Course material

Course:

ENG606 / PHYS 442

Video:

DOE_lesson13_part1_StatisticalTests

Concepts (extracted from automatically generated subtitles):

Mean square of the air. Square tau. Statistical test. Lot of time. First things. Small tau. Different levels. Normal distribution. Different categories. Different values. Adequate mean. Given degree of freedom. Square of random variable. Blue surface. Effects.

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

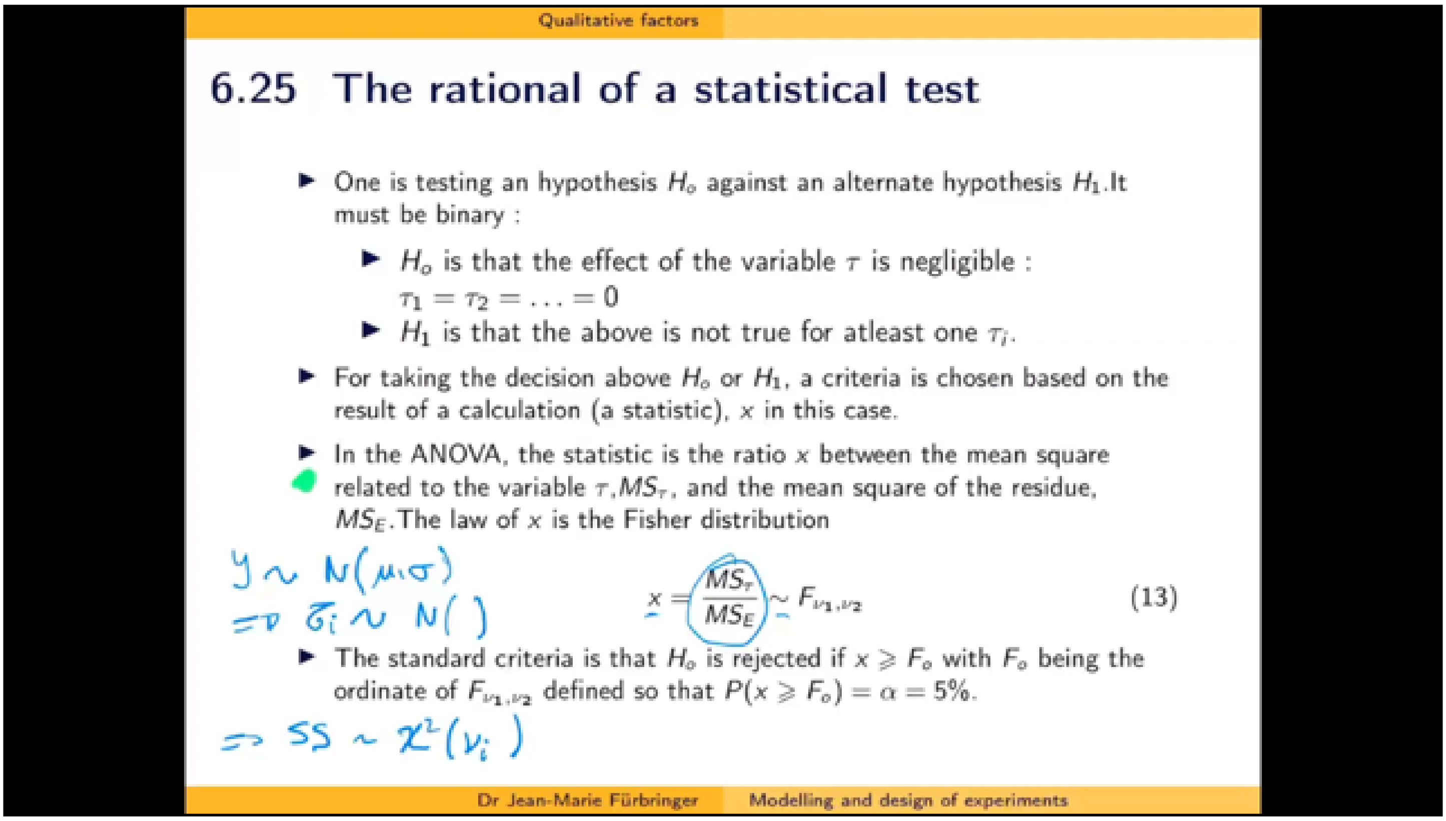
to video

...

