Course material

Course:

ENG606 / PHYS 442

Video:

DOE_lesson12_part1_CCM

Concepts (extracted from automatically generated subtitles):

Constant coefficient model. Main effect. Interesting case. Type of model. Types of machines. Different categories. Qualitative factors. Grand mean. Important factors. Different types of machines. First strategy. Different strategies. New type of model. Possible experiments. Best machine.

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

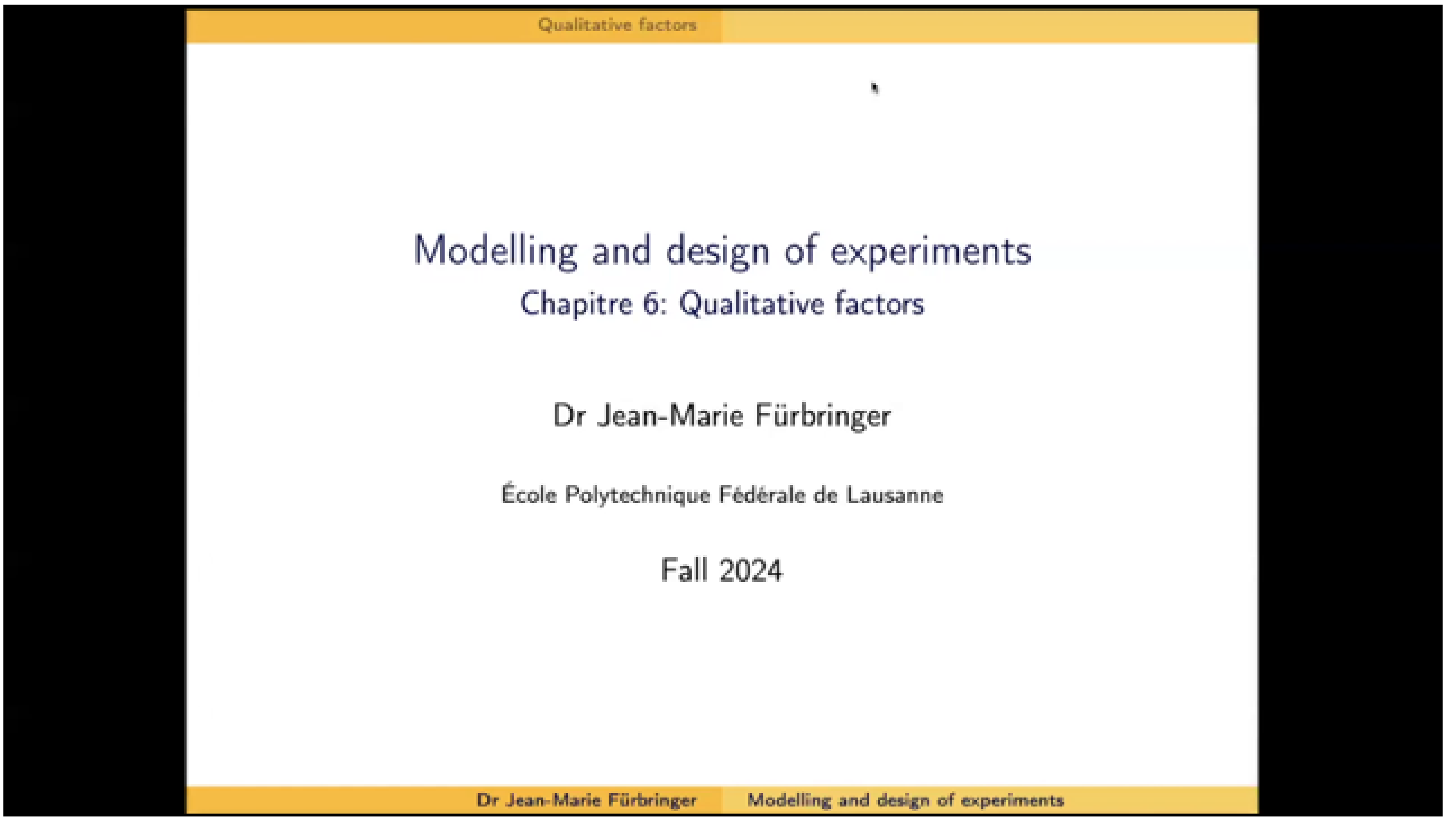
to video

...

notes

summary

0m 0s

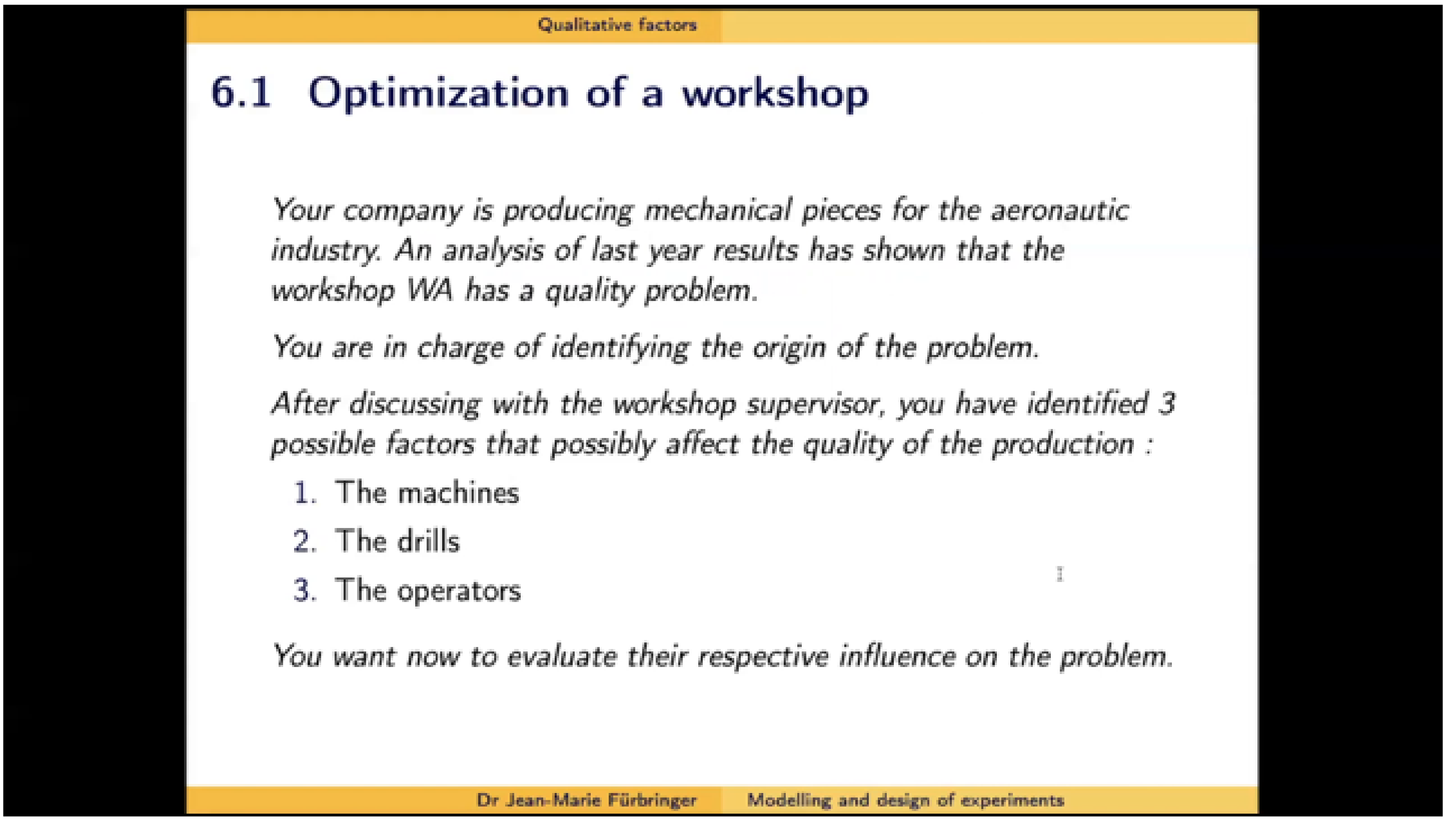


These subtitles have been generated automatically Okay, so the lesson of this week is about the qualitative factors.

notes

summary

0m 1s

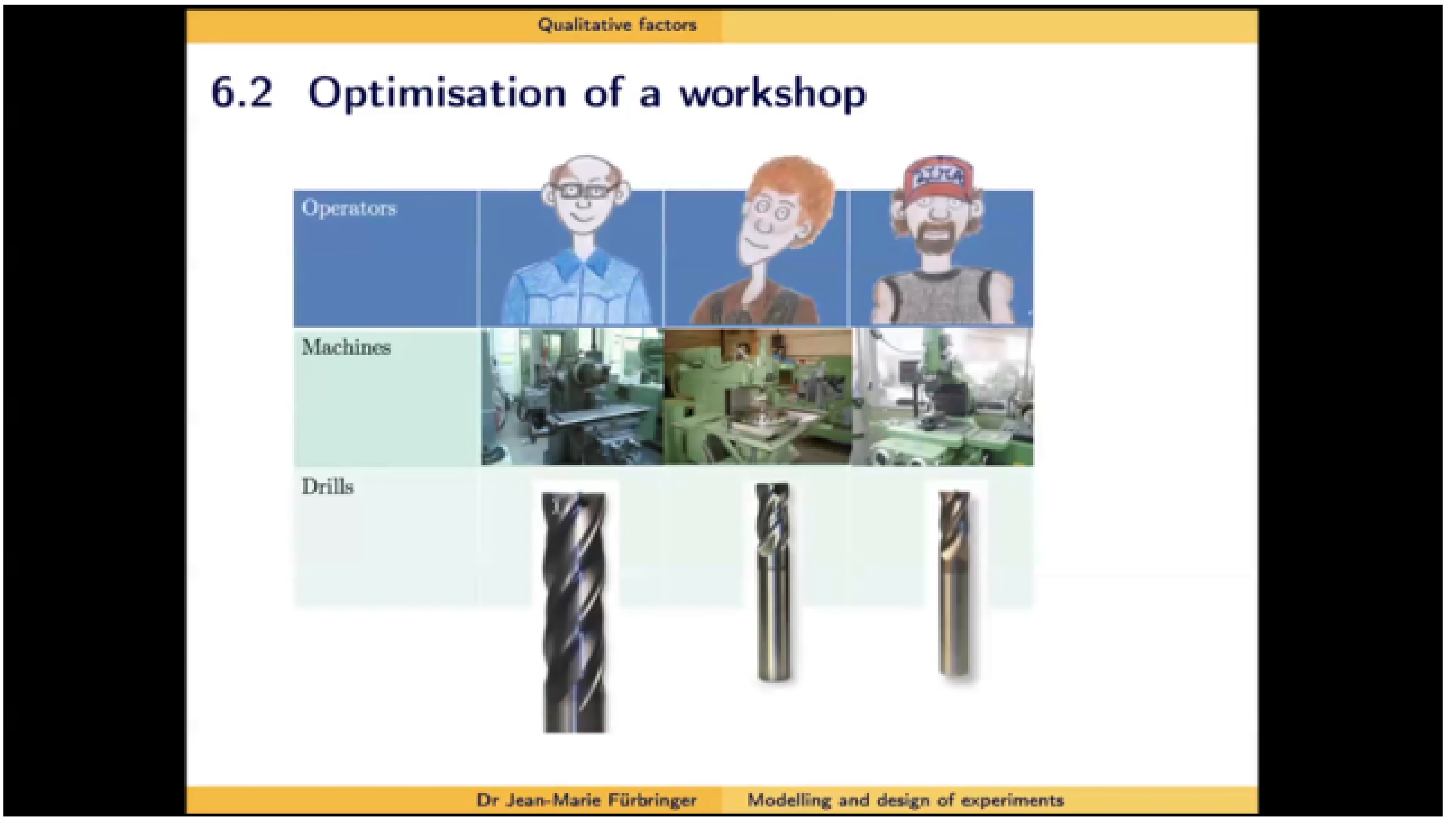


That means we mention qualitative factors when your factors are not represented by a series of numbers when you have different categories. For example, if you are testing some glues,

notes

summary

0m 5s

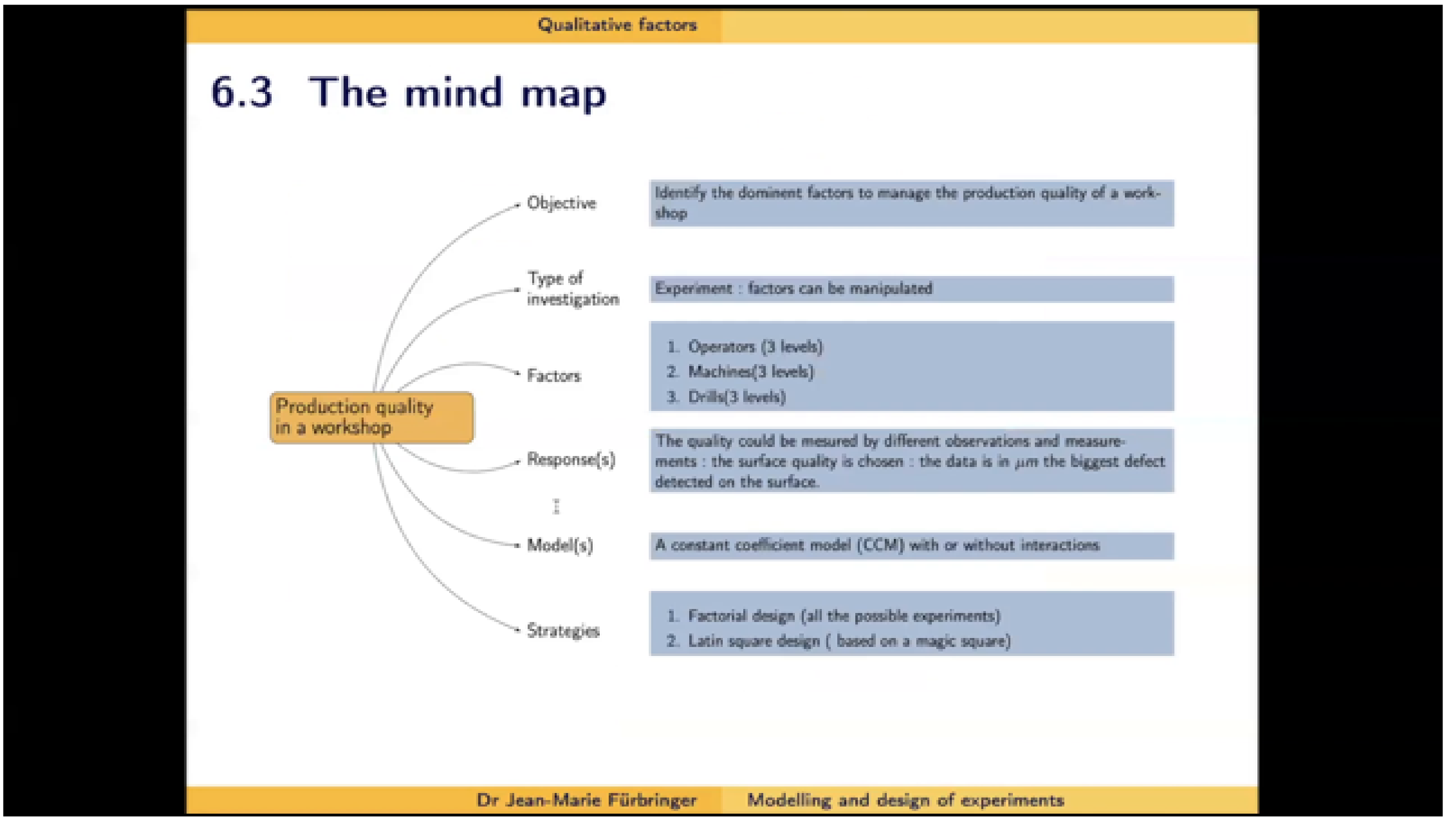


you have the glues A, the glues B, the glues C. So we will treat this new chapter with an example, an example of a workshop. Imagine that you are working in a workshop which is producing mechanical pieces for the industry and it has been detected that one of the workshops are problems. So you have discussed with the specialist and you arrive with the conclusion that in fact the problem can come from the machines, from the drills, and from the operators. And so you are interested to investigate to see where the problem is coming from.

