Course material

Course:

ENG606 / PHYS 442

Video:

DOE_lesson12_part2_CCM

Concepts (extracted from automatically generated subtitles):

Degrees of freedom. Different functions. Marginal means. Sort of magic square. First variable. Constant coefficient model. Numbers of level. Magic square. Degree of freedom. Big change. Second variable. Big difference of distance. Dot plots. Grand mean. Different marginal means.

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

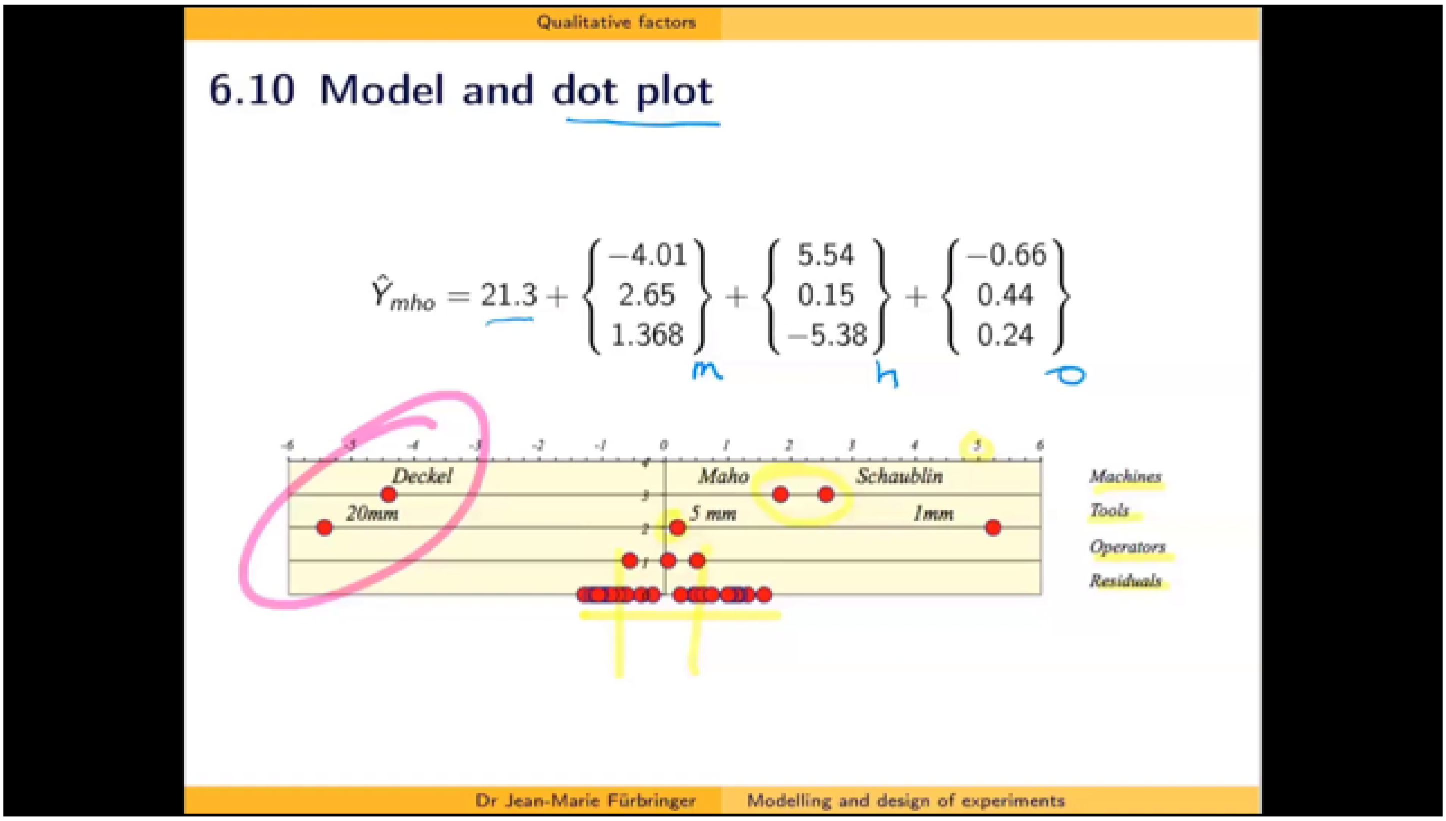
to video

...

notes

summary

0m 0s

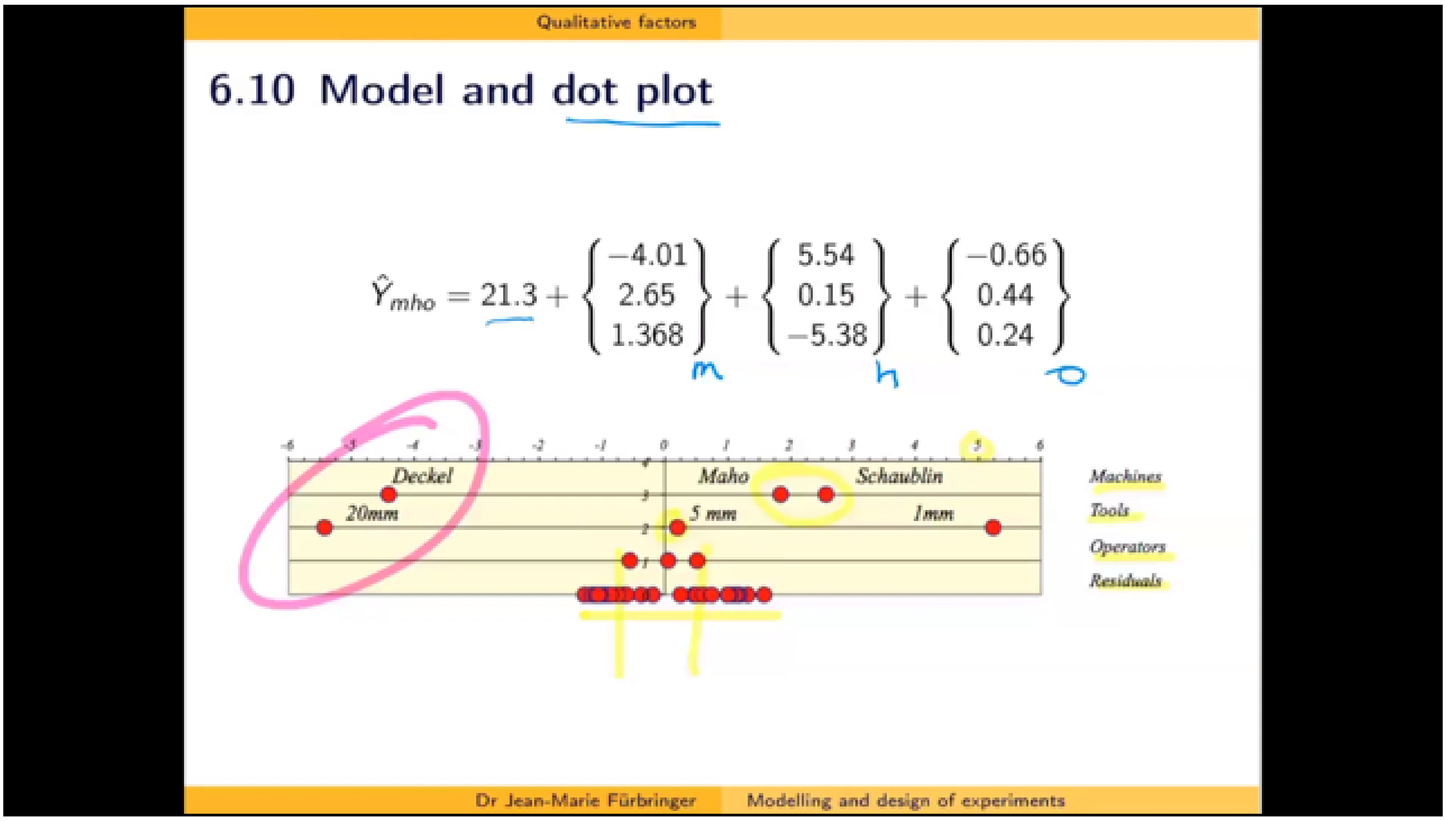


These subtitles have been generated automatically So now we have seen what is a constant coefficient model.

notes

summary

0m 1s

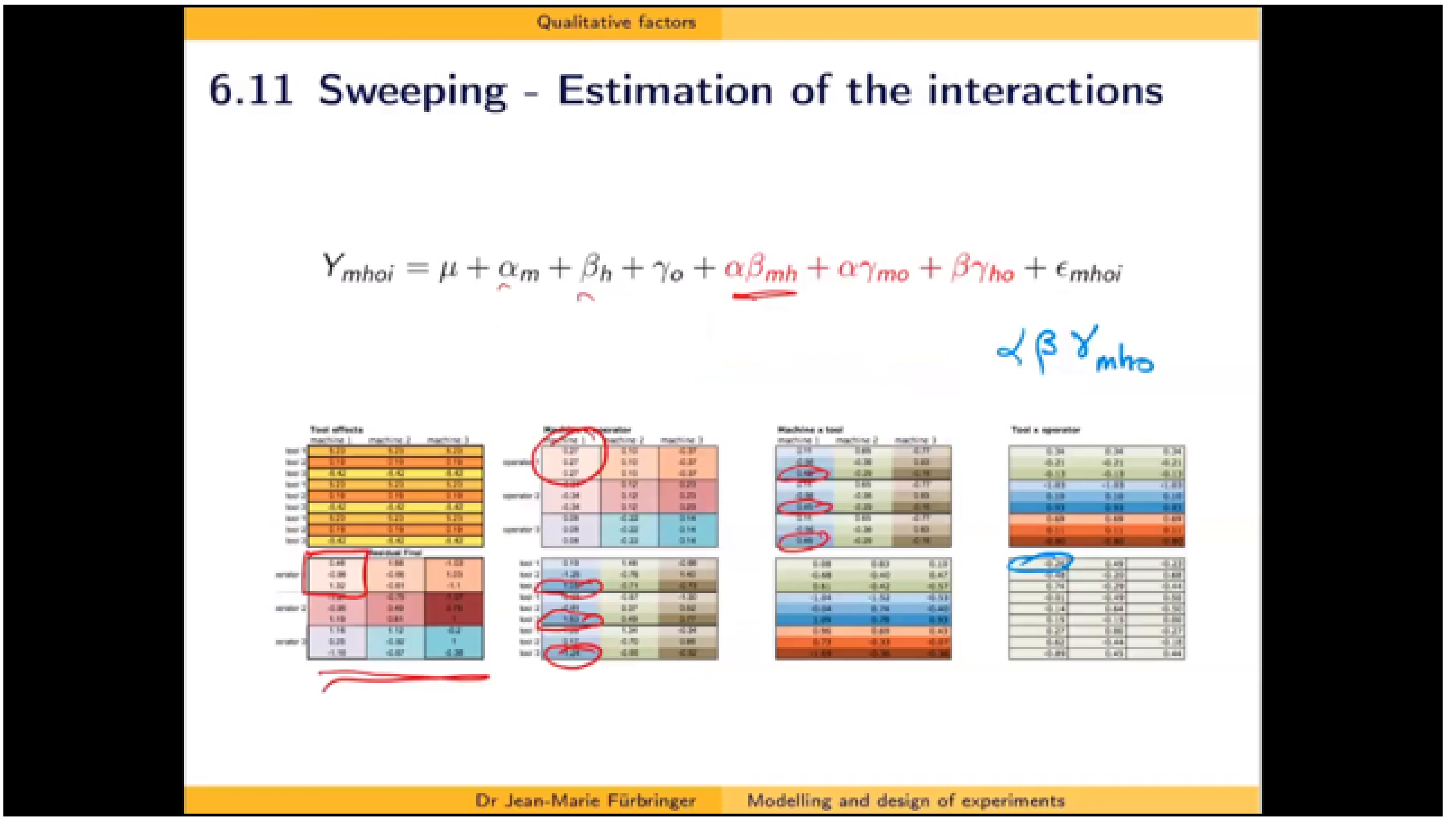


We have seen what is sweeping this way of decomposing a frame of data in orthogonal components.

notes

summary

0m 5s

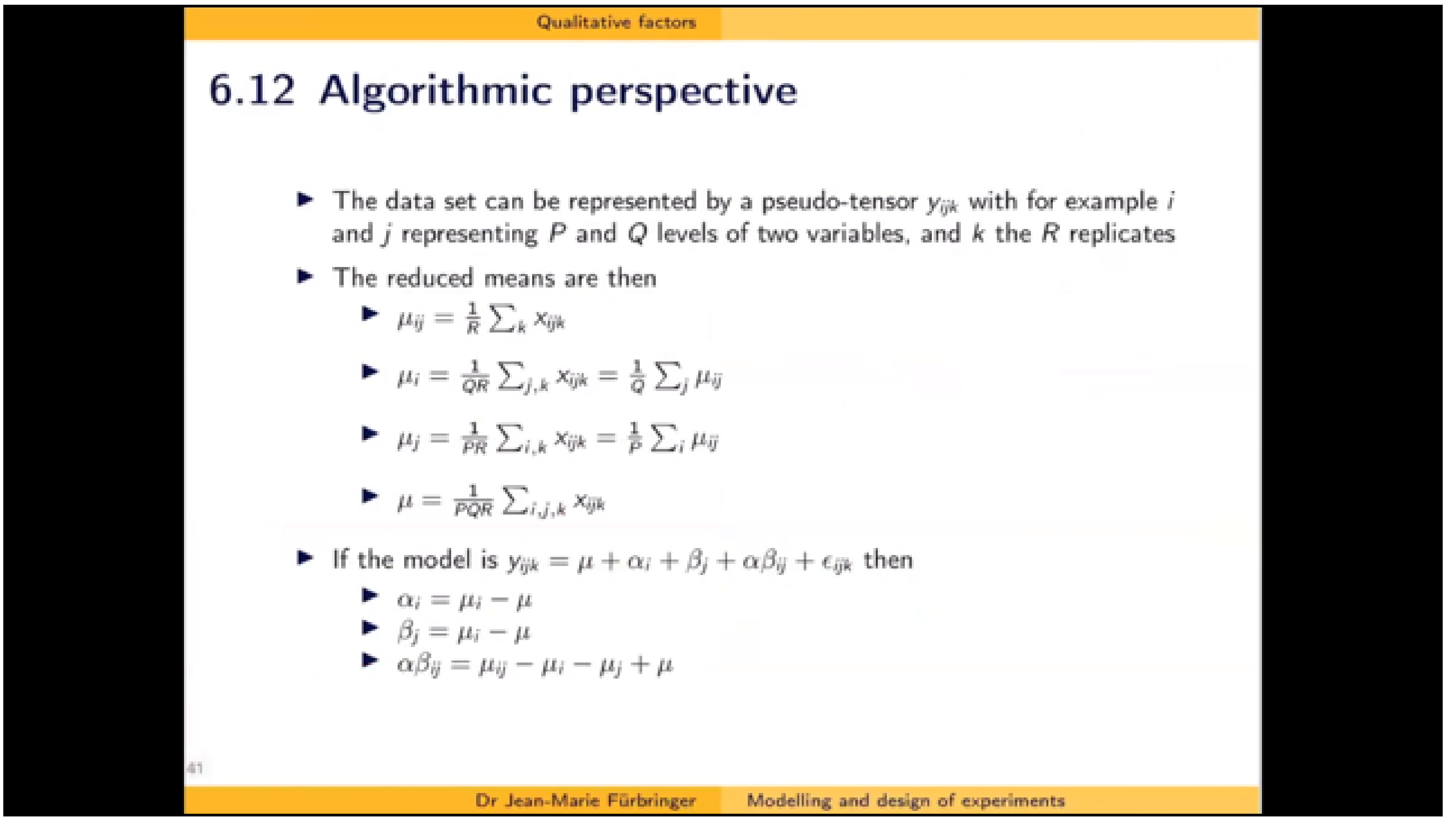


So how to represent that model with this coefficient and also how to make dot plots. The dot plots in Excel, in MATLAB, in Python, as far as I know, you have to do it by yourself. You have to make the program by yourself. It doesn't exist as a possible plot. A few things approaching, but not really a dot plot like that.

notes

summary

0m 15s

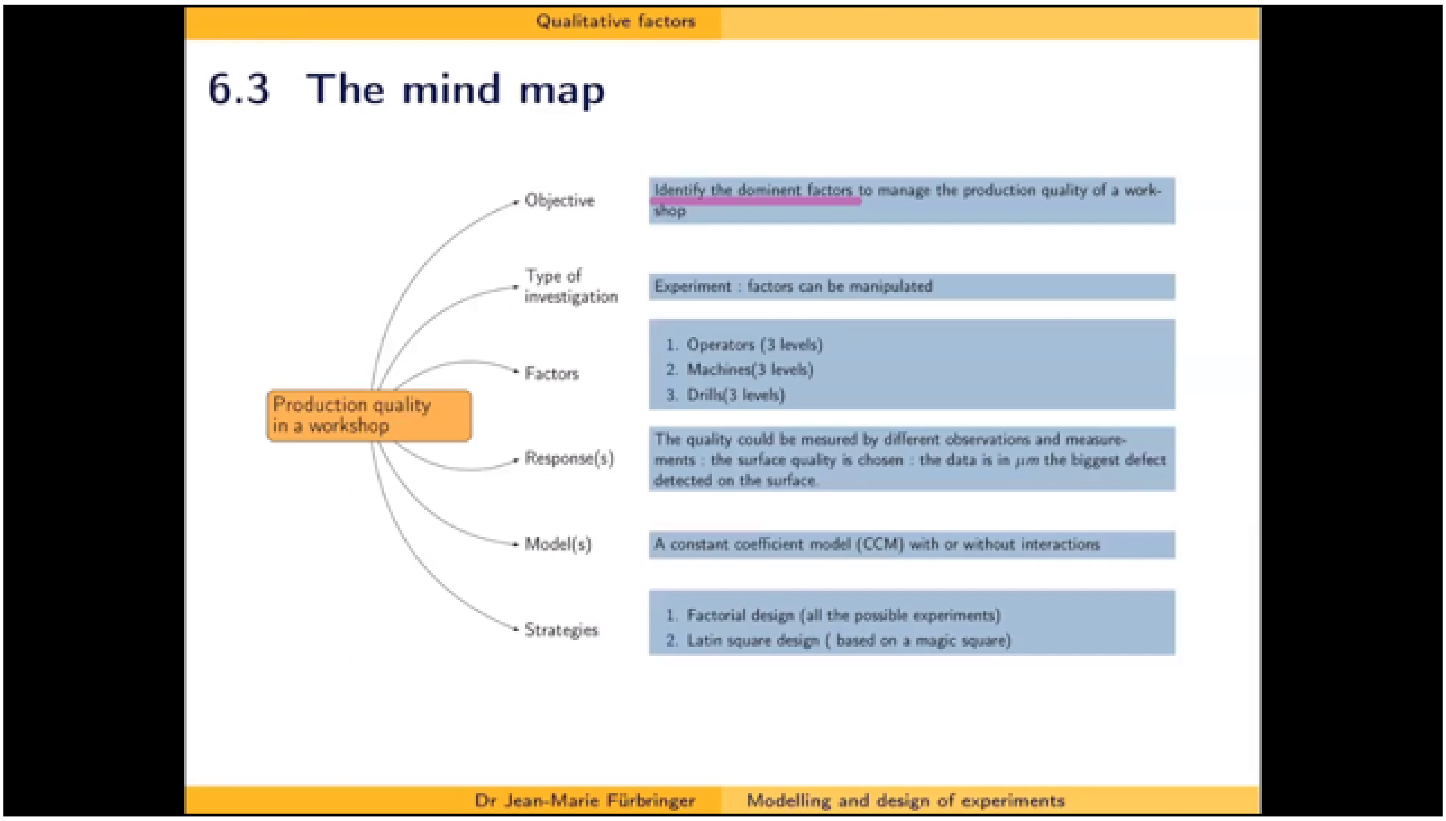


And I also show you the instruction. Now, what is important is to bring insight on the statistical significance of those things.

