Course material

Course:

ENG606 / PHYS 442

Video:

DOE_lesson10_part2_LackOfFit

Concepts (extracted from automatically generated subtitles):

Complex model. Best point. Design of experiment. Linear model. Small line. Second degree. Variance of the point. Average of those points. Type of model. Root b-bowl. Surface response design. Factorial design. First degree. Border of the domain. Center.

to video sequence search
(within ENG606 / PHYS 442.)

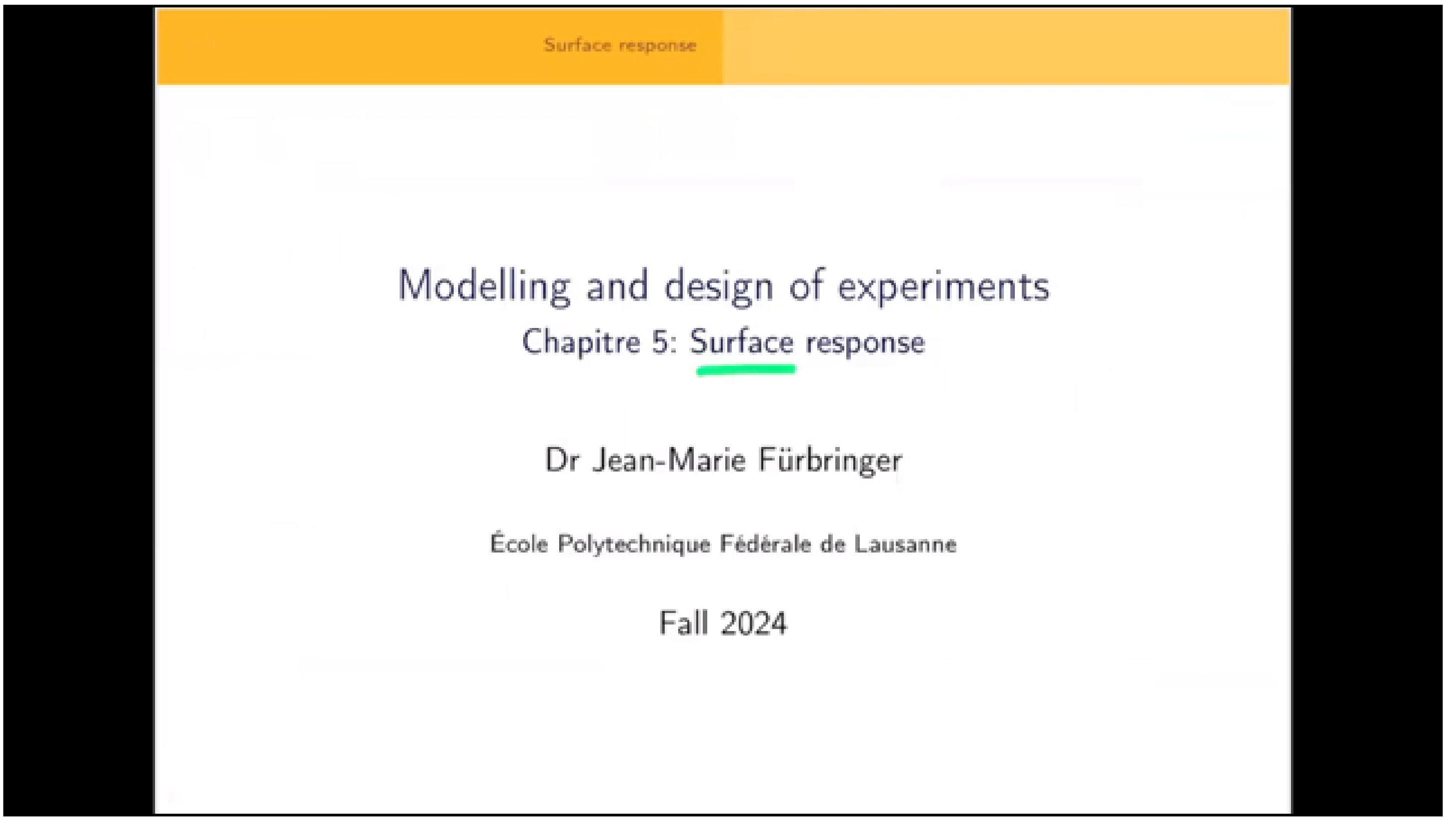
to video

...

