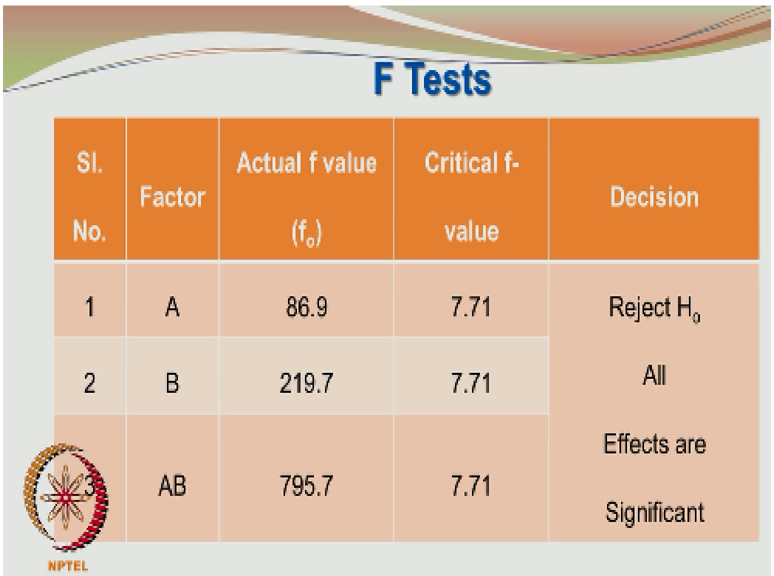
(Refer Slide Time: 23:38)



So what we have to do is we have chosen alpha the level of significance at 0.05 which is a common value by default we can take it as 0.05 unless and otherwise specified AB*n-1. 2 repeats a*b is 4 and hence we have the denominator degrees of freedom in the F test to be 4 and the numerator degrees of freedom would be A-1 the levels of A is 2 and so we have 2-1. So numerator degrees of freedom is 1 for all the factors and the denominator degrees of freedom is 4.

And we calculate all the F alpha values and if F actual is greater than F critical then we reject the null hypothesis and these are all the critical F values.

(Refer Slide Time: 24:41)



And the critical F values comes to be 7.71 you may verify this using the F distribution charts and the appropriate degrees of freedom. In our case now it is 1 and 4 numerator and denominator degrees of freedom respectively and we are looking at that table corresponding to alpha=0.05 and this value comes as 7.71, but the actual value is much higher than this. So the F statistics is firmly lying in the rejection region.

And so we can reject the null hypothesis that factor A is insignificant factor B is insignificant and the interaction between the 2 factors is also insignificant. So all the 3 hypothesis are rejected and hence we have to conclude that all the factor A and B as well as the interaction between them A B are significant.

(Refer Slide Time: 25:42)

ANOVA TABLE

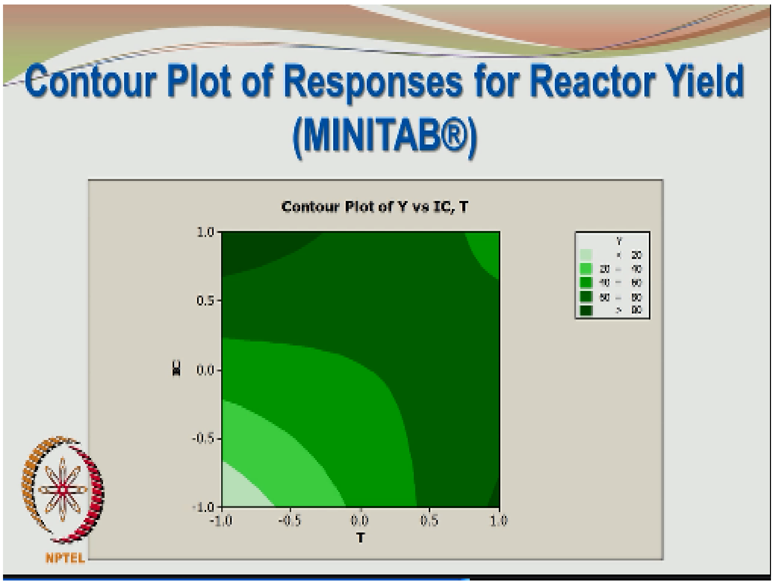
Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	f_o	$f_{critical}$
T	760.5	1	760.5	86.91	7.71
IC	1922	1	1922	219.66	7.71
T*IC	6962	1	6962	795.66	7.71
Error	35	4	8.75		7.71
Total	9679.5	7			

So now we can construct the ANOVA table. Here you have the source of variation from temperature initial concentration and the interaction between the 2 and then you also have the error source of variation and you have the sum of square 760.5 1922 6962 35 and this adds up 9679.5 just checking it 2+2 4+5 9 6+2 8 14, 17, 26, 29 wait a second. Then you have 3+6 9 that is fine.

