Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

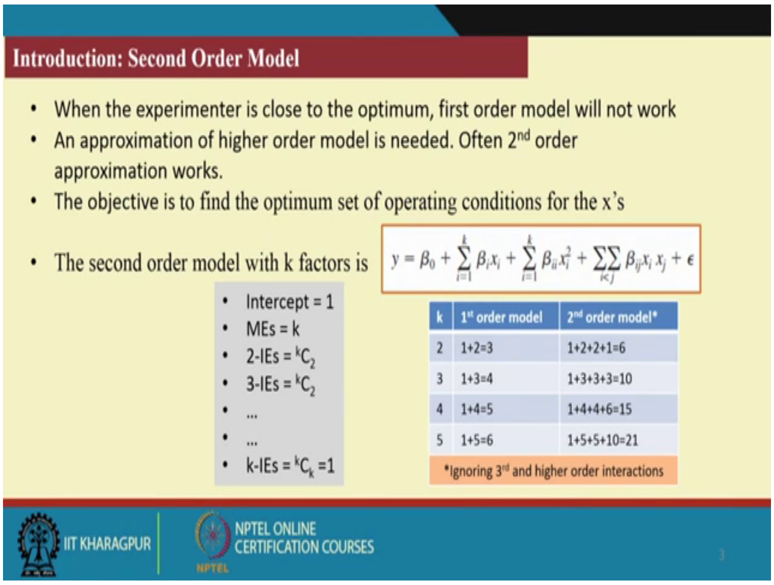
Lecture - 55
Response Surface Methodology (RSM) – Fitting Second Order Model

Hello welcome to the class on Response Surface Methodology. Today we will discuss second order model primarily how to feed second order model; the experimental designs required for data collection through experimentation and then how do you estimate the parameters and check the model adequacy, followed by the second lecture will be on analyzing the second order response surface model.

(Refer Slide Time: 00:57)

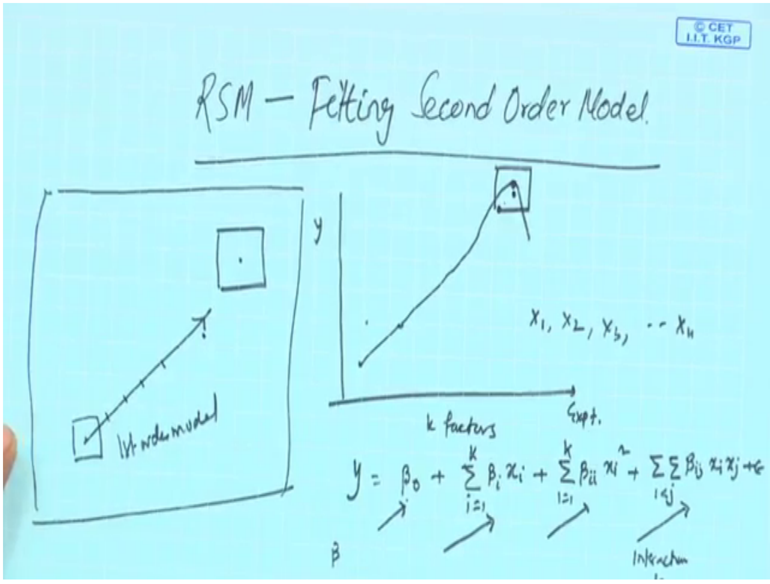
So, content of this class we will introduce the second order model then I will show you the second order response surface designs which we have elaborately discussed in the last class. And then I will show you the how to fit the second order response surface whilst we will end this lecture with an example which we will continue in next class also.

(Refer Slide Time: 01:30)



So, if you recall the last, but one where I have discussed on the sequential experiments.

(Refer Slide Time: 01:40)



That means, so we have the operating zone and we have started with the experimenting somewhere here and the first order experimentation and this will give you that you follow this direction and maybe your that optimum zone lies somewhere here. So, when using the first order model here; first order model here you get the direction of improvement and then use you start doing experiment one after another at different points.

And accordingly if you plot the yield value for example, the y and the your different experiments runs. So, last class we have shown you that the what way the yield value is changing and in this manner we come to some point like this. And then here we say that this is this as curvature; so, in at this point for perhaps this is the zone where the optimizer optimum point lies.