notes

summary

0m 0s

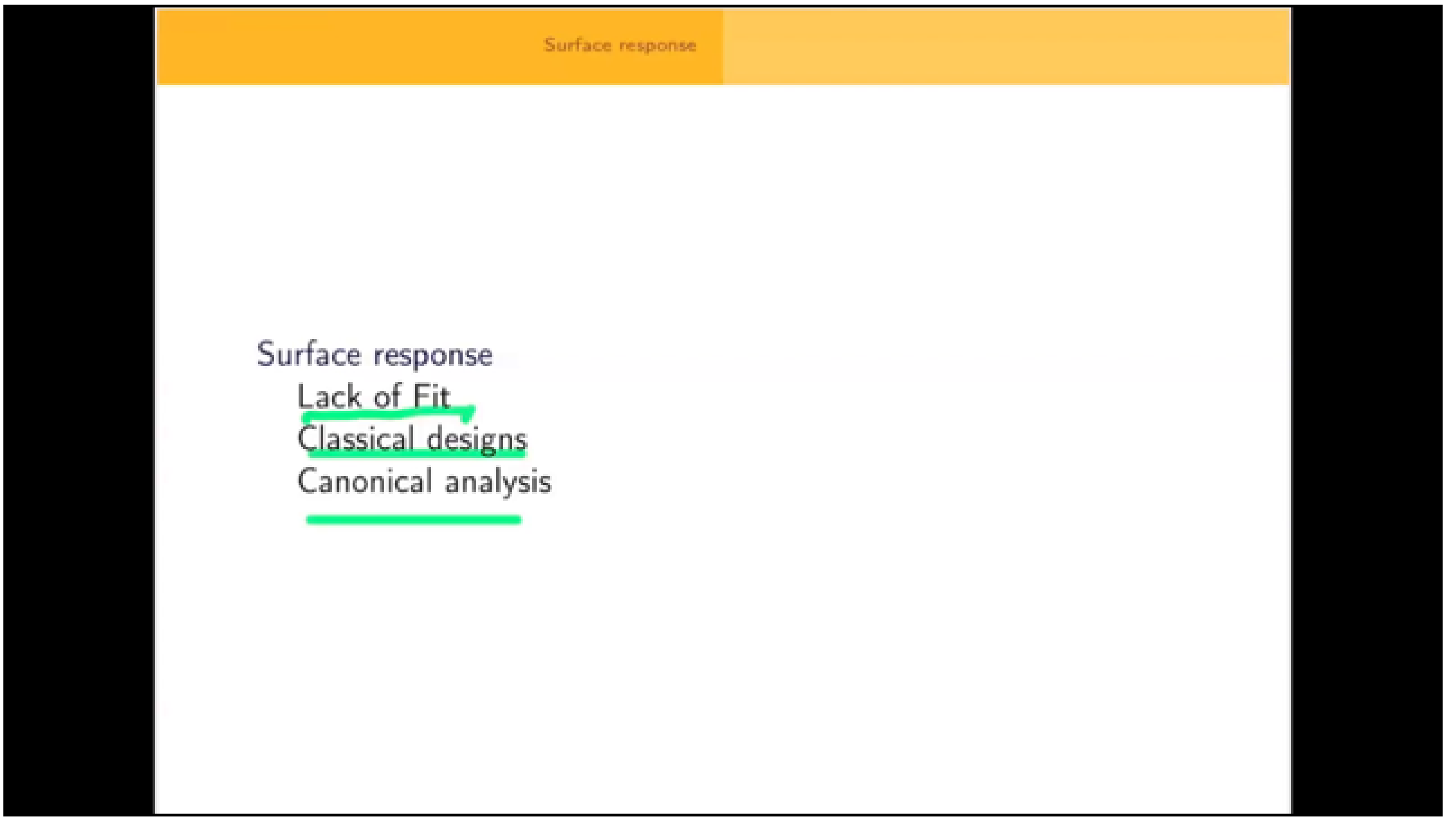


These subtitles have been generated automatically So now we talk about surface responses. That means that you imagine even if you are in

notes

summary

0m 1s

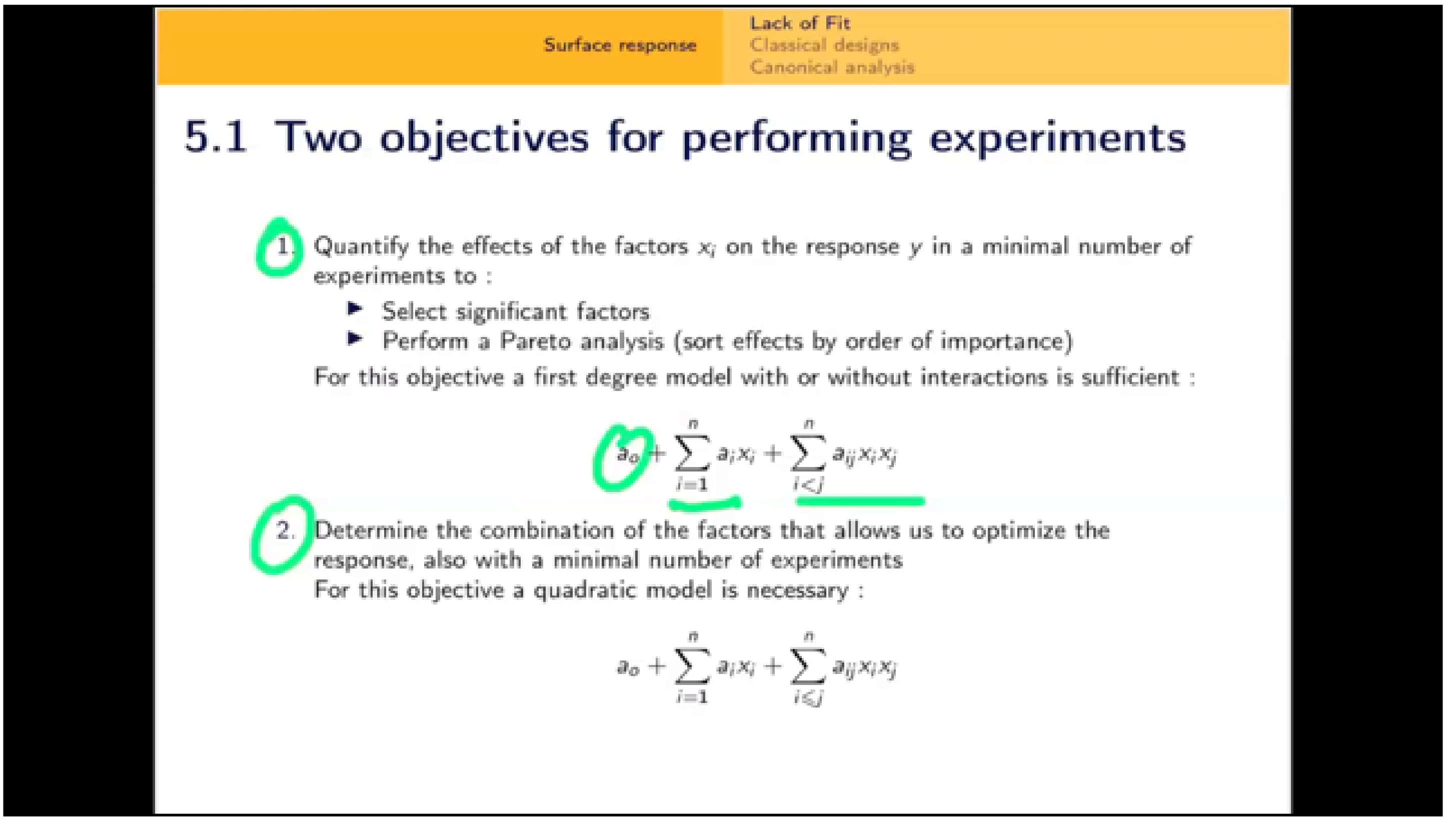


more than two factors, in two factors it's evident you have two factors and you are looking for a surface. Now we are in the upper surface, we can have three, four, five, seven factors and we are looking to the equivalent in the n plus one dimension what could be the answer and that could we could visualize this answer as a surface. We would like to understand where we have the minimum, where we have the maximum, where we have the stability is mainly what we are checking when we are looking surface response design. So we will see let's say three elements. I will show you that sometimes that it's most costly, sometimes we would like to be sure that we need to go for a second degree. So we have one element for that we call it lack of fit. It's a possibility to evaluate if we really need to go from the first degree with instructions to a second degree. The tool work also between main effects and instruction. It could also work for second degree and third degree etc. But it's a way of evaluating if within the data that you have you have some elements that telling you you better have a more complex model. After I will present you some classical design if you and I will finish the chapter with something we call the canonical analysis. Okay, when you have two factors even three factors you can make plots and you can understand where where is your maximum, where is your minimum in which direction you better go for maximizing, minimizing, stabilizing your processes. But when you have five factors, six factors impossible to really understand and just looking at the coefficients