notes

summary

0m 37s

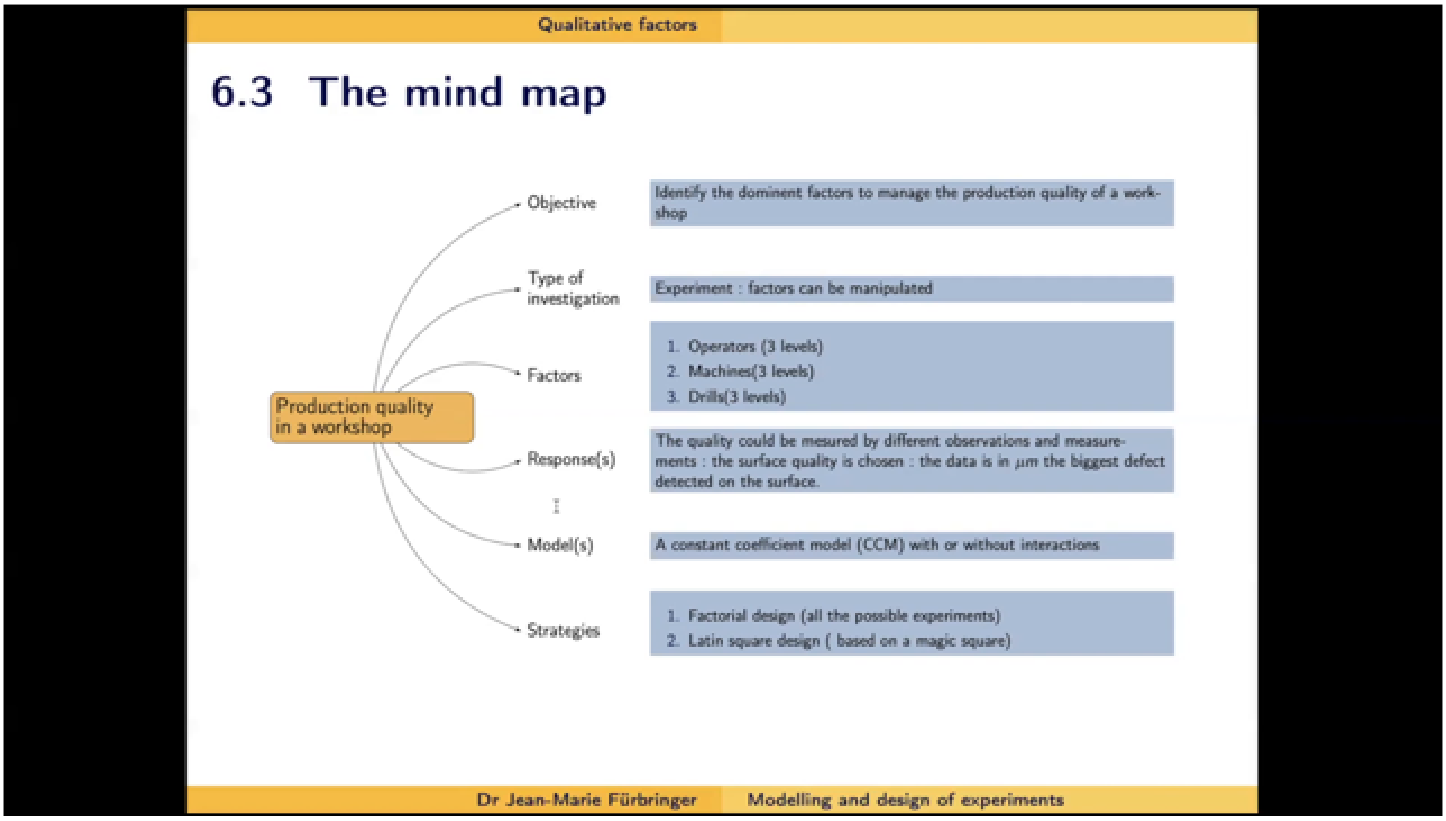


So just for the sake of having an interesting case and enlarging also this case for the following of the chapter, let's imagine that in this workshop we have three operators, Peter, Lewis and Charlie. You have three types of machines, three different types of machines. And for producing the pieces that have a problem, it's possible to use three different drills. So you see that we have three factors, each one was three levels.

notes

summary

1m 25s



So now this is the mind map presenting the gays and presenting the factors and presenting the

notes

summary

2m 12s

6.4 Factorial strategy

Table – Biggest defect detected on the surface in μm.

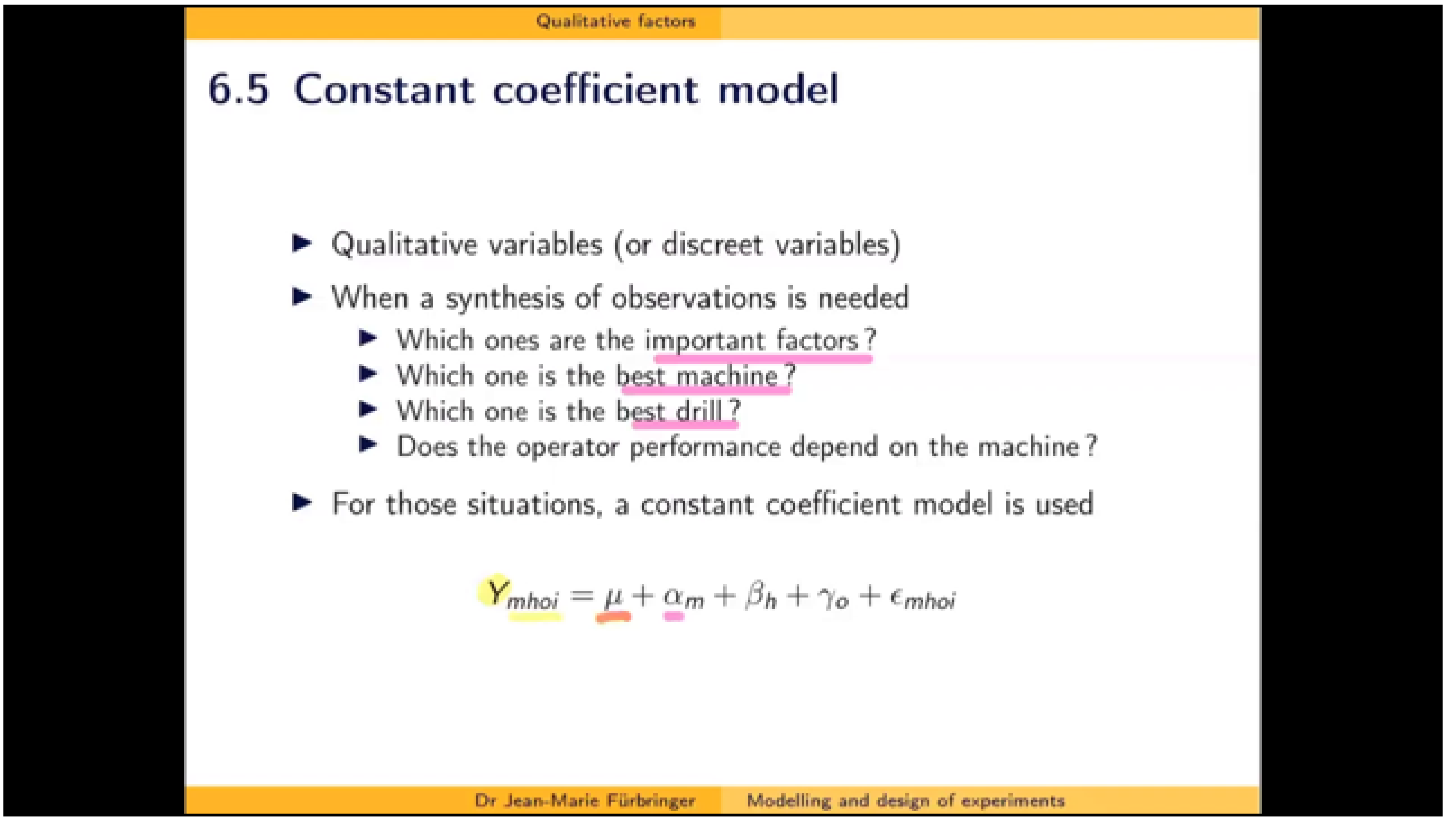
Operator	Drill	Machine		
		Deckel	Schaublin	Maho
Charlie	1mm	23.46	28.46	27.06
	5mm	17.66	24,07	23.79
	20mm	12.33	19.00	16.63
Pierre	1mm	24.22	28.56	28.06
	5mm	18.87	25.32	22.24
	20mm	10.85	19.21	17.75
Louis	1mm	22.55	29.69	28.50
	5mm	16.93	24.92	24.99
	20mm	12.25	17.24	16.72

different strategies that we wanted to analyze. We are interested in evaluating, identifying the dominant factors. We could eventually also be interested to know if they have instructions. Do some operators work better with some of the tools or better with some of the machines? Do some tools are more adapted to some of the machines? You see all the possibility. But in this case we are really interested in not only discovering each factor, which factor is important, we are also interested to understand which level of the factor are the most interesting or are the ones to avoid. See, this will make a small difference. So for that, the type of model that we will work with are a new type of model. It's what we call constant coefficient model. And I will present you two strategies. We will try to do things with what we call the factorial design. It's not a factorial design at two level of three level. It's just a factorial design. That means we will make all the possible experiments with all the levels that we have. And I will have to show you an optimized design, which is called a Latin square design, quite very interesting design to discover. You will see that you have the Latin square design. We have the Greek, Latin, Greco-Latin design. We have the hyper Greco-Latin design. We have different theory, but the philosophy is the same as the Soudoku. It's exactly the same mathematical structure than the Soudoku. And we will see how we can use that in design of experiment.