notes

summary

0m 56s

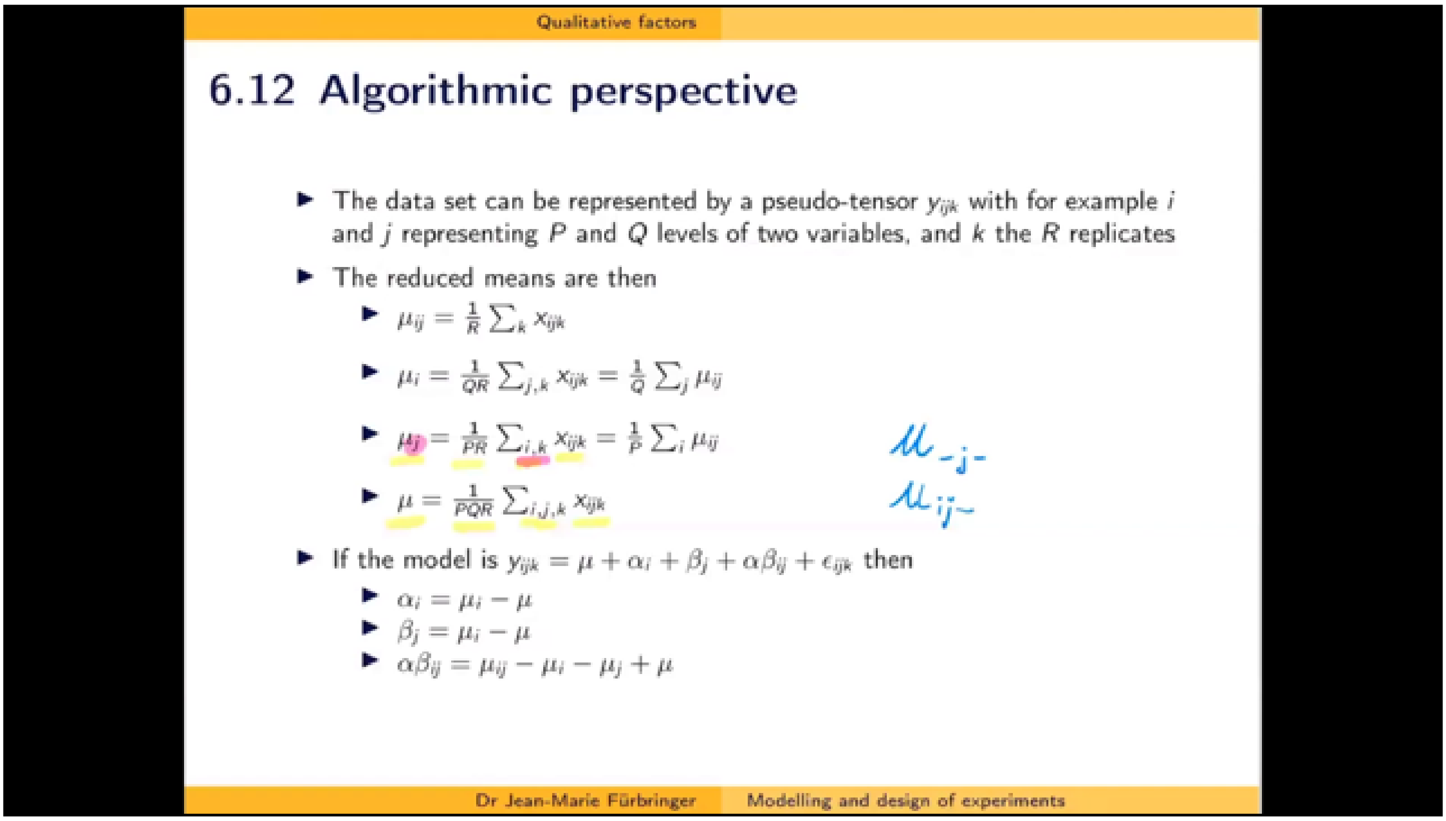


Before that, here is the algorithmic part. So we work when the different functions in MATLAB, in Python, are not making the frames. I present to you this is just a demonstrative aspect. They are working with average and marginal means. So if you have to program it that. So the grand mean is the sum of all your data. You sum on, imagine we have two variables and replicates. So you sum on EGK, AJ being the two variable and K being the replicates. And then you divide by the sum of all your data, which is the numbers of level in each factor, multiplied by the numbers of replicates. So in this case, one divided by PQR. You can make also what is called marginal mean. You sum your data on one factor and the numbers of replicates. And you divide by the numbers of level in that factor and the numbers of replicates. They give you a marginal mean for this factor. In this case, the factor J. And if you see, I call it J and I'm making the summation on the other variable. It's sometimes written also in some papers.

notes

summary

1m 8s

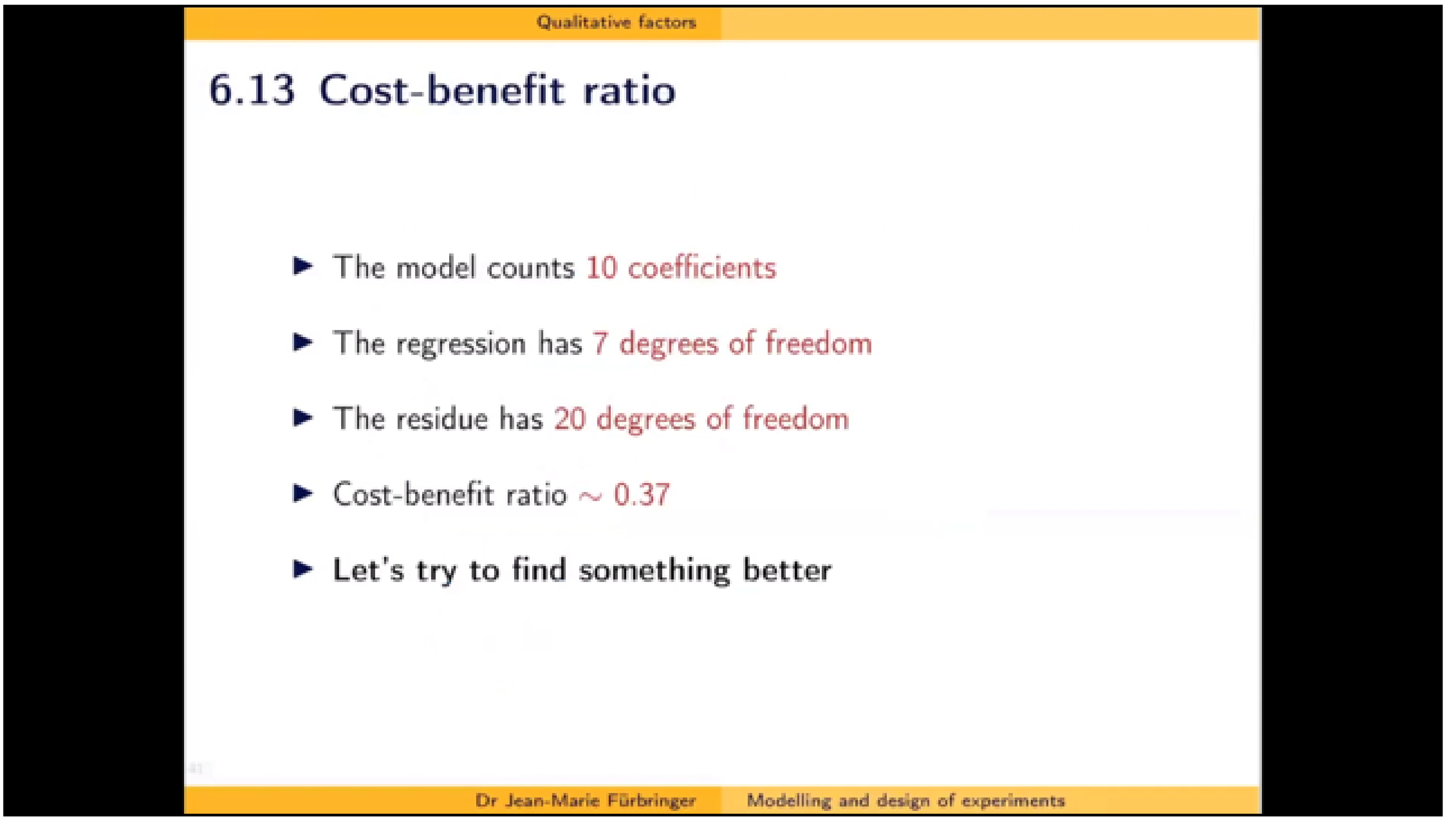


We could eventually also be interested to know if they have interactions. It could be written like that. The iPhone representing I and K. And so you would have things like that. I, J, so it's a way of writing it. If you look through the papers about that. And so marginal means for the variable, the first variable, the second variable,

notes

summary

2m 59s

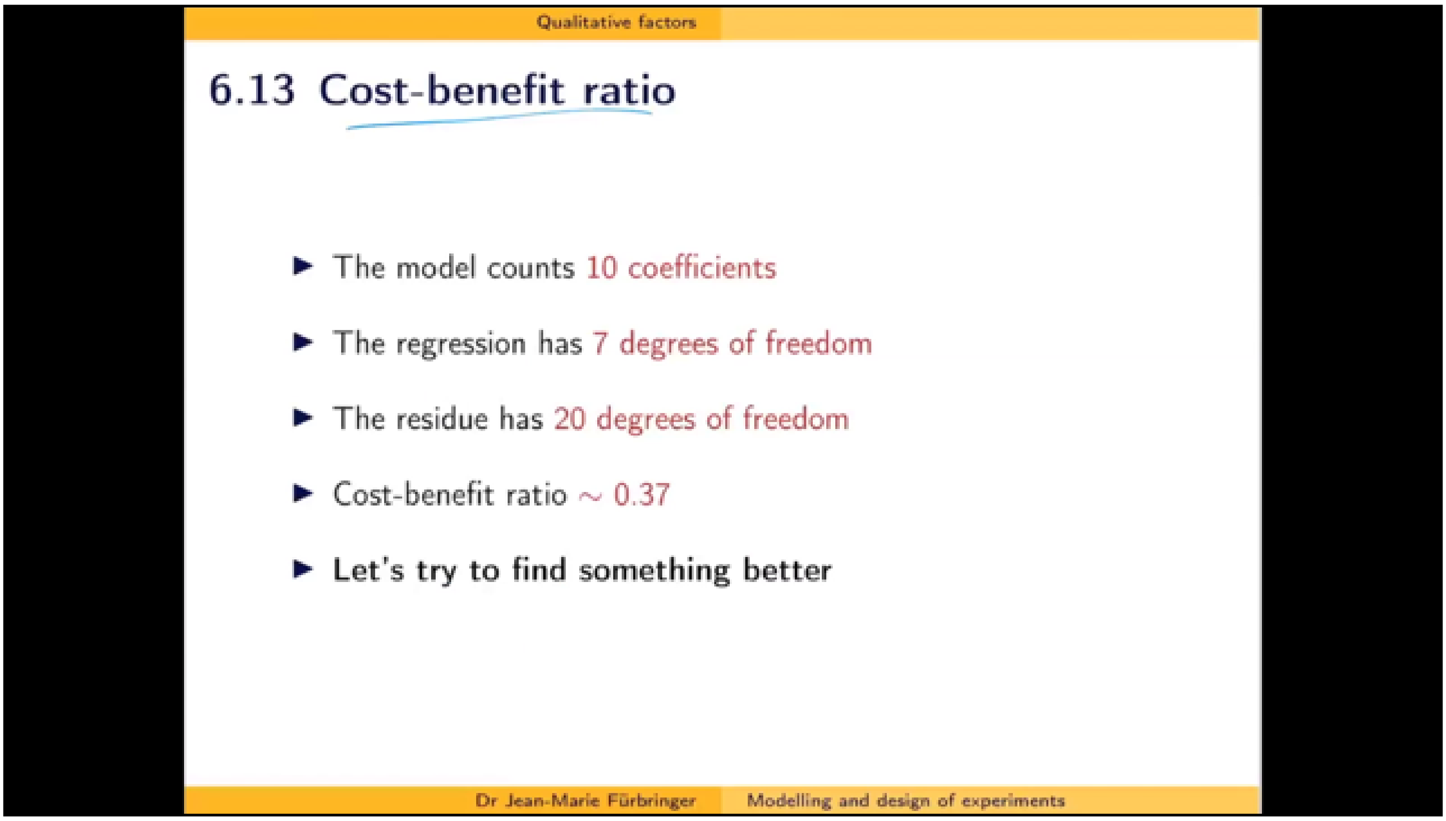


and this would be the marginal mean taking into account the possible interactions. I make the sum and I just divide by the numbers of replicates. And after when I would like to calculate my model, which will be a grand mean plus an effect for the first factor plus an effect for the second factor plus an interaction. You see that the effect of the first factor will be the different marginal means for the first factor minus the grand mean. The effect for the second factor are the marginal means. The, which one is a J? Yeah, this should be a J minus the grand mean. And what is different is for the interactions, the effect of interaction. It will be the marginal mean for the case of interaction, some value of A and J, minus the marginal mean for the first factor, minus the marginal mean for the second factor, plus the grand mean.

notes

summary

3m 25s

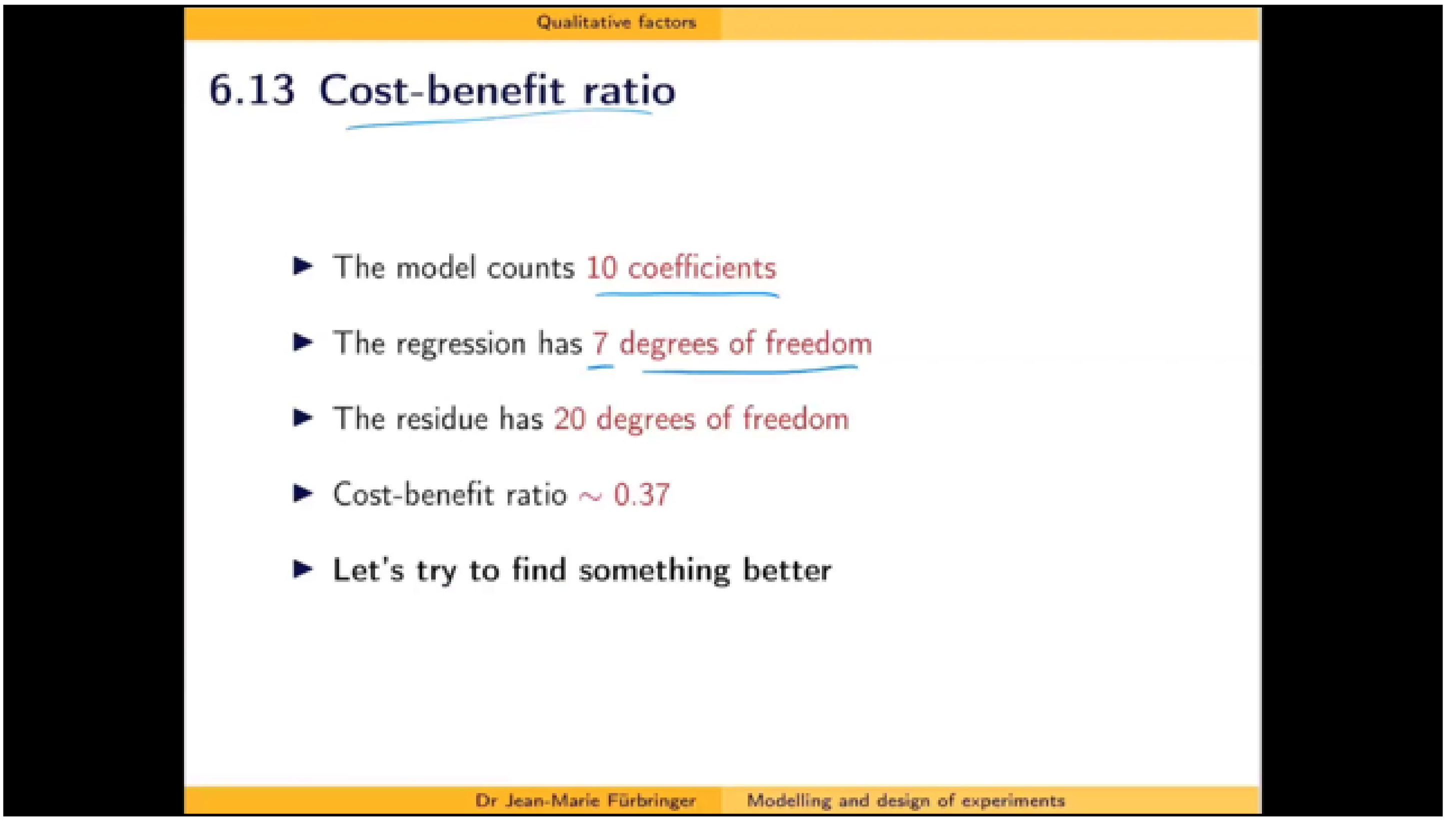


So let's just think a little bit about the cost-benefit ratio.