notes

summary

0m 5s



it's not sufficient for understanding what is the shape you are confronted with. So the canonical analysis will do that will tell you what is the geometry that we have and in the second degree we just have two geometry. We have the ellipsoid so the root b-bowl and you would like to understand what is the orientation of the axis of your root b-bowl or you have the hyperboloid. So I used to say like China's bowls and you are was the same thing you will try to understand what are the directions in which so with hyperboloid some directions you are increasing your value and some other you are decreasing. In the root b-bowl the expansion is already provoking the same thing. Also increasing imagine that you have, well it's more decreasing is more easy you have the gravity which is diminishing when you are going far away of the mass or you use the heat if you have a center which is with a lot of heat when you are going away of this center you can have the temperature decreasing. You can also imagine very difficult to imagine if you have heat at the outside you can see are going to the center you are decreasing. So going to outside you are increasing temperature. When we are performing design of experiment we are used of having two types of objectives. One is that we would like to understand what are the main effects what is important to manage. We are with this idea of the Pareto analysis. So the type of model on which it's very interesting to fit our data is having a constant for taking out everything which is the same for all the factors having the main effect and having instructions and after you could decide how many level of instructions you want but usually with one level of instructions, instructions 2x2 is

notes

summary

2m 25s

5.1 Two objectives for performing experiments

1. Quantify the effects of the factors x_i on the response y in a minimal number of experiments to :
   - Select significant factors
   - Perform a Pareto analysis (sort effects by order of importance)

   For this objective a first degree model with or without interactions is sufficient :

$$a_o + \sum_{i=1}^{n} a_i x_i + \sum_{i<j}^{n} a_{ij} x_i x_j$$

2. Determine the combination of the factors that allows us to optimize the response, also with a minimal number of experiments

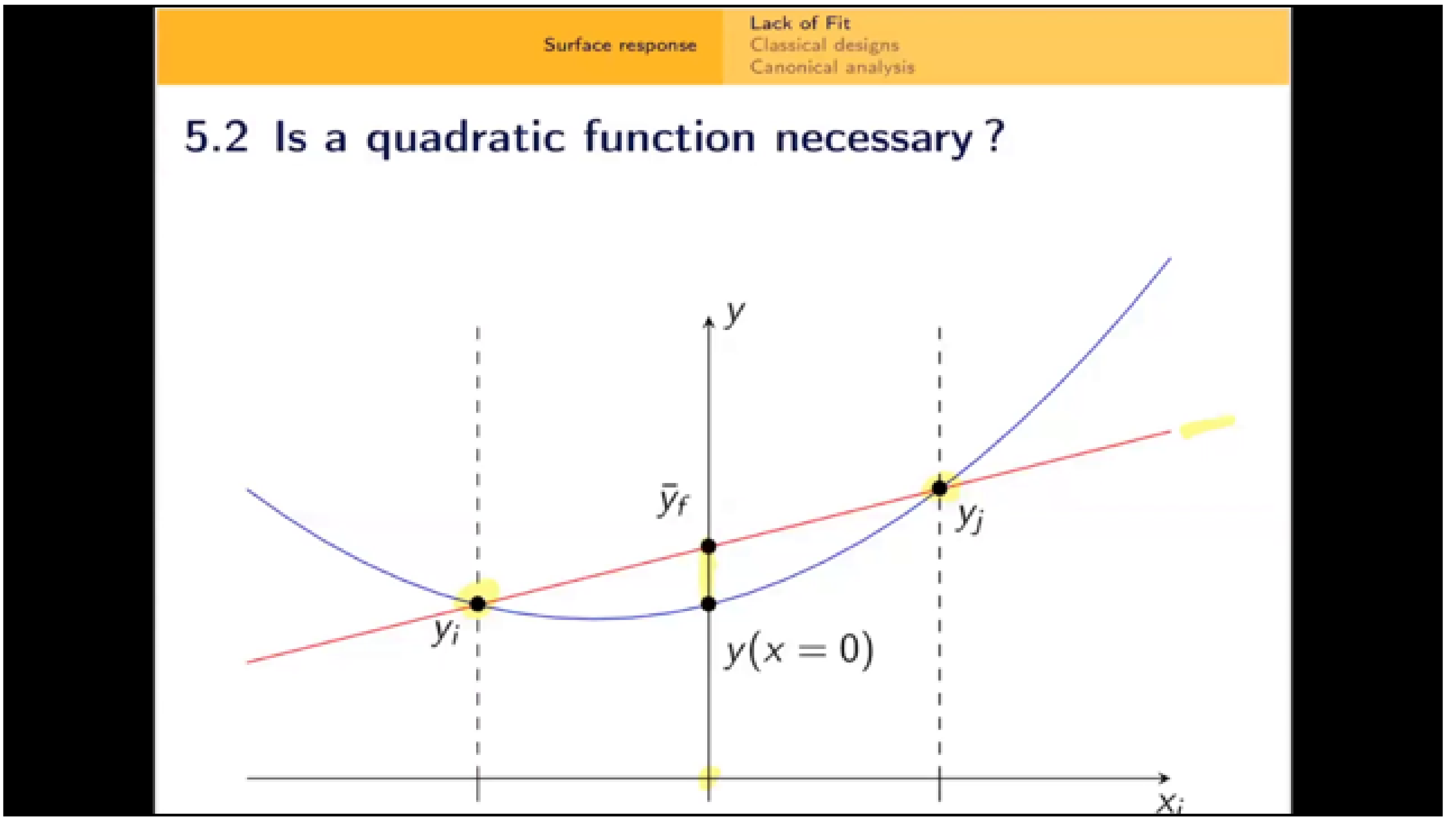
   For this objective a quadratic model is necessary :

$$a_o + \sum_{i=1}^{n} a_i x_i + \sum_{i \leqslant j}^{n} a_{ij} x_i x_j$$

quite sufficient for understanding what's happen and taking the most important things with you but sometimes you are more interesting by