Please check all the calculations at every intermediate step there is a likelihood that there may be a small error here and there. Anyway these values are correct as we saw 9679.5 and that is also the same value which we saw earlier 9679.5 so there is no problem here. And you can see the degrees of freedom also listed. We have 4 degrees of freedom for the error and it would be AB*n-1 A is 2 and B is 2 so 2*2 4 n-1 is 1. 4*1=4.

So we have totally 7 degrees of freedom and then what we do is we divide the sum of squares by the degrees of freedom we get the mean squares. We also get the F 0 value and the F0 is much higher than the F critical value so we reject the null hypothesis. And this ANOVA table tells how we got all these numbers. So if you are stuck you may kindly refer to this table and see that the calculation have been done according to the formulae given here.

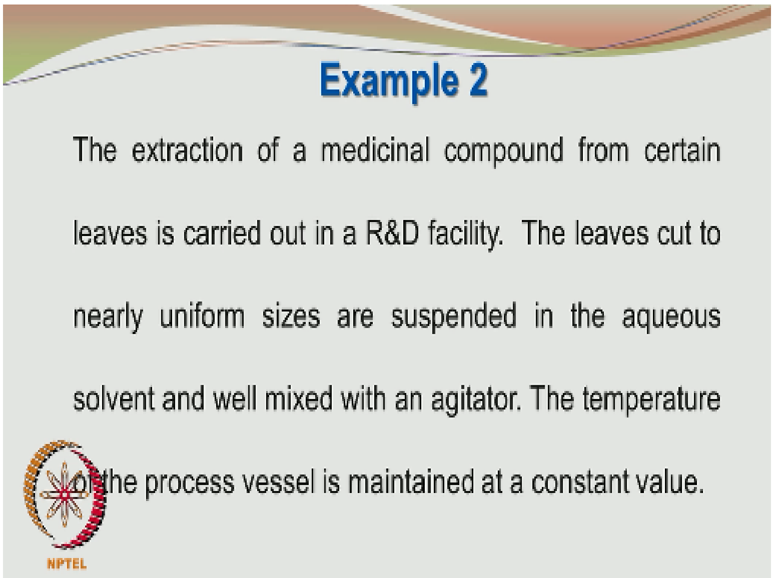
(Refer Slide Time: 27:41)



So these are nice contour plots and this really tells that you get pretty high conversion when you are having high levels of initial concentration and low levels of temperature. And also the contour plots are not showing any straight lines. They show considerable curvature which is also indicative of considerable interaction between the 2 factors. We will be now looking at example 2.

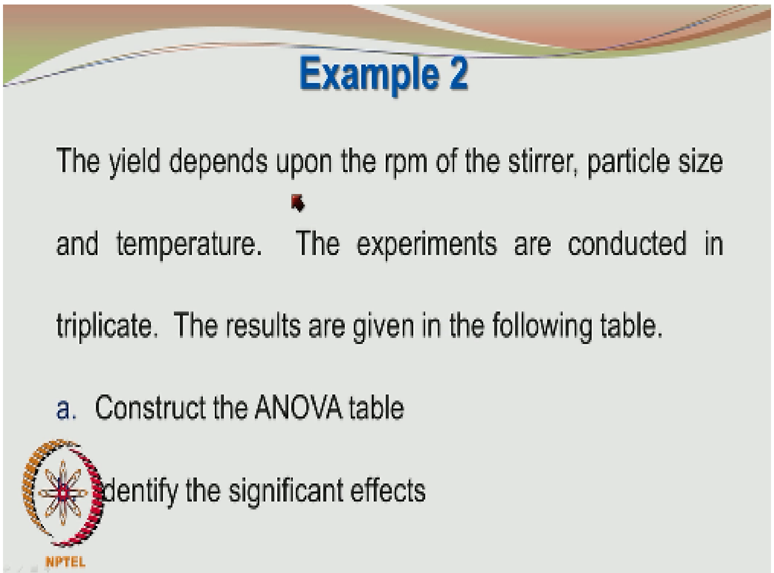
In this example, we will be looking at design involving 3 factors. 2 level factorial design with 3 factors. So the basic set would be 2 power 3 which is 8 runs, but we also need repeats to get an estimate of the experimental error.