notes

summary

0m 0s

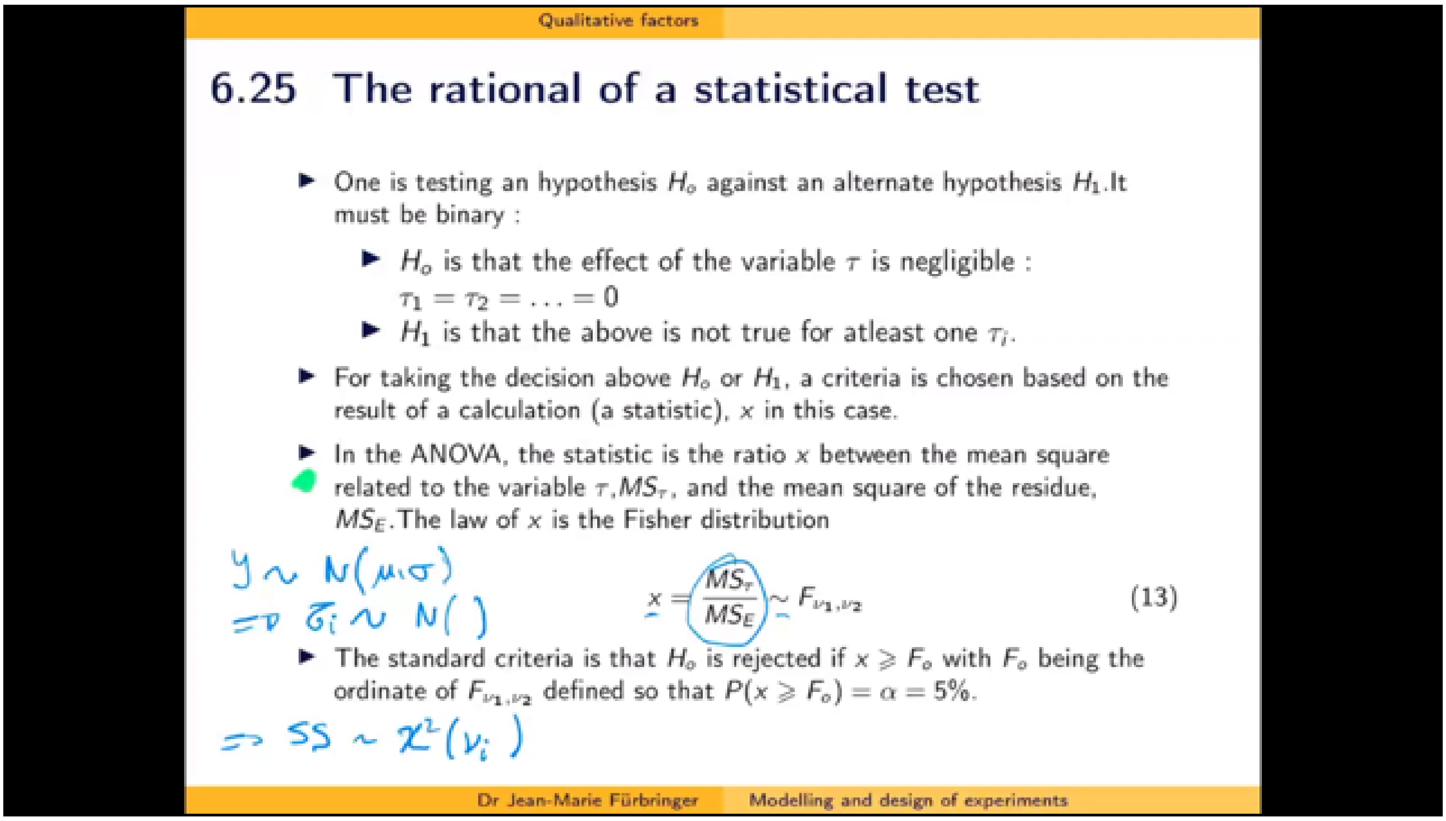


These subtitles have been generated automatically So I have a few slides that are quite, I find them complicated. I hope you find them very easy after I try to simplify them, but it's nevertheless quite complex concepts. Understanding something about that probably is important. So try to understand. So it's about the rational which is behind the statistical test. And what makes it complicated is because in statistics what is good is a no. A yes is not so good in statistics. What is clear is a no. When you say something is not working or something is a test is reject, this is really what matters. But it's why so we have a lot of time with double negative and human brain after a few negation of negations became quite complicated to follow. So it's about the rationale of the statistical test. This test we have been made for the ANOVA for already a few months, but it was in this chapter that I wanted to develop a little bit the idea. So when we are making a test in statistics, in fact, we are trying to make the separation between two hypotheses and to see which one is the most probable. And when the probability are 50-50, so you cannot decide. And so standardly we decide when we start to have 95 on one side and 5% on the other side. And you will see, okay, it's not so easy in fact. And there are an argument, a discussion about where to put this 5% limit. So first things to have in mind is that when you make a test, you need to have a binary situation. It's one thing or the other. It could not be between. If it's in between, you cannot decide. So you really need to be in a binary situation. So usually we have a pro hypothesis. We call H0 and you