notes

summary



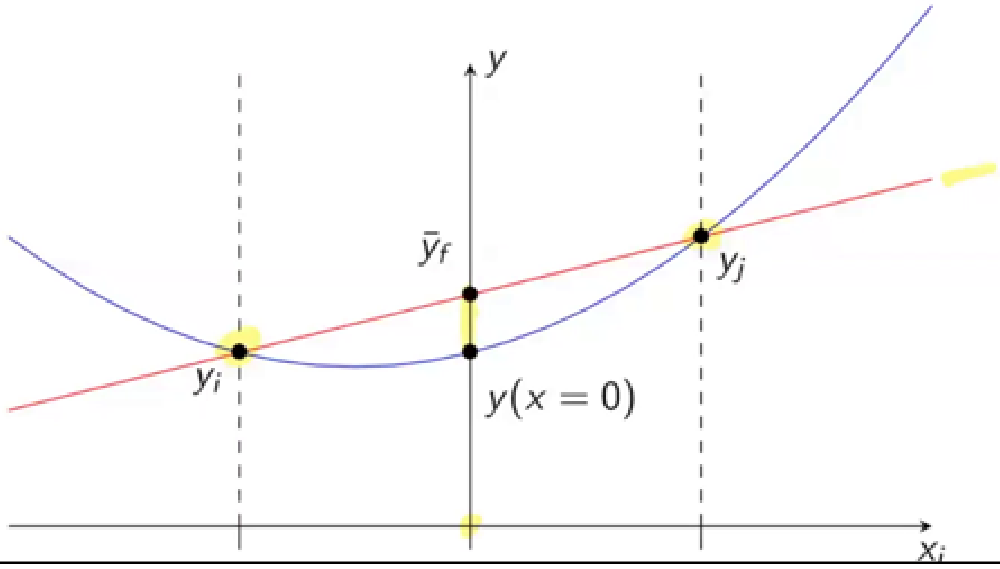
optimizing you are managing a process and you would like to use the process in this situation for your factor where you have the maximum yield. So for that it's just a small line which is no different here you have bigger or smaller a smaller than j now it's just one different smaller or equal and this makes all the difference. When you have this difference in your model so you start having heels and having valleys you can have a series of heel you can have a series of valleys. So the design we are using are for now detecting where are the heel where are the valleys. So first question is necessary because it's more expensive to work like that. Now we have been working making experiment only at the border of the domain and it was sufficient for understanding what was the interest of going in one direction on the other thinking that in between the evolution was linear. If it's a quadratic answer and even more if it's more degrees in your model now it's different what's happened within the domain became critical and you need to enter the domain to understanding really what's what's happening. So you see with those two models the red one which is a linear model and the blue one which is a quadratic model which we just do measurements at the extreme we are not able to differentiate these two models. So we really need to make a measurement and the best point for making a measurement for checking that is at the center and what will tell you if you prefer the linear model linear response model to the quadratic model is in fact this distance between the value at the center and your a zero value if you have used balanced design. So the first criteria for saying you are happy or not with that is your

notes

summary

4m 49s

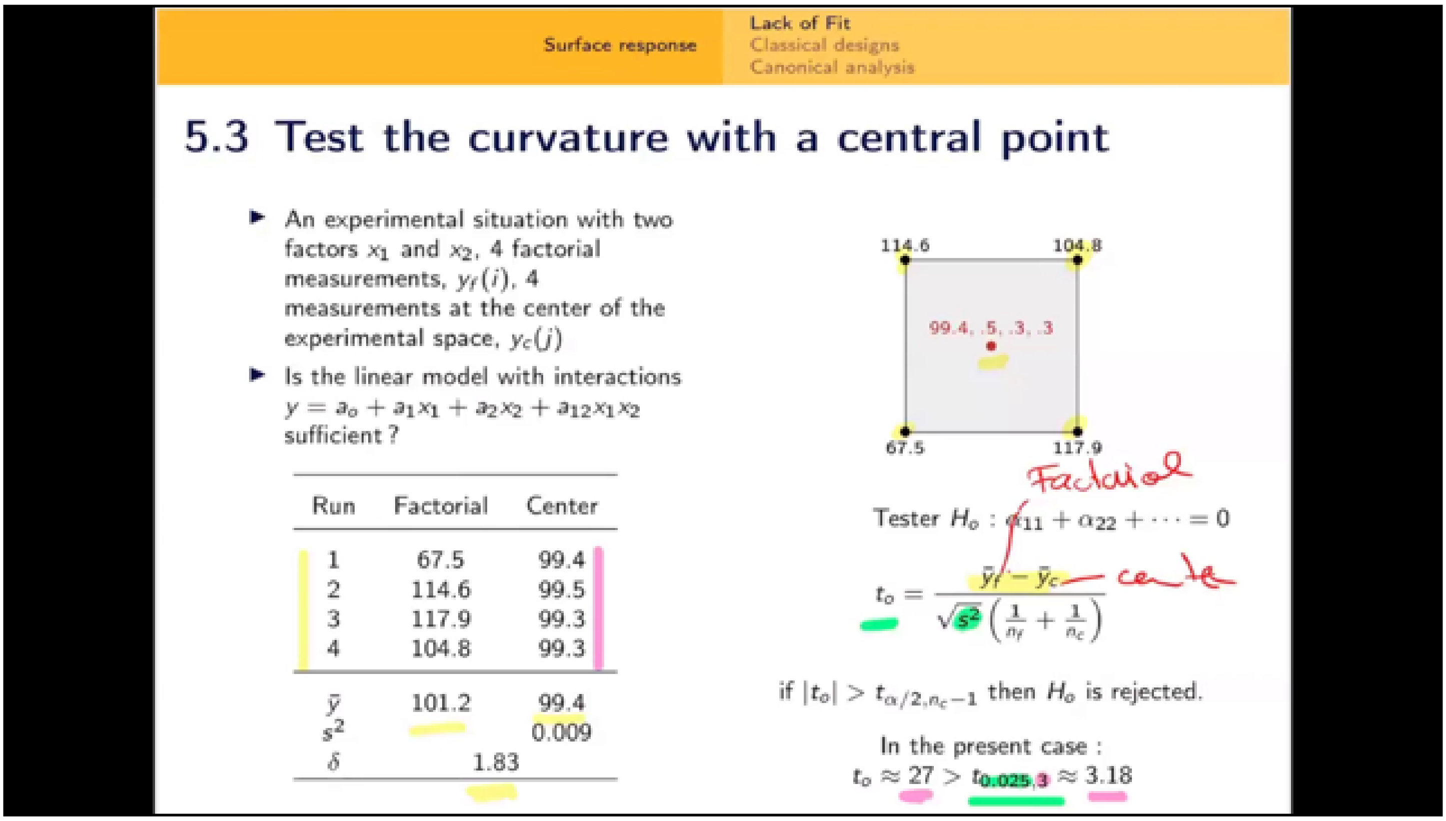
5.2 Is a quadratic function necessary ?