notes

summary

2m 29s

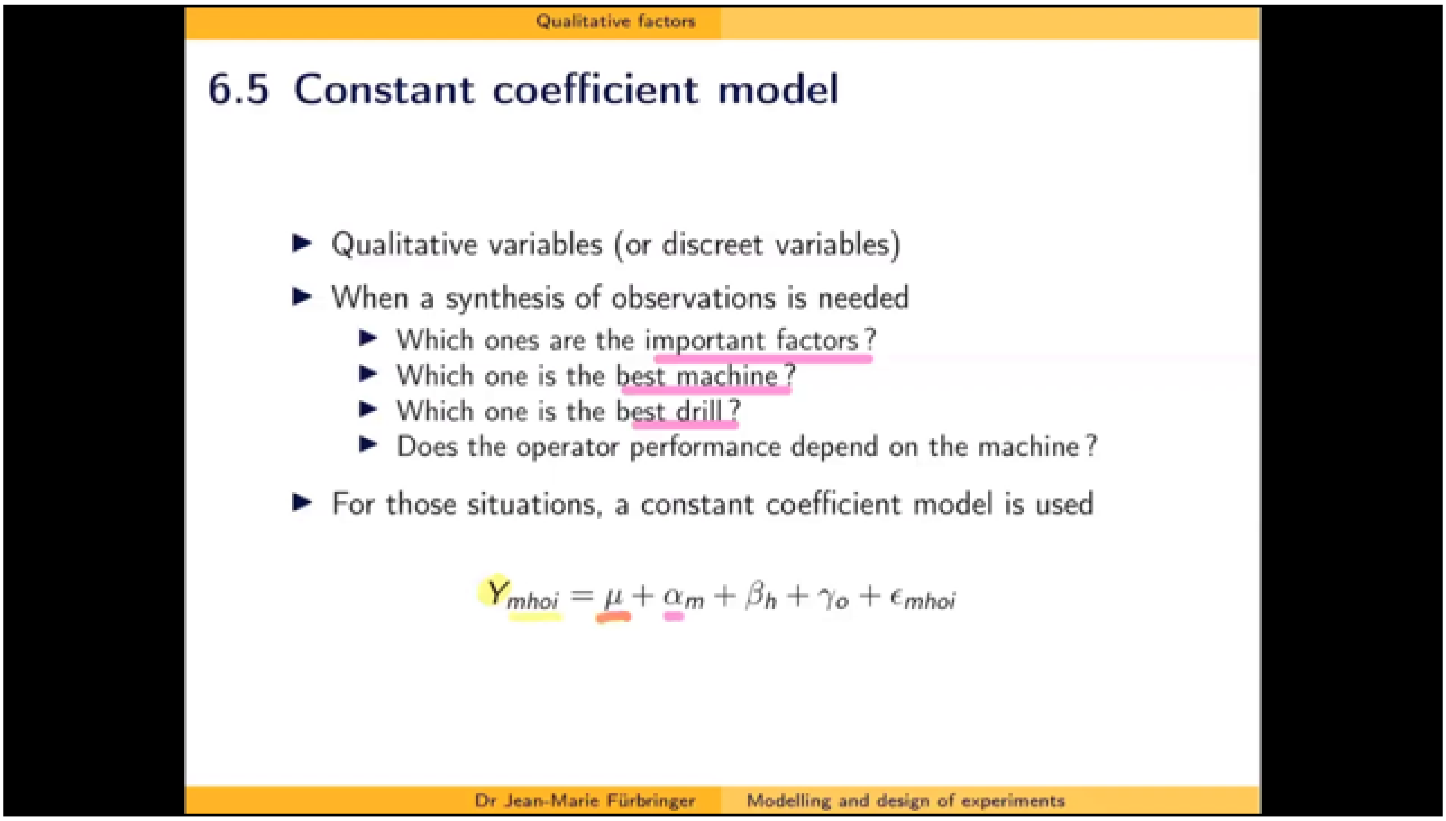


So the first strategy I have proposed you to do is to make a factorial strategy. So in that sense, factorials mean that we do all the possible triplets. We want that all the operators work with all the machines and with all the drills. So you see that here you have one result. 23 microns is what have been measured by the quality units. They look at pieces produced that way by Charlie with the one millimeter drill and the decal and discover that the highest defect on the piece was of 23 microns, 23.46. And like that, you see that Charlie have worked with all the drill and all the machine. So for Charlie, it makes nine measurements. So it's again, it's an experiment because we have asked Charlie to do these pieces, etc. It could be observation if we just have take within the work done samples of the work. Now it's an experiment because we have planned to do exactly this by random. It's better, etc. So you see now that you have the data is now here. This is my data. And what was in the chapter before matrix of experiments, the experiment is described here, but it's not in the form of a matrix. I don't know. When you just look at the data, would you be able to see which factor is important, which factor is not important? Okay, you can see something because you just have three levels. And even you are not choosing the instruction, you cannot see them. So in fact, this type of model is exactly for that is for analyzing your data is also have predictive capacity, but mainly here's about analyzing your data. So as already mentioned, is the type of question we would be interested to answer would be what are the important factors are the three factor important are one factor more important than the

notes

summary

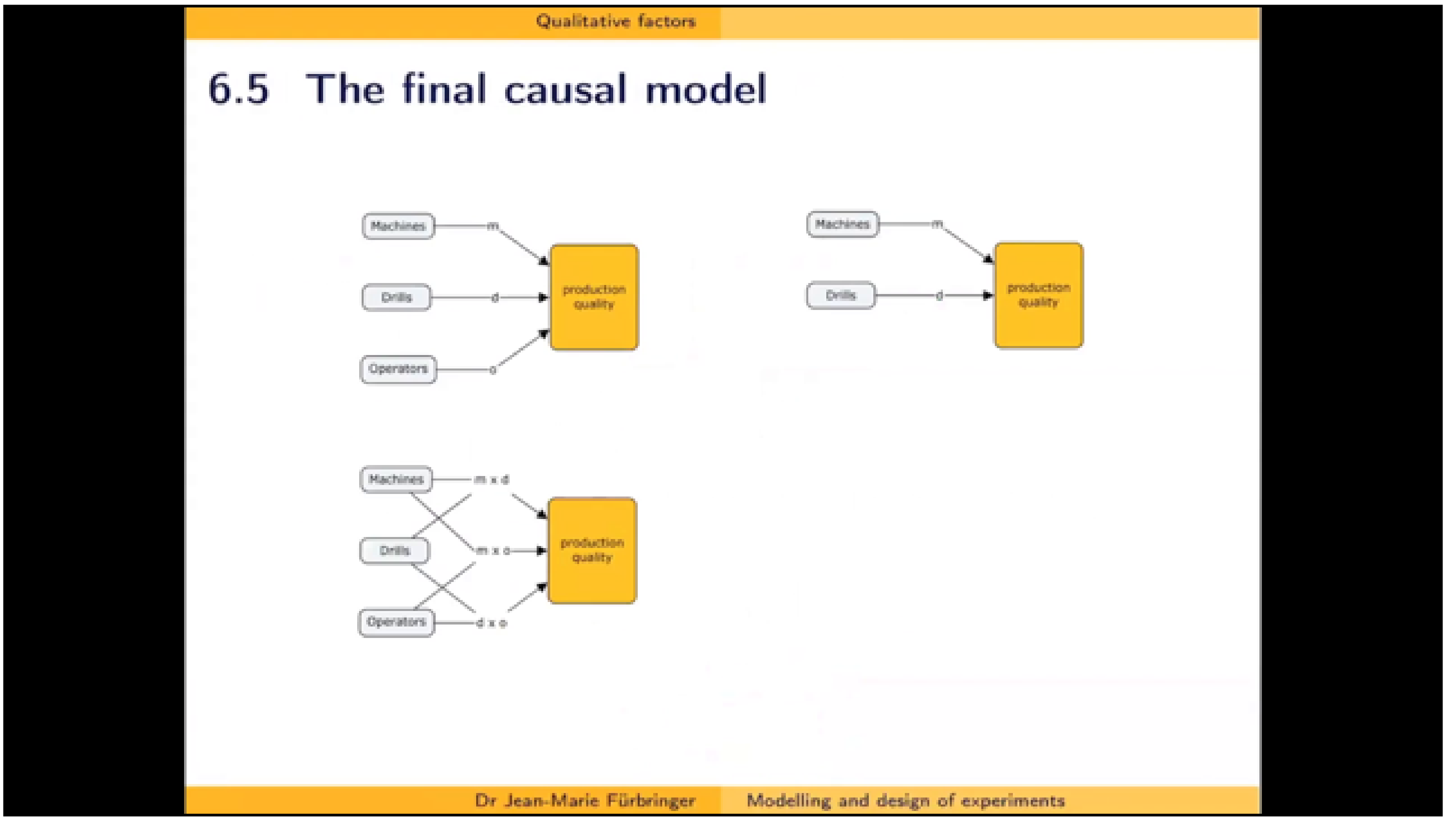
4m 37s



other? Eventually, do we have a best machine? We are perhaps interested to invest in the workshop and choose one of the machines that probably produce better or perhaps is the drill that make that we have to make some recommendation that for this type of pieces, it's better to use this type of drill. The influence of your analysis could be quite diverse. And we will answer to these questions making a constant coefficient model and you have here a constant coefficient model. So we say that we have one answer. In this case, it's the defect, the highest defect that we can detect on a piece. It could be another thing. And we have four indices, one M for the machine, H for the tool, because I start doing this exercise in Spanish. So it was herramienta, O for operator and E because you can have replicates, so it represents the numbers of replicates. And so we will represent the data was a grand mean, mu, which is an information which is the same for all the cases. And after we will have one coefficient, it's why we say constant coefficient, we will have one coefficient for the machine. So we will have three level in this

notes

summary

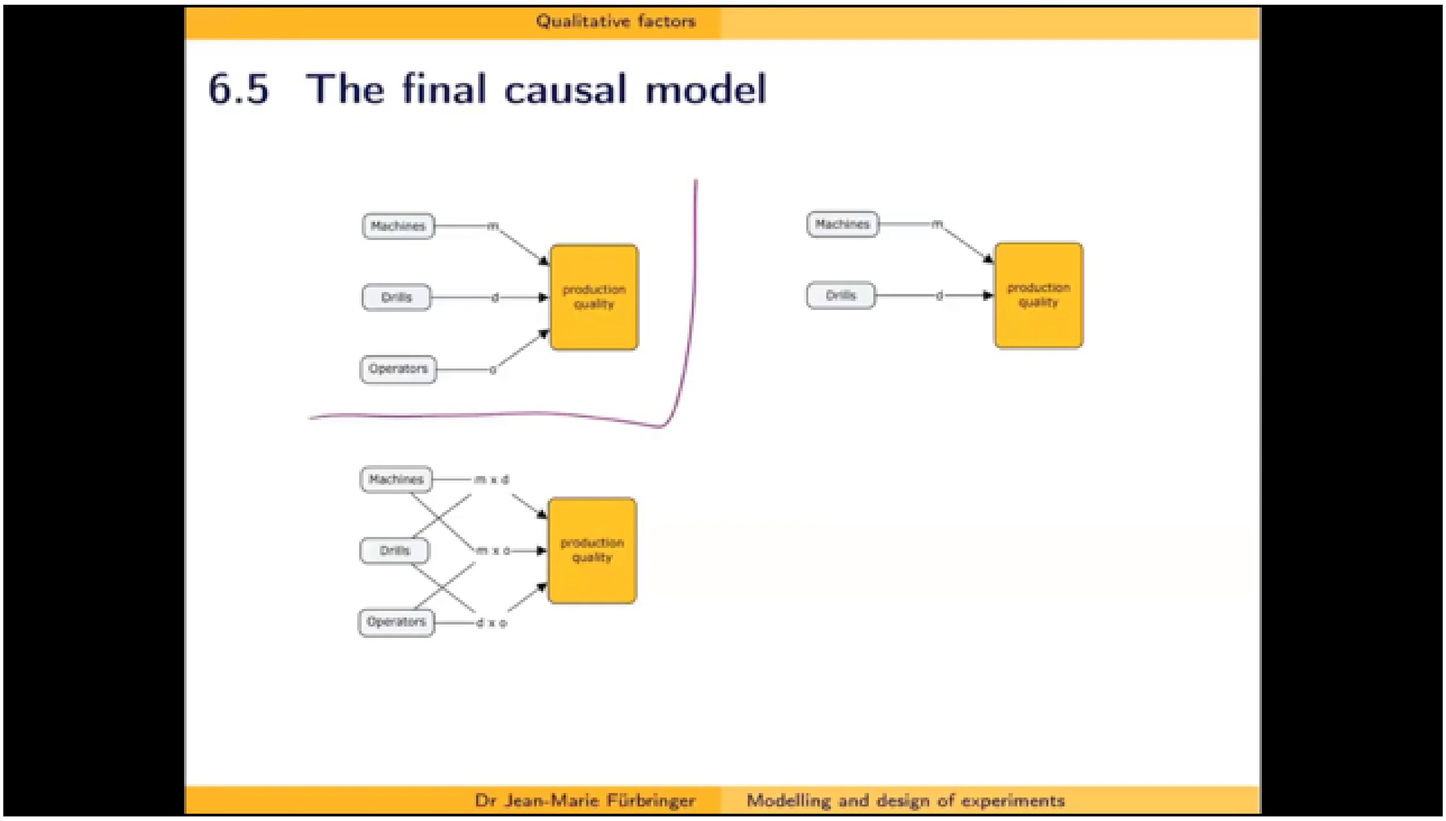


coefficient, we will have one coefficient for the machine one, one coefficient for the machine two, and one coefficient for the machine three. So M is equal one, two, three. And it's why we call that constant coefficient. The same thing for the coefficients for the tool. So we have an index. H and this index again could be 123 if you have four tools, it could be four. So you have as many coefficients as you have levels in your variable. So H is also equal 123 in my case. And the same thing with gamma with O, which could be equal 123. And after I have a residue and the residue will be different for each level, for each cell that I have in my table, and also for each replicate. So it's why my residue has four indices, one representing the level of the machine, one representing the level of the tool, one representing the level of the operator, and one for the replicate.

notes

summary

9m 25s

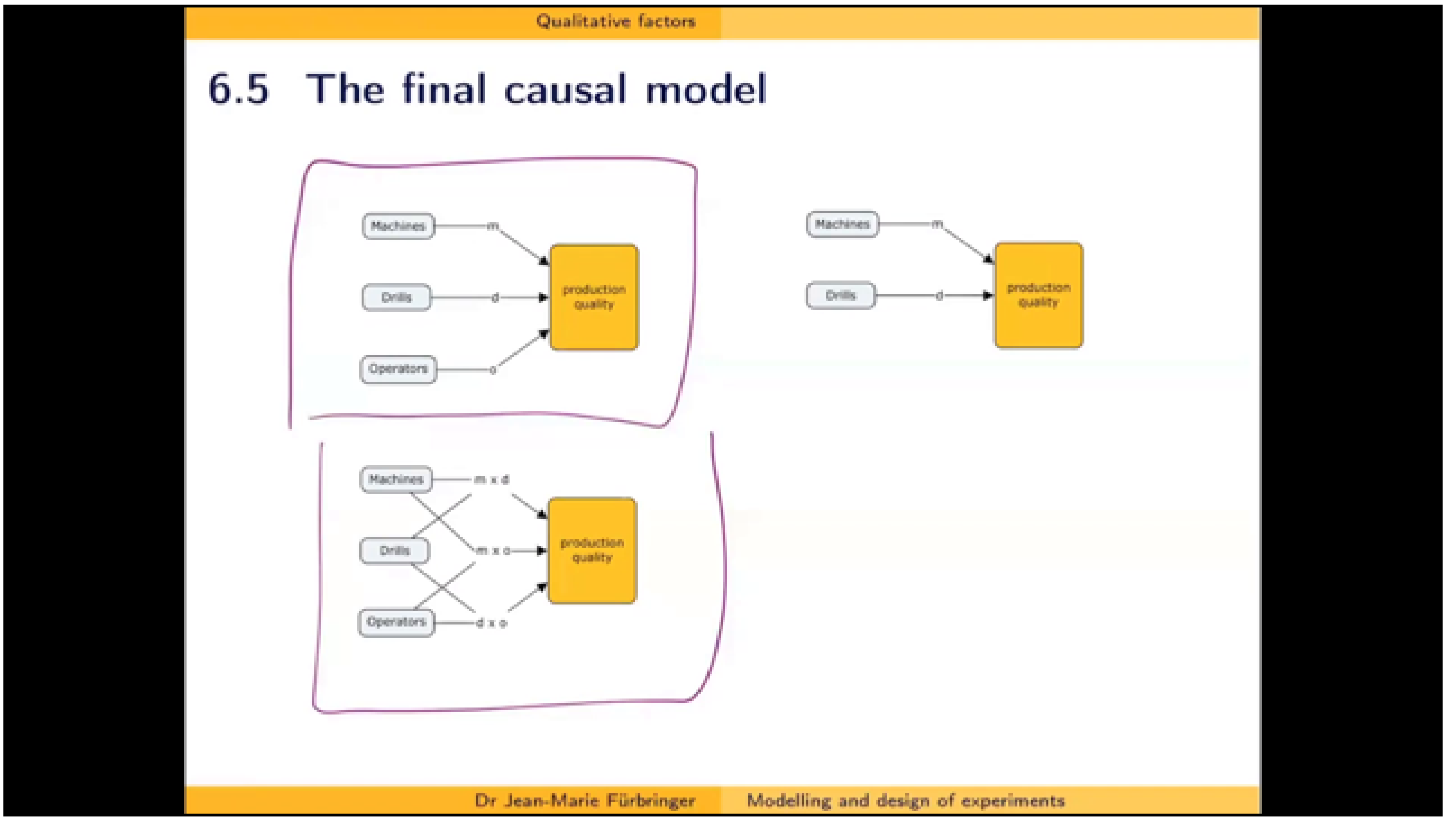


So now the question is how do we get those numbers from the data. The interest doing that is that I will be able to model. So also explain you that it's as we have done with the other cases, it's also modeling because I will be able to choose between different models. I could have the idea that I