(Refer Slide Time: 28:33)



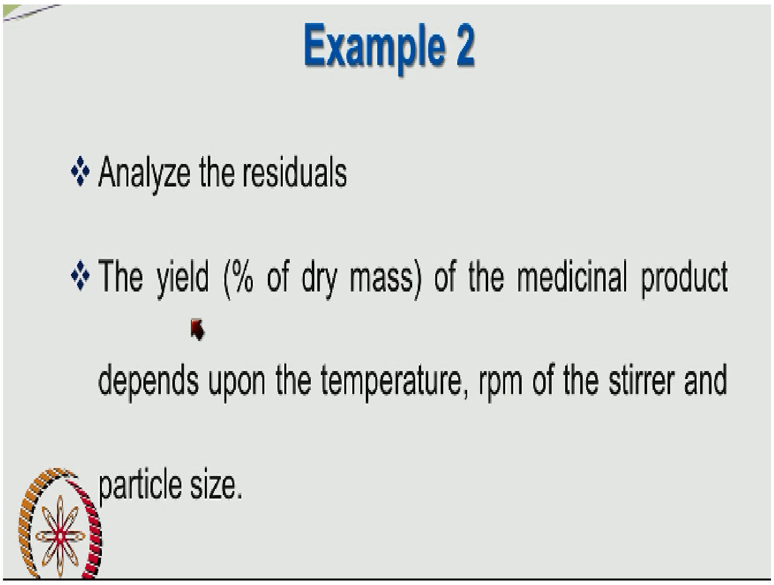
So the problem statement is the extraction of a medicinal compound from a certain leaves is carried out in an R&D facility. The leaves cut to nearly uniform sizes are suspended in the aqueous solvent and well mixed with an agitator. The temperature of the process vessel is maintained at a constant value.

(Refer Slide Time: 28:59)



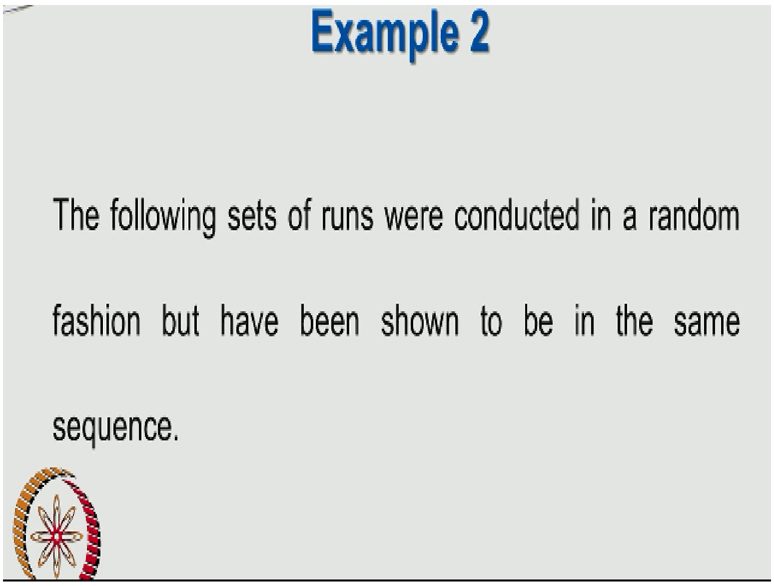
The yield of the medicinal substance depends upon the rpm revolution per minute of the stirrer particle size and temperature. The experiments are conducted in triplicate that means we have a total of 8*3 24 runs. The results are given in the table which will be shown in the next slide. Construct the ANOVA table and identify the significant effects.

(Refer Slide Time: 29:28)



Analyze the residuals and the yield percentage of dry mass of the medicinal products depends upon the temperature rpm of the stirrer and the particle size. So these will represent the 3 factors for our experiment.

(Refer Slide Time: 29:50)



Even though the results are reported in a same sequence. In reality they were performed in a randomized fashion.

(Refer Slide Time: 30:03)

First Set of Runs

Sl. No.	Temperature (°C)	Stirrer Speed (RPM)	Particle Size (mm)	Yield (%)
1	40	200	5	3.355
2	60	200	5	4.425
3	40	400	5	3.245
4	60	400	5	5.535
5	40	200	20	2.225
6	60	200	20	3.115
7	40	400	20	2.015
8	60	400	20	4.085

So these are the results the temperature was either kept at 40 degree centigrade or at 60 degrees centigrade. The RPM of the stirrer was at 200 RPM or at 400 RPM. Well you can go for higher RPMs, but there may be a possibility of vertex formation in the stirred vessel and for removing the vertices you may have to go for (()) (30:32). Apparently in this study the RPM was restricted to 400.