experience

notes

summary

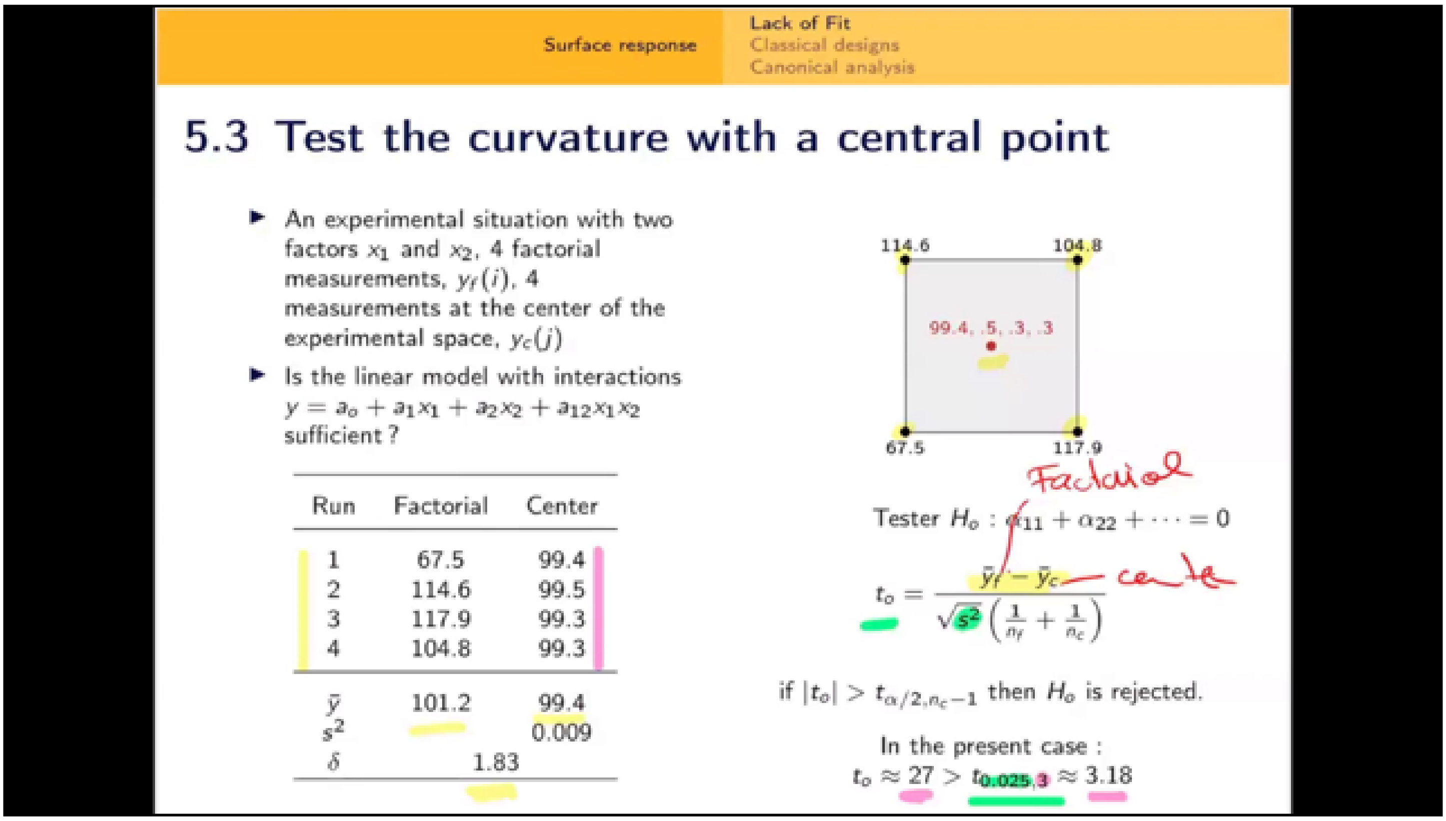


your competence in your domain is this difference at the center of my domain sufficient for staying in the sufficient small for staying with linear response model because the change is very the curvature is very low for me or do I need the blue model do I need the more expensive model. So a good check if you have made already experiment for a linear response you make an experiment at the center and you decide what is the difference between a zero and this measurement at the center this could be a first criteria making an experiment at the center. So usually when I'm performing factorial design or fractional factor design all the time I keep time and money for making an experiment at the center because I just would like to check if I have a curvature and this curvature is sufficient for let's say invalidating my linear approach and I need to go through for a second degree. Nevertheless there are more statistical way of doing that. So imagine the situation we have made measurement at the corner of my domain I have a very simple situation with two factor only and I have made four experiment at the center it's interesting why to make four experiments at the center so I have repetition so it's give me an information about the stability of my measurements the repetition that I start making an experiment at the center I start making a few experiments a factor experiment I redo an experiment at the center like that I can also check if my process is constant since I have a repeatability I'm finishing my design and I'm finishing my point at the center. So now which is interesting today is the same number I have four points on the border and I have four points at the center so I'm able to calculate the average of those points