notes

summary

11m 6s

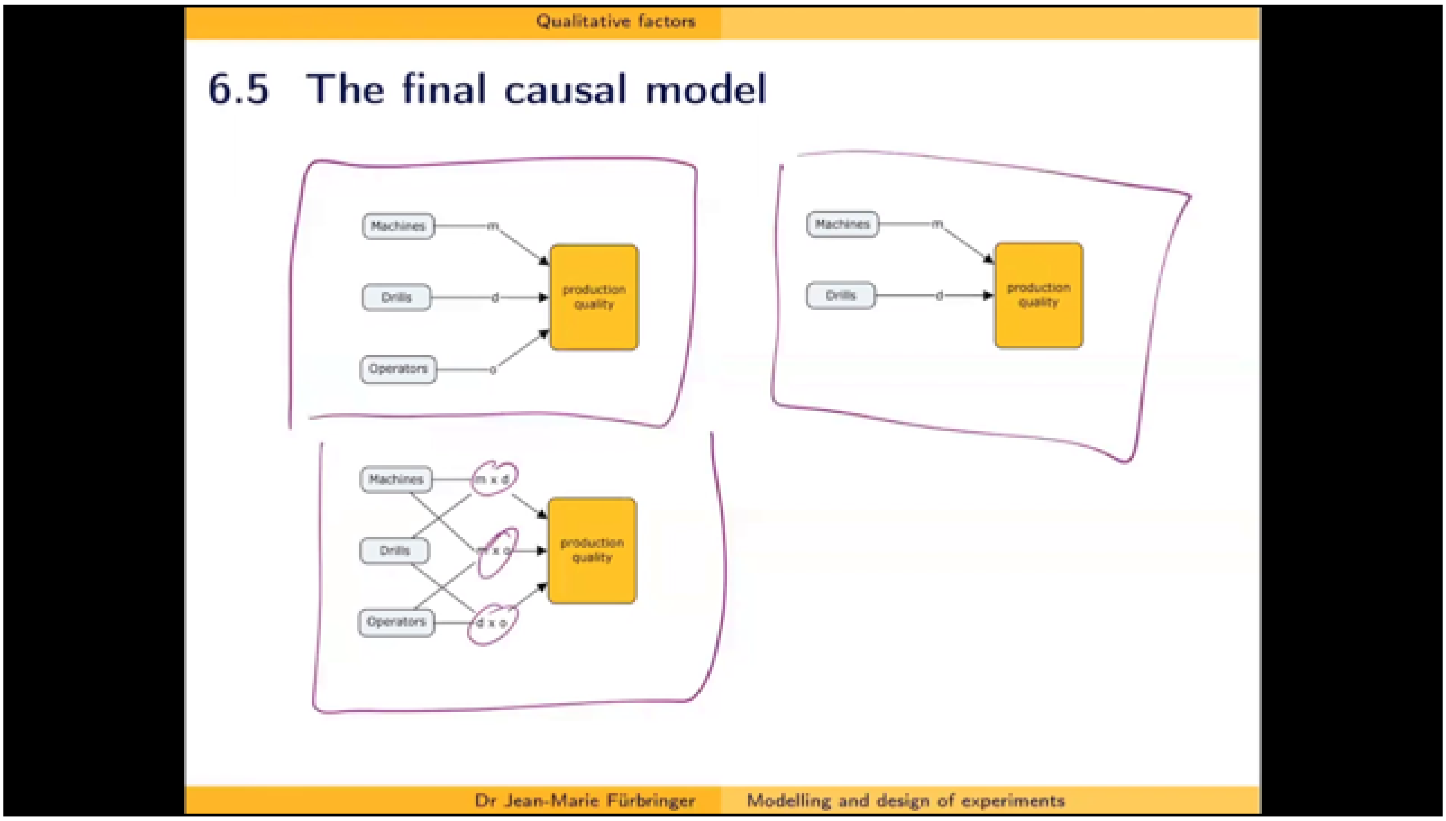


have a model where each factor is acting directly on the outputs. So it will be a linear model. It's the equivalent of the linear model without interaction. Or eventually I have some interactions and I

notes

summary

11m 25s

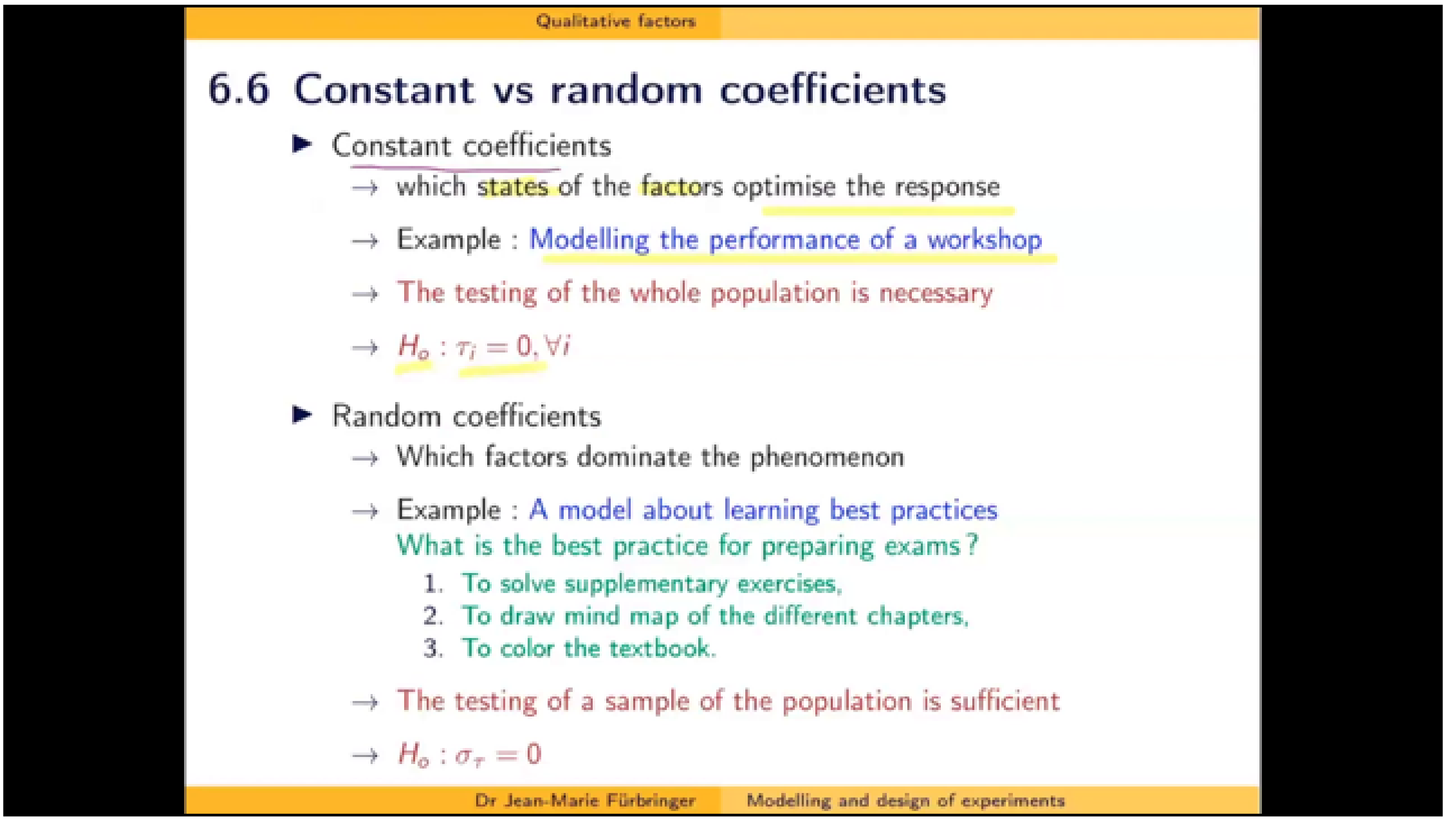


also to show you how will be the model if I want to introduce instructions. So in this case, you see what could be eventually a model where you have the machine and the drills that have interaction. And so you are managing more your reality taking into account instructions or not. And

notes

summary

11m 43s

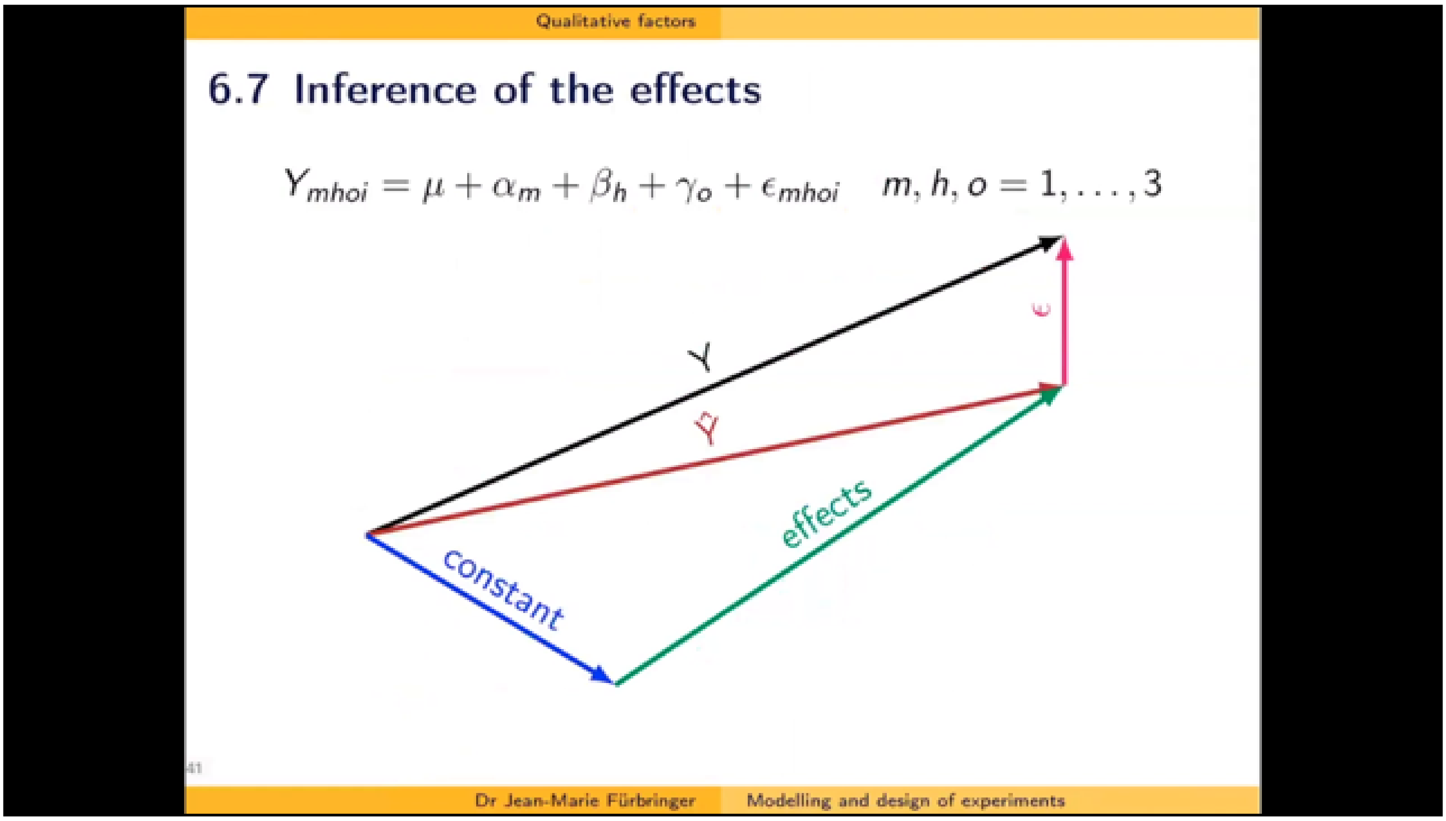


you can also have a type of model where you have eliminated some of the factors. And finally you are able to say, okay, if you want to manage your workshop, you just have main effect, but only of those and those factors and the other factor are random and will not have a different influence. Nevertheless, was the same mathematical structure. So also the same way of evaluating the effect, etc. that I will show you later. We have two types of different models, two types of models. One is the one I just present to you. We called constant coefficient. And in fact, you are interested by discovering which states of the factors optimize the response or minimize the response or maximize the response. You really want to know which operator, which tool, which machine. So the example that you will have in mind is modeling the performance of a workshop. And for that, you have to test all the level of all the factors. And if you are talking about the test, if you are talking about the ANOVA that we will perform at the end, we have an hypothesis that we like to call H0. The hypothesis that we would like to test. And the typical hypothesis is the effect of the factors are all zero or not. It's typical, the test that we are performing when we are working with constant coefficient model. With the same mathematical