notes

summary

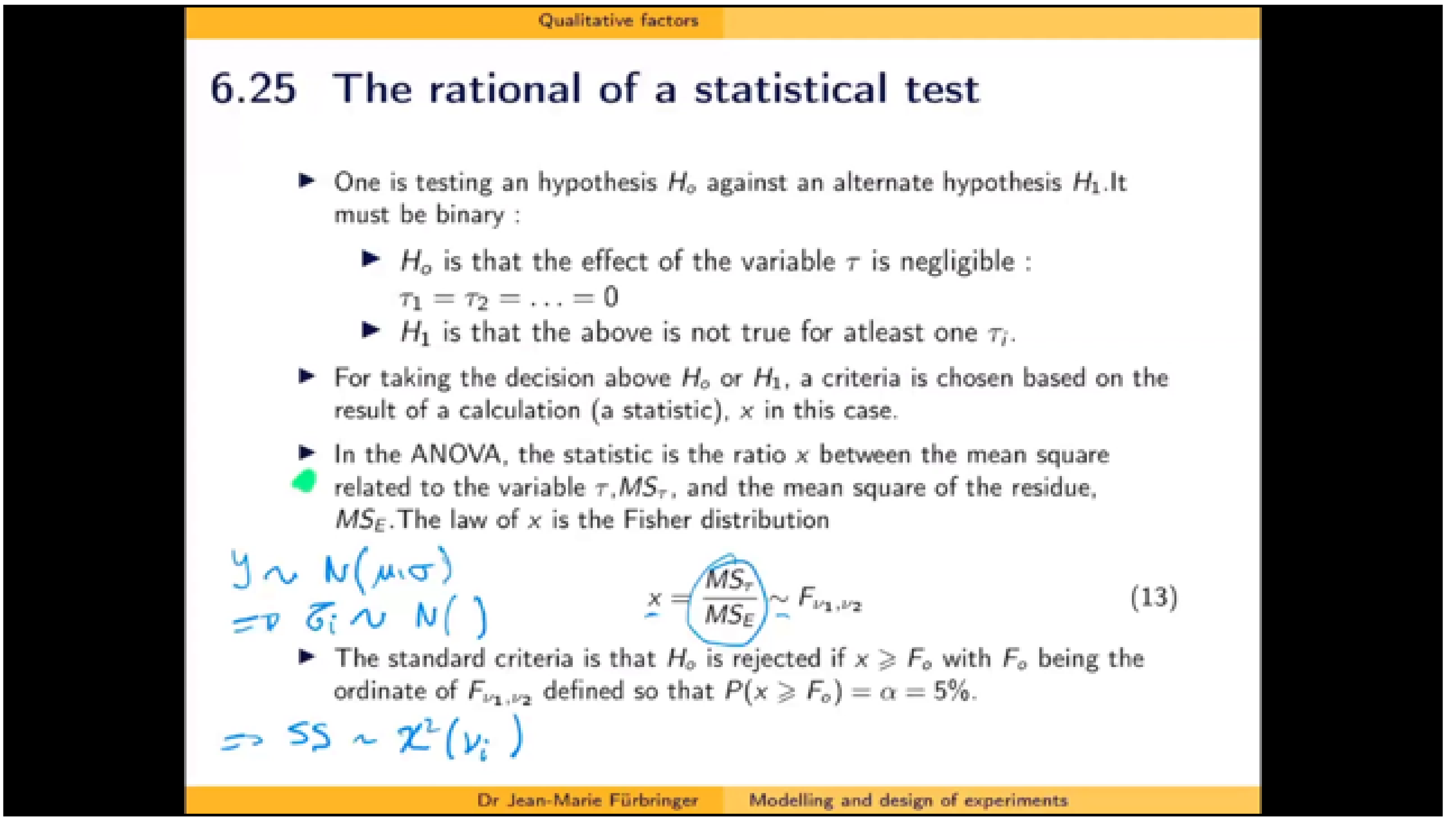
0m 1s



have its contradictions. It's opposites that you call H1. So the standard test we are performing when we are making a nanova is to know if we have effects. This is the, so H0 is there are no effects. And H1 is eventually one of the situation is an effect. So we have been making measurements with different categories. So when we are making the test again, H0, no one of the categories have an effect. Or H1, at least one of the category has an effect. So when we were testing tools, it was the tools have no importance. It was H0 and H1 is at least one tool have an importance. It could be more than one. So this also introduced the last, quite the last slides of this chapter that it's not sufficient to only make a nanova. After we will have to test the different effects of the categories that we are testing. And so with, I will introduce something we call contrast. And all this I will be saying today, it's in fact, an introduction for the necessity of that. So H0, all the effects are negligible and H1, at least one. So for making the decision between these two situations, we are calculating something. And it's what has a name of statistic. We calculate a statistic and based on this thing that we have calculated, based on this statistic, we can take our decision checking the probability to be in one situation or in the other situation. In H0, I'm now here. In H0, the statistic is about the ratio between the sum of square related to a category that can have different levels for the tools, for the operators, for the machines. If you remember the case we are treating from last week. So we are calculating a statistic, which is the ratio mean square. You see a small tau, mean square tau.