notes

summary

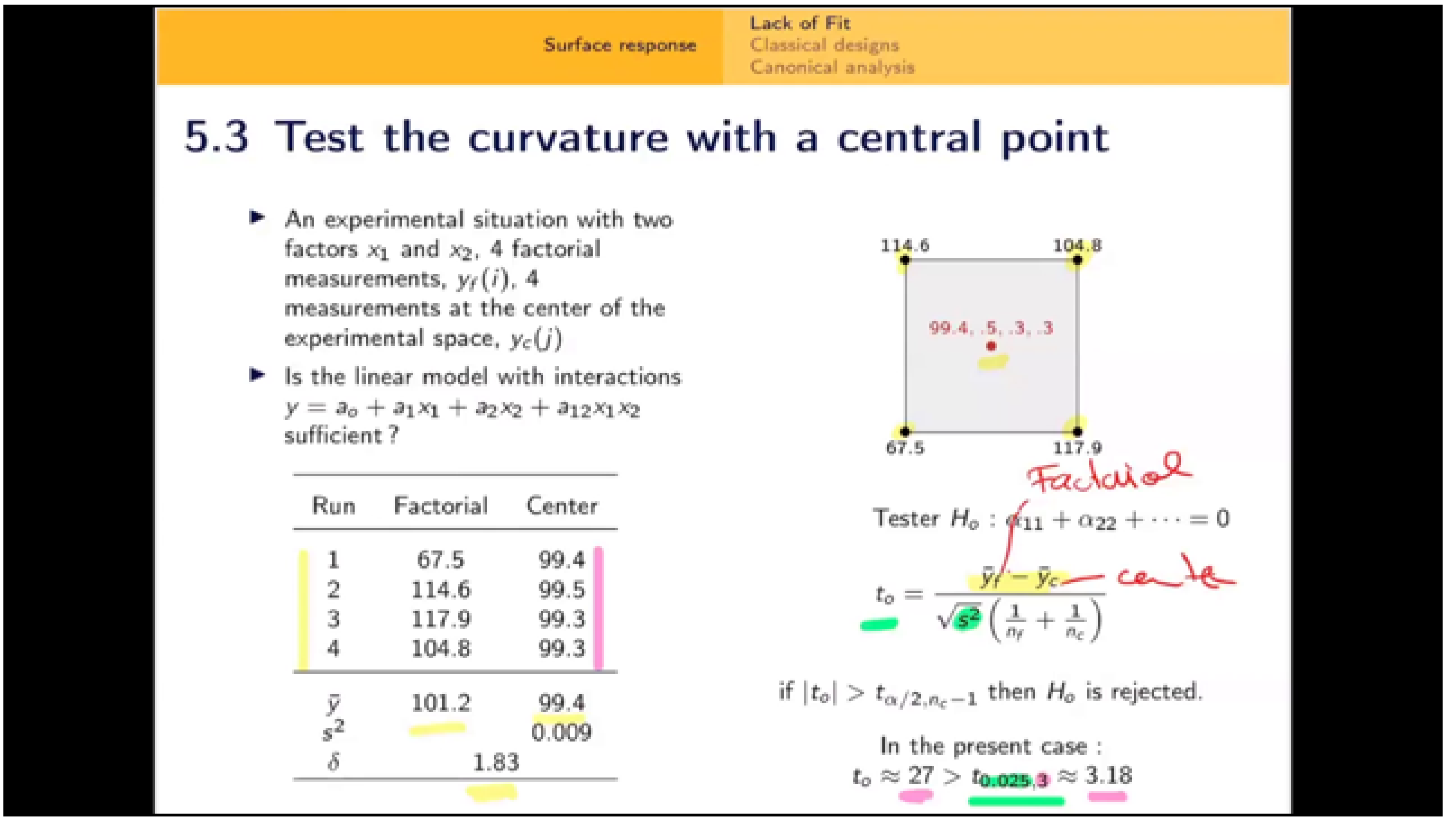
7m 13s



and comparing those average because I have a center point and my factorial point are at the center so if it's a linear response model fit well the average should stay the same if I have a difference it's because I have a curvature or some measurement error so it's all times the problem that I have to balance between the possibility of having errors in my measurements and having a non-linear behavior so in this case if you do the average between the factorial points you get 101.2 and if you get the average of the center point to get 99.4 is the variance of the point that I have the center and the difference between my two average sufficient for telling me if I have a curvature or not so we can make a student test with that so the bar means average so y bar f is for factorial so this was for factorial and this was for center so I have the difference and I divided by the roots of the variance the variance so is s square not sigma square because I do not have information on the real experimental variance but I can have what we call the sample variance it's what I'm looking the the latin letter and not the Greek letter and this s square I get it by the repetition by the measurement at the center so when you want to make this check what you need you need two things all the time is the the rational of this test you cannot have only factorial points because the factorial points will never contradict you a linear response model so you need some points I prefer the center but it could be a part of the center it could be even at the center of of one side but I need some points it could contradict the linear answer and I

notes

summary



need repetition I need at least that some measurements have been made two times the minimum for being able to calculate an s square so I need a difference and I need a variance is the two things I need for making this calculation and after the rest is statistical calculation I have to divide by the sum of the inverse of the numbers of points so it's it's the management of the degree of freedom I will not enter too much in the explanation so this give you a statistic the t statistic that you can test with the student distribution and so you can decide that you would like to take the risk to accept a more sophisticated model with 95 percent so you remember it was the same calculation we have done for the confidence interval so I'm calculating the t value for half of 90 1 minus 95 percent okay so it was 5 percent divided by two so this is this this value is the level of the risk that I'm taking so I'm telling you that I would like to have 95 confidence in my decision so 95 1 minus 95 makes 5 percent and half of it makes 2.5 percent and the degree of freedom is 3 because I have calculated the degrees of freedom of my variance calculation so I had made four repetition of the measurement at the center so for the calculation of s square of the sample variance I have three degrees of freedom and this I'm able to calculate a t value and it was so when I'm calculating the t zero I get the 27 as an answer and when I get reference t value it was three something so my 27 is bigger than three so it's worse to go for a second degree model nevertheless you do not have sufficient data yet because with one

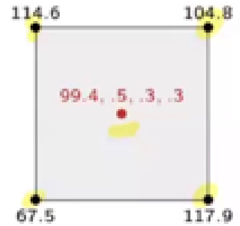
notes

summary

5.3 Test the curvature with a central point

- An experimental situation with two factors x_1 and x_2, 4 factorial measurements, $y_f(i)$, 4 measurements at the center of the experimental space, $y_c(j)$
- Is the linear model with interactions $y = a_o + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2$ sufficient ?

Run	Factorial	Center
1	67.5	99.4
2	114.6	99.5
3	117.9	99.3
4	104.8	99.3
$\bar{y}$	101.2	99.4
s^2		0.009
δ	1.83	



Factorial

Tester $H_o : \alpha_{11} + \alpha_{22} + \cdots = 0$

$$t_o = \frac{\bar{y}_f - \bar{y}_c}{\sqrt{s^2}\left(\frac{1}{n_f} + \frac{1}{n_c}\right)}$$