notes

summary

12m 11s

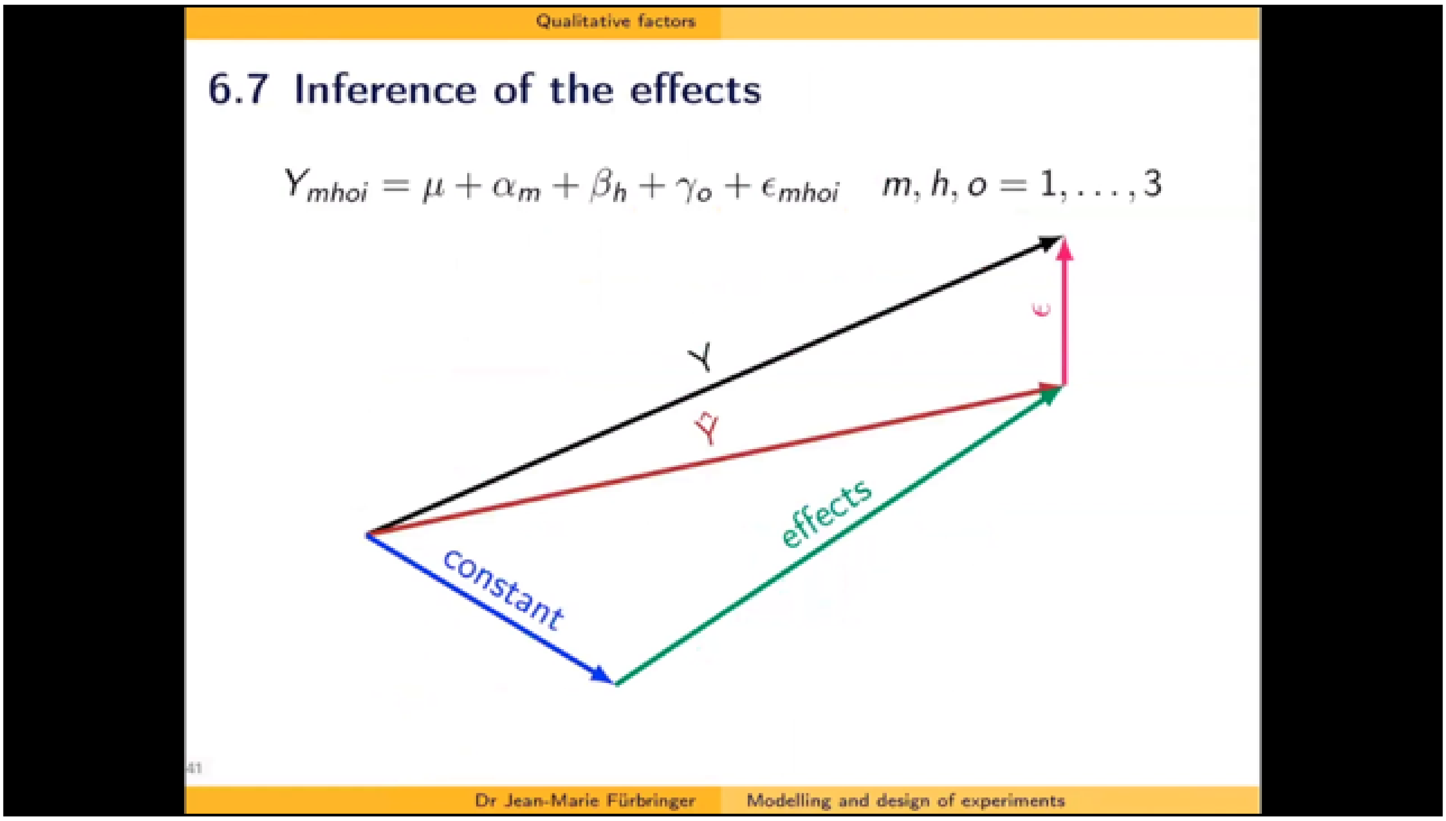


structure, we have another type of model that we call random coefficient. Imagine that you are in you are doing you are working with the cap here in the PFL, you are trying to discover what is the activities that the students have to prefer for getting success in their study. And so they observe some students, some are going every day in the class, other are following the course on internet, other are just looking at the videos afterwards. Some are doing the exercises very precisely, some other are not performing the exercises or only some exercises, etc. And things like that. And they would like to finish with some guidance for the student and say, no, if you want to succeed in your study, could be also different type of courses and different courses. They would say, okay, no, it's very important, it's not important to be on internet or look at the video afterwards or be in the class. This is not an important factor, but making the exercise is very, is very important. Is where I don't know very important from the start of the semester and during all the semester, etc. But they are not interesting or saying, this students do that and not do that, they are not interesting of selecting students. They are just interested of selecting factors. They are not also perhaps interesting to say exactly which strategy is important, but they will say if it's a strategy of attendance to the course is important or if the attention to be put through the exercise is important. So when we do that, it's okay to just have a sample of the students. It's not necessary to test all the students of EPFL for saying something about the best strategy for today. And we will finish with another type of hypothesis. I will not test if all the effects are there or not.

notes

summary

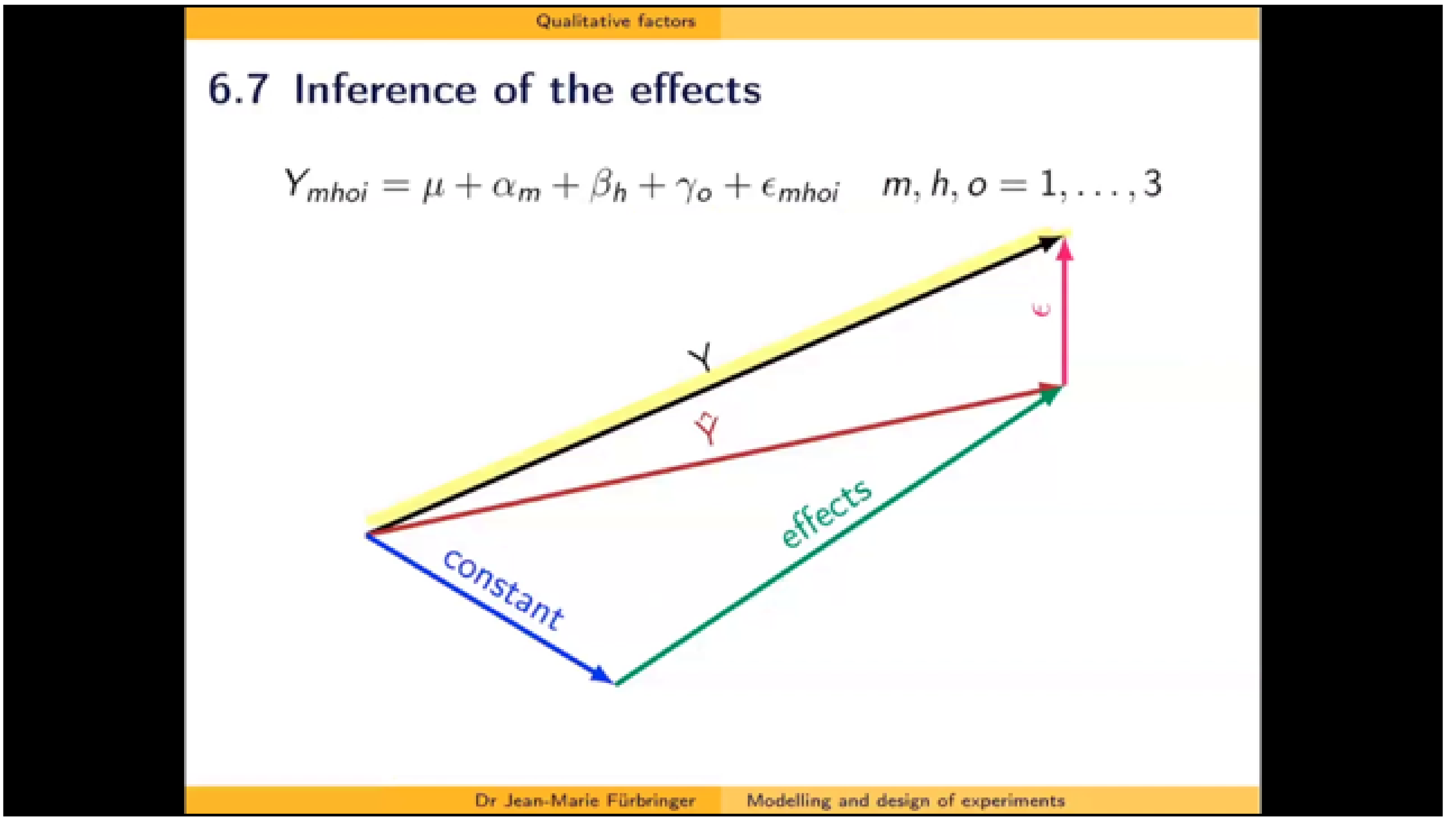
14m 1s



I'm more interested to say if all variable are more important. And for doing that, I will test the variance of my factors and not the level of the real effect of my factors. And when we do that, we call that random coefficient. And it's possible to have things in between. Some factors you would like to have them as random. Other you really would like to test each thing. Perhaps for the attendance, you really want to say if it's important to be in the room, to be in internet or to look at the video afterwards because the recommendation must be precise on that. But for other aspects, you just are sufficient to say this is an important factor that you have to put attention on it. So same structure, two different cognitive approach. One is called constant coefficient and constant coefficient imply that you need to test all the level of the considered factor. Random you are not, you can have samples. Now if we go to the

notes

summary

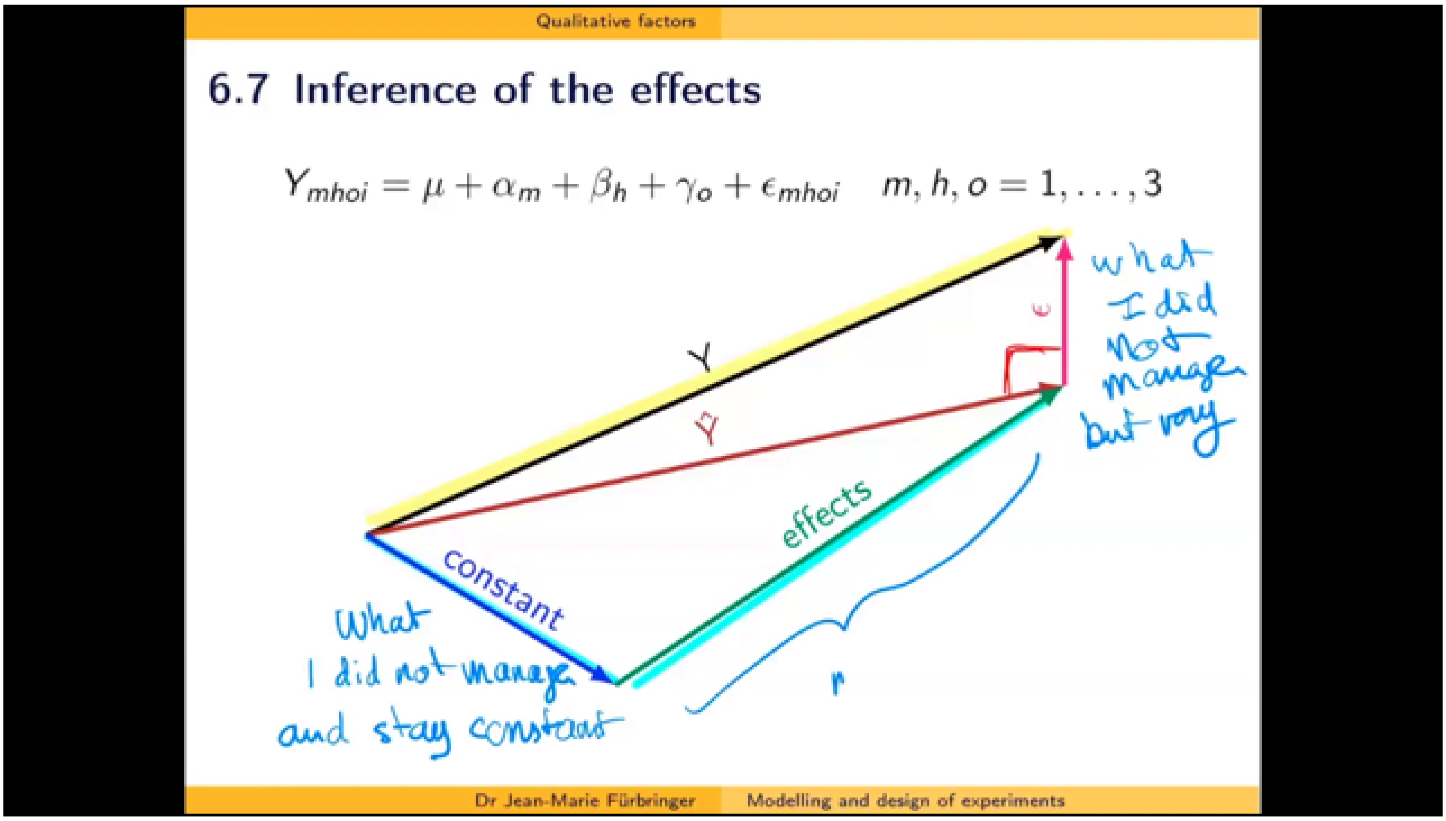


modeling, to the structure, it will be model. It's all the time geometrically speaking, it's all the time the same thing. In fact, we will have an answer. This was our table. In fact, the 27 number in my table represents a vector in a space of results. And each one is a coordinate of my vector. Each answer is a coordinate of my vector. So in the example, the dimension in which I'm playing is 27 dimension because I have 27 experiments. And I will calculate the average of all my data