Perhaps to minimize or avoid the formation of vertices and the particle size was either kept at 5 millimeter or at 20 millimeters and of course you can cut the particles to a smaller size, but that would require more energy. You cut the particle sizes into smaller and smaller bits to increase or improve the surface area for mass transfer. The percentage yield is pretty low, but the probably the medicine is of rare type and hence expensive and hence even though the yield is at very low levels you may want to concentrate it further.

And then sell it off with due processing. The yield values are reported here in percentage of dry mass of the sample taken.

(Refer Slide Time: 31:45)

Second Repeat

Sl. No.	Temperature (°C)	Stirrer Speed (RPM)	Particle Size (mm)	Yield (%)
1	40	200	5	3.562
2	60	200	5	4.278
3	40	400	5	3.288
4	60	400	5	5.632
5	40	200	20	2.298
6	60	200	20	2.862
7	40	400	20	1.972
8	60	400	20	4.108

NPTEL

This is the second set of repeats.

(Refer Slide Time: 31:48)

Third Repeat

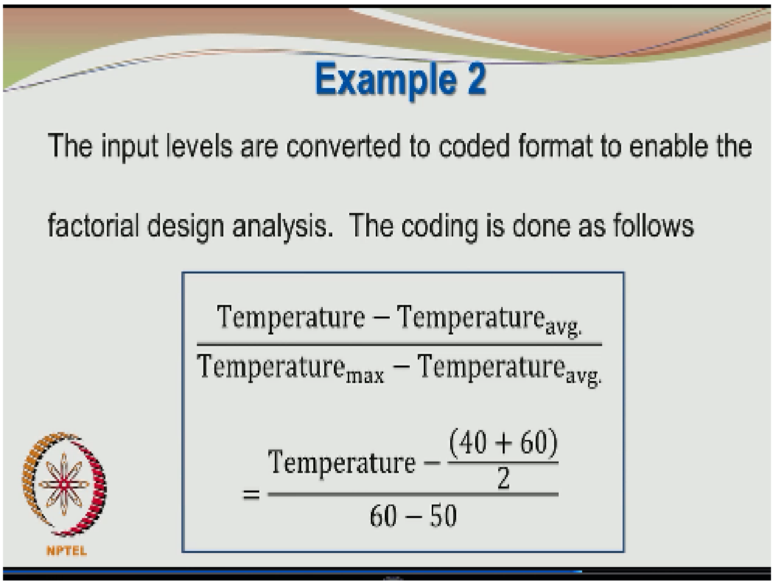
Sl. No.	Temperature (°C)	Stirrer Speed (RPM)	Particle Size (mm)	Yield (%)
1	40	200	5	3.367
2	60	200	5	4.493
3	40	400	5	3.173
4	60	400	5	5.727
5	40	200	20	2.233
6	60	200	20	3.067
7	40	400	20	1.867
8	60	400	20	4.073

NPTEL

And this is the third set of repeats. The experimental conditions are the same, but they have been repeated 3 times. So as before we have to convert the variables values into the coded form. Here the particle size is ranging from 5 to 20 and the RPM is ranging from 200 to 400 a factor of 40 to 80 times and here the temperature is changing from 40 to 60. So there is a considerable difference between the values taken by the factors.

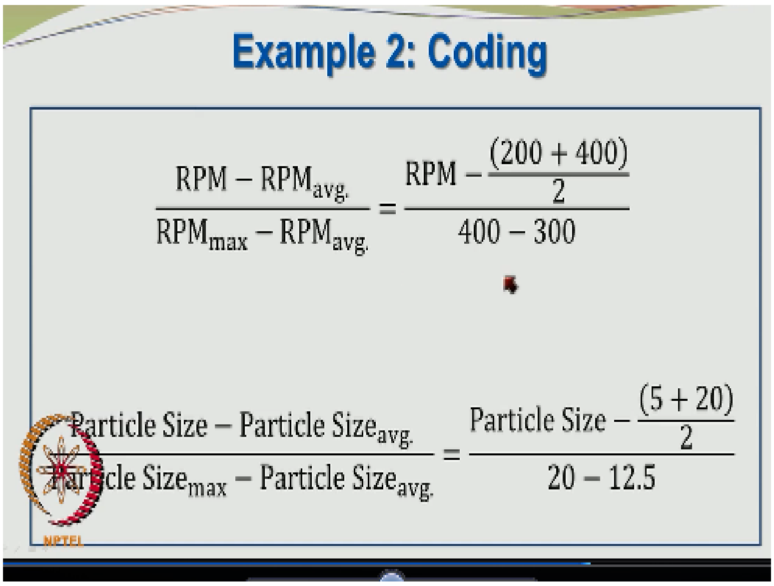
So to put them on the same basis we need to scale them and quote them and quote them.