center

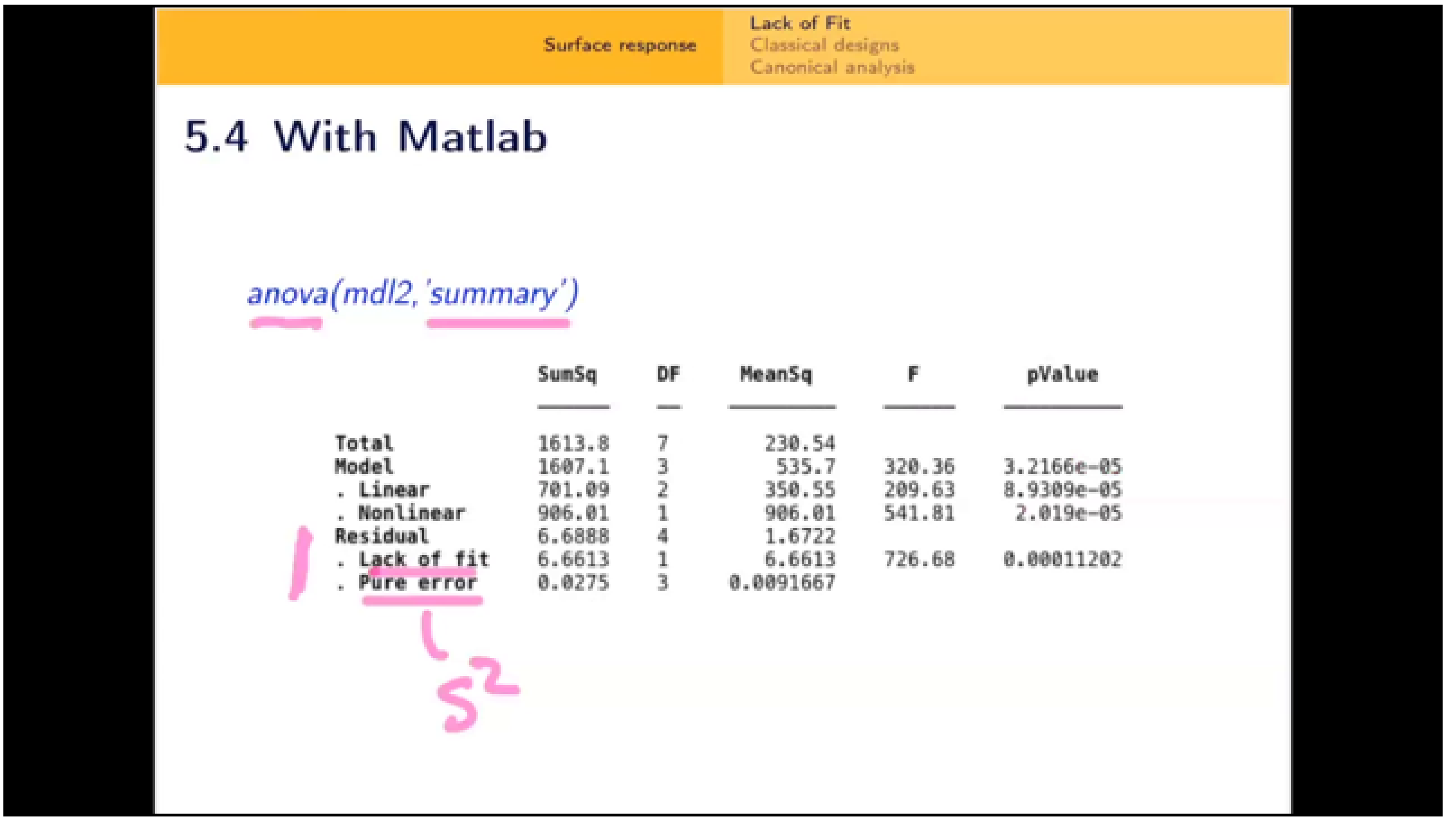
if $|t_o| > t_{\alpha/2, n_c - 1}$ then H_o is rejected.

In the present case : $t_o \approx 27 > t_{0.025,3} \approx 3.18$

point at the center is not sufficient you need at least because the difference between the center and the factual

notes

summary

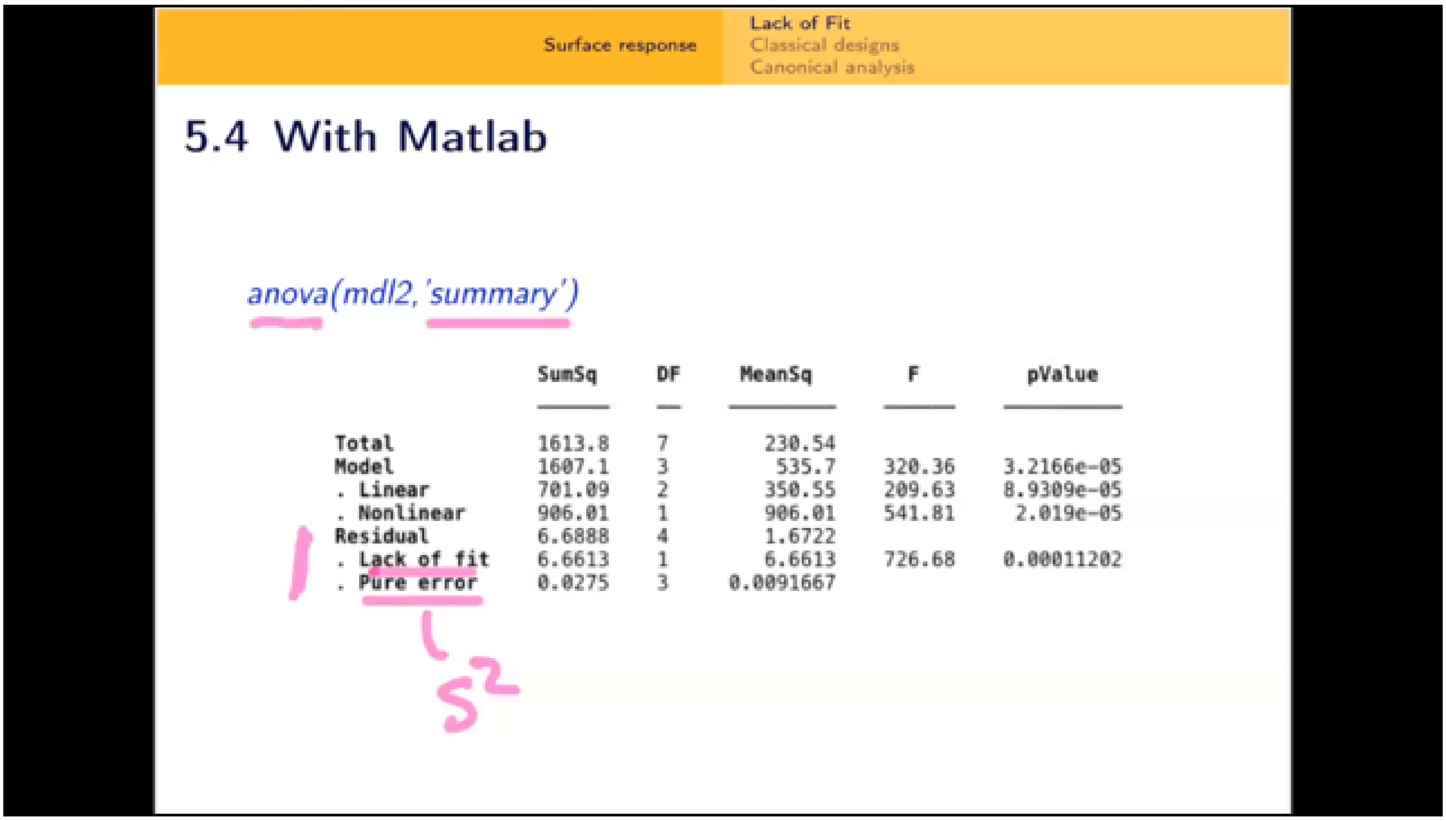


points is just the sum of the quadratic terms is not each one so you know that you have a curvature but you don't know in which direction if you have curvature in both directions so you need more points so the test is worse because you have made more points that's the minimum the factorial but if you need it to go further you need to make more experiments so it's a test before making more experiments so here is a test of the curvature using the student test you don't need to make it by yourself matlab or python can help you so if you use the function of matlab anova so you have used fitlm for fitting your linear model on your on your data but you need to put the factorial data and the center point you have to put the keywords summary and when you do that you will observe that matlab separates the residual in two parts it will separate and I will explain it better after the pause between two elements once coming from the repetition which is a pure error there is an estimation of the pure error this was typically what we get was the air square in the previous slide and the lack of fit is in the your difference between your measurement and your model what cannot be taken into account by this sample variance we call it lack of fit so now the residual of a fit can be separated in two parts but for that you need more than minimum points for your model that mean if you have a linear model with instructions you need more than factorial points you need point in the center point in the center of the side etc you go all the faces you are obliged to have more than the minimum and you need to have repetition if you do