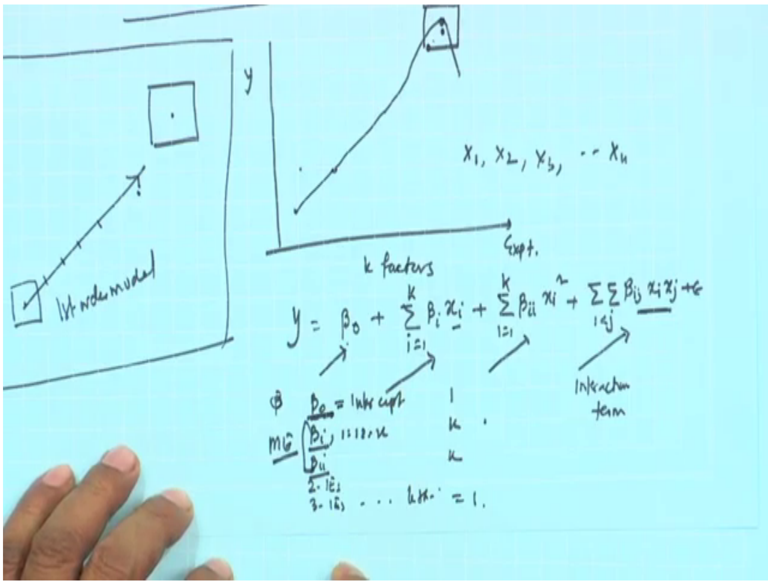
So, this zone is a non-linear zone which is here it is probably quadratic 1. So, at this at this particular operating zone like this one equivalent to this; so, here your first order model will not fit you require a second order model. And second order model then the if I consider the same example that where we have two x independent variable or factors x 1 and x 2 then the second order a first order model we have explained like ok.

Now, let us do the general one suppose that we have k number of factors k factors. So, X 1, X 2, X 3 like X k then your second order model looks like this beta 0 plus sum of I equal to 1 to k beta i x i plus sum of i equal to 1 to k; beta i i x i square. So, this is our main effect; this is the intercept and this is the quadratic effect plus there will be interaction term i less than j. So, beta i j x x i x j plus error term will be there plus error. So, this is your interaction term.

So, by second order model we mean this where there will be intercept, there will be that benefit or the first order effect there will be second order effect there will be interaction effect. Today class is for how to estimate this all beta and then in order to estimate the all the beta parameter parameters what should be the experimental design so, that those many number of parameters can be estimated. So, the because if we should require a experiment where the number of experimental runs should be sufficient enough to estimate all the parameters.

At the same time will not will be we should be cost effective means we will not go for a large number of experiment which is practically impossible.

(Refer Slide Time: 05:34)



So, here then how many what is the what is the number of parameters to be estimated one is beta 0; this is intercept. So, it is number 1 now you will be beta i. So, i equal to 1 to k. So, that mean there are k number of your first order coefficients first order effect basically or other way we can say main effect first order main effect.

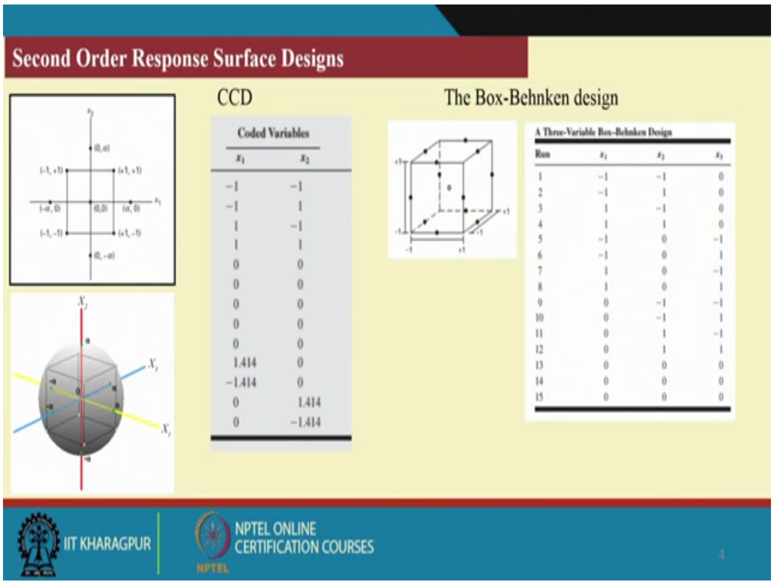
Then beta i i that will be k in number and then you see that you have the interactions interaction will be 2 interactions, 3 way interactions and in this manner there will be k a interaction. So, last one will be 1; so, you will be a large number of effect parameters. So, if I say that beta 0 is intercept beta a beta these are all main effects this is main effects. So, then there will be so, many interaction also there will be so, many interaction effects.

So, let us see that the difference in first order and second order model with reference to different factors. If you go for first order model which is beta 0 plus beta x i then when k equal to 2 number of factor coefficients is 3 or number of parameter to be estimated is 3 and in that time what will happen in the second order model. So, there will be one intercept main first order effect, quadratic effect plus interaction effect. So, you require to estimate 6 number of parameters if k equal to 3; in first order model number of parameters to be estimated is 4 whereas, number of parameters to be estimated in case of second order model it is 10.