(Refer Slide Time: 32:42)



That is what we have here. I have already told you the methods for coding in the previous example following exactly the same methodology we can represent the code form of temperature RPM and particle size as shown in these slides.

(Refer Slide Time: 33:01)



So RPM would be RPM-RPM average which should be 300 RMP max-RPM average. So 400-300 divided by 400-300 would be +1. If it is 200, 200-300 is -100, -100/+100 would be -1. Similarly, you have the coded form for the particle size.

(Refer Slide Time: 33:32)

Runs in Coded Format

Symbol	T	RPM	D	R1	R2	R3
(1)	-1	-1	-1	3.355	3.562	3.367
a	1	-1	-1	4.425	4.278	4.493
b	-1	1	-1	3.245	3.288	3.173
ab	1	1	-1	5.535	5.632	5.727
c	-1	-1	1	2.225	2.298	2.233
ac	1	-1	1	3.115	2.862	3.067
bc	-1	1	1	2.015	1.972	1.867
abc	1	1	1	4.085	4.108	4.073

So you can now report the values in the coded form and this is the lowest level setting for all the 3 factors and that is represented by the symbol 1. A is represented by only A at a high level so you have +1 here -1 –1. B is at -1 +1 -1 and so on. And here you have the variability because of the 3 repeats. So you can see that there is some variability when you repeat the experiment. Obviously when anyone reports the values in as identical then there is some reason to doubt those data.

There should be some variability in the reported results when they are repeated.

(Refer Slide Time: 34:30)

2^3 Design Matrix : A = T, B = RPM , C = D

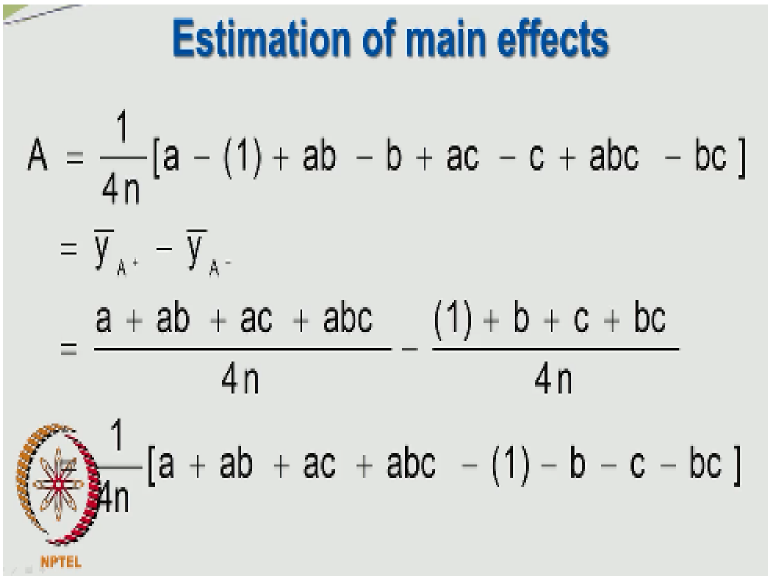
Treatment Combination	I	A	B	AB	C	AC	BC	ABC
(1)	+1	-1	-1	+1	-1	+1	+1	-1
a	+1	+1	-1	-1	-1	-1	+1	+1
b	+1	-1	+1	-1	-1	+1	-1	+1
ab	+1	+1	+1	+1	-1	-1	-1	-1
c	+1	-1	-1	+1	+1	-1	-1	+1
ac	+1	+1	-1	-1	+1	+1	-1	-1
bc	+1	-1	+1	-1	+1	-1	+1	-1
abc	+1	+1	+1	+1	+1	+1	+1	+1

So we construct the Design Matrix A and temperature B= RPM C= particle diameter. So we have I which is the column where all the entities are +1 A for 1 it will be -1 for small A it will be +1 and so on. For AB A is a higher level so it will be +1 for A. We have come across this several times in the previous lectures, but just for the sake of completion let me do for factor A.

All the factors are at the lower levels so you have -1 here factor A is at a higher level so you have +1 factor B is at a higher level so A is at a lower level so it is -1. So AB both A and B are at their higher levels. So A will be having +1. C A and B would be at the lower levels so A would be having -1. AC both A and C would be at the higher level so A would be having +1 B and C except A both B and C would be at their higher level so A would be having -1.

ABC all the 3 factors would be at their higher level so A would have +1. Similarly, you can fill up the values for the other columns. So you have in addition to AB you have AB which is nothing, but the product of the elements in each of the columns -1*-1 you get +1 for AB for 1. So AB would be having +1 here. And then you have ABC which is A * B * C -1*-1 +1*-1 -1. ABC obviously is also equal to AB*C. So setting up this table is quite straight forward.