notes

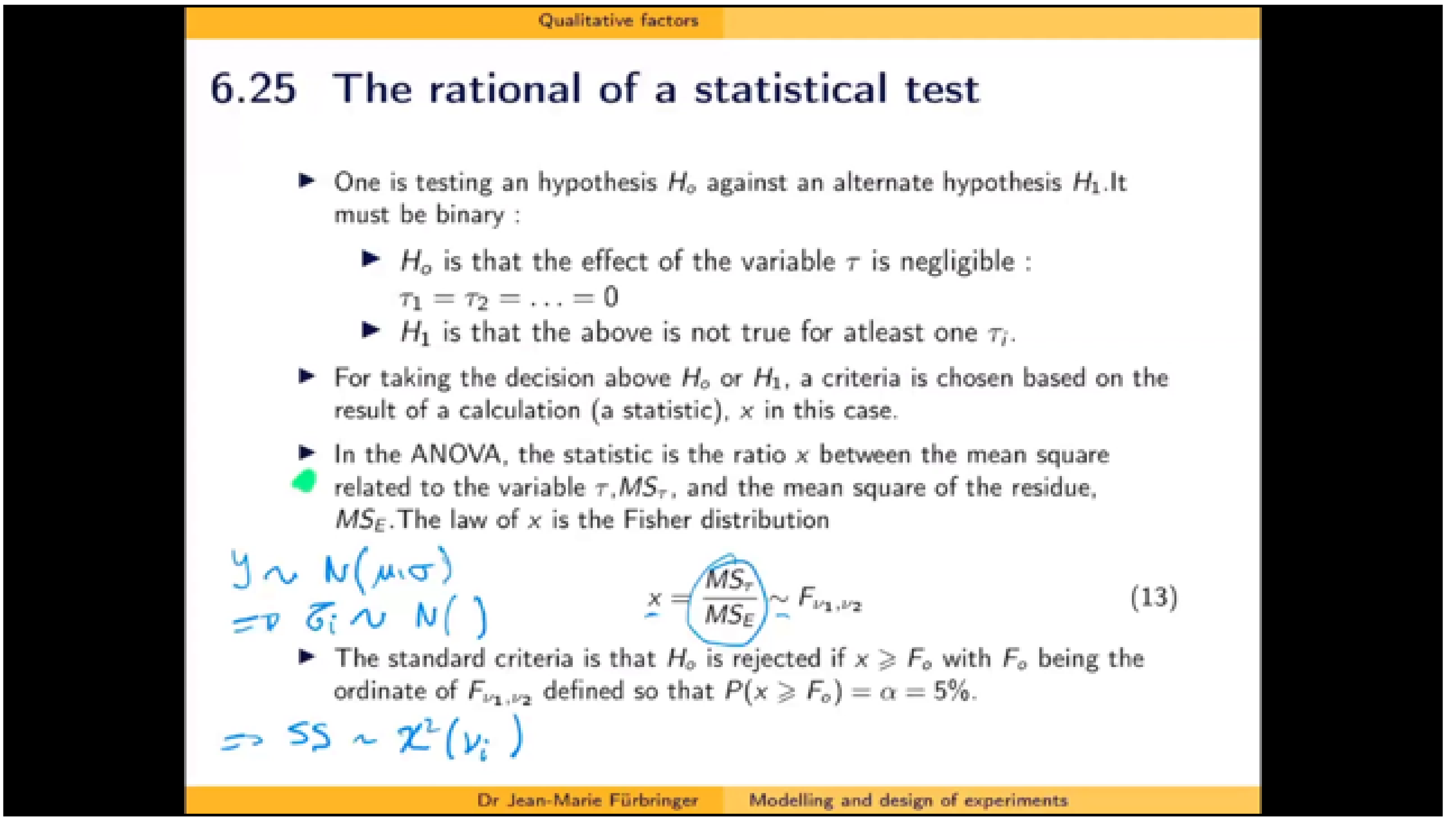
summary



So this is the sum of square divided by the degree of freedom, divided by ms epsilon. So the mean square of the air. We are comparing the ratio between the mean square coming from the variation, from the variance around one of the categories. And we relate it to the residue. And it's clear that when the ratio is something as one, it's not good. That means that you have an effect that is of the same magnitude as the error. And you say, okay, so I'm not sure that I have an effect because my effect is the same magnitude of the noise. So if you want to listen to something, if you want to catch an element of the model, it's better that this element of the model is above the noise. So we write, so we call it x. If you look at some of the slides, I call it x. It was what is below the column f, so the Fisher ratio in the ANOVA table. And so we make the hypothesis where the statistics enter before it was just playing with the number. The statistic enter with this tilde sign saying, okay, we make the hypothesis that this ratio is following a distribution. And we will see what is the probability to get this ratio. So the hypothesis is that our data, if it's H0, if we have no effect, in fact, we change something, it's a button which is not connected to nothing to the reality and change nothing to your reality, to your experiment. The hypothesis is that if the effect is unrealistic, is negligible, so it's behaved like nothing and like the noise. And what behaves like the noise is a normal distribution. So we make the hypothesis that my movement is just, I just have different values when I'm making an experiment, but it's just noise. It's just a variation

notes

summary



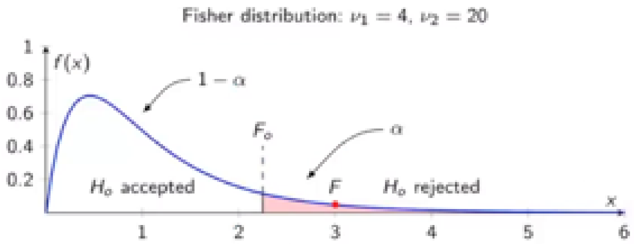
of my installation of, I don't know, of things I'm not managing, but it's not the result of my experiments. So if my data is following the normal distribution, then if I calculate some of squares of those, so if I calculate effects, it's linear combination of a normal distribution. So if it's a linear combination of normal distribution, it's follow also a normal distribution. So if I have my y's, I can try to write that if I have my y's that are following a normal distribution with a mean and a standard deviation, then my effects, I call them tau. Tau E are also following a normal distribution with the adequate mean, it's another mean, and another standard deviation that I can calculate, but I'm still following a normal distribution. So when I'm calculating the square of a random variable, so this concept is random variable, when I'm calculating the square of random variable, I could say that the sum of square will have to follow a key square distribution with a given degree of freedom. This is a distribution of the square of a normal random variable, and if I'm calculating a mean square, I'm just dividing by a coefficient, it doesn't change nothing. And when I'm calculating