notes

summary

4m 49s

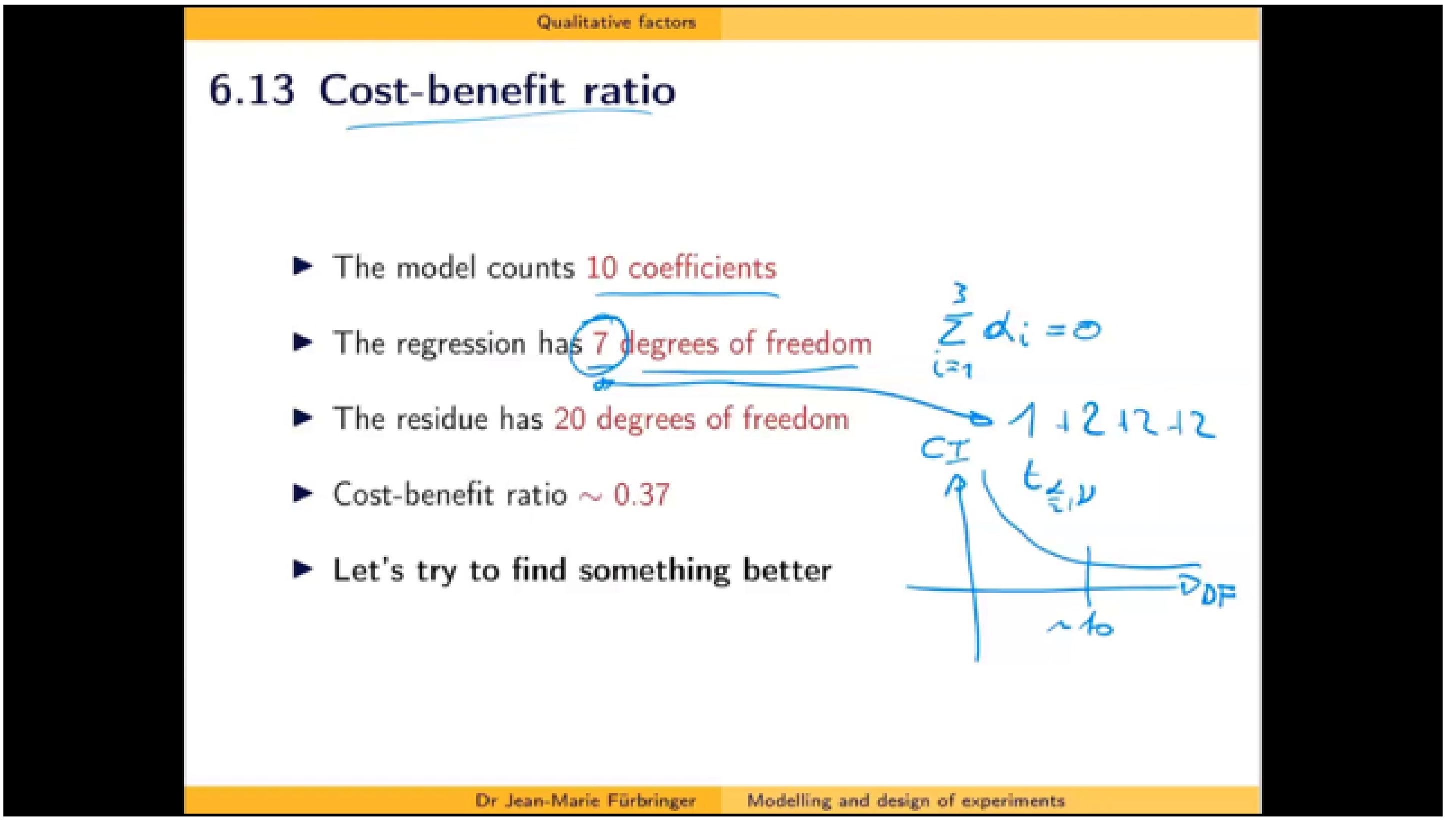


Our model for our workshop, so three variables, three levels for each one, we had 10 coefficients, the grand mean and three coefficients for each of the variables.

notes

summary

4m 53s

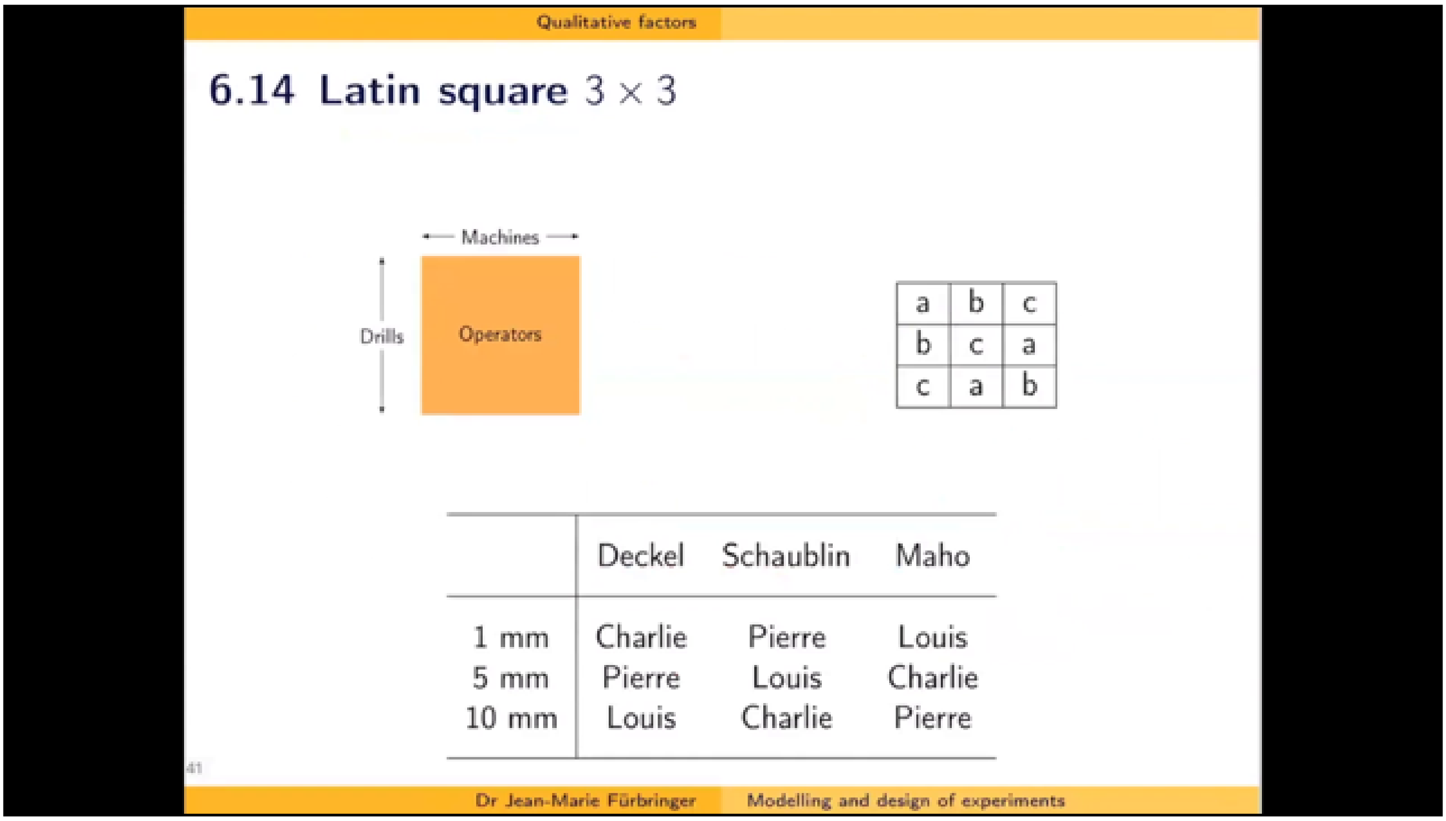


In fact, we have only seven degrees of freedom because I show you that, for example, the sum of the alpha J or E equals 1, 2, 3 equals 0. The sum of my effect of one of the factors equals 0. So I do not, this three number do not have three degrees of freedom. I cannot choose a three number independently. If you know two of them, you have the third one. So I have one degree of freedom for the constant plus two degrees of freedom for the machine, plus two degrees of freedom for the tools, plus two degrees of freedom for the operator, which makes seven degrees of freedom. So I have 27 experiments. So my residue will be 20 degrees of freedom because it will be 20 minus seven, 27 minus seven. So it makes 20 degrees of freedom. And if you remember a graphic I present you on the calculation of the confidence interval, we are here at around 10 degrees of freedom. If here you have degrees of freedom and here you have the importance of your confidence interval

notes

summary

5m 14s

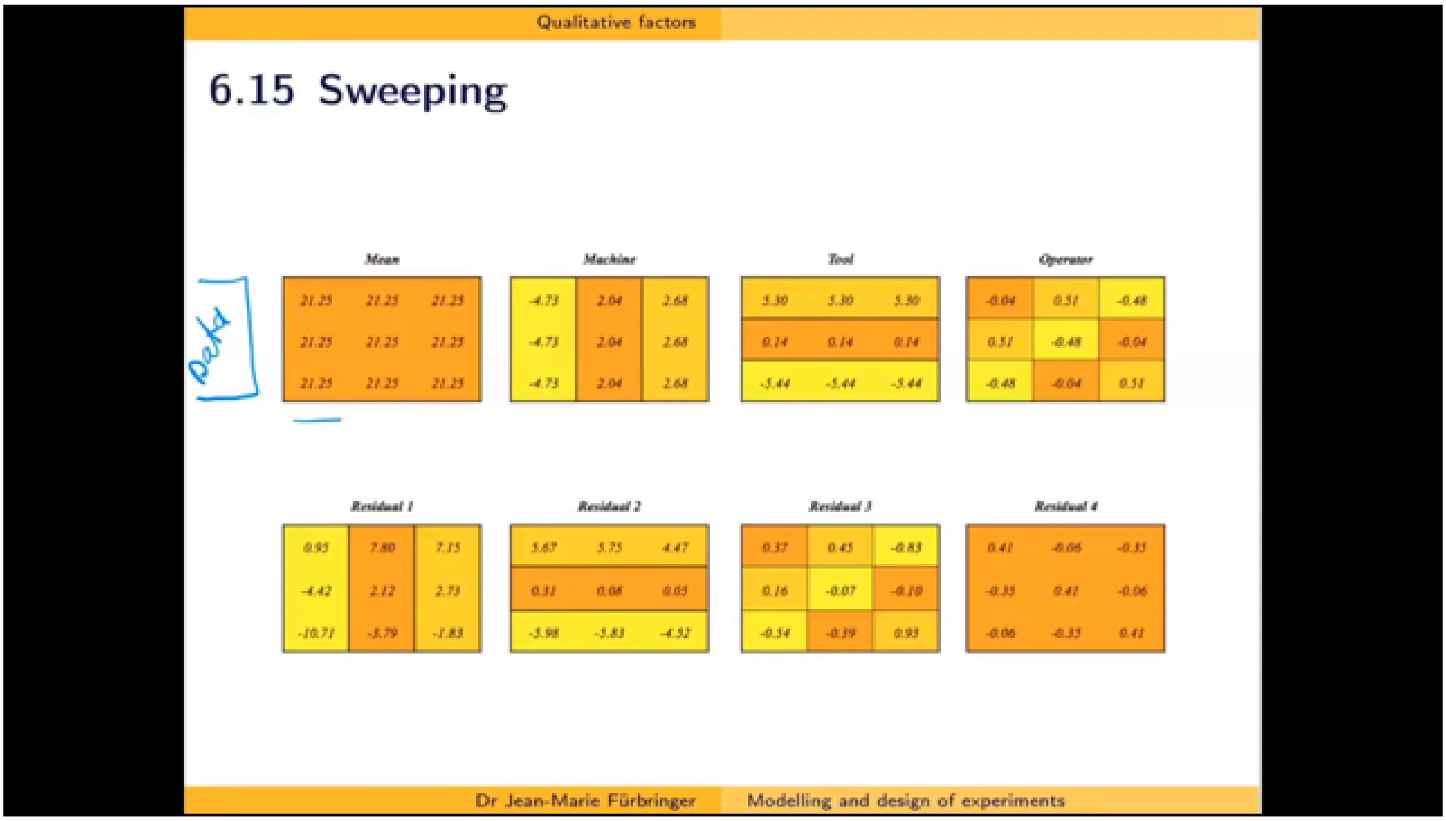


or the value of the T alpha divided by 200 degrees of freedom, the student's distribution. We see that we have a hill here and that we start having something, okay, around 10, 12 degrees of freedom. 20 is perhaps not so much interesting. 200s could be interesting. 2000s could be interesting, but between two and 20, I'm not improving a lot my results. If I understand that I have 27 experiments and seven degrees of freedom, I can calculate the Jean-Marie Febringer benefit ratio coefficient. About one third, I would be interested to find something better where the ratio between your efforts, the experimental effort on what you get is degrees of freedom in your model is smaller because 20 degrees of freedom is not necessary.

notes

summary

7m 1s

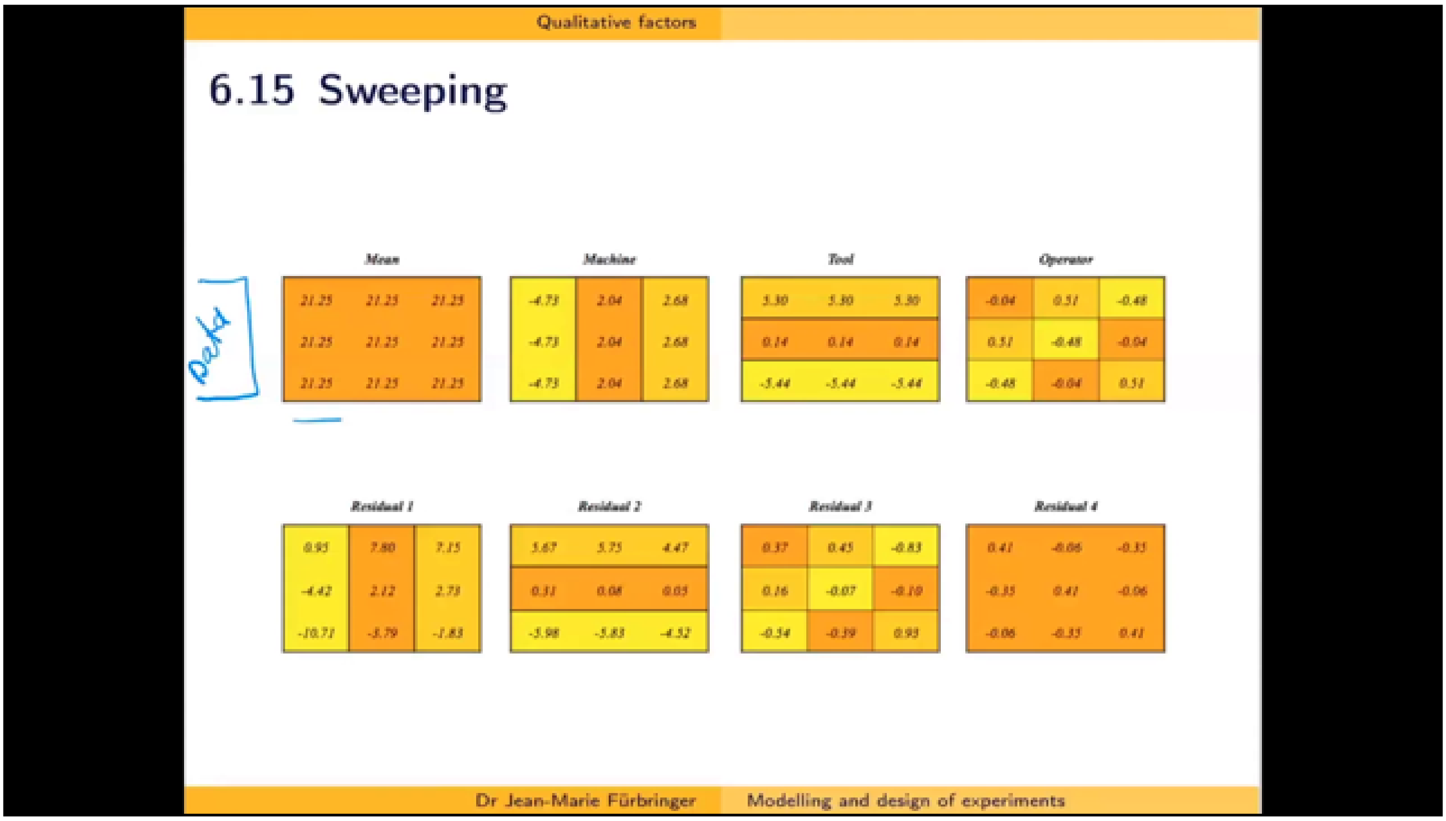


And there's something which is called Latin square. And for years, I was teaching that it was earlier or invented it. I just read that in fact, Arabs have already worked on this type of problem. But as often in the Occidental science, we are so proud that we think that only Europeans are inventing things, but in fact, the Arabs have already worked on this type of problem before we start to work on it in the Occidental. Okay, so what is Latin square is a sort of magic square, like I was mentioning for the Sodoku. You see here a magic square. You see that in each row and each column, I have all the possibility, A, B, C, D. And it exists in fact, there are two possibilities of different. You can have this one, A, B, C, D, B, C, A and C, A, B, or you can just change the two last one and it's make another magic square, which is orthogonal to the first one. So how we can use that for making experiments? How we can use Sodoku for making experiments? So if you say that the columns correspond to the machine and the row correspond to the tool, you can distribute the operator following this magic square. And so you see Charlie Pierre-Louis and Charlie Pierre-Louis and Charlie Pierre-Louis. In fact, in my experiment, I have distributed the operator according to my magic square. And you see here a scheme. The scheme, machine are organized horizontally, tools are organized vertically, and operator are organized following the magic square. If I do that, I do those experiments, I will be able to evaluate the main effects. So again, I have my data that I'm not representing. I have now nine experiments only before I have 27. So I have divided by three the numbers of experiments that I'm doing. So I have calculated