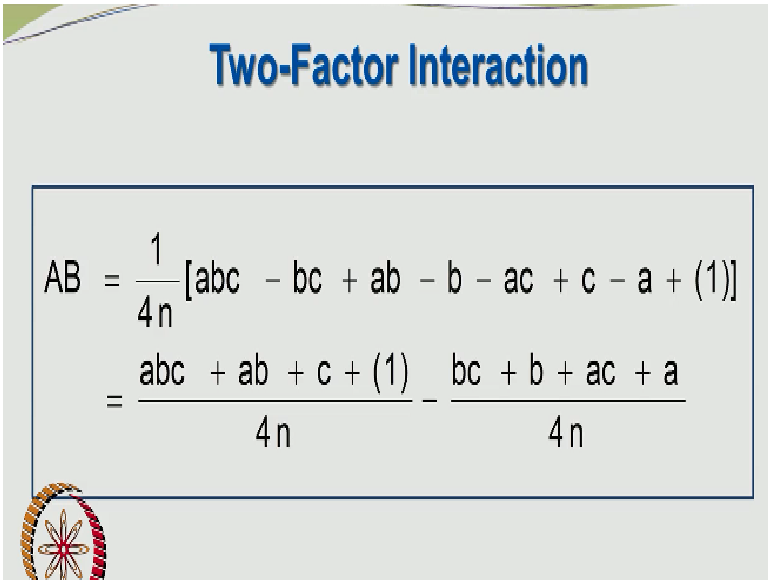
(Refer Slide Time: 36:30)



Now to estimate the main effects you adopt the same procedure you identify the table of contrast for each factor and their respective interactions and here you have a-1+ab-b+ac-C+abc-bc. This exactly corresponds to the table we saw earlier. For A it will be -1 +A. Then you have –b+ab and then you have –c +ac. Then you have –bc+ abc. So setting up the factor A would be quite straight forward and that represents the average value of A at the higher level of A-the average value at the lower level of A.

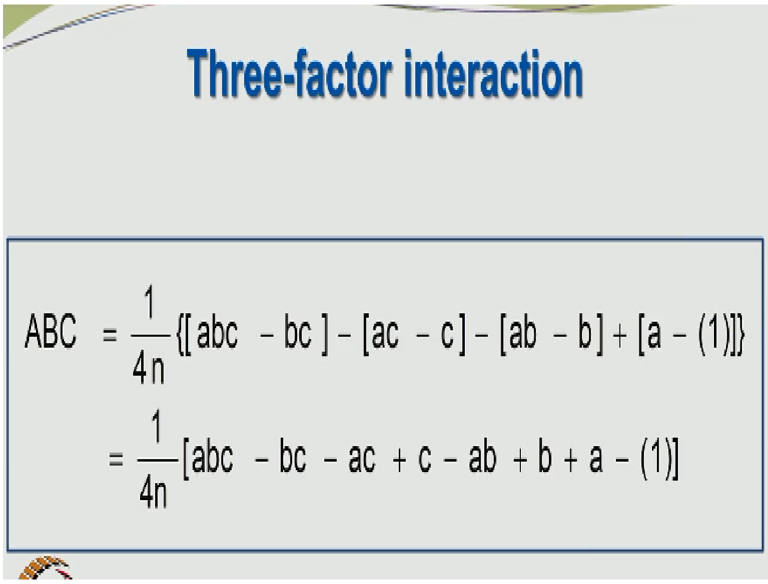
And so you can calculate the effect of A pretty easily. Similarly, you can use the same methodology to calculate the effects of factors B and C.

(Refer Slide Time: 37:50)



AB also the same procedure is involved please look at AB here you have abc- bc same as what we saw in that equation. And then you have –ac +c and so on. I think you got the point –a-b+1 so pretty straight forward and easy to calculate the effects. If you recollect it was n*2 power k-1. So to calculate the effects you had to divide it by n*2 power k-1. K here is 3 2 power k-1 is 2 power 3-1 which is 2 power 2 4 *n so that is what you have here.

(Refer Slide Time: 39:05)



Similarly, you can calculate the 3 factor interaction ABC and you can find the effect.

(Refer Slide Time: 39:15)

Sum of squares: SS = (Contrast)²/8n

$$SS_A := \frac{1}{8n}[(a + ab + ac + abc - 1 - b - c - bc)]^2$$

$$SS_B := \frac{1}{8n}[(b + ab + bc + abc - 1 - a - c - ac)]^2$$

$$SS_C := \frac{1}{8n}[(c + ac + bc + abc - 1 - a - b - ab)]^2$$

$$SS_{AB} := \frac{1}{8n}[(1 + ab + c + abc - a - b - ac - bc)]^2$$

Sum of squares of ABC and AB may be also found from the table of contrast the same contrast is used and squared divided by 8*N where n represents the number of repeats. You can do it for a b c ab and also bc ac and abc. So that is what you have here. Every place the contrast is squared and then divided by 8n as shown in this slide.