notes

summary

17m 53s

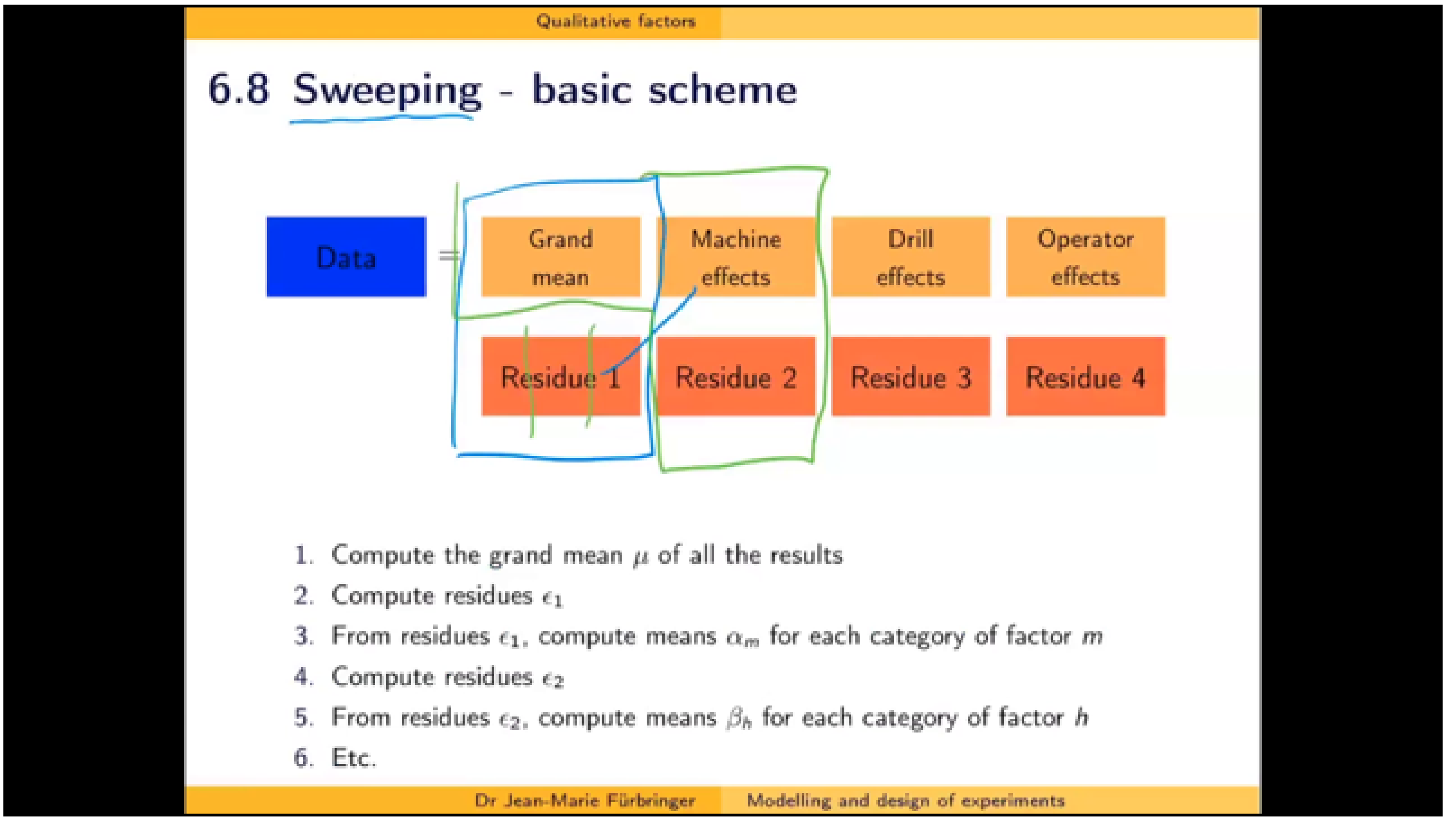


for taking away the information which is the same for all the things. Again, the constant is all the time the information which is common for, so in fact, is what I'm not studying. The fact that we have success or not in the study of the fact that we have defect or not in my pieces depend on the numbers of factors. The type of pieces, the type of material, any other things can influence. But we have not vary those elements during the experiment. So I'm not able to say something on those factors. They are in the constant. After we have the effect in our case in the workshop, they are the operator, the machine and the tool. And we have, for each one, we have a specific number. So it represents some vectors. And the different effect we will detect in a moment, we will calculate in a moment, represents the coordinate of those vectors. And I have a residue. And each time we have the same hypothesis that we have all the time a right angle between my model, my estimate, which has my constant plus my effect. And all the time I have my residue, which is by the rule of making an average, it's each time orthogonal to my model. For why hat is the model, why is my data, I have the constant, I have the residue and I have the effect. And again, I repeat the perhaps point. Right, the constant is what I did not manage and stay constant during my experiment. And this is what I did not manage, but vary during my experiment.

notes

summary

18m 49s

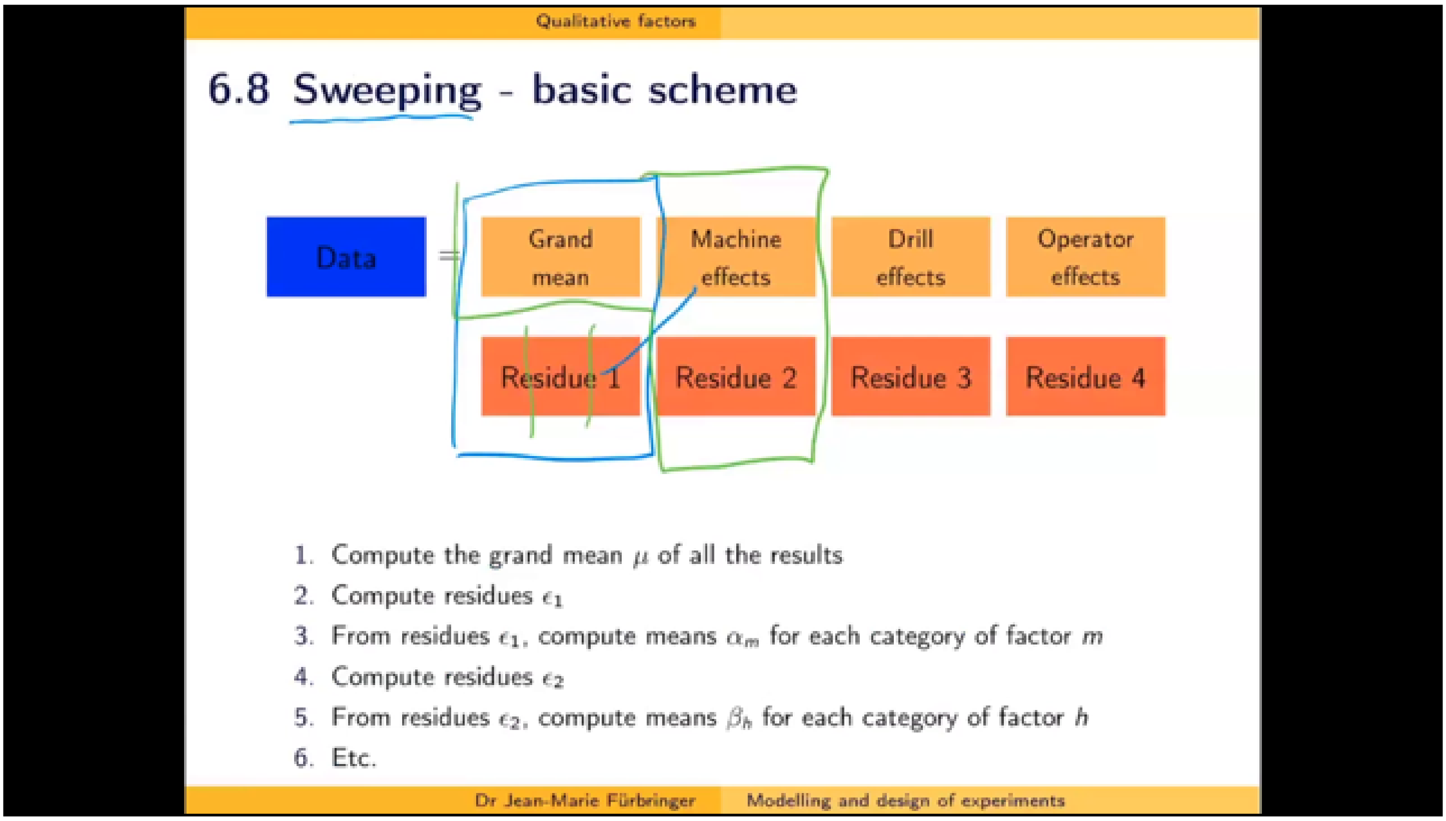


And this is my work is where I would like to bring insights is where I would like to understand something. So for discovering the different coefficients of my model, I have to do what we call the sweeping. It's what we can do for explaining in reality, the algorithms that Matlab, Python are using, they are not doing exactly as I presented to you. I will also present you afterwards the different calculations and formulas that is used usually. So it's a basic scheme. So if I have my data, I will first calculate the average of all my data, I call it ground mean. And I will calculate the residue of that. And if you see the size of my frames are each time the same because it's represent a table or it's represent a vector, you can have the different representation. If I have 27 data, my vector of the average is 27 times the average of my data with the same value. It's why I telling you that this vector of average is the vector which is going from down up through my space, a diagonal of my space. Because it's one value that 70 times, if I want to represent geometrically, represented, copied 70 times. If I have 27 times, if I have 27 experiments. And my residue, I have 27 residues because each measurement is a residue to the average. And now from the residue, I will decompose the residue to get the effects of the machine. And now I'm able to decompose my data with this. I will have a ground mean plus an effect of the machine plus the second residue. And what is the effect of my machine? It will be the average within my residue for each level of the factor. So if the machine, I will take in the first residue, or the residue concerning machine one, and I

notes

summary

20m 57s



will make the average of this residue.

notes

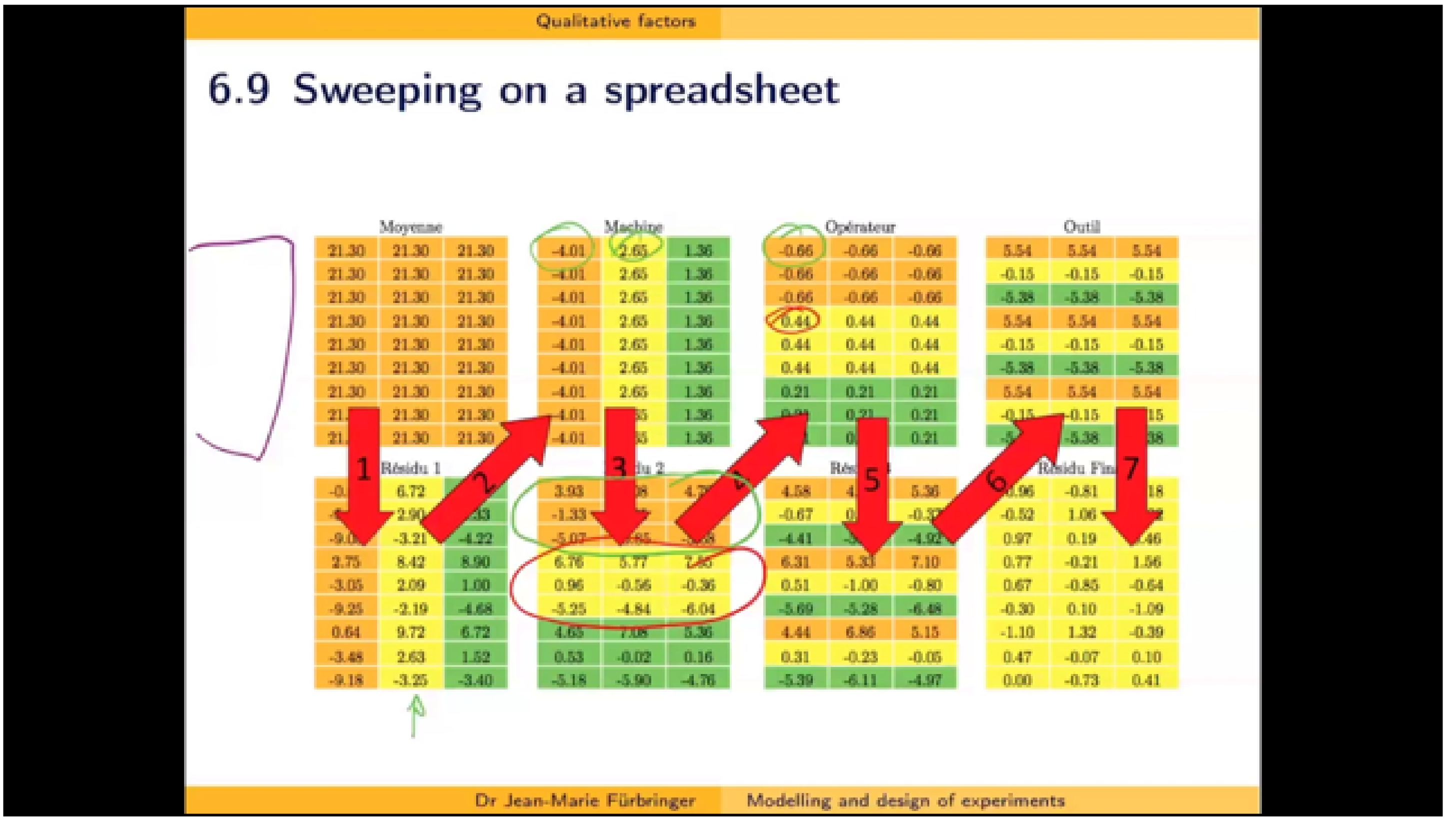
summary

I will then take all the residue concerning machine two, and it will became my effect of machine two. And the same thing with the third one. And after I calculate the difference between the frame residue one and the frame machine effect, and it gives me the residue two. And after I will do the same with the residue two, I will calculate the average in my residue two that are concerning the drill one, the drill two, and the drill three. And after I will have a model that will be a ground mean plus a machine effect plus a drill effect plus a residue two. So it's why we are talking about sweeping because we will sweep, so balayage in French, we will strip the information through the residue. So each time you have a complete model, you can stop when you want. And you can make things in the order that you want. It's not obliged to when you have main effect, you can order your main effect in the order that you want, it doesn't change nothing. So here I have written the different step. You compute the ground mean. After you compute the first residue, you compute from the second residue, the mean, we call that a marginal mean