notes

summary

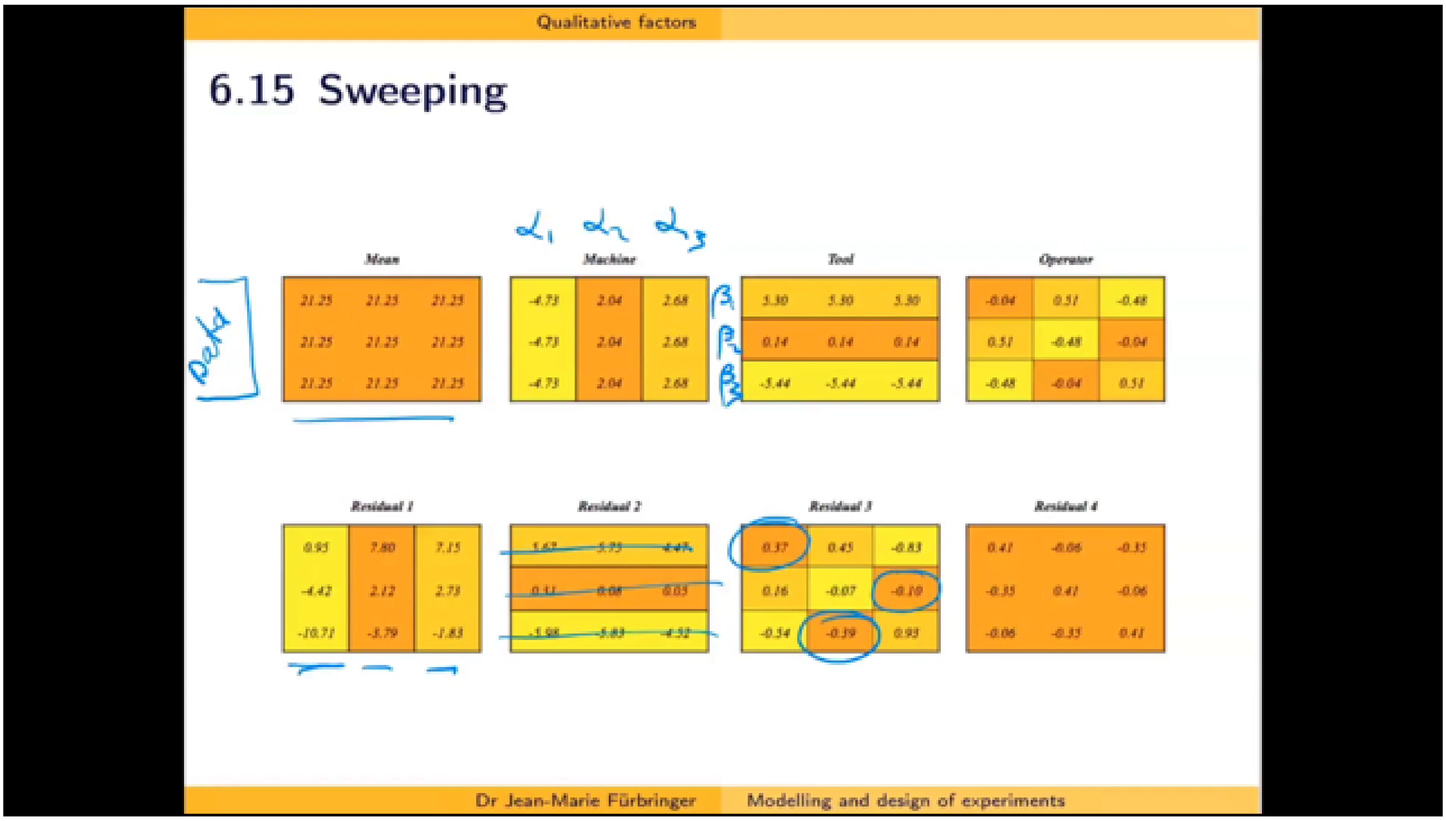
7m 59s



the ground mean.

notes

summary

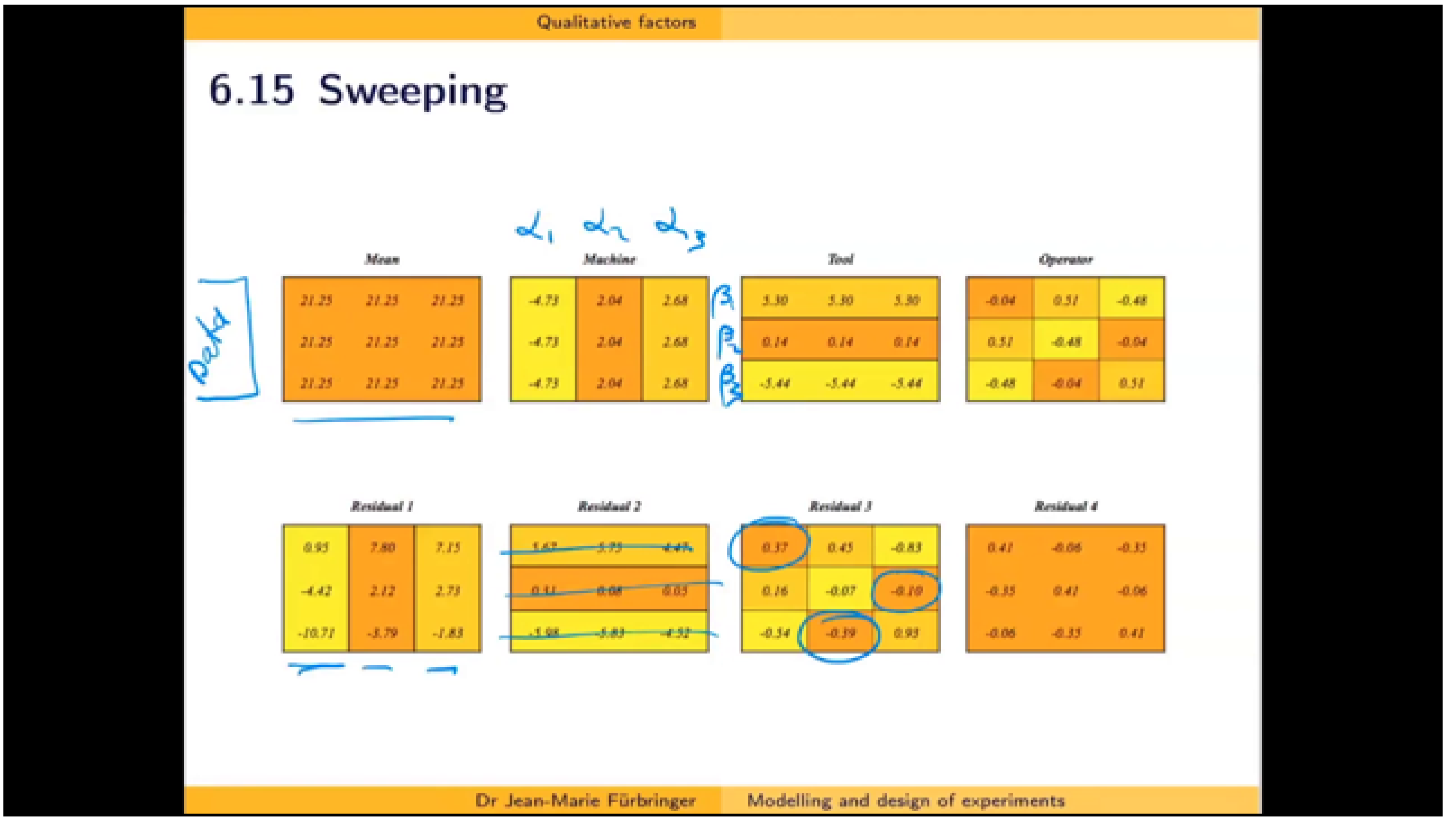


So the average of my nine experiments, I get a residue. And in my residue, I'm able to calculate a sub-residue concerning only machine one or only machine two and only machine three. And so I'm able of having, I call it alpha one, alpha two, alpha three, which are the effect of my machines. And I see minus 4.73, 2.04, and 2.68. And again, I have been in a residue and I'm able to analyze my residue by raw, representing my different tools. And I'm able to calculate my different effect of tool, beta one, beta two, beta three. And I get a residue. And now it's become just a little bit more complicated because my cells are not all together. But I can understand which cells represent my first operator,

notes

summary

10m 37s

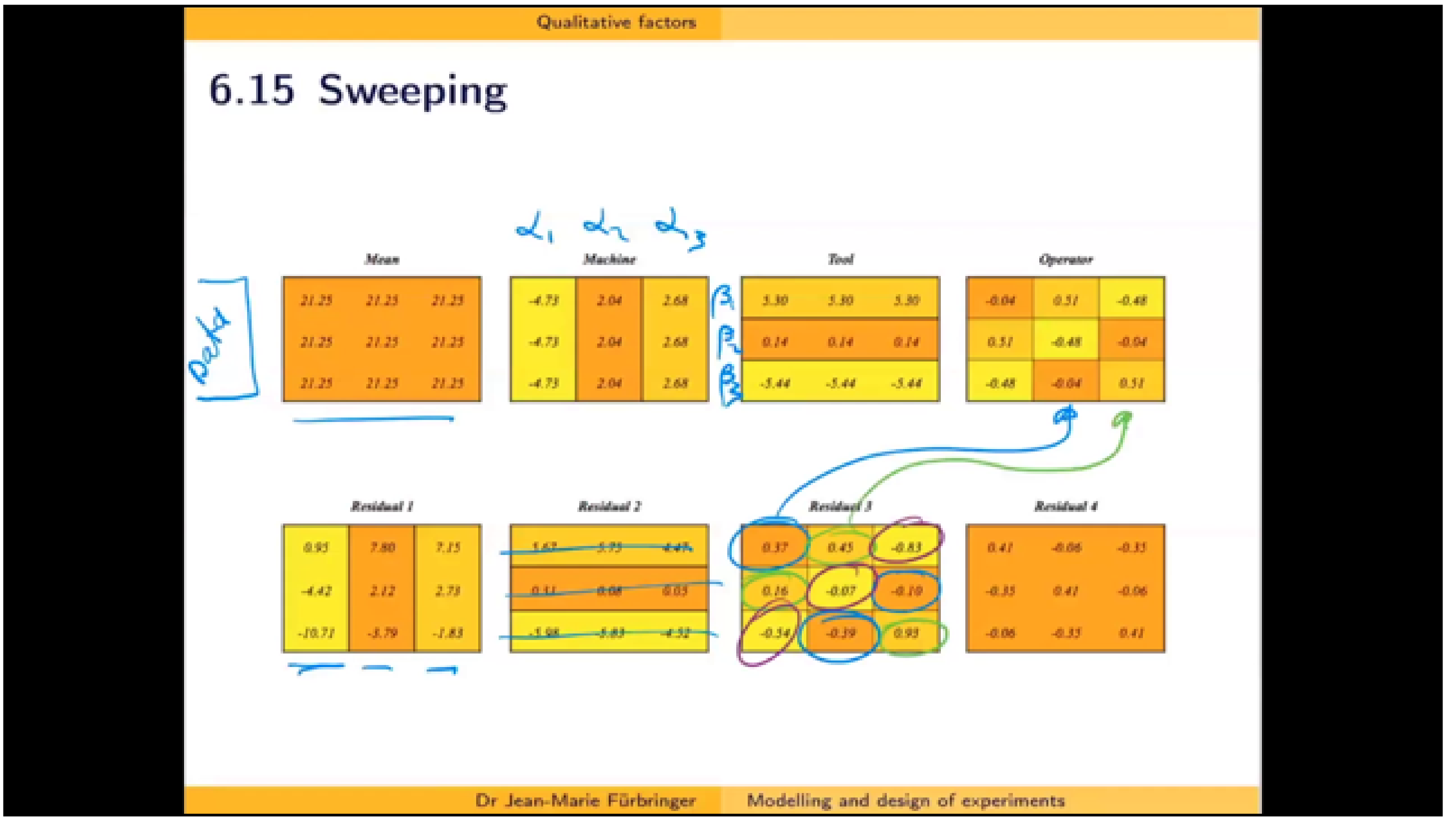


which other cells represent my second operator, and which cells are representing my third operator.

notes

summary

11m 49s

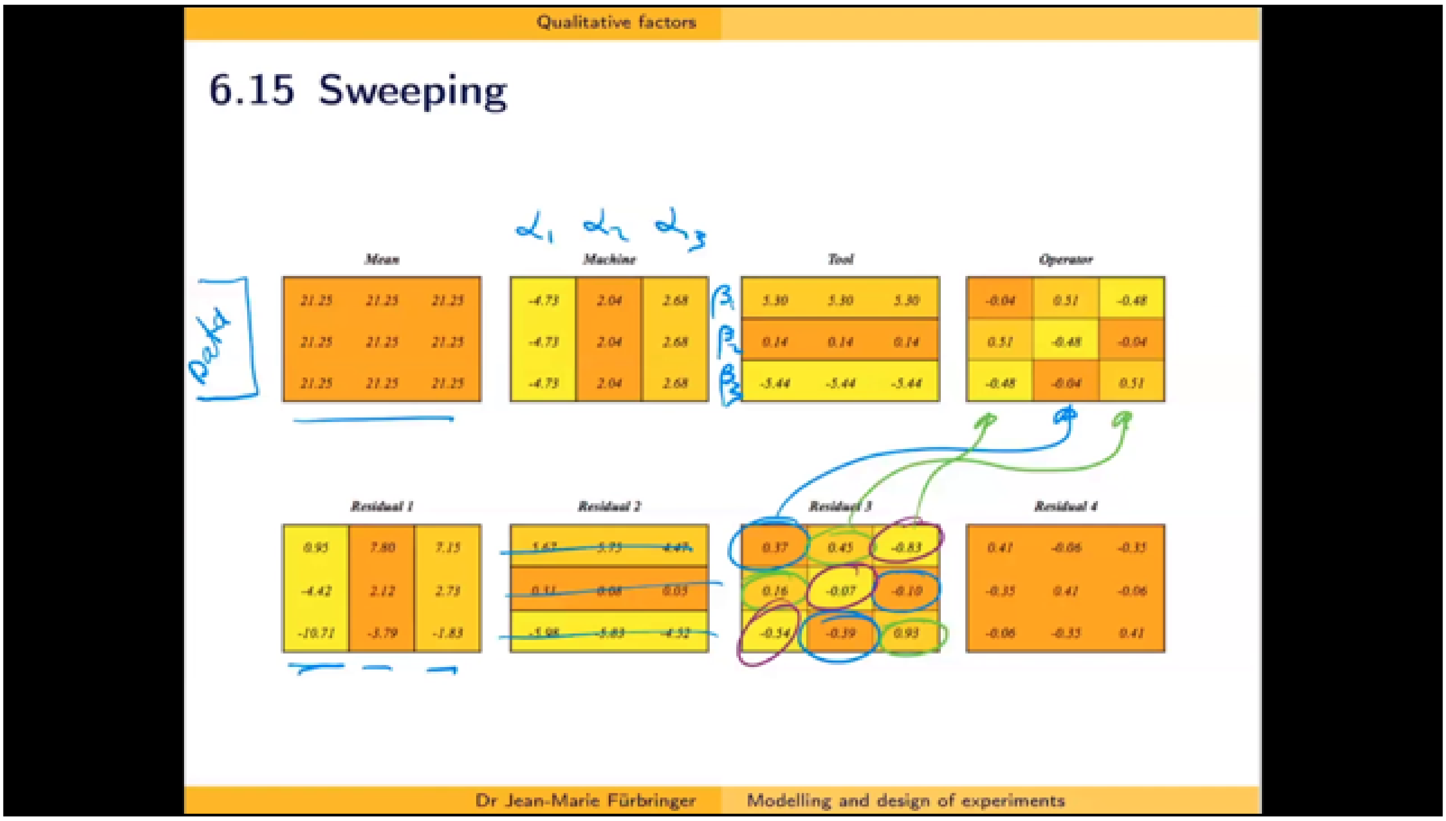


And I make the average and then I'm able to get an effect for the first operator,

notes

summary

11m 51s



another effect for the second operator, and a third effect for my third operator. Just don't mix what is coming from what. You have to do things properly.

notes

summary

12m 15s

6.16 Comparison of the two models

Model infered with 27 data points (factorial design)

$$\hat{Y}_{mho} = 21.3 + \left\{ \begin{matrix} -4.01 \\ 2.65 \\ 1.368 \end{matrix} \right\}_{m} + \left\{ \begin{matrix} 5.54 \\ 0.15 \\ -5.38 \end{matrix} \right\} + \left\{ \begin{matrix} -0.66 \\ 0.44 \\ 0.24 \end{matrix} \right\}$$

Model infered with 9 data points (Latin square)