notes

summary

6.26 Type I error : rejection of a true H_0

- When fixing a threshold α and then rejecting hypothesis H_o, there is a risk α that H_o is, in fact, correct.
- Example : a test is done to determine if the choice of the tool has an effect on the quality of the production.
  - H_o : "The choice of the tool is negligible"
  - H_1 "At least one tool has a detectable effect"
  - If $\alpha = 5\%$, it determines a limit value F_o to reject H_o.
  - When performing the ANOVA, if $x > F_o$, then H_o is rejected.
  - If the tools have, in fact, no effect, it would be a **false positive**.

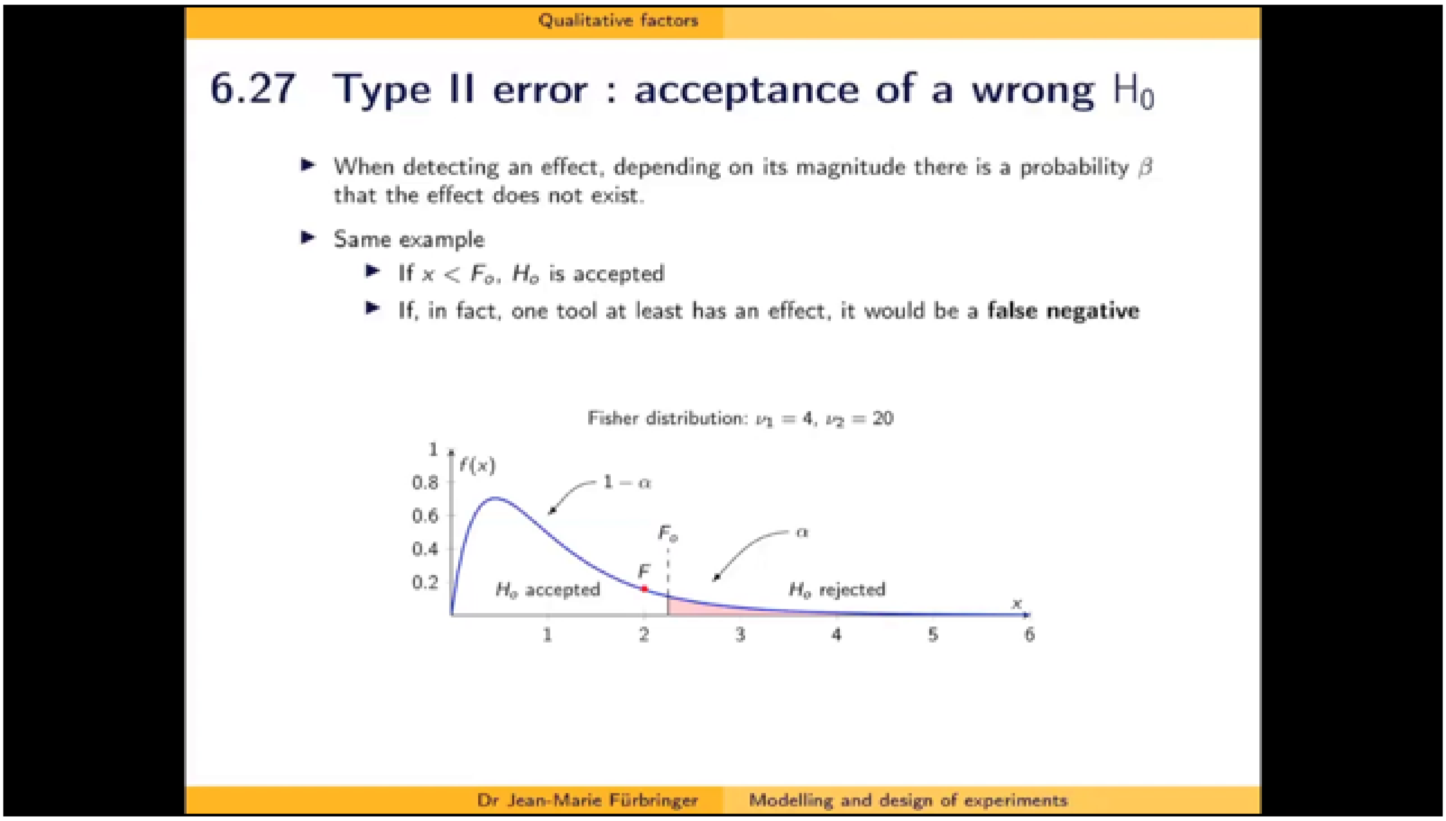


the ratio of two mean squares, then I follow a Fisher distribution. So it's why we are testing with Fisher distribution. So if you remember, usually when you use fitlm or this type of or another one, probably the first thing that is giving you the algorithm is a t-test, is a student's distribution, because it's just calculating on the average of the effect. Now we are calculating on the ratio, we are making the statistic on the ratio of mean square, and we follow the Fisher distribution. So what I don't want is I don't want to resolve by chance. So I say that this is the probability to get this number, the ratio of my mean square is below than a number. So it's between one and something, but very close to one. The probability is not so nice. And it's when I have a big number for this ratio, that means that the mean square of tau is a lot bigger than the mean square of the residue, that I could not get things by chance. And the way of deciding, if they're okay, I decide of a probability, usually is 5%, bigger than alpha, I'm not so confident. Eventually it could be a noise that when this probability is smaller than 5%, smaller than alpha, then the probability of getting it by chance