(Refer Slide Time: 39:54)

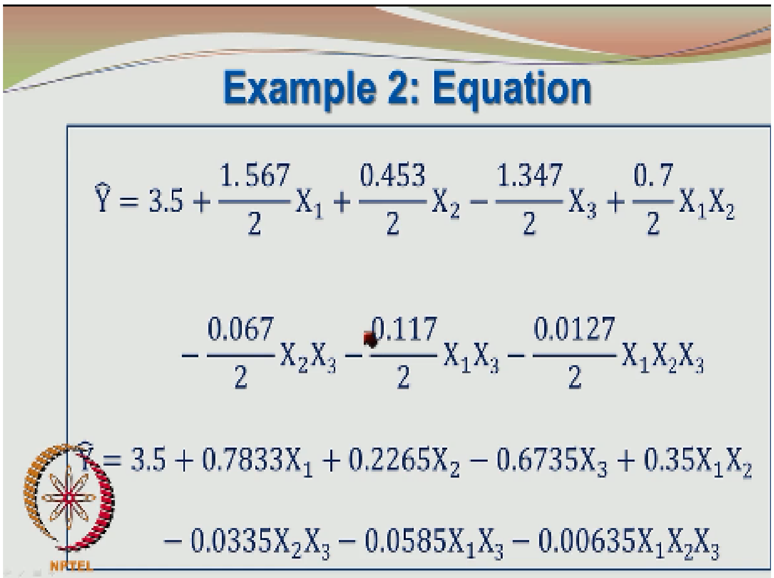
Example 2: Equation

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_{12} X_1 X_2 + \hat{\beta}_{13} X_1 X_3 + \hat{\beta}_{23} X_2 X_2 + \hat{\beta}_{123} X_1 X_2 X_3$$

So finally after having written all these things we can write down the model. This is the intercept so called intercept in the multidimensional plane and this would be the coefficient associated with the factor 1 or factor A coefficient associated with factor 2 or factor b coefficient associated with factor 3 or factor C binary interaction between A and B factors binary interaction between a and c factors. Binary interaction between b and c factors and then you also have the ternary interaction between a b and c given by beta 1 2 3.

As before the hats here indicate that these are predicted. This is the predicted value using this equation involving the predictors beta 0 hat beta 1 hat so on to beta 1, 2, 3 hat.

(Refer Slide Time: 41:02)



Then you have the calculations carried out and you can find that the coefficients are 1.567/2. The effect of factor A is 1.567, effect of factor B is 0.453 effect of factor C is -1.347 and effect of factors interaction between factors A and B was 0.7 and as I said earlier since you are jumping from -1 level to +1 level we have to divide the effects by 2 to get the coefficients in this equation.

So finally after all the calculations (()) (41:45) has settled we get y hat= 3.5+ 0.7833 X1+0.2265 X2-0.6735 X 3 and so on. You can see that the interactions are of the same magnitude order of magnitude has the main effects for X1 X2 lower for X2 and X3 and 2 orders of magnitude lesser for the third level interaction between A B And C.

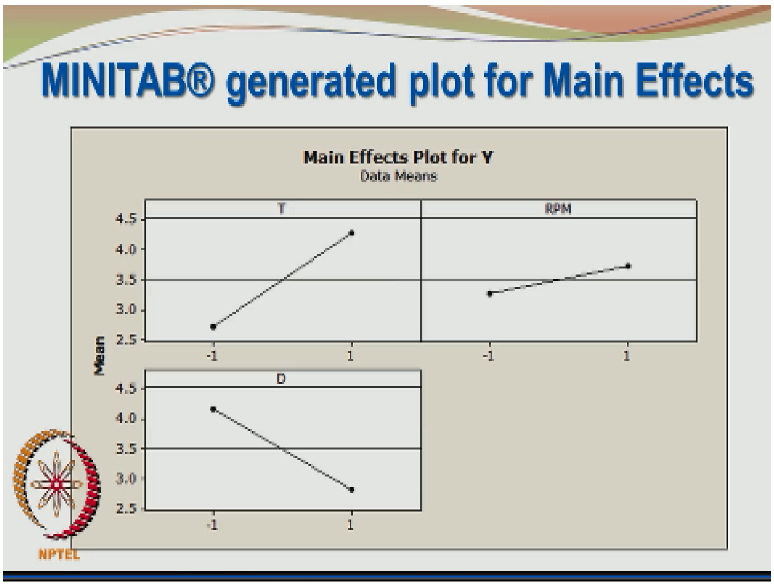
(Refer Slide Time: 42:23)