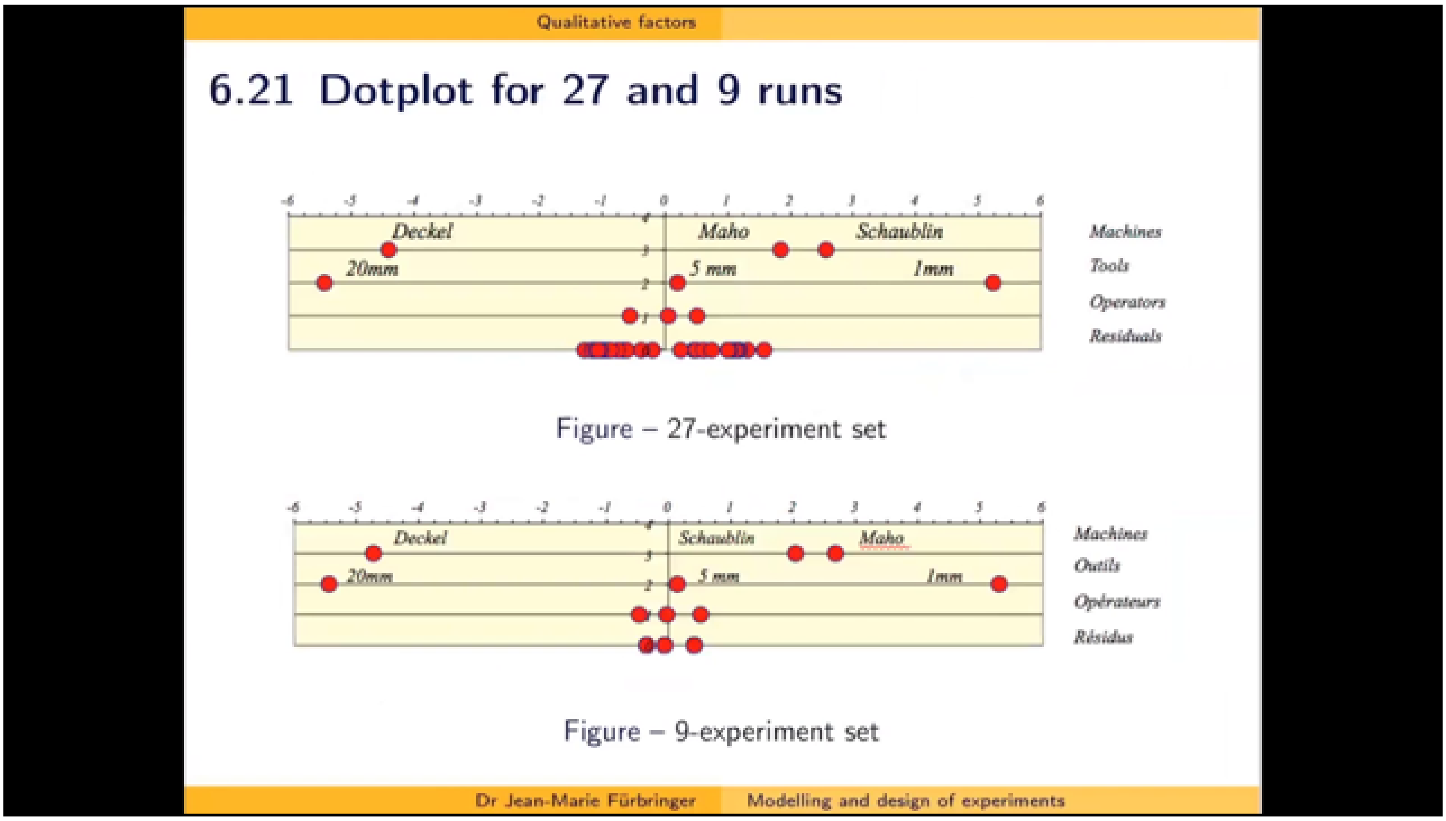
$$\hat{Y}_{mho} = 21.90 + \left\{ \begin{matrix} -3.71 \\ 2.26 \\ 1.45 \end{matrix} \right\}_{m} + \left\{ \begin{matrix} 4.94 \\ 0.63 \\ -5.57 \end{matrix} \right\} + \left\{ \begin{matrix} 0.18 \\ -0.17 \\ -0.01 \end{matrix} \right\}$$

That is still possible. And I have a final residual. But now I'm unable to go further. I cannot calculate instructions because I only have one case that have a couple of operator and tool or machine and tool when I have three in the other case. But I can have a model and I can now compare the model. So I have a first model that I have obtained with 27 data points. And I have another model which is quite equivalent, which is with nine data points. So it costs three times less experimentally. If you look for the different values, so 21.3, 21.9, okay. If it's just microns, it's not make a big difference of distance. What is interesting is what is good and what is bad. So if I go for the first one, it was the machine. And so the first one was the best. And here the first one is still the best. The second one was the third one with one microns, 1.4 microns error, more than the average. And it's still the third one, which is also in second place. And if I would like to see the third one is still the third one.

notes

summary

12m 29s

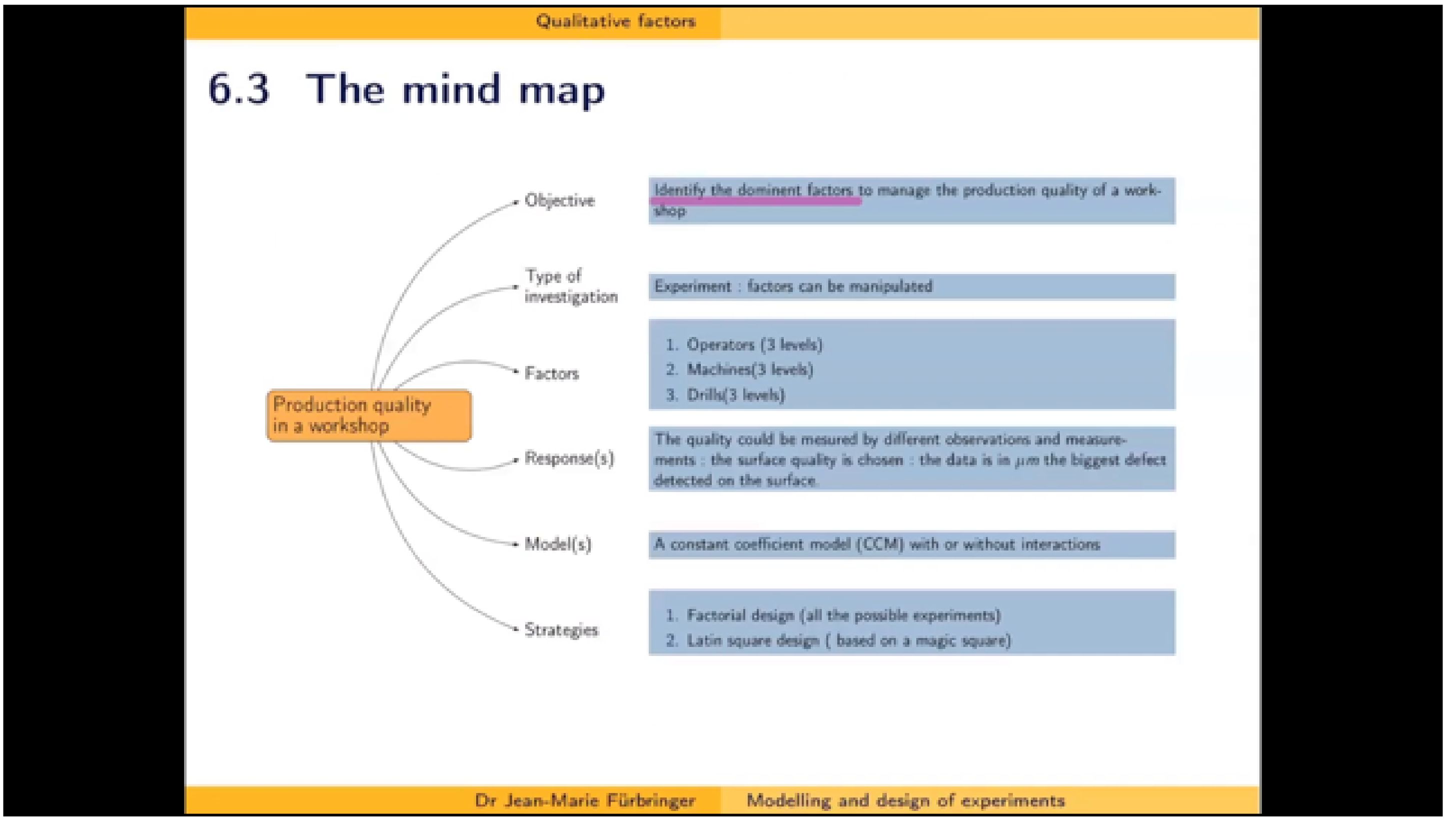


So at least for the case of my machine, the reduction of price is not changing what I have. The value is not exactly the same, okay. But which one is okay, which one is not okay, which one is the best one is quite okay. The same thing for the tool and not at all the same thing for the operator.

notes

summary

14m 13s

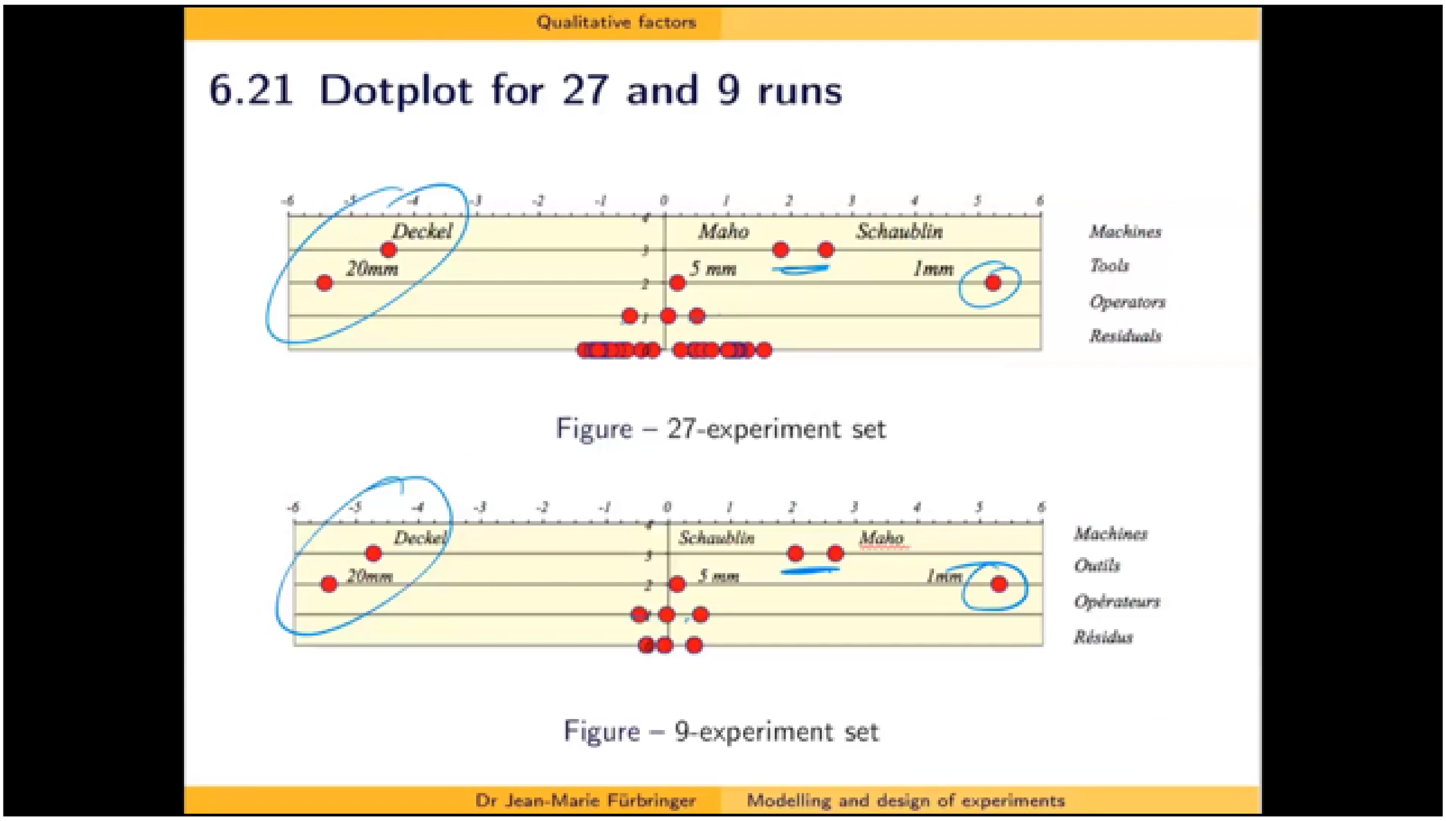


So if you look at the dot plot, it's evident in the two dot plots that the best situation is okay. And the different operator are not in the same order, but we have seen that the operator is very probably not an interesting effect, it's not an efficient effect. So within the noise, it's not a problem that in fact things are mixed. Now something is interesting if for the same data, thinking that the result was not the best defect, but a good quality and not a bad quality of my data and what I was interested was the maximum. Okay, in this case, it changed something. So if the best is the maximum, so the best would be one millimeter operator, and it's okay in both situations.

notes

summary

14m 45s

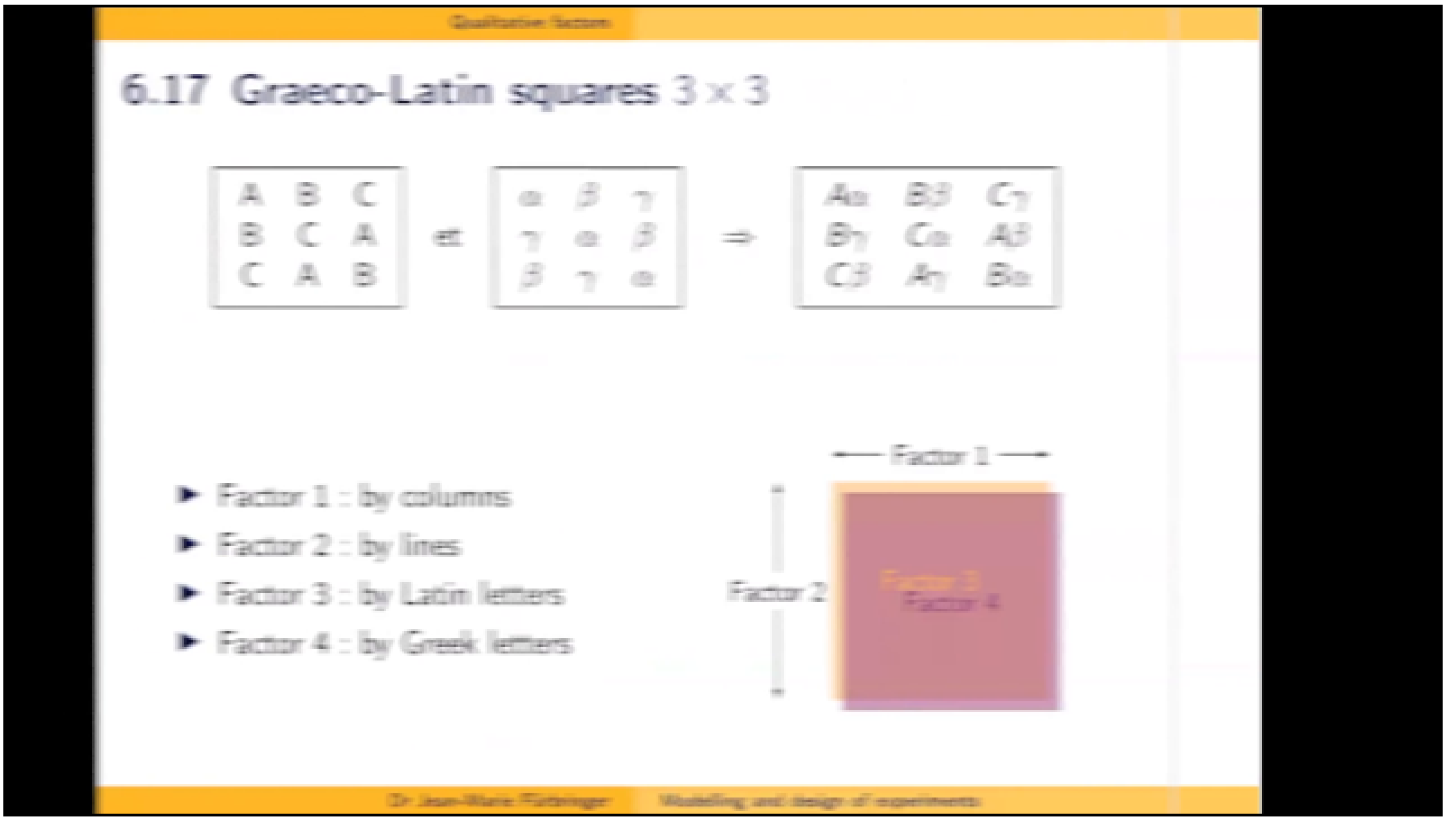


But now if you see, we could eventually also be interested to know if there are interactions, the effects for the different machine have changed.

notes

summary

15m 46s

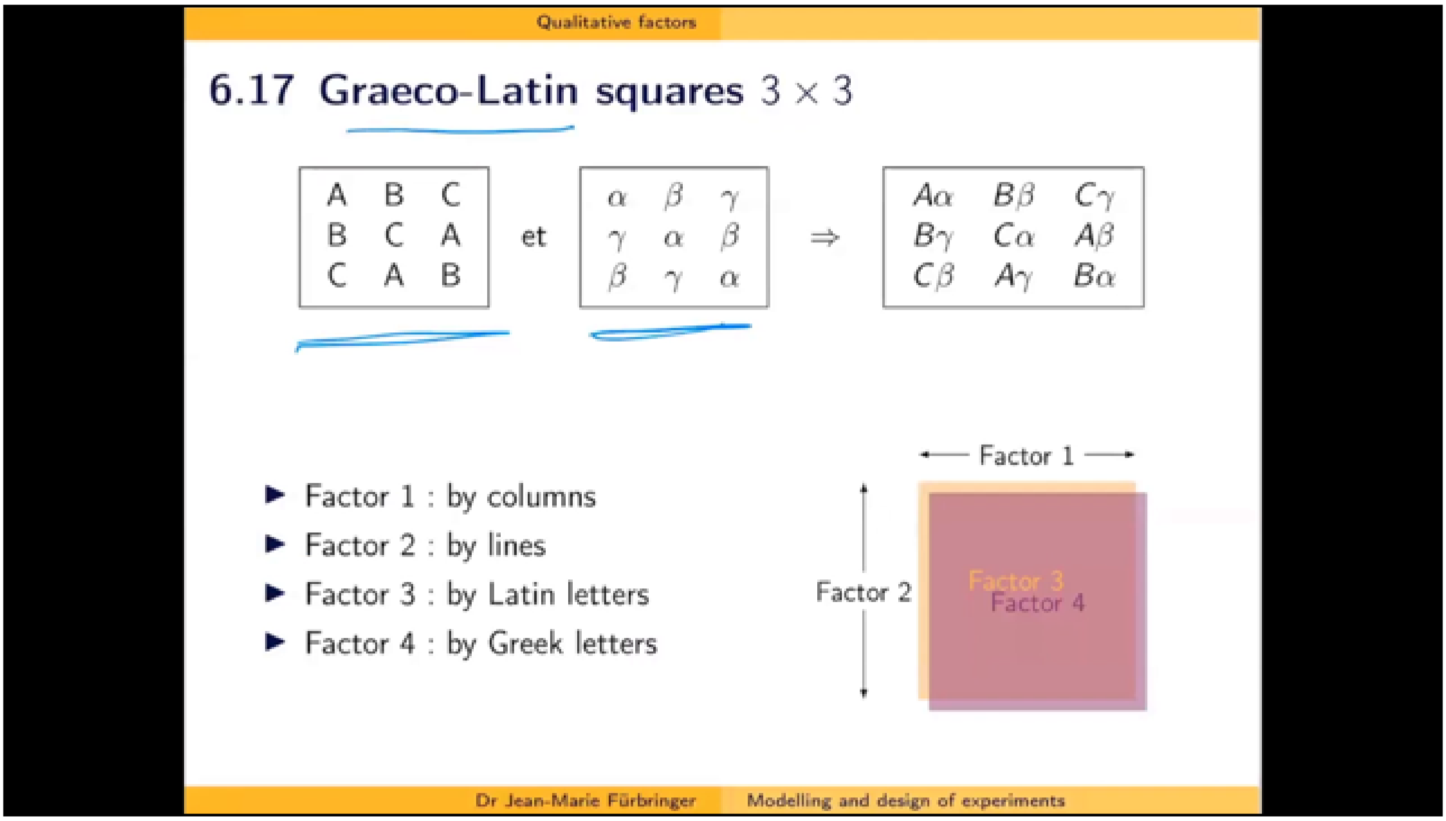


It was very close and it was not evident which one must be the best. So when you see effects that are very close, you have to make some analysis for understanding what is really the best because they are very close. Because they are very close. When you have distance, one is one side of the average, one the other side of the average is quite in the distance is bigger and bigger than the residue. Something different also that you can see with only nine experiments, my degrees of freedom for my residue is no small. I just have two degrees of freedom in my residue because I have nine experiments, seven degrees of freedom in my model. So I just have two degrees of freedom stay for my residue. That's why you have a lot of points. I have nine residues, but I have only three different values and one value is the opposite of the two other value.