Now, if you are number after number of factors is 5, then you see in first order model you require 6 number of parameters to be estimated, but in the second order model 21 number of parameters to be estimated please remember we are ignoring third and higher order interactions for the second order model. That means, only up to check to a second order interaction if we consider this is the number of parameters to be estimated.

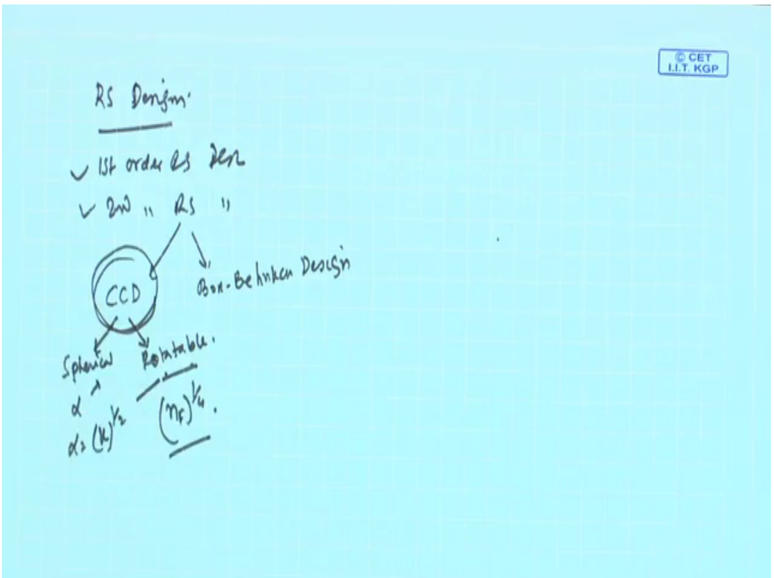
So, as the number of parameters to be estimated gross significance grows significantly; what is required? You require a different kind of design so, that you will have sufficient number of information and accordingly you will be able to estimate beta 0, beta i, beta i i and beta i j.

(Refer Slide Time: 08:40)



So, if you if you recall my last lecture there I said that in the response surface design response surface design.

(Refer Slide Time: 08:50)



So, we have given you the first order design first order response surface design and also we have given you the second order response surface response surface design. So, in second order response surface design we primarily say that central composite design and Box Behnken design.

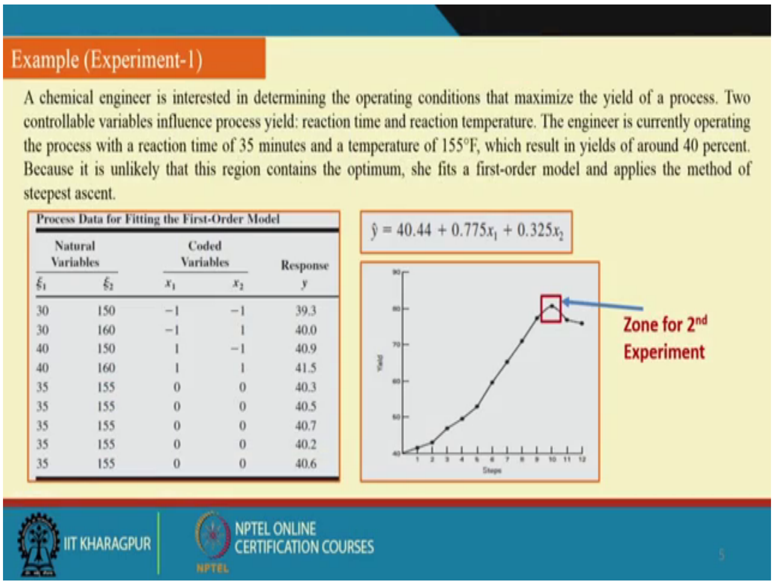
And again in central composite design we say that is spherical and rotate rotatable. So, depending on the value of alpha; so, alpha will determine that whether it is face spherical or rotatable. And in spherical case we say alpha equal to k to the power half and whereas, here in rotatable gave n f to the power 1 by 4 and we I also shown you that that there is not much difference in case of spherical and rotatable; particularly for different case and they are more or less equivalent in the sense a spherical one is also a minimum equal variance ad from the as from at a distance equal distance from the center of the surface anyhow those things are known to you.

Now, let us say see that which one let us see the central composite design with two factors and that part is that one looks like this. So, central composite design here it has factorial points; your central point and axial points. And then depending on the value of alpha it will be CCD or editable, but here I have showing you a 2 to the power basically two factorial central composite design where there are 4 factorial points 1, 2, 3, 4, 5; 5 central points and 4 axial point where alpha equal to root over 2. So, that plus minus root over 1 ok. So, if you adopt this design; so, how many experiments you are conducting?

1, 2, 3, 4 4 plus 5; 9 9 plus 4 13 number of experiments you are conducting. So, if you go for a two factor case. So, what is the second order model requirement 1 2 to put 6 number of parameters to be estimated.