notes

summary

9m 1s



is quite low and accepting that the effect. This is the ratio of a test. But when we are making this test and taking a decision, we are interested to understand what is the risk that we are wrong. Okay, we have put 5% as a threshold. What is the risk that we accept an effect, but in fact, we have no effect. It's just random that the numbers just align, but in fact, they don't exist. So when we are fixing the thresholds, they are all time a risk of the same amount. So the risk, if I fix 5% is a limit, I still have, you see in the graphic below that represents the distribution, the fissure distribution, I have fixed at the level, I call it F05%, it's the absciss at which the integral from minus infinity, in this case from zero, because some of square are positive, so it's from zero to, in this case, it's an x. So it would be 1 minus alpha, and what is in red, it would correspond to alpha. And I put this f0. So when I'm calculating my, you see where I have an f, I put a red dot. So imagine that finally you get this value of 3 for the fissure ratio. So you say, okay, I have an effect. There are still probabilities that in fact, you do not have an effect. It's what we are evaluating. And what is the probability? The probability is alpha. The probability is the same as the limit, the thresholds that you put. It's the risk that you say it's an effect, but you don't finally, you don't have an effect. In fact, you have decided that your result is outside of your distribution, no effect around your no effect distribution. In fact, do you wear in the queue of this distribution? And in fact, you don't have an effect.

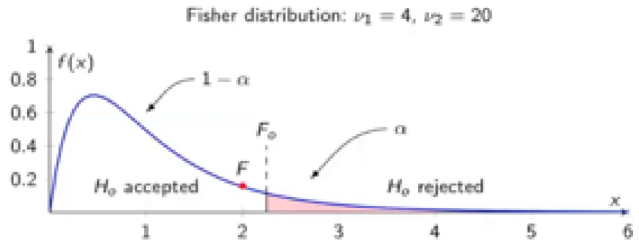
notes

summary

10m 42s

6.27 Type II error : acceptance of a wrong H_0

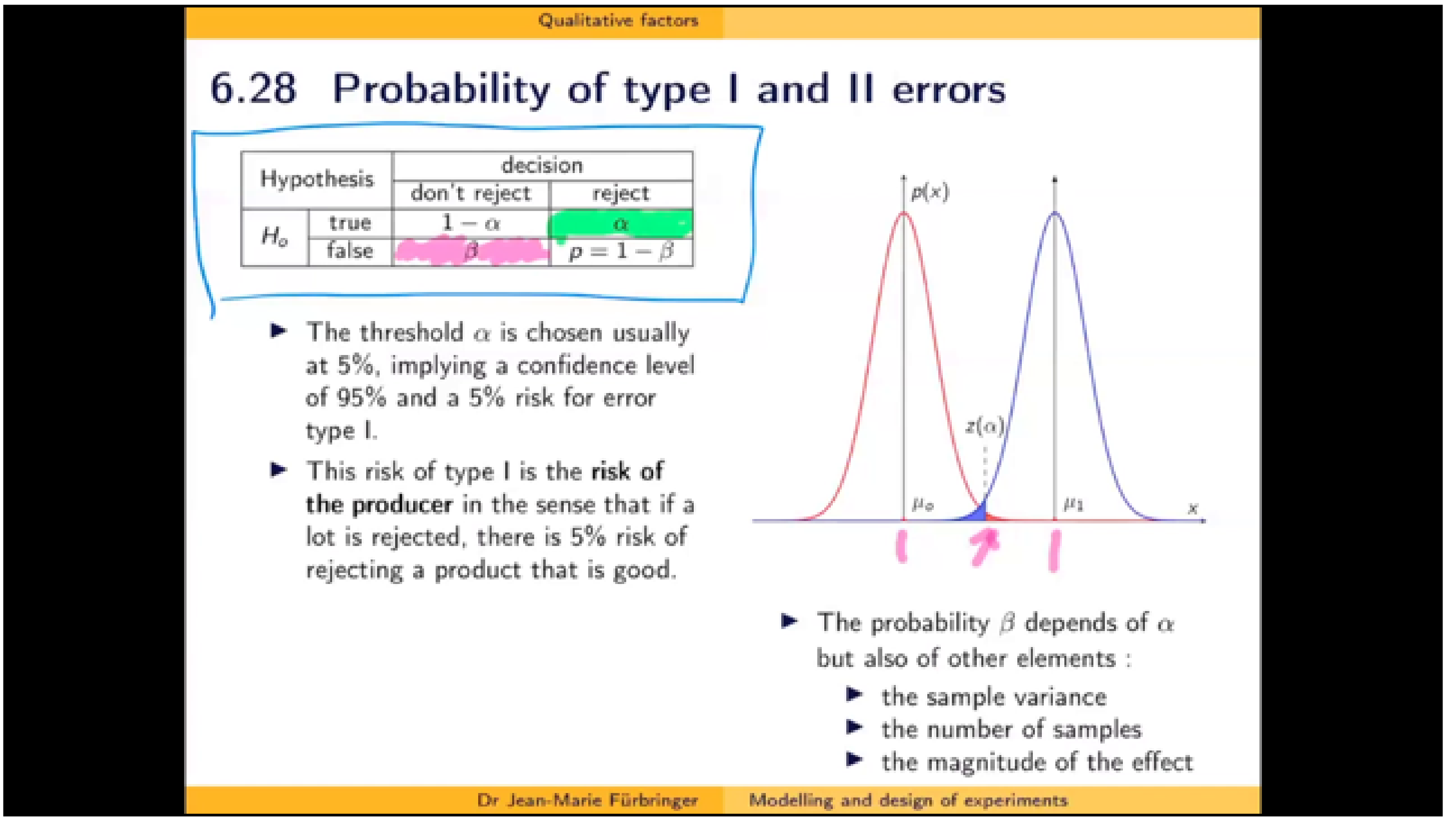
- When detecting an effect, depending on its magnitude there is a probability β that the effect does not exist.
- Same example
  - If $x < F_o$, H_o is accepted
  - If, in fact, one tool at least has an effect, it would be a **false negative**