notes

summary

15m 52s

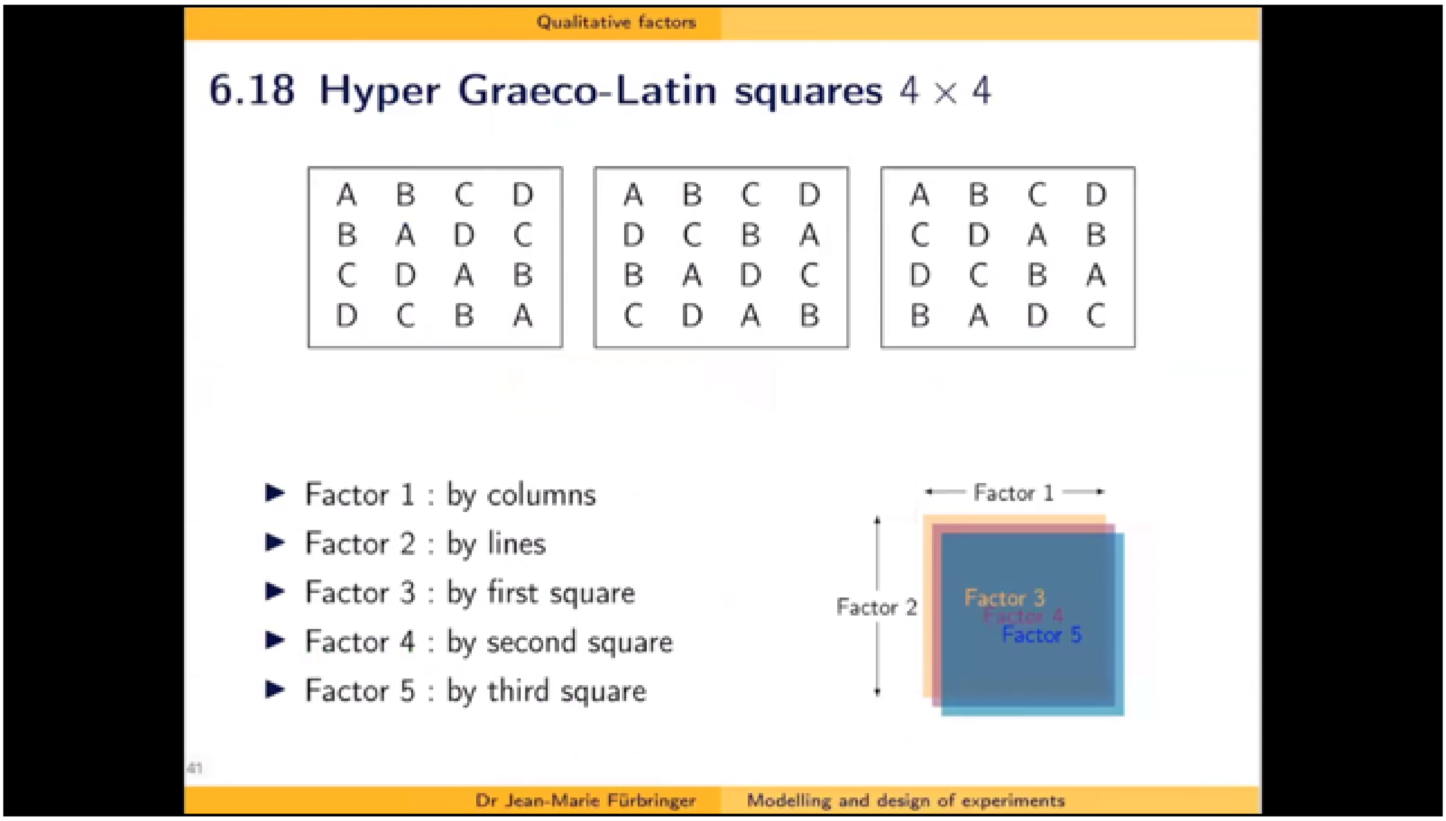


So first, I show you when we use one of these frames. In fact, I tell you that it exists two orthogonal frames. If you change two of the line or two of the row, I could be able to go two of the columns. It will do the same, but let's talk about the row. So if you inverse two of the row, it gives you also another magic square. So one is called Latin because when Euler makes his study, he uses Latin letters. And for the second one, he uses Greek letters. So it's why we are talking about Greco-Latin square. Just remembering that Euler have used Greco. There are nothing related with the people from Rome and people from Athens. So it's just for differentiating our two different frames.

notes

summary

16m 54s

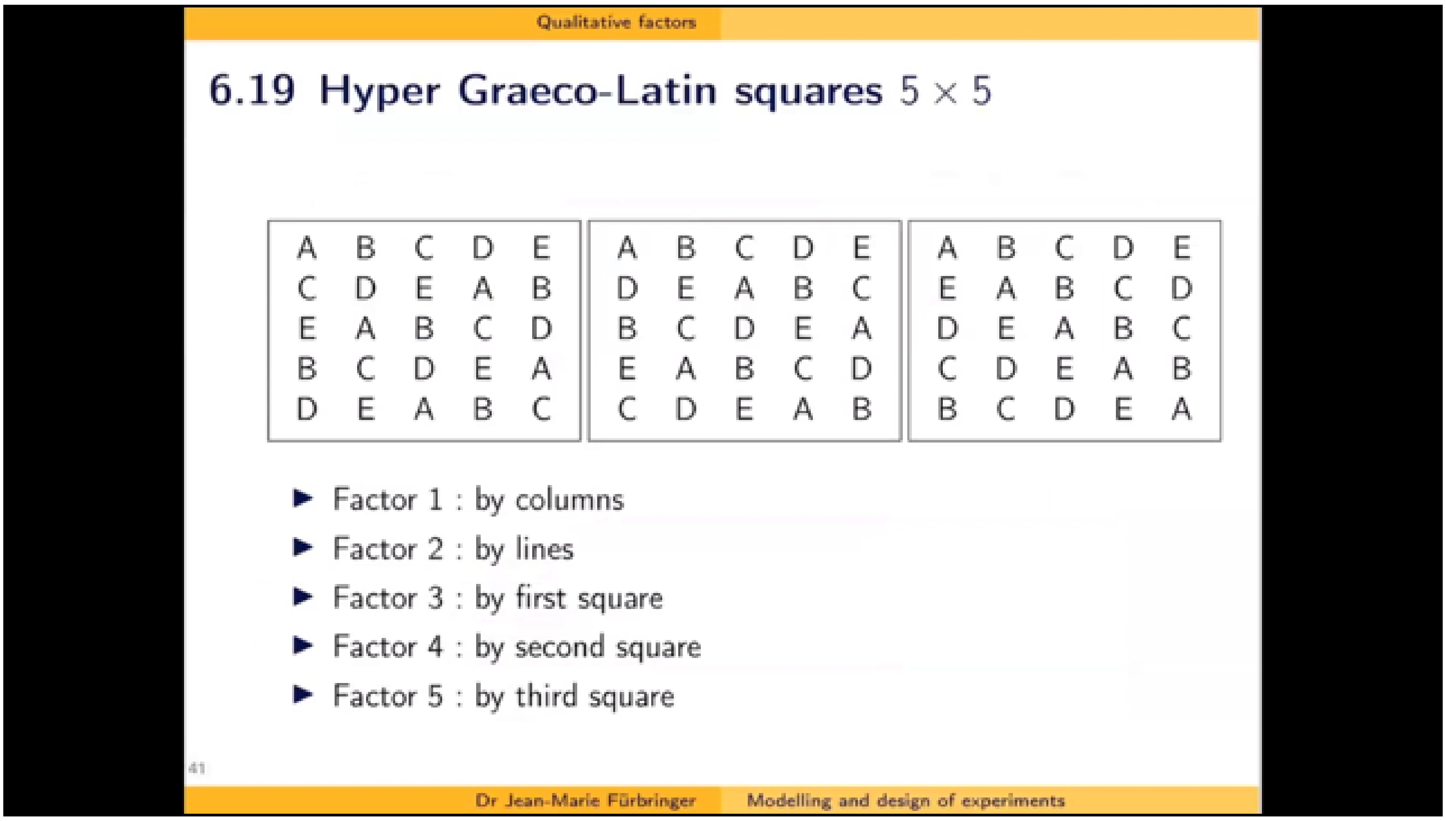


So that means that I would be able to analyze with nine experiments. I would also be able to analyze more factors. I could have, for example, different type of material, one in copper, one in steel, and one is aluminum, for example. So with nine experiments, I can estimate for three levels, I can estimate the main effects till four factors. One by column, one by lines, one by Latin's letter, and one by Greek letter. And so this is a Greco-Latin square.

notes

summary

17m 49s



But in this Greco-Latin square, the numbers of levels are important. So if you have four levels, you can do the same, but even you have more choice. If you have four levels, you can have, in the combinator, is more interesting. We have, in fact, three different frames that are orthogonal. You see usually the first lines is the same. And after, we have three other rows, and you have then three different ways of organizing the other rows. And each one is orthogonal to the other. So that means that I can have five factors. I can have one factor by columns, one factor by lines, and after, I can have one factor for each of my frames. One for the first square, one for the second square, and we used to call that Ypres Greco-Latin square. And you see that usually in this situation, we forget the Greek letters because it became too complicated. Or you can have another alphabet. You can use Arabic alphabet or Chinese alphabet for the third one, but it became quite complicated. So usually we forget the, what we called, Ypres Greco-Latin square. In this case, with 16 experiments, because four by four makes 16 experiments, I can use still five factors. And I can go up.

notes

summary

18m 29s

6.20 ANOVA for 27 and 9 runs

27-experiment set :

Source	SS	DF	MS	F	p
Constant	12'465.7	1	12'465.70		
Machine	264.5	2	132.23	109.3	0.000%
Drill	511.0	2	255.50	211.19	0.000%
Operator	5.2	2	2.62	2.2	14.1%
Residue	24.2	20	1.21	1	
Total	13'270.6	27			

9-experiment set :

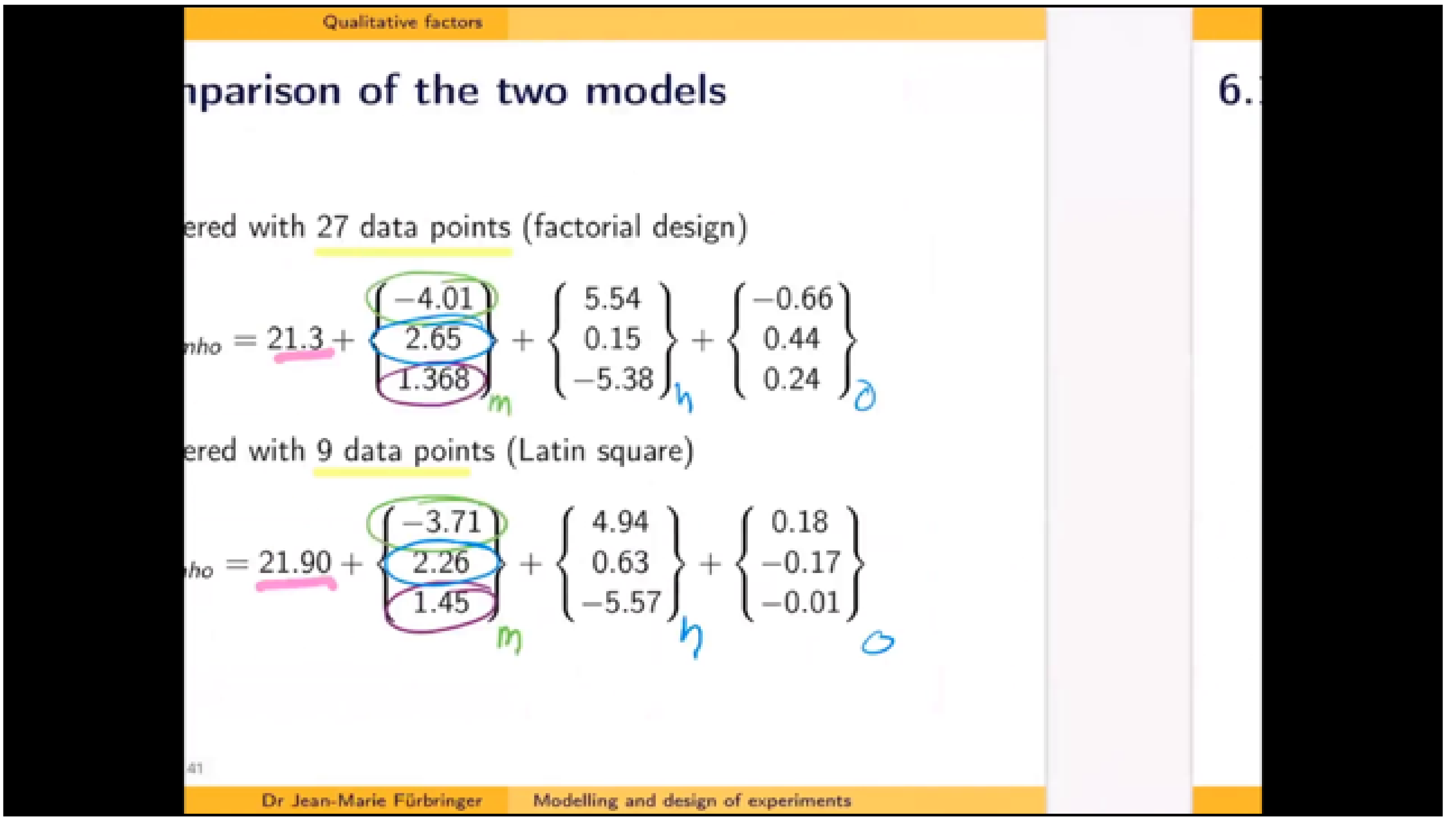
Source	SS	DF	MS	F	p
Constant	4'064.9	1	4'064.91		
Machine	101.1	2	50.55	113.7	0.000%
Drill	173.2	2	86.61	194.7	0.000%
Operator	1.5	2	0.74	1.7	21.5%
Residue	0.9	2	0.44	1	
Total	4'341.6	9			

With five, you can go also to five factors. But to six and seven, you can even go higher. So you have magic square of three by three, four by four, five by five, six by five, seven by seven.

notes

summary

20m 18s



Now it's important to make an ANOVA with this data for qualifying correctly the statistical quality of our, because it's nice to make a dot plot, but the dot plot is just a representation. It's not definitive. So we need to make an ANOVA of that.

notes

summary

20m 38s

6.20 ANOVA for 27 and 9 runs

27-experiment set :

Source	SS	DF	MS	F	p
Constant	12'465.7	1	12'465.70		
Machine	264.5	2	132.23	109.3	0.000%
Drill	511.0	2	255.50	211.19	0.000%
Operator	5.2	2	2.62	2.2	14.1%
Residue	24.2	20	1.21	1	
Total	13'270.6	27			

N-1

7

9-experiment set :

Source	SS	DF	MS	F	p
Constant	4'064.9	1	4'064.91		
Machine	101.1	2	50.55	113.7	0.000%
Drill	173.2	2	86.61	194.7	0.000%
Operator	1.5	2	0.74	1.7	21.5%
Residue	0.9	2	0.44	1	
Total	4'341.6	9			