So, here if we consider see that what are the distinct point; distinct points are here 4 factorial point, one center point another 4 axial points. So, from this figure you see 1, 2, 3, 4, 5, 6, 7, 8 plus central point 9. So, 9 unique points are there we require 6 number of parameters to be estimated. So, number of observation is sufficient to estimate the parameters including errors. So, in case of 3 factor the case you may go for box Behnken design that also we have discussed in last class ok.

(Refer Slide Time: 12:44)



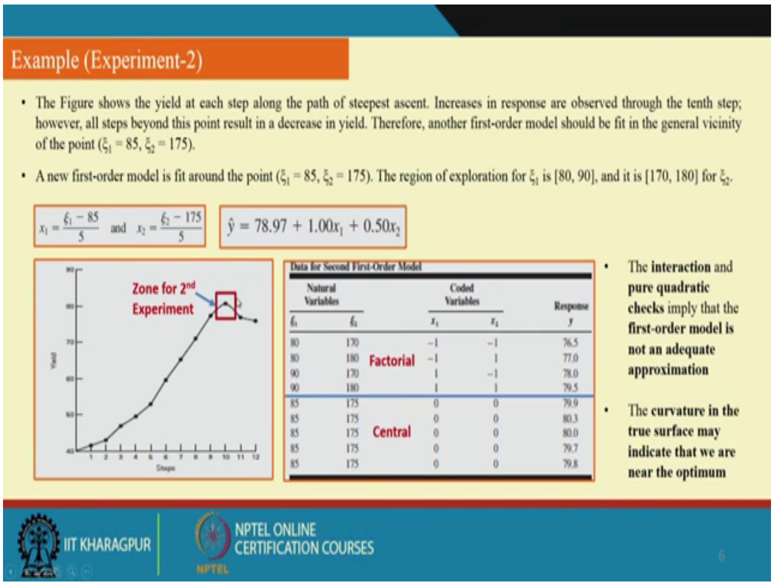
So, for the time being what we will do we will use the central composite design and we will bring the same example what we have discussed in last burst one lecture when I when we were dealing with first order response surface.

So, let me repeat a chemical engineer is interested in determining the operating condition that maximize the yield of a process. Two controllable variables you know influence process yield or reaction time and reaction temperature and that is the natural variable reaction temperature and then the engineer is currently operating with the process of reaction time 35 minutes and temperature 155 degree Fahrenheit which result in yields of around 40 percent because it is unlikely that this region countries the optimums if it is the first total model that applies the method of steepest ascent.

In fact, the entire thing though all the steps procedures with reference to this example we have discussed in last, but one class. And I have also shown you that that the new zone of experiment will be this one and early and the experiment what we have done with this operating zone and we got this degree first order regression equation. And then we have followed the path of steepest ascent and really we landed here that this is the place zone where the second explore second set of experiment will be will be conducted.

So, let us see that here at this point if we conduct a second order second experiment. So, what we will do ?

(Refer Slide Time: 14:35)

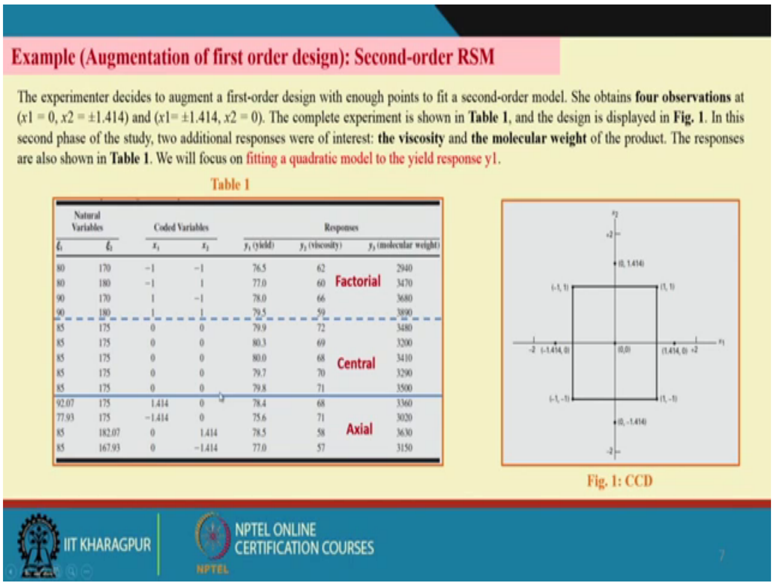


Here what happened at this point again; we thought that there may be there may be your first order relationship. So, as a result at this point with new your factorial points and new central point the experiment was conducted you see that at this place the factorial points are 80 and 90; 80 and 90 and 170 and 180, 170 and 180 and the central point is 85, 175, 85, 175.