You say you have an effect, and we call also that, you perhaps hear sometimes, sometimes in the news, during the COVID, we heard a lot about those words because statistic was at war. And it's also called a false positive. So you make a test, you reject the hypothesis I0. So it would be the very positive aspect. In fact, you don't have an effect. This would be the type one error. It exists another possibility that in fact, you say that you don't have an effect, but you have one during the COVID, it was terrible. It was a big discussion. You make the test, the test that you don't have the virus, and in fact, you have it. So you have accept H0, and it's not correct. And this probability

notes

summary

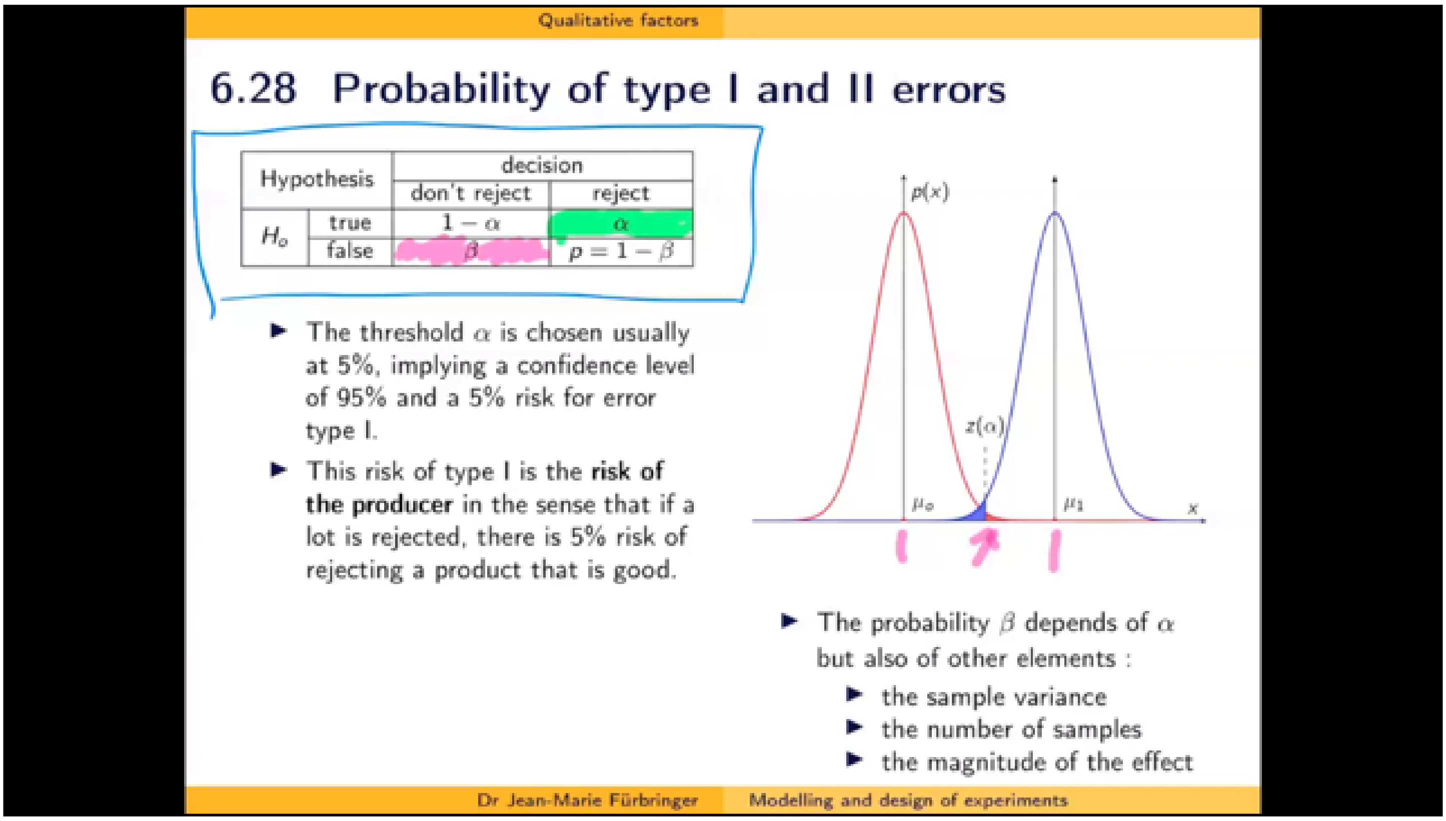


is quite complicated to calculate. So when you have these types, you absolutely to look at this type of table, you have to compare the reality that you don't know perfectly, and what is your decision. So you see that vertically, you have your decision. Don't reject, reject. And horizontally, you have the reality to which you do not have access directly, but you can have the probability related to all of these cells. So the H0 can be true or H0 can be wrong. And so first we have looked at the first line. We have seen that we have taken a decision to reject H0 with a probability alpha. So that means that we have a probability to not rejecting it one minus alpha. And the error is when you are in the green zone that you have reject, and it was true. H0 was true, and we have rejected. You have this probability. So it's nice. You can never say I have 95% of chance that if I'm accepting H0, H0 was true. So the first lines represent the error alpha is the error type one, type one error, false positive. The other situation in fact, H0 is false. That means that the tools have an effect, but you take a decision. So the error would be to take the decision to do not reject. So you have in fact an effect. You should reject a H0, but you didn't do it. This is the error type two. And there are the cells aside. P, we talk about the power of a test is just the complementary of the beta. So why it's so complicated? Look at my two one in red represents the H0 probability. So a zero that we have nothing, let's imagine that nothing is around zero, but you have a probability around that because everything are not certain. You have a probability around

notes

summary

13m 37s



that, that it's still the same situation. You have no effect. And you see here, I have put my stress holds alpha. So you see a part at which you decide if I get this result, the result is the value horizontal. If I get this result, I will reject H0. I will say when it's good, you see, it could belong to this population which is around zero. It's still possible to have things like that. So you put the stress hold and you decide to cut at that level. And then it's why it's complicated in this one of the things complicated in this explanation that to make the opposite probability, you need to make an hypothesis. So