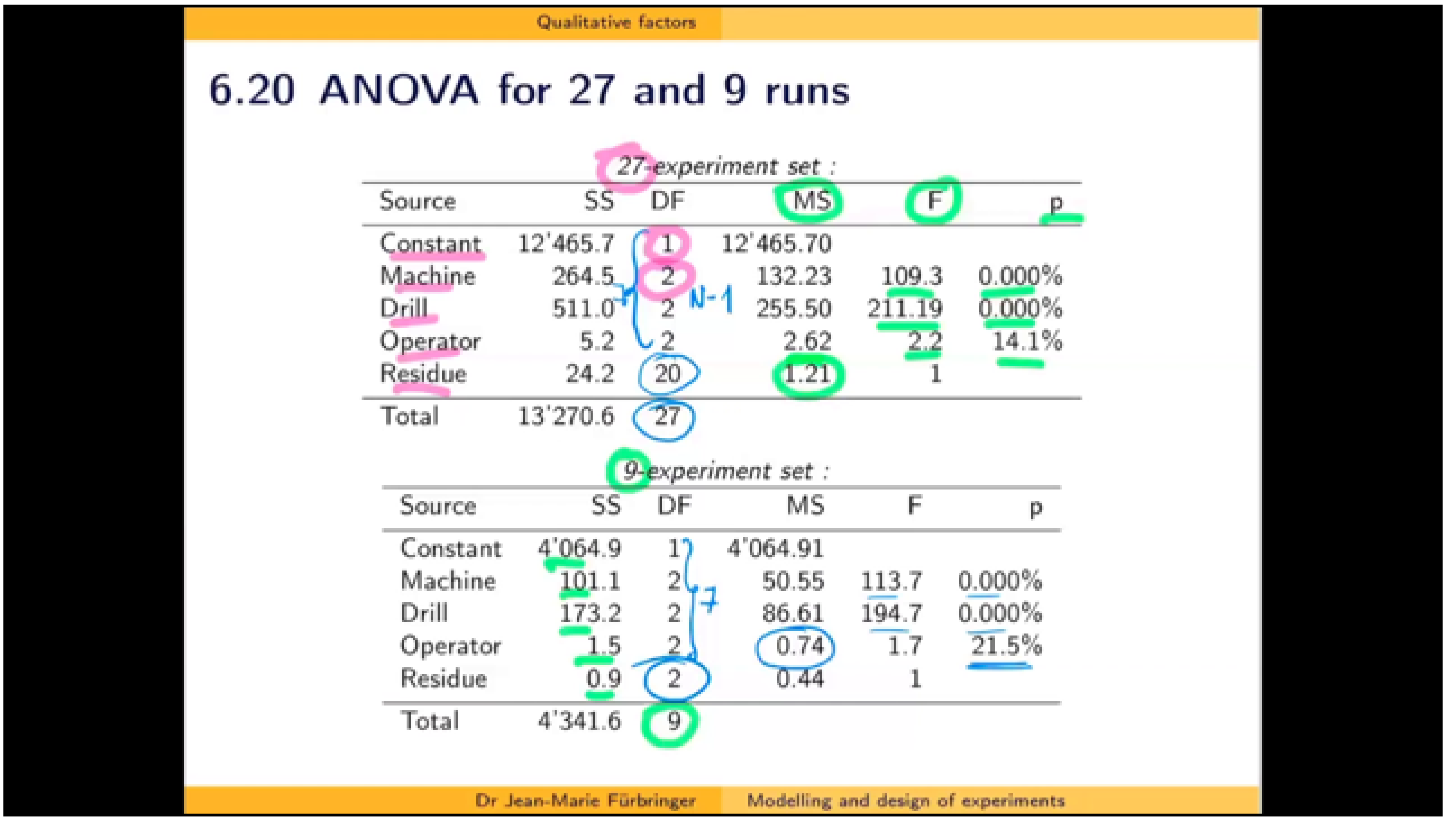
7

So the ANOVA is quite easy because when you have your decomposition, in fact, the ANOVA the square will be the square. So the square will be the sum of the square of all the numbers that you have in one of those frames. So if we do the work for the 27 experiments, we will get one line for the constant, one line for the machine, one for the drill, one for the operator, and one for the residue. And it's exactly the sum of the square of each of the corresponding frames. And the residue being the last residue that you calculate. And the degrees of freedom. So in this case, I have one degree of freedom for my constant. I tell you it's a vector which direction is fixed. It's just the length of this vector, the ground mean that can change, but the orientation of this vector cannot change. The degrees of freedom for my factor is the numbers of level minus one. Because I tell you that if I have three effects, one effect is dependent of the other because the sum of the effect is zero. If you remember in the linear parametric models, it was one per line, the degree of freedom, because I have one coefficient. Now I have three numbers. It's not three. It's three minus one. And so the same structure, sum of square, degrees of freedom. And so the degree of freedom for the residue is the degree of freedom of my model, seven, which is subtracted to the total number of data, total degree of freedom. So it's make 20. And after you can calculate the mean square and you see rapidly that you have some big number and some numbers that became very, very small. And after we have the Fisher ratio, which is the ratio was the mean square of the residue. And then I can

notes

summary

21m 3s



6.20 ANOVA for 27 and 9 runs

27-experiment set :

Source	SS	DF	MS	F	p
Constant	12'465.7	1	12'465.70		
Machine	264.5	2	132.23	109.3	0.000%
Drill	511.0	2	255.50	211.19	0.000%
Operator	5.2	2	2.62	2.2	14.1%
Residue	24.2	20	1.21	1	
Total	13'270.6	27			

N-1

9-experiment set :

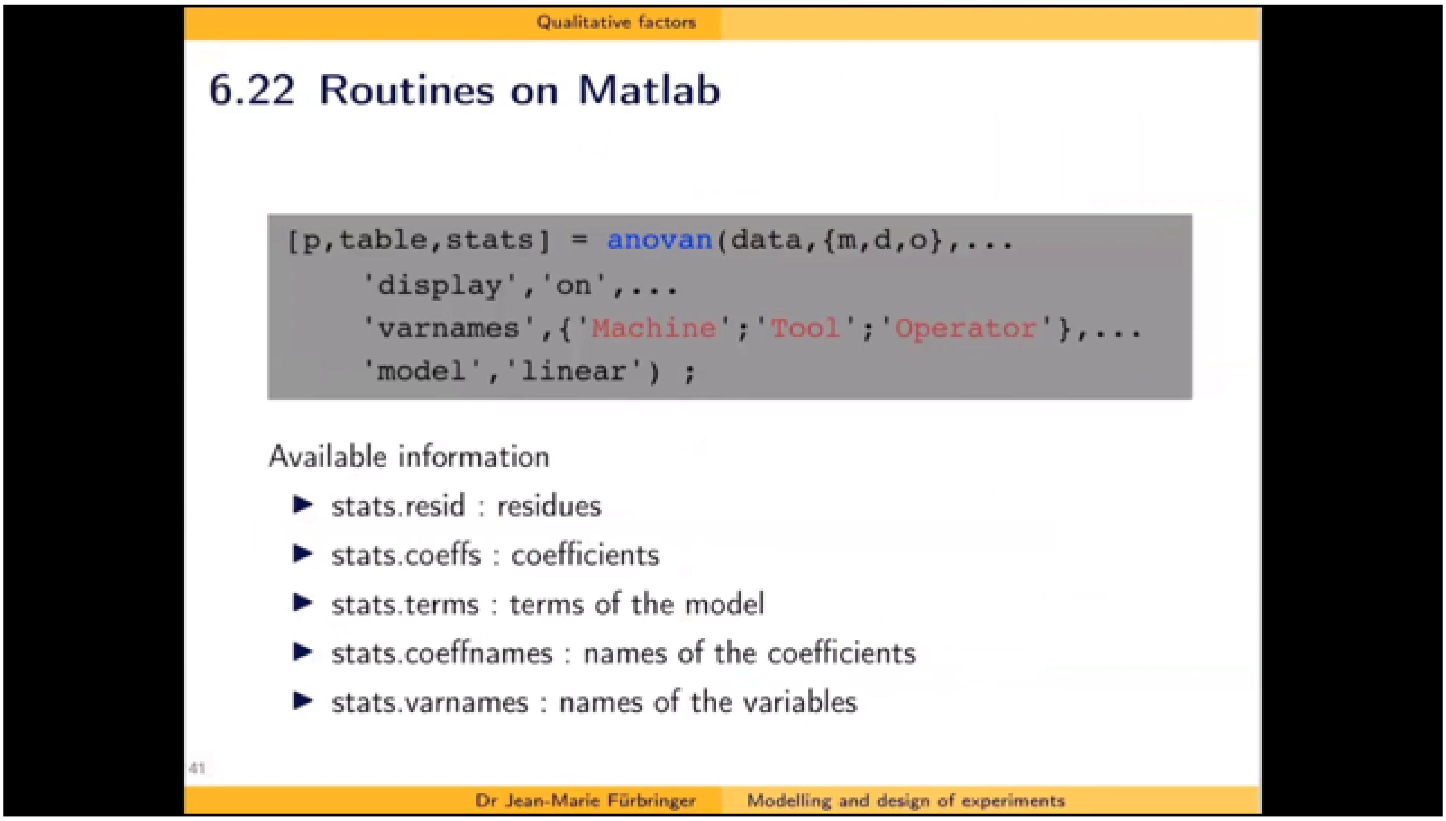
Source	SS	DF	MS	F	p
Constant	4'064.9	1	4'064.91		
Machine	101.1	2	50.55	113.7	0.000%
Drill	173.2	2	86.61	194.7	0.000%
Operator	1.5	2	0.74	1.7	21.5%
Residue	0.9	2	0.44	1	
Total	4'341.6	9			

7

calculate my Fisher ratio. And then I can calculate my p values. And you see that I have p values very small for my two first effects. And I have p values of 14 for my operator. Definitely the effect of my operator is not significant. Good. No problem with the union. Now we can see the difference when I have nine experiments. So sum of square are all smaller because I have less data. I have three times less data. The total degrees of freedom is no nine. For my model is exactly the same as seven. I'm not changing my model. Where I have a big change now is in my degree of freedom for my residue, which is only two. I can calculate the mean square. You can observe that the mean square for the operator is still worse than before. And I'm able to calculate Fisher ratio quite the same level as before. But as the degrees of freedom finally doesn't change nothing for the p values of the significant. It will change something but not in the accuracy I was looking. And I can see that definitely my operator is not good. So the first frame is more precise than the other. But you see that I'm arriving to the same conclusion with one third of the effort.

notes

summary

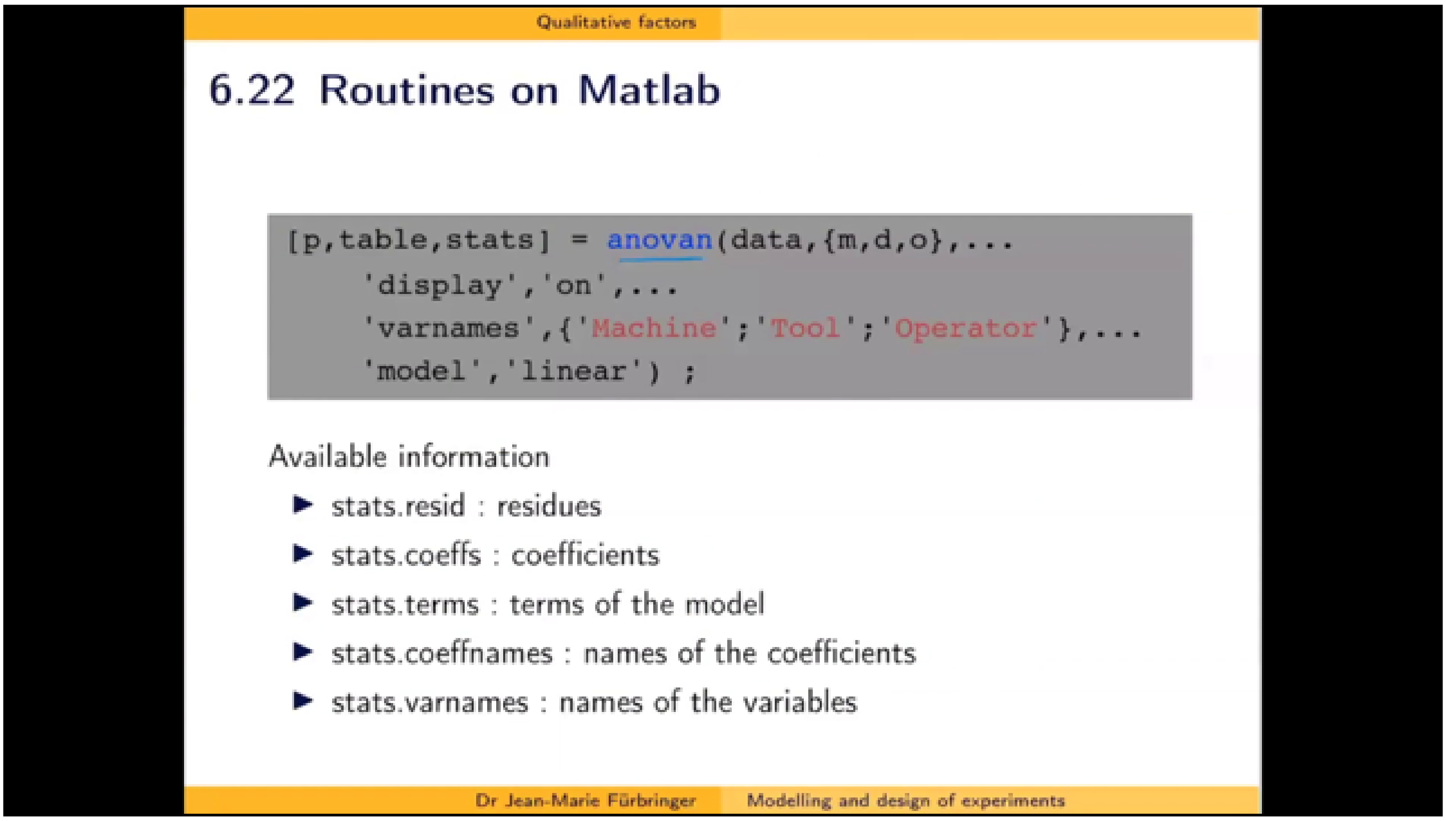


The big difference is that here I could try with 27 experiments, I could try to see if I have interaction with the nine experiment. I cannot check if I have interaction or not. So it's a little bit the same thing as the difference between Adamar and factorial design. Okay.

notes

summary

25m 1s

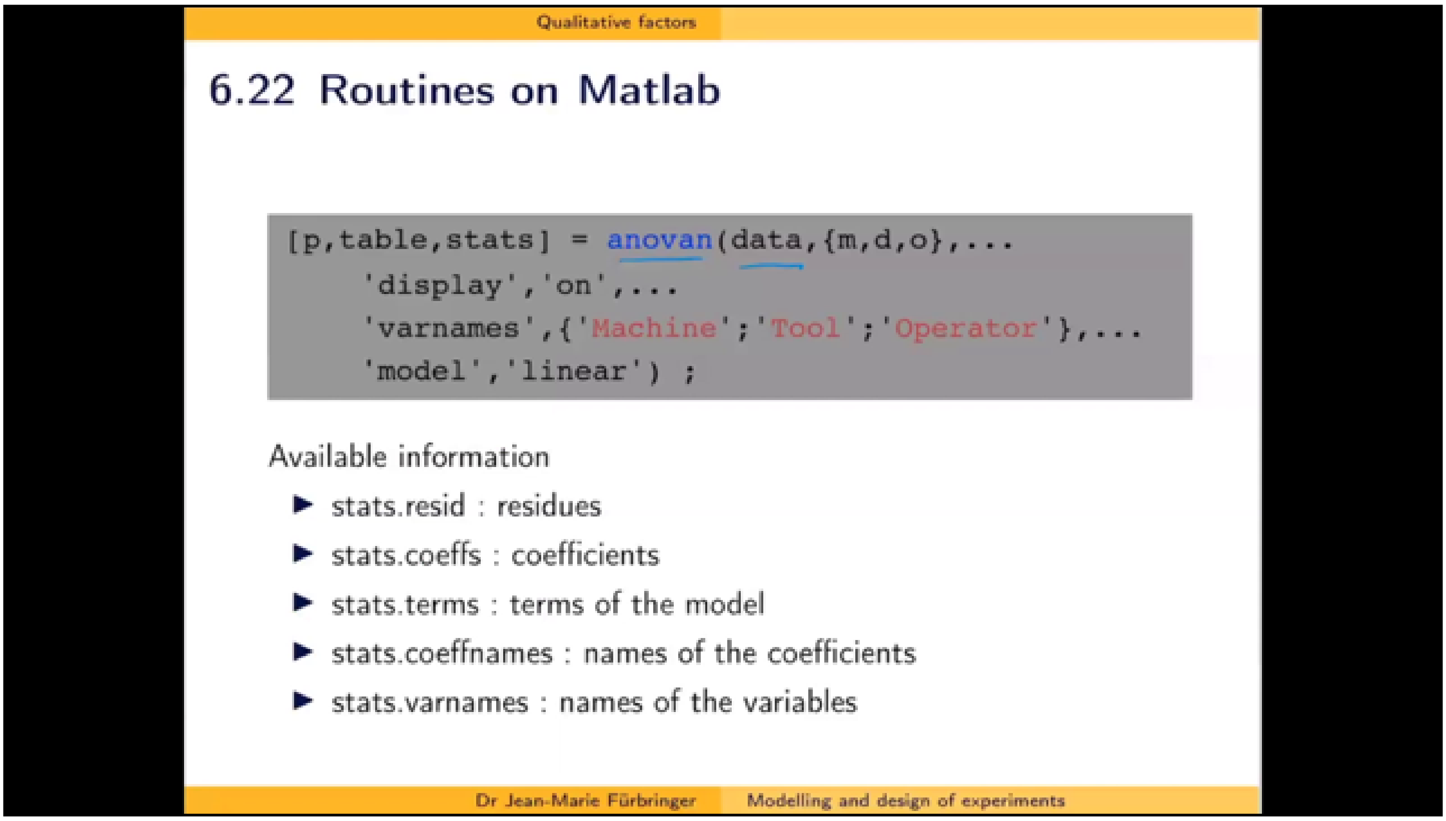


So in MATLAB you have some routines that make the work for you. So there are three routines. I show you the ANOVA N. So usually you have in all libraries in Python you will find the same.

notes

summary

25m 23s

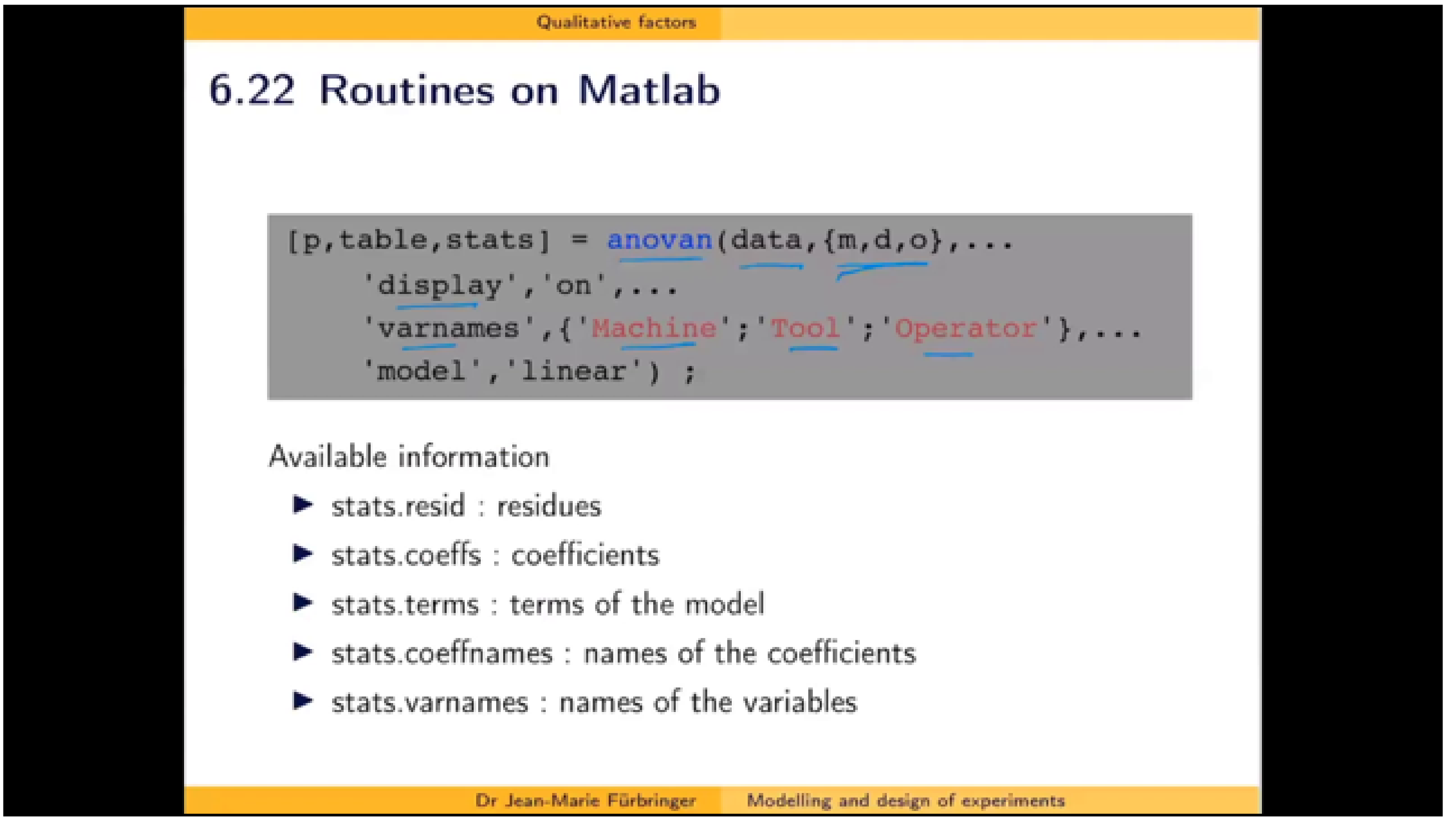


So you have ANOVA 1 for doing the ANOVA of one variable, ANOVA 2 when you have two variables, and ANOVA N when you have more than two variables. So for three variables we have to choose three variables. We have to choose ANOVA 1.