notes

summary

23m 37s

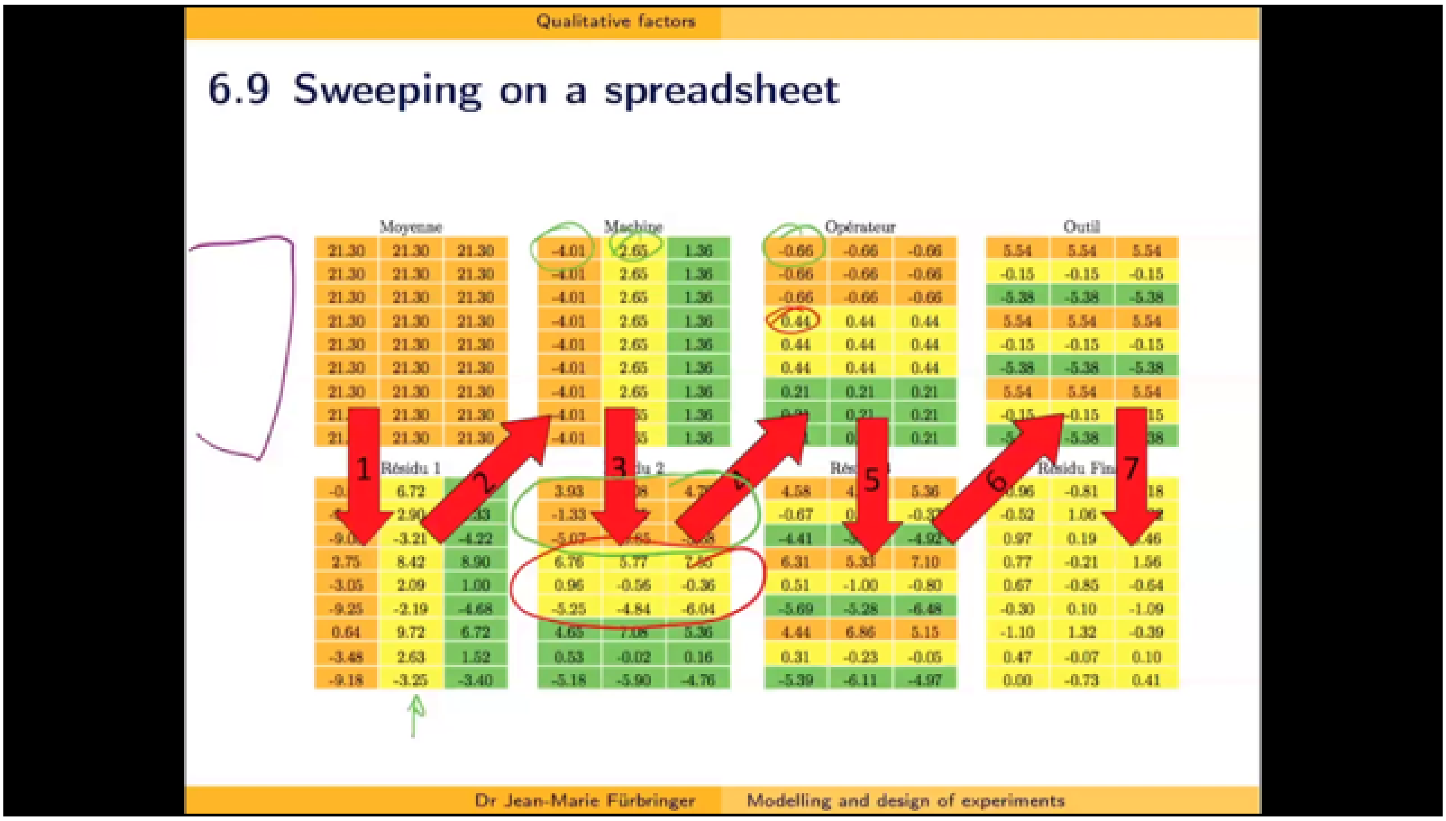


for the different level of the first effect that you choose and after etc etc. So here it's the exercise done in Excel. So it works very well to represent what we are doing with Excel. So you don't see the data. The data was my 27 values that you have two slides before. And I have calculated the ground mean. So the ground mean was 21.3. And so I have reproduced in this case for demonstration. Usually you don't do it, you forget it. But I have reproduced 27 times 21.3. And if you see, so it's a table, but it's also a vector of 27 coordinates with the same coordinate at each cell. And after I have my residue one. And in my residue one, I have made the average of the first columns that correspond to the machine one. And I have obtained minus 4.01. And I reproduce it nine times for all the cells corresponding to machine one. And after I calculate the average in the second column of residue one, and I obtain 2.65. It represents the effect of the machine two. And so I reproduce this effect in all the table where machine two. So it's okay. Things are organized in column row. It could be anywhere. You can do this the same operation. But if it's anywhere after you have to remember where you have to put the different things. And after I've calculated the residue two. You see operation three, the residue two, that was a difference between residue one and machine one. So if you make, okay, let's say minus nine, point 18, minus, minus four, so plus 4.01, make 5.18 probably a problem of rounding. Should be 17 and not 18. Et cetera. So in the second residue, I have analyzed things for the operator. So the operator are organized by block of three rows. So I have then calculated

notes

summary

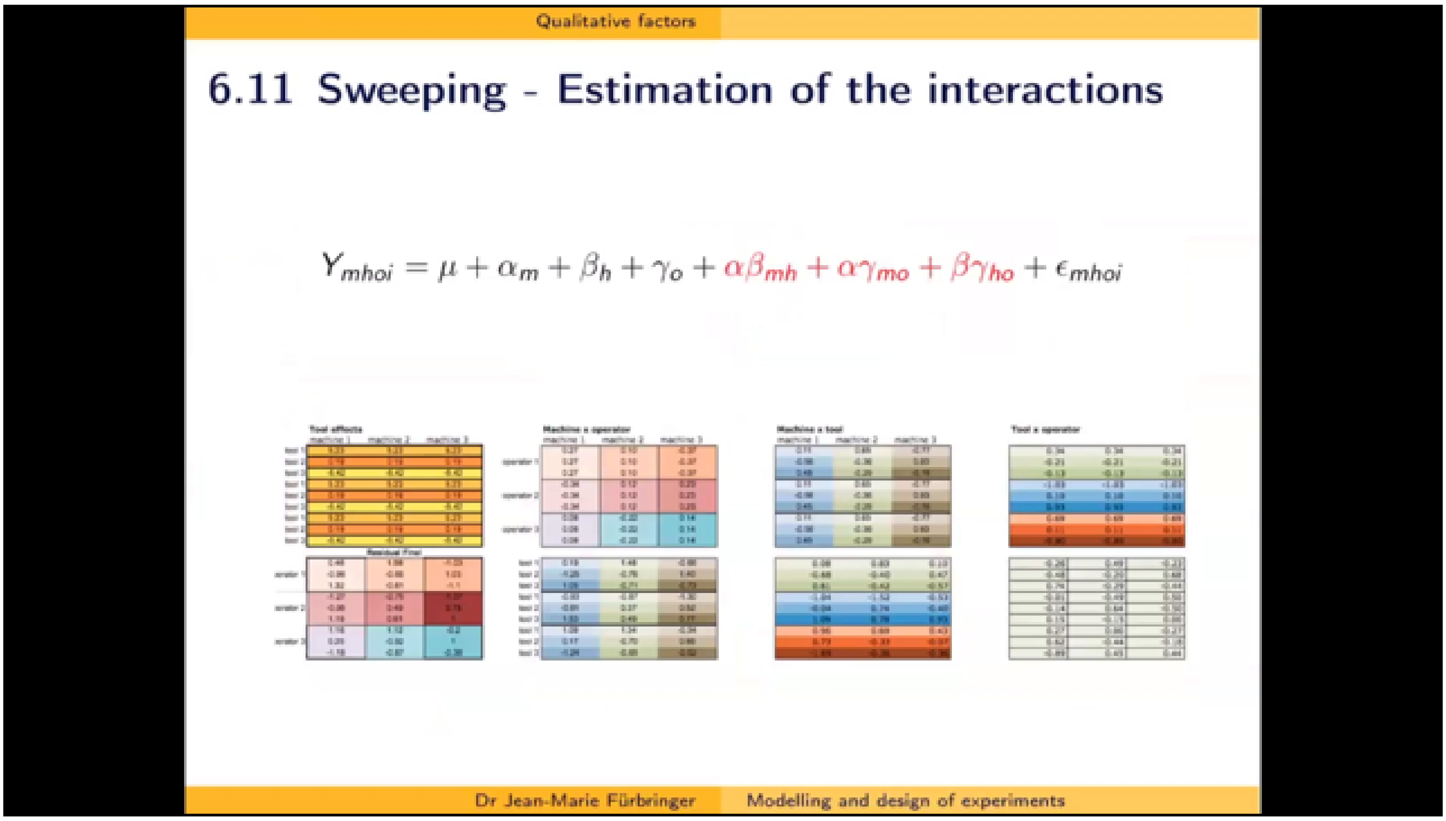
25m 13s



the average of the three rows and obtain minus 0.66. And I put this value in the three row. And after I'm calculating the average of this

notes

summary

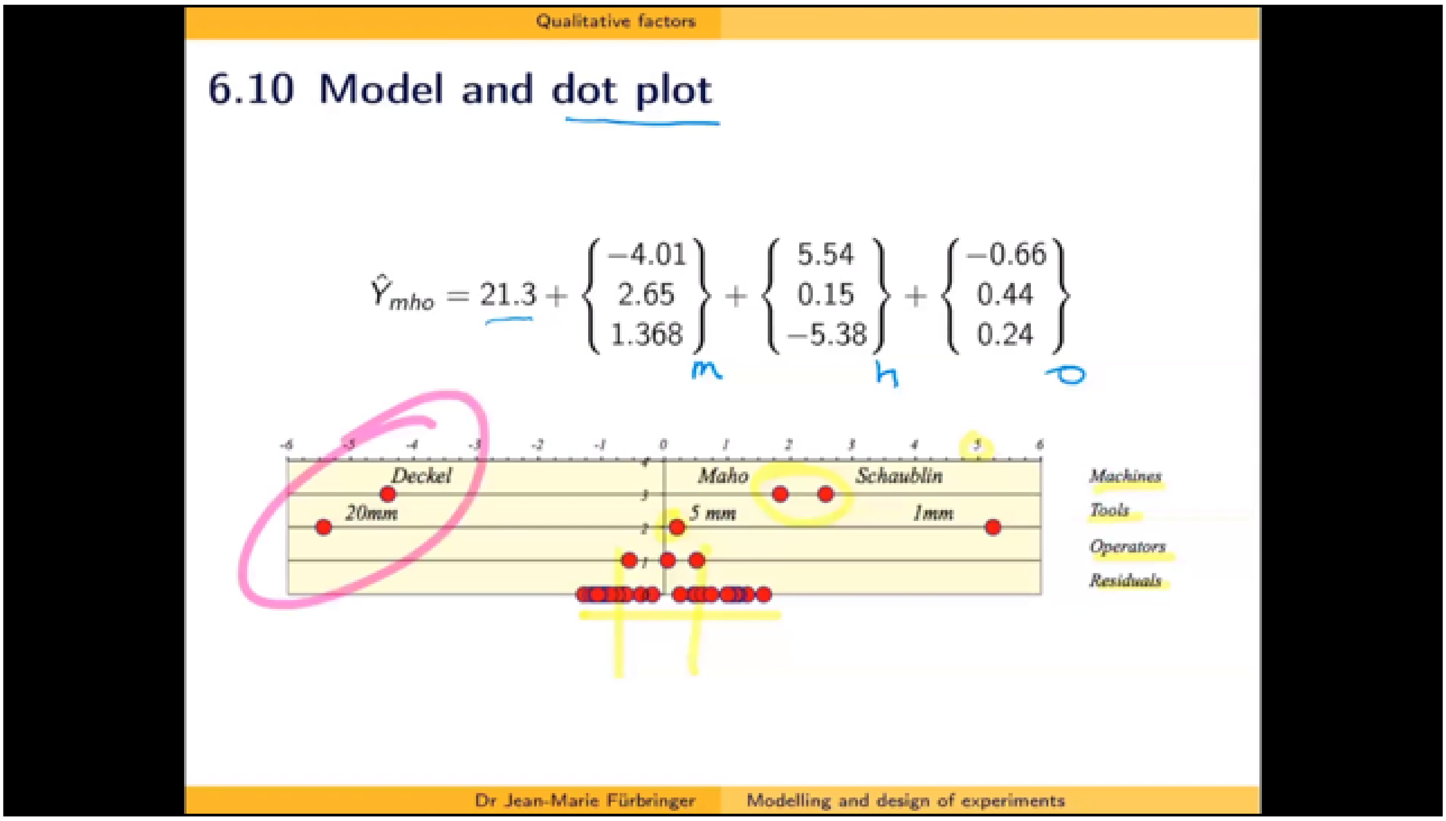


fourth, fifth, and sixth row, and obtain 0.44. And I put this 44 for the cells corresponding to this operator. And after what you can observe that if you, it's very evident here, if you take the sum of the different values of the effects, it's sum to zero. It's normal because the things that we are building around the mean. So if the sum would not be zero, the mean would not be the mean. Because the sum of the residue is zero. In fact, without you see it, if you consider vectors, in this case, the vectors are orthogonal. So in fact, is without you can see it, it's a least square fit. So we finish here with the residue for the main effects. And after we have calculated