So, this is one 80; 170, this one is 90; 170 and this one is your then 80; 180 and 90, 180 and this is 85, 175 like this with reference to this with reference to this diagram not with reference to x and y axis here. We have adopted factorial and central point here and in the same manner, we have fitted the first order model and the first order model feet when we have done we found that the interaction and pure quadratic chicks imply that the first order model is not an adequate approximation.

So, I am not explaining here further that why first order model is not fit the calculation we are not giving. Because the similar calculation we have we have given you in the in the first order model you just repeat the same procedure and with this data what you will find out when you check the first order into that is pure quadrating and the interaction effect; you found out that they are they are significant and as a result what happened first order model will not work. So, the curvature in the two surface may indicate that we are near the optimum like here the curvature is there so; that means, you require a second order response surface.

(Refer Slide Time: 16:42)



Natural Variables		Coded Variables		Responses		
ξ_1	ξ_2	x_1	x_2	y_1 (yield)	y_2 (viscosity)	y_3 (molecular weight)
80	170	−1	−1	76.5	62	2940
80	180	−1	1	77.0	60	3470
90	170	1	−1	78.0	66	3680
90	180	1	1	79.5	59	3890
85	175	0	0	79.9	72	3480
85	175	0	0	80.3	69	3200
85	175	0	0	80.0	68	3410
85	175	0	0	79.7	70	3290
85	175	0	0	79.8	71	3500
92.07	175	1.414	0	78.4	68	3360
77.93	175	−1.414	0	75.6	71	3020
85	182.07	0	1.414	78.5	58	3630
85	167.93	0	−1.414	77.0	57	3150

Now, to get the second order response surface means by saying to get the second order response surface means to feed the second order response surface; what do you require? You require more number of observations, otherwise you will not be able to fit estimate all the parameters.

For example, if you do these 4 factorial points and one central point then what happened there are 6 distinct observations; 6 distinct of observations in the says at the central point whatever 5 observation there at the same point we are treating them as a single distinct observation. But you require 6 number of parameter to be estimated; so, here number of parameters all parameters cannot be estimated also error cannot be estimated. So, what we have adopted here that we have adopted the central composite design with factorial point, central point and axial points. Please note that this ultimately this example I have taken from the montgomery design analysis of experiment book and in fact, here 3 responses are included.

But we are interested in the yield only for this explanation purpose here; the viscosity and molecular weight these two also other two responsibility variable which was of interest for the experimenter. But for understanding the second order model point of view we are not interested to discuss anything about viscosity and molecular weight. So, as a result what I mean to say that I if I write that this part a natural part coded variable for an yield is y; then you see here there are 4 factorial points and 4 axial points and one central point.

So, as I told you that 9 distinct observations are there impact at the central point there are 5 number of observations which helps us to estimate the error. So, this is what is the data you got and this is what is the experimental design; based on this experimental design this coded variable values are like this and your y value response value the yield value like this.

So, with this data we can fit a second order response surface; to fit the second order response surface we are using the coded values, we are not using the natural variables values. So, but keep in mind that when you conduct the experiment you say the process with the natural values, when we analyze the experimental data instead of natural values we have taken the coded values and app and then we fit the models ok.

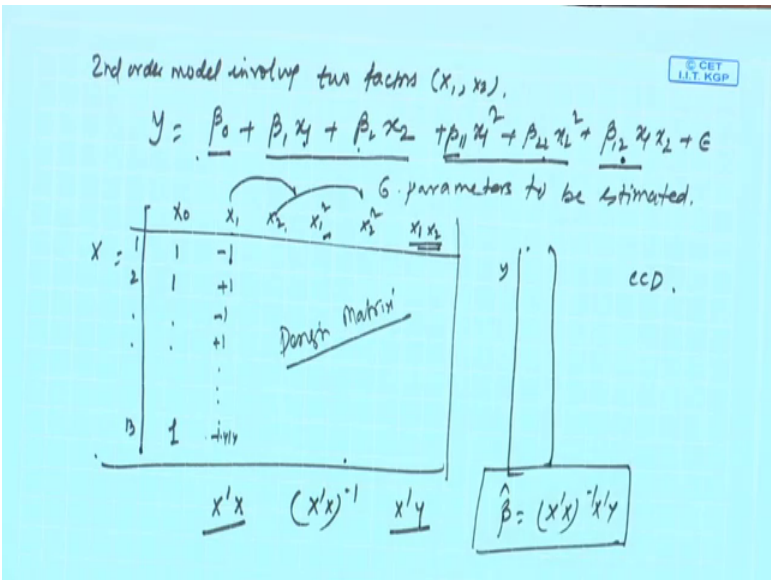
(Refer Slide Time: 19:35)

x_1	x_2	x_1^2	x_2^2	x_1x_2	yield
-1	-1	1	1	1	76.5
-1	1	1	1	-1	77
1	-1	1	1	-1	78
1	1	1	1	1	79.5
0	0	0	0	0	79.9
0	0	0	0	0	80.3
0	0	0	0	0	80
0	0	0	0	0	79.7
0	0	0	0	0	79.8
1.414	0	2	0	0	78.4
-1.414	0	2	0	0	75.6
0	1.414	0	2	0	78.5
0	-1.414	0	2	0	77