notes

summary

14m 25s

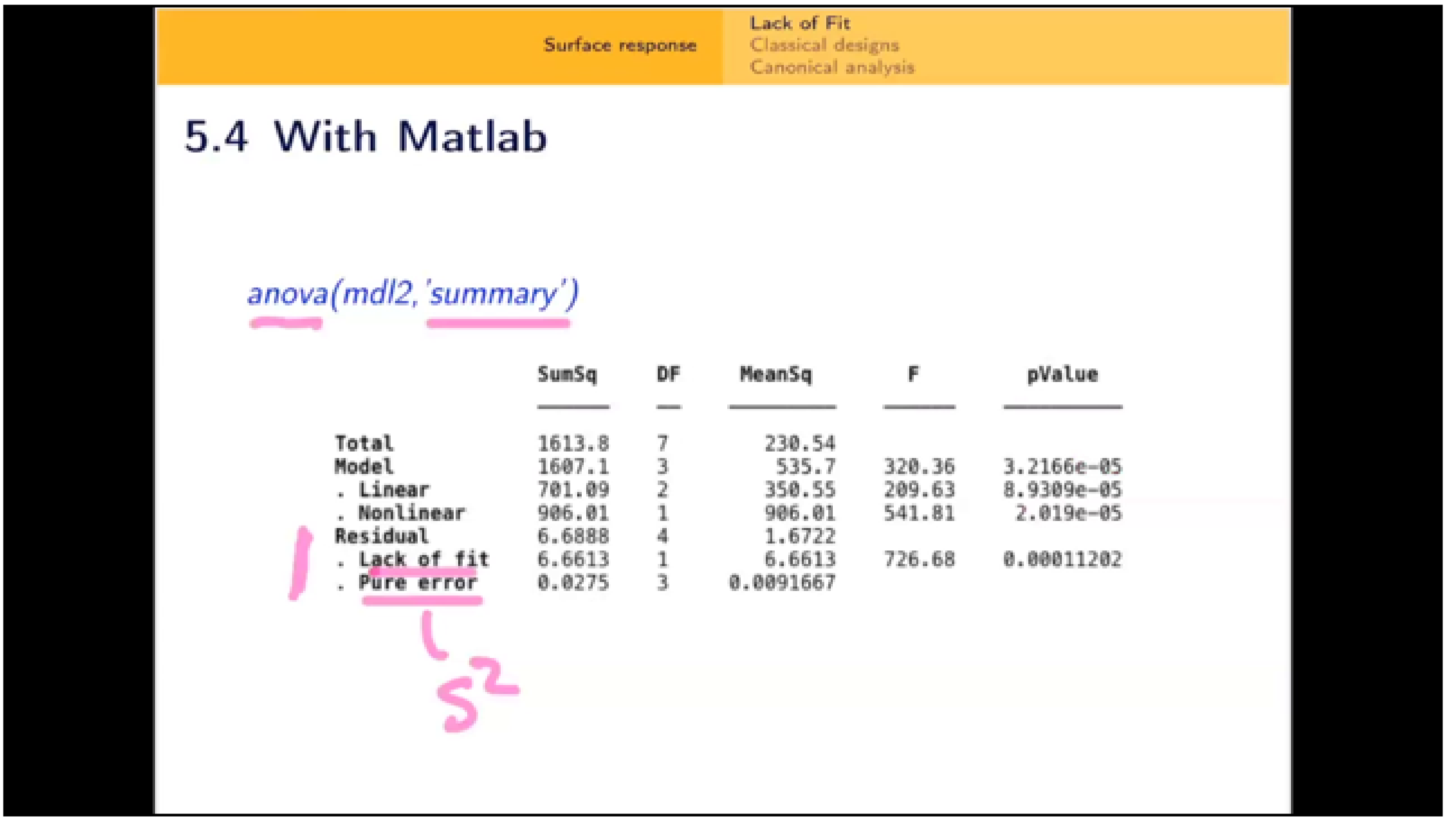


	SumSq	DF	MeanSq	F	pValue
Total	1613.8	7	230.54		
Model	1607.1	3	535.7	320.36	3.2166e-05
. Linear	701.09	2	350.55	209.63	8.9309e-05
. Nonlinear	906.01	1	906.01	541.81	2.019e-05
Residual	6.6888	4	1.6722		
. Lack of fit	6.6613	1	6.6613	726.68	0.00011202
. Pure error	0.0275	3	0.0091667		

not have these two things you cannot calculate the lack of fit because you will not be able to separate the discrepancies between your data and your model due to the variance and due to the the your model lack of fit you understand lack of fit so that means on your feet there's something which is not sufficient and after you have to be careful how to analyze this lack of fit so if you remember when you make a nanova you check the the p-value and when the p-value is small smaller than 5% you consider that your source is valid okay so as for my model I have a value of 3 10 minus 5 smaller than 5% this is 30 ppm so I say my model is okay my model is the origin of the variance after you can separate that I have separated between the linear and the non-linear part of my model so it could be the main effect and the instructions the non-linear in this case is the instructions and you see that the values are also smaller than 5% so that means that the linear part is acceptable and the non-linear part is acceptable know the difficulty that we have a double negation a double negation because lack of fit is a negation so it's something missing lack of fit is something missing so now I have a p-value for the lack of fit but I'm testing I'm testing this lack of fit against the pure error and I have a value which is smaller than 5% so I'm telling that the lack of fit was an origin of the variance it's an ideal variance so was a source of variance so I'm accepting that I have a lack of fit so I need to go further okay so when I have a nice p-value for the coefficients it's like a stop

notes

summary



sign it's okay I'm happy with that when I have a nice p-value with the lack of fit is a go sign you have to go further okay let's have a pause

notes

summary