notes

summary

28m 13s

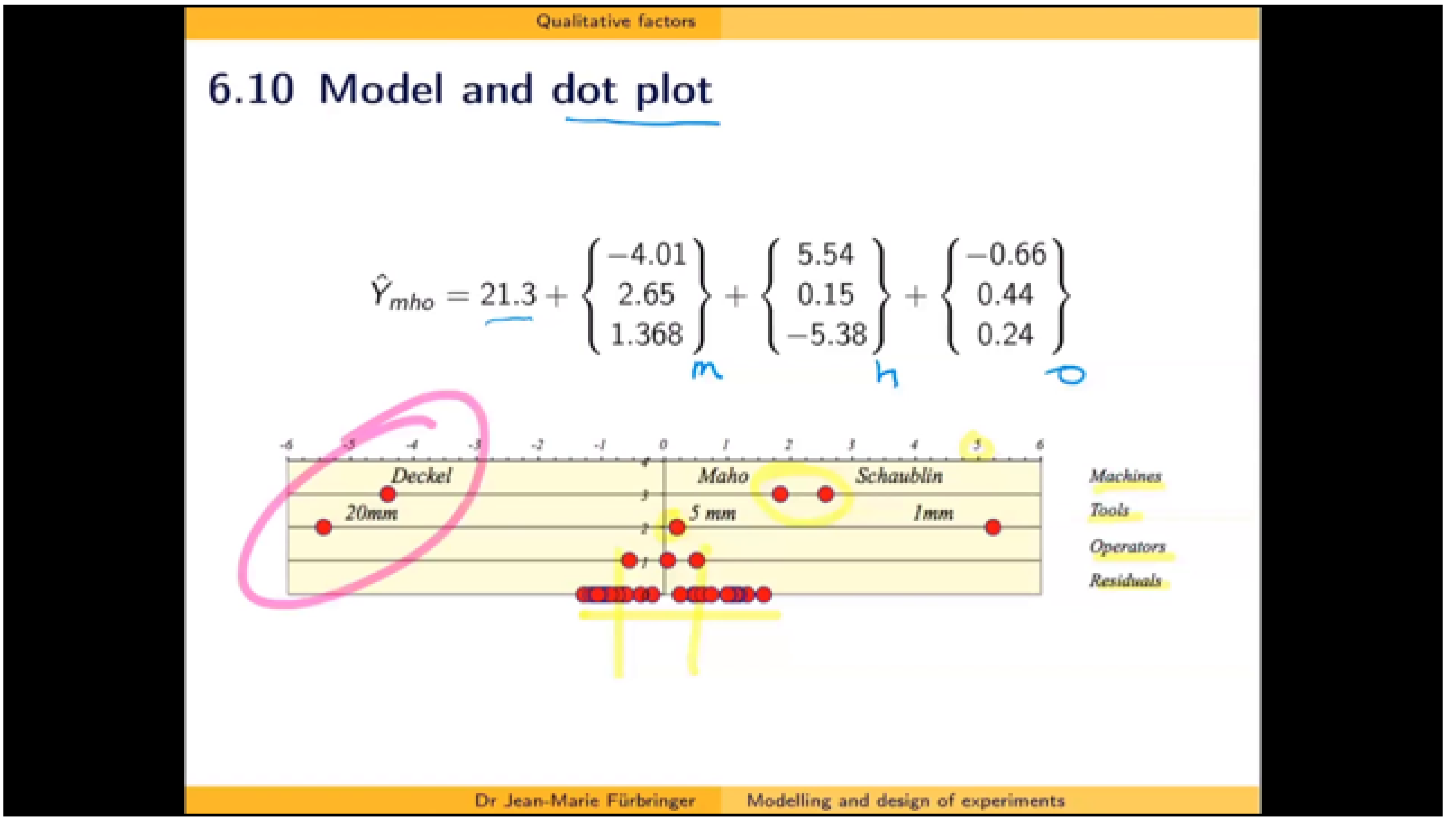


main effects, if we have sufficient experiment, you can calculate instructions. And so instructions are coefficient like that. I call it alpha beta because the effect was alpha beta. And but they have two indices. One, if I'm interaction between machine and tool, I have a number for each couple, a machine and a tool. We do that the same way. So this was the last frame of the previous slide. And now I'm analyzing the couple. So what are the cells with the same machine and the same tool? Well, it's this data have the same machine and the same tool. And I calculate the average of it. And then it makes me an effect of interaction between the machine and the tool. And I can do that nine times because I have nine couples. And the same thing after, but after the cells are not aside. So I can calculate the interaction between. So it was sorry, it was machine and operator. And now it's machine and tool. And so it's the color. So you see that the different case with the same machine and the same tool are those cells and not cells that are contiguous. And I make the average and I obtain in this case, 0.045 that I put at the right place. And I have then the possibility to calculate the three different frame of instructions, machine and operator, machine and tool and tool and operator. I could not go further. It exists the possibility of an instruction three by three. It would be a coefficient that I would write as alpha beta gamma m H O. So it will be a pseudo matrix with three dimension. But I cannot because I just have one data with the characteristic one, one, one. For example, first machine, first operator, first tool, I just have one occurrence of this triplet. So if I would like to

notes

summary

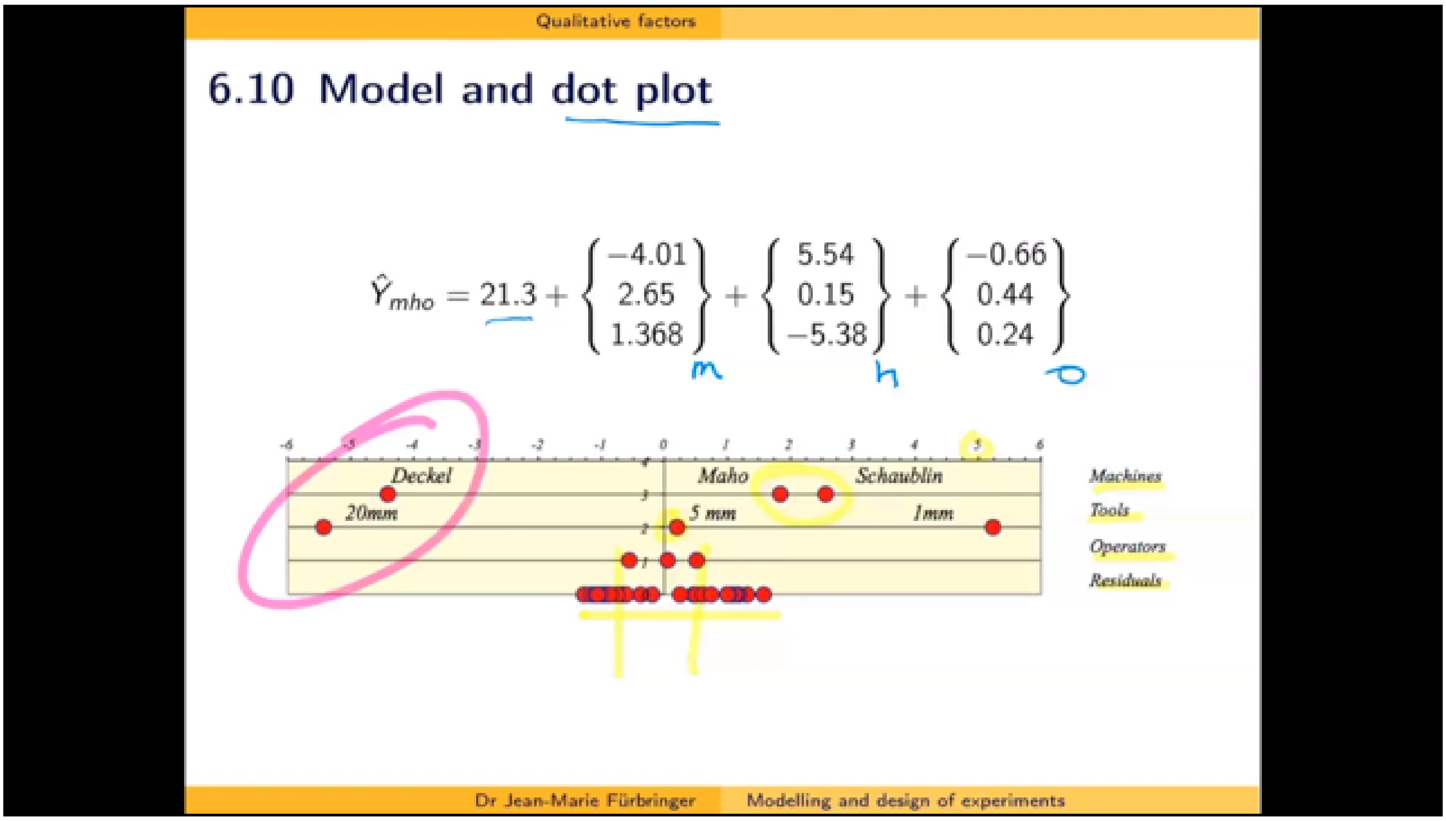
29m 19s



calculate an instruction three by three, I would need to make replicates in my experiment. And in that case, I will make the average of my replicates and have a residue for each replicate. And I would be able to calculate effect of interaction three by three. It is the same thing, exactly the same thing for the linear model we have analyzed. We have a lot of information for the main effect and after the quantity of information is diminishing. So the probability that the instruction three by three is interesting is lower than the probability of interaction two by two, which is itself lower than the probability of having effects. So how do we represent the results of such a model? We can write things like that. Sometimes we have one curly bracket. Sometimes we have two curly brackets are not very definitive way of writing. And what I really appreciate is also indicating which effect is which one. Because when you write like that, is you do not have the x value that you have in parametric, you don't remember. So very probably this was the machine, this was the tool, and this was the operator. And so you have your main effect, your ground mean, you have your different constant coefficient with the name of this type of model, constant vector model. So minus four plus 2.6 and 1.4 for the effect of the machine. Because it's you don't want to represent things like that. It's not very practical and not understandable easily. This is a better way of representing your model after you have made your calculations. And a graphic way of representing things is making what we called dot plots. So you have one line for the residual and after you have one line for each of the effect. And if you have interaction, if you want to represent interaction, you can do the

notes

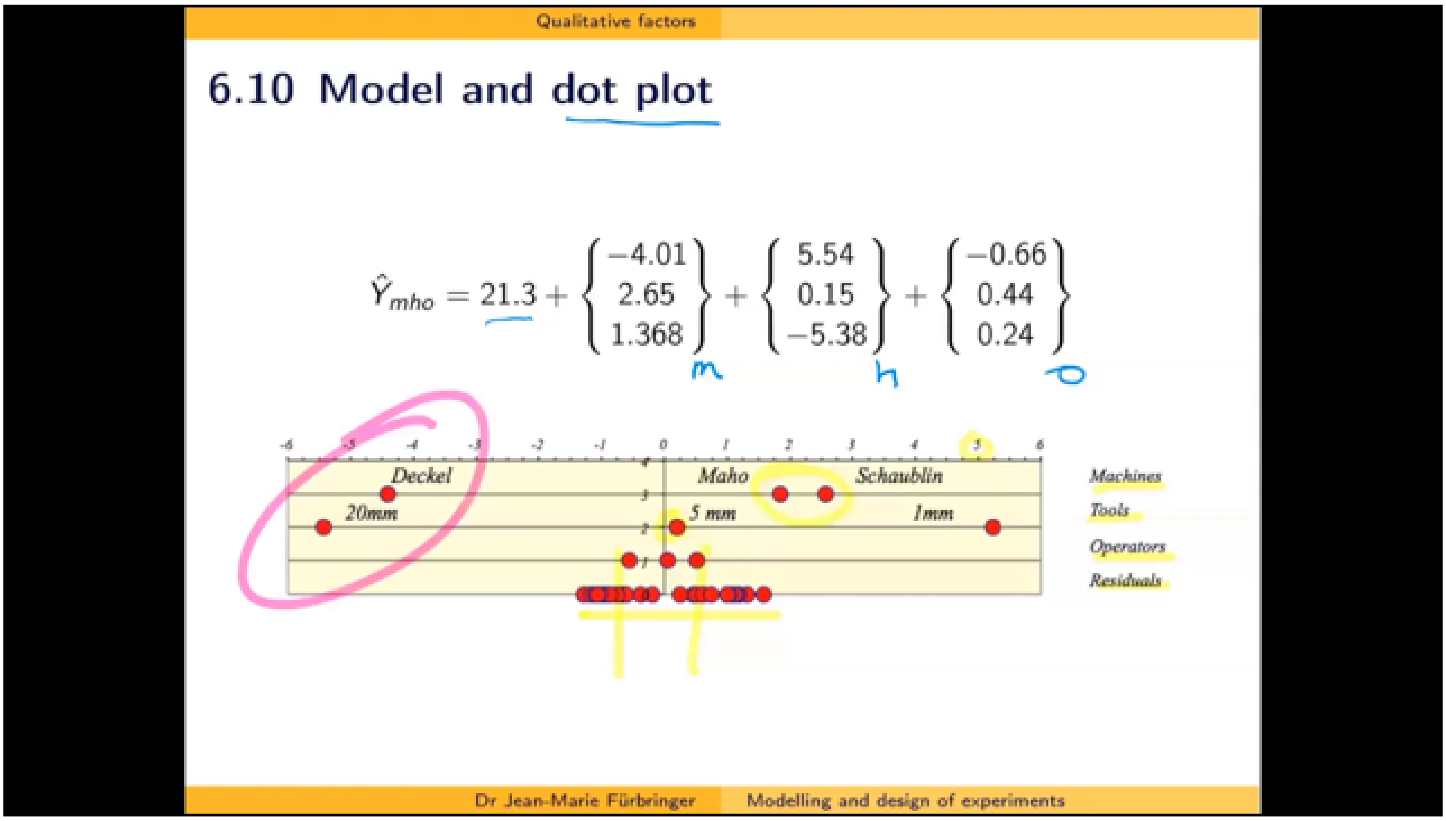
summary



same one for each of the block of my decomposition. So you see here the residual, the values of my final residual distributed in one line. After the effect of the operator, you can observe that the range of variation is smaller than the range of my variation in my residue, which is already an indication that probably my effect for my operators are not significant. It's just a rule of thumb. But what usually I appreciate is having a range which is at least three times bigger than my residue. It's just a rule of thumb. It's not, there are a lot of exceptions, but it could be a first. After you see that you have the effect of the tool, you see one dot for each of the tool. And we see that we have the tool with 20 millimeter, which is producing something as five microns or five point five microns less defect in average on the object. When I have the one millimeter, which is producing something as five microns more defect on my pieces, and it seems that my five millimeter is quite the center is really at the average. And on the machine, I see that the Dekel machine seems to be the best machine because it's producing an average for microns less defect when my machine's Mao and Sherblings are producing respectively two microns more and three microns more. I also observe that in the tools, it's very clear that we have one good tool, one bad tool, and one in the average. Honestly, when you have this distance here, you are not sure. You should be perhaps more experiment. The effects, they are very close. So we need to analyze a little more to put a little bit more statistic for understanding what distance between two effects matter or doesn't matter because they are too close and you are not really sure

notes

summary



that they are distinct. So in this case, we were interested to finding the best solution. So just with this, I would probably say that forget the operator. The all good or bad operators are not a lot of difference between them, but typically produce with Dekel and 20 millimeter drill would probably produce less defect on your pieces and probably solve your problem of quality. Okay, so I let you 15 minutes of pause and we will come back to that.

notes

summary