	1	-1	-1	1	1	1
	1	-1	1	1	1	-1
	1	1	-1	1	1	-1
	1	1	1	1	1	1
	1	0	0	0	0	0
	1	0	0	0	0	0
X=	1	0	0	0	0	0
	1	0	0	0	0	0
	1	0	0	0	0	0
	1	1.414	0	2	0	0
	1	-1.414	0	2	0	0
	1	0	1.414	0	2	0
	1	0	-1.414	0	2	0

So, this is what is our model.

(Refer Slide Time: 19:42)



Our model is y equal to beta 0 plus beta 1 x 1 plus beta 2 x 2 plus beta 1 1 x 1 square, beta 2 2 x 2 square plus beta 1 2 x 1 x 2 plus epsilon. So, this is our second order model involving two variables; two factors X 1 and X 2.

So, this is for first order main effect, this is for second order main effect, this is for interaction; so, this is for intercept. So, 1, 2, 3, 4, 5, 6 parameters to be estimated to be estimated; so, what is your design matrix? Design matrix is X; So, what you require? You have you have to write it first X 0 that is your per beta 1, then your X 1 for beta 1, X 2 for beta 2, then X 1 square for beta 1 1, X 2 square for beta 2 2 and X 1, X 2 for beta 1 2 ok.

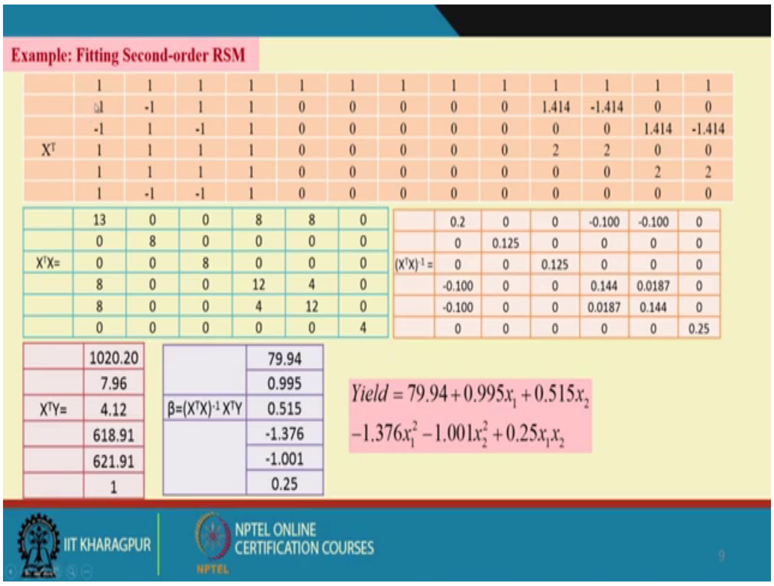
So, this now how many observations we have? We have you CCD we have you CCD here CCD with 4 factorial points 4 axial points and 1, 2, 3, 4, 5 central points so; that means, 4 plus 4 8 plus 5 13. So, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13; so, you have 1, 2 like this the 13 number of observations. So, for x 0 all will be 1 for X 1 that is minus 1, then plus 1, then minus 1, then plus 1.

So, like this the last one is your minus 1.414 similarly X 2 similar then you square the X 1 row column you will get X 1 square; square X 2 square you will get X 2 square. So, like this you will be you will be and then multiply X 1 and X 2 you will be getting X 1 and X 2 column. So, then what is the resultant column? Resultant column is this. So, first column will be all 1, second column will be minus 4 minus 1 minus 1 minus 1, plus 1, minus 1, plus 1. So, like this what will happen ultimately you will get this kind of this kind of a design matrix. So, this is known as design matrix ok.

So, you just verify and if there is any mistake you point out in the in the discussion forum , but this is the procedure first you find out the design matrix; once you have the design matrix then you also have the y values y the 13 y values you have.

So, what do you require? You require X transpose X to be computed X transpose X inverse to be computed X transpose Y to be computed; then finally, beat out cap will be X transpose X inverse X transpose Y that is what we have we have seen in regression lectures.