Effects	Value	Coefficient	Sum of Squares
A	1.567	0.783	14.727
B	0.453	0.227	1.233
C	-1.347	-0.673	10.881
AB	0.7	0.350	2.94
BC	-0.067	-0.033	0.0267
AC	-0.117	-0.058	0.0817
ABC	-0.0127	-0.0063	0.00096
Total			29.890

So I have summarized the effects the coefficient on the sum of squares. I request you to do the calculation on your own using the formula given in these slides and ensure that the values you get are matching with my values. I am reasonably sure or even pretty sure about my calculations, but who knows even I might have made a mistake and if I had done so please bring into my notice and I will correct it.

So you have the value coefficient and sum of squares and the sum of squares are written down here. Obviously the sum of squares cannot be negative whereas the value and the effect value and the coefficient which is nothing but the effect value divided by 2. They may be either positive or negative.

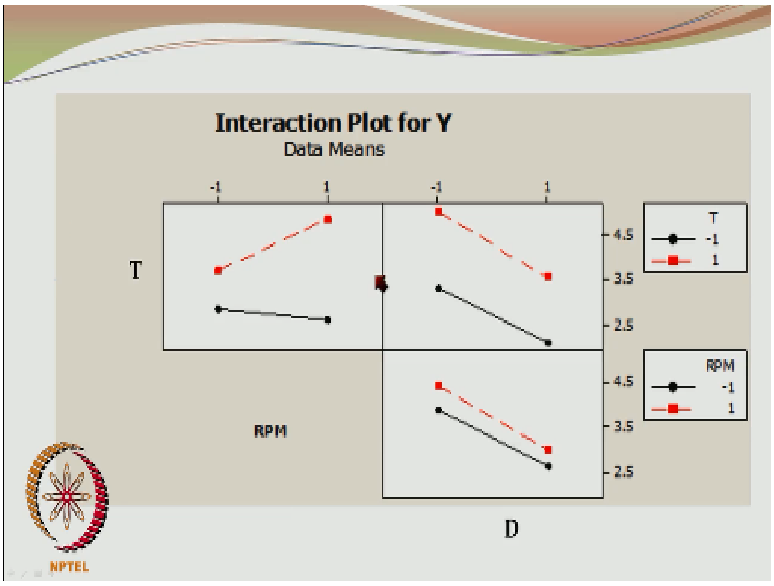
(Refer Slide Time: 43:23)



So this is where the main effects are plotted and they show a pretty consistent and monotonic trend temperature improves the yield RPM improves the yield and particle diameter or particle size reduces the yield here. So if you increase the particle size the yield reduces that is what graph is telling you which is also pretty logical when you increase the temperature the mass transfer kinetics are improved.

And hence you have higher yield when you have more RMP you have higher turbulence and hence the extraction is improved and when you decrease the particle size your interfacial area for mass transfers improves and hence the yield is also increased. So the value seems to be pretty logical here.

(Refer Slide Time: 44:24)



When you look at the interactions this is the interaction between temperature and RPM this is quite interesting at low levels of temperature actually increasing the RPM brings down the yield. At high levels of temperature increasing the RPM increases the yield. This is quite interesting. So depending upon the level of temperature RPM is having a negative effect or a positive effect and this was not picked up by the main factor plots.

Whereas from the other hand the interaction between temperature and particle diameter or particle size and the interaction between RPM and particle diameter or particle size is pretty much negligible. The interactions are not there because the lines are parallel. So when I am increasing the particle diameter irrespective of the level of temperature when I am increasing the particle diameter the yield decreases by the same amount.

So there is not much interaction in these 2 plots. So the error sum of squares. Okay before I go to that let me also tell you something these are only for illustration purposes. Actually the interactions may become a bit complicated to understand when you have more factors for qualitative interpretation these kind of graph are useful, but they seem to be a really so beyond a certain point beyond a certain number of variables.

So when you have more than 3 factors you have to consider the interaction effects between A and B. With the other factors like C and D kept at their average values. So even the computing of the interaction becomes a cumbersome task. S what we need to do in such situation is to rely on mathematical or statistical software and get those results and interpret them.

We can also do the analysis of variance and quantify strictly the effects of the interactions and the main factors. So these are only for qualitative interpretation we have to do the analysis of variance and the F test to really see which interactions are significant which main factors are significant. So this completes the first part of our lecture. We will continue shortly for the second part. Thanks for your attention.