notes

summary

25m 37s

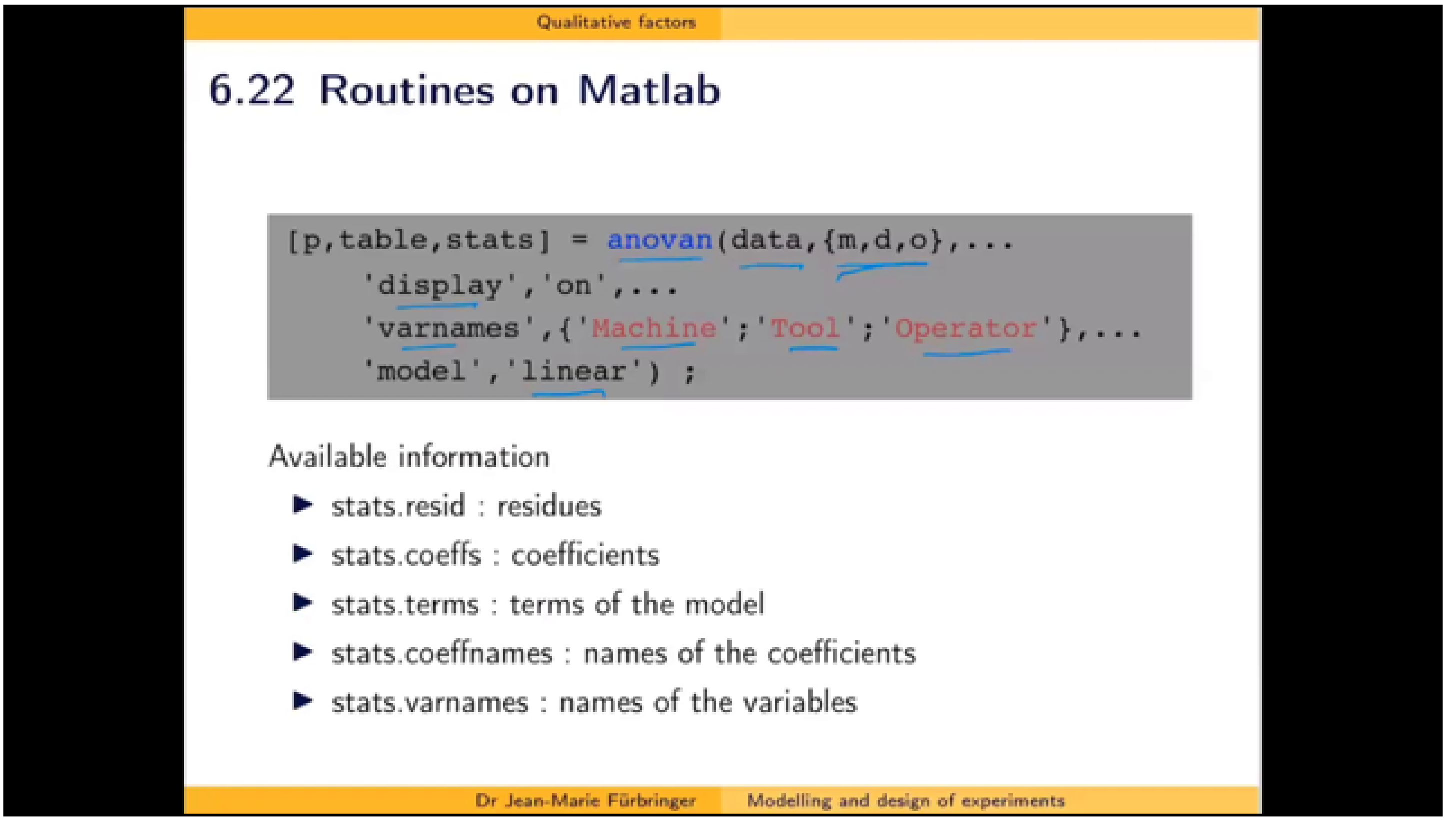


And so the ANOVA N as input you have to give your data and your data can is a table or it could be just one vector. You have to tell the algorithm where is what. So it's the meaning of this description of machine drill and operator. You have to indicate with eventually characters of level where what is the data, what is what. And after you can have the display on or off if you want to have the results print on the screen or not.

notes

summary

25m 58s

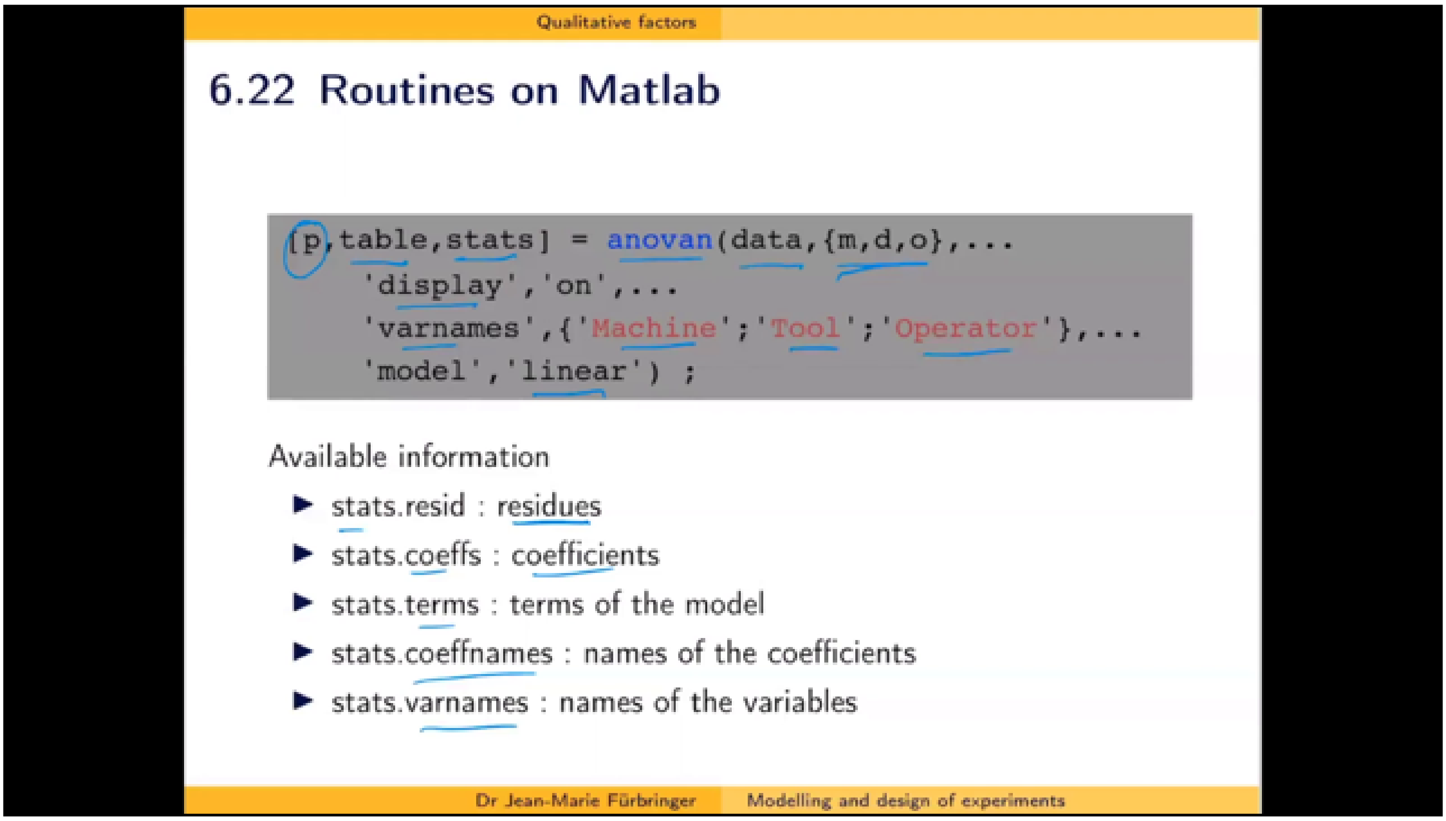


And you can define your variable names and you can decide the type of model that you want. You have just the choice between linear and interactions.

notes

summary

26m 40s

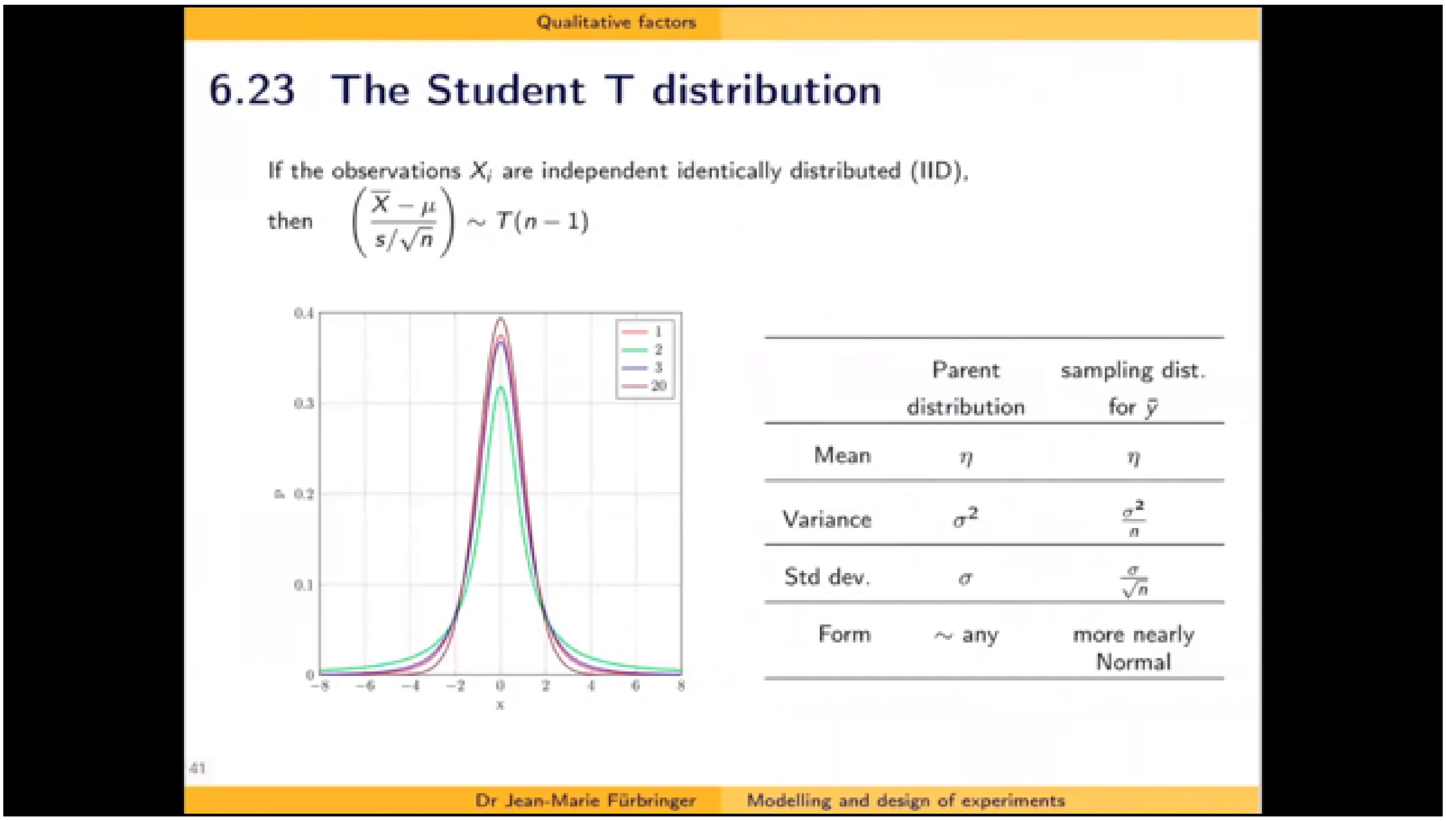


You do not have second degree when you have this type of model. And as output you have the p-value, you have a table. I will show you the form of the table and you have a small database with the different calculations that the algorithm have done on your data. And after you can question. So you have the residue in the data. So you get it by start dot residue give you the residue. You have the coefficients. You have the terms of your model. You have the names and you have the name of the coefficient and the name of the variable as output.

notes

summary

26m 53s

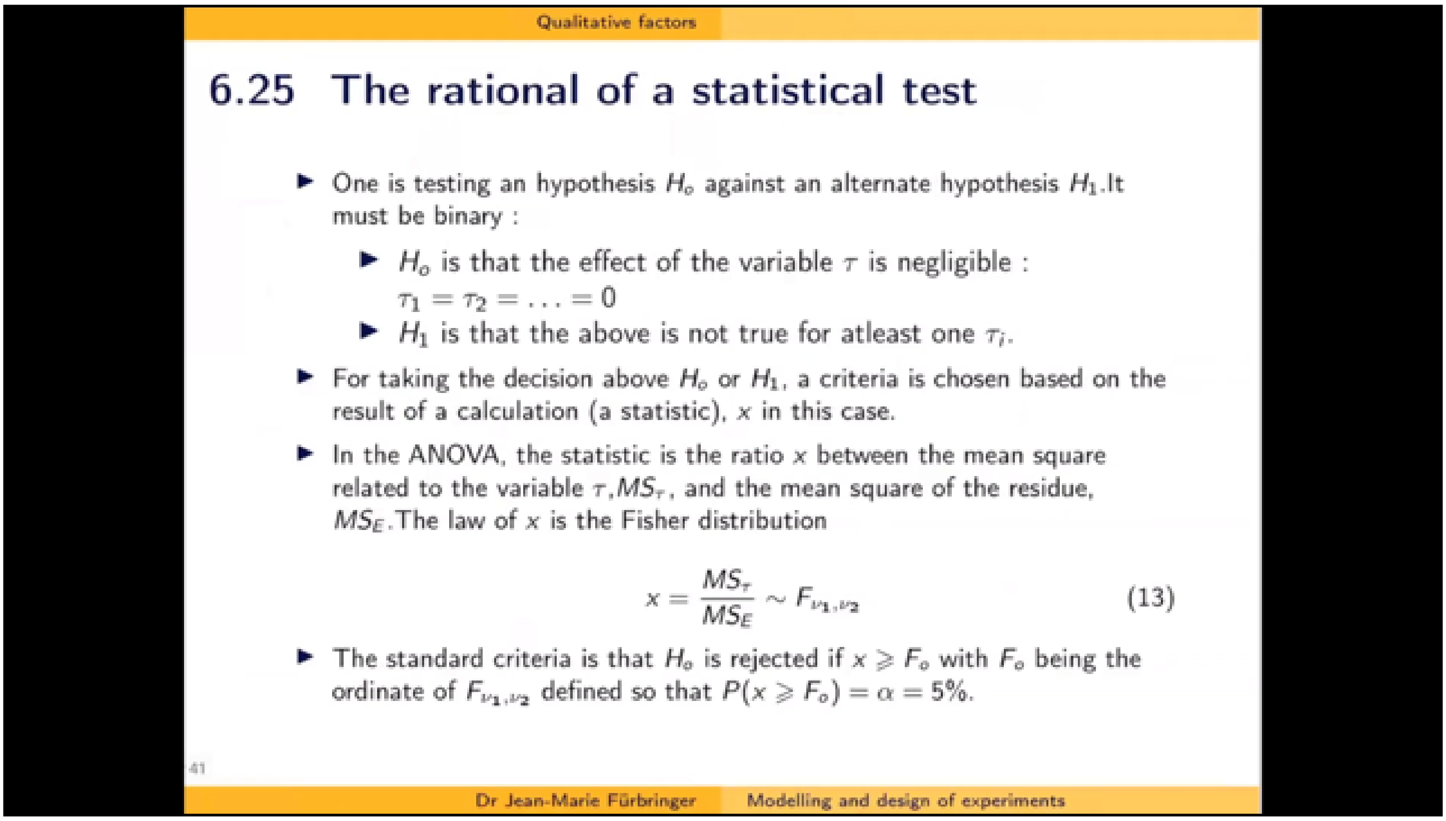


So do I have.

notes

summary

27m 37s



So this is typically what you have as an output. It give you an sort of ANOVA table. It's an ANOVA table with the different sources, machine, tool, operator error. So again the constant is forgotten. You have the sum of squares. You have the degrees of freedom. So here you have 26 as degrees of freedom because you have subtracted the degrees of freedom for the constant. And you have the mean square and you have the results corresponding to what I present to. And here you have the structure of the stats database. So I will stop here.

notes

summary

27m 39s