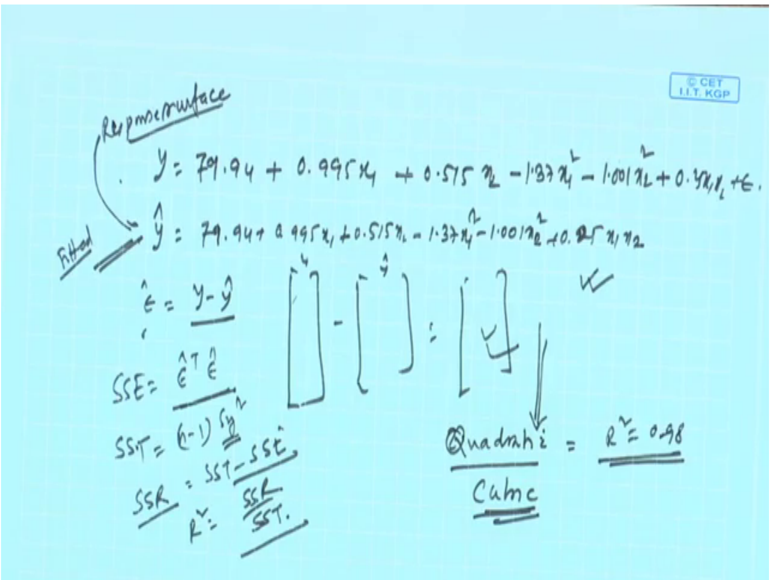
(Refer Slide Time: 23:49)



Exactly the same thing we have done here first found out the X; that X transpose then X transpose X, then you see that X transpose X inverse, then X transpose Y and beta is X transpose X inverse; X transpose y this will give you this value. So, intercept is 79.94, x 1 with 0.995; x 2 0.515 then x 1 square minus 1.376, x 2 square in that is minus 1.001 and with x 1, x 2; 0.025 what do I mean? I am saying these are the beta estimate we will write beta 0 plus beta 1 x 1. So, beta 1 in 0.995; beta 2 is 0.515, beta 1 1 is minus point 1.37, beta 2 2 is minus 1.007 and beta 1 2 is point 0.025.

So, this is your predicted pretty fitting value or fitted regression line.

(Refer Slide Time: 25:11)



So, if you one the that regression equation then it is y equal to 79.94 plus 0.995 x 1 plus 0.515 x 2 minus 1. 37 x 1 square minus 1.001 x 2 square plus 0.25 x 1 x 2 plus error will be there.

So, this is your regressions response surface. So, you view one fitted value it will be like this 7 only error term will not be there 0.995 x 1 plus 0.515 x 2 minus 1.37 x 1 square minus 1.001 x 2 square plus 0.25 x 1 x two. So, this is my responsible this is my this is what it is fitted one fitted response this is fitted response surface ok.

So, now when you fit the second order model your immediate work is whether the model is adequate or not and next is whether the parameters estimated are significant or not. So, let us do this.

(Refer Slide Time: 26:40)

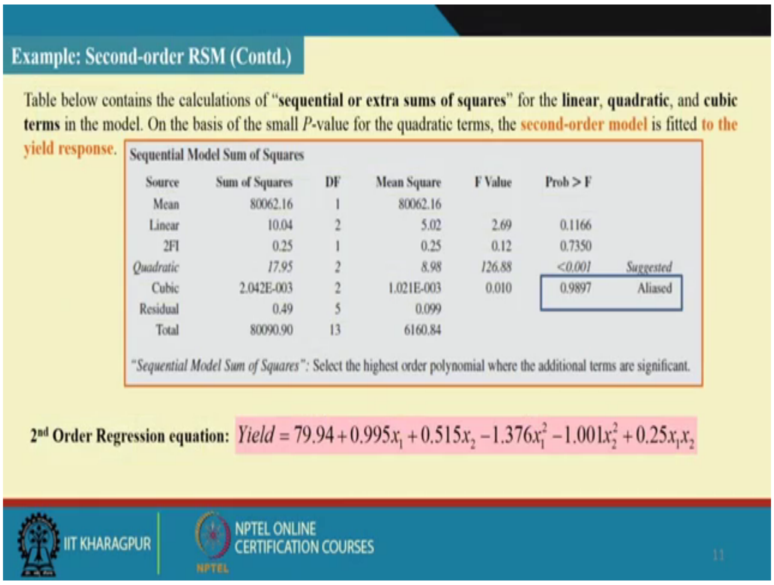
Regression Statistics								
Multiple R	0.991327734							
R Square	0.982730677							
Adjusted R Square	0.970395445							
Standard Error	0.266290253							
Observations	13							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	5	28.24670343	5.649341	79.66860702	5.14703E-06			
Residual	7	0.496373494	0.07091					
Total	12	28.74307692						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	79.93995461	0.11908862	671.2644	4.3003E-18	79.65835477	80.22155444	79.65835477	80.22155444
X_1	0.995050253	0.094154931	10.56822	1.48449E-05	0.772409219	1.217691286	0.772409219	1.217691286
X_2	0.515202796	0.094154931	5.471862	0.000934011	0.292561762	0.737843829	0.292561762	0.737843829
X_1^2	-1.376449283	0.100984169	-13.6303	2.693E-06	-1.615238897	-1.137659668	-1.615238897	-1.137659668
X_2^2	-1.001335998	0.100984169	-9.91577	2.26204E-05	-1.240125613	-0.762546384	-1.240125613	-0.762546384
X_1X_2	0.25	0.133145127	1.87765	0.102519191	-0.064838196	0.564838196	-0.064838196	0.564838196

So, using the that regression procedure what you require to do; here you first find out epsilon cap which is y minus y cap so; that means, you have the 13 observation for y minus 13 observation for y cap then it will give you the 13 epsilon cap value then you find out SSE, SSE will be epsilon cap transpose epsilon cap when this one it will give you a SSE; then you will you will get SST from y only this is nothing, but n minus 1, is y square 1 is s y say square is the variance of this that you know how to compute.

So, then SSR that is basically the regression or model part you will be getting s s t minus SSE. So, in this manner then r square is basically SSR by SST. So, in this manner when you compute you will get this one you see that R square value is 0.98 and adjusted R square is 0.97 and it is quite significantly high; so, that in model is fit.

Now, if you see the ANOVA and ANOVA also the F test talks that that is a very high value 79.66. So, that we model is perfectly fit and now if we see the interest that the parameter estimated value that is 79.94, 0.995, 0.515 minus 1.37 minus 1.001 all those things with the standard error and if you see that t value; t value is quite high except X 1 X 2, but even then it is also 1.87 ok. So, these are very high values so; that means, they are significant and if you see the P value if you see the P value you see the P value is very low and only X 1 X 2 a significant at point one probability label. So, a point 0 probability label we can say the interesting term is not significant

(Refer Slide Time: 28:57)

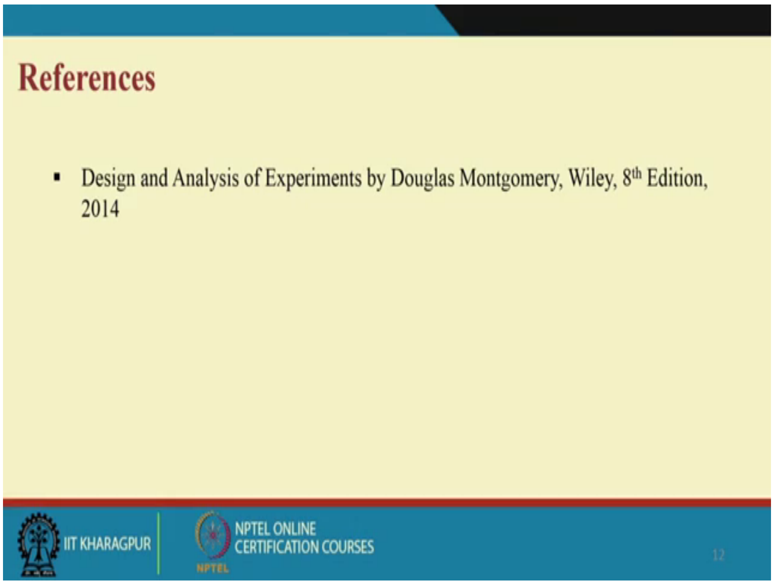


So, we can say that the second order response surface is a feet one here. Now question is that then you may say that why will not go for y will go for only second order why not cubic ? So, we have gone for quadratic why not cubic? It may give even better result now quadratic R square is 0.98. So, it is already quite large very high. So, cubic I think we do not require , but what happened if you use cubic one then your number of parameters to be estimated will be even more so; that means, you require cubic part X 1 cube X 2 cube.

So, ultimately what will happen if you go for cubic one you will find out that there will be some alias structure means some parameters cannot be estimated completely. So, and with your if I say the probability value is also you see that cubic one 0.98 F value is very low. So, that mean cubic part is not there.

So, as a result; so, we can say that second order model is a fit model and this is what is for today.

(Refer Slide Time: 30:26)



And that I want to say; what I mean what that what I wanted to tell you it is clear to me I hope that you understood that what when; what you will do when you want to fit a second order response surface, what kind of designs you employ for experimentation once you get the data how do you analyze the data aging regression and get the parameter estimated; you check the model adequacy, check the parameters of the models whether they are significant or not.

Once it is done your response surface is ready second order response surface is ready. Now this response surface will be used or will be analyzed properly and to see that the point where we have done the second order experimentation; for example, for example, here whether at this zone each your optimum zone or not that will be known through analyzing the response surface. So, there are two modes or two ways we can analyze using contour plot and using canonical analysis.

Next class I will discuss contour plot based analysis and canonical analysis base inter analysis as well as interpretation.

Thank you very much.