imagine that in fact, the reality is that my average is not at zero here. In fact, I have an effect and my population is in mu one. In fact, it's where my reality is. So when you decide to fix at alpha, in fact, you cut also this distribution and it's the blue surface that represents the probability beta, the probability to be wrong when you say that you have an effect when you reject H0. And you see that this surface depends on a few things. It's depend of alpha. It's clear that it depends on alpha. You see that if you move alpha, this surface, if you make alpha smaller, you have a bigger probability beta. If you make alpha smaller, you have a better surface in blue. But it's also the position of mu. If mu would be 10 times more

notes

summary

6.29 The concept of contrast

- Often the standard hypothesis $H_o : \mu_1 = \mu_i = 0$ is not answering the question of the investigator
- What is important is the comparison between treatments such as $H_o : \mu_3 = \mu_4$
- It is equivalent to $H_o : \mu_3 - \mu_4 = 0$
- A contrast is defined as (a is the nb of treatments)

$$\Gamma = \sum_{i=1}^{a} c_i \; \mu_i \qquad (14)$$

- The t-statistics is then

$$t_o = \frac{\sum_{i=1}^{a} c_i \; \bar{y}_{i.}}{\sqrt{\frac{MS_E}{n} \sum_{i=1}^{a} c_i^2}} \qquad (15)$$

at right, so the surface would be really, really, really small. Why is this beta is important? Because alpha represents the risk of the provider, the risk of the one taking the decision. And beta represents the risk for the one supporting the decision. Come in your plant, I make the analysis, I decide that you have an effect of the tool. You invest a lot of money for solving this problem and changing the tool. But it was not the origin of the problem, in fact. It's why it's the probability of the consumer. And what we call the P is the probability to not have an error, which is also, let's say, the quality. The beta is the defect. P is the quality. So it's why many times you will see when you talk about quality, about what is the power of your task and you understand it's not so easy. It depends on the sample variance. You understand that it depends on how these curves are narrow or larger. Larger they are, more of those probabilities are important. It depends on the numbers of samples because it's reduced the noise making more measurement. And it depends on the magnitude of the effects. That means the difference between mu zero and mu one. You have mu zero here, mu one here. If you are very distant, relatively, this is probably

notes

summary

18m 13s

6.30 Contrast confidence interval and LSD

- The confidence interval (CI) of a contrasts can be evaluated by

$$\Delta = t_{\alpha/2,\nu} \sqrt{\frac{MS_E}{n} \sum_{i=1}^{a} c_i^2} \qquad (16)$$

ν being the DF of the model and then

$$\sum_{i=1}^{a} c_i \, \bar{y}_{i.} - \Delta \leqslant \Gamma \leqslant \sum_{i=1}^{a} c_i \, \bar{y}_{i.} + \Delta \qquad (17)$$

n being the number of samples for each treatment, N being the total number of observations

- The least significant difference (LSD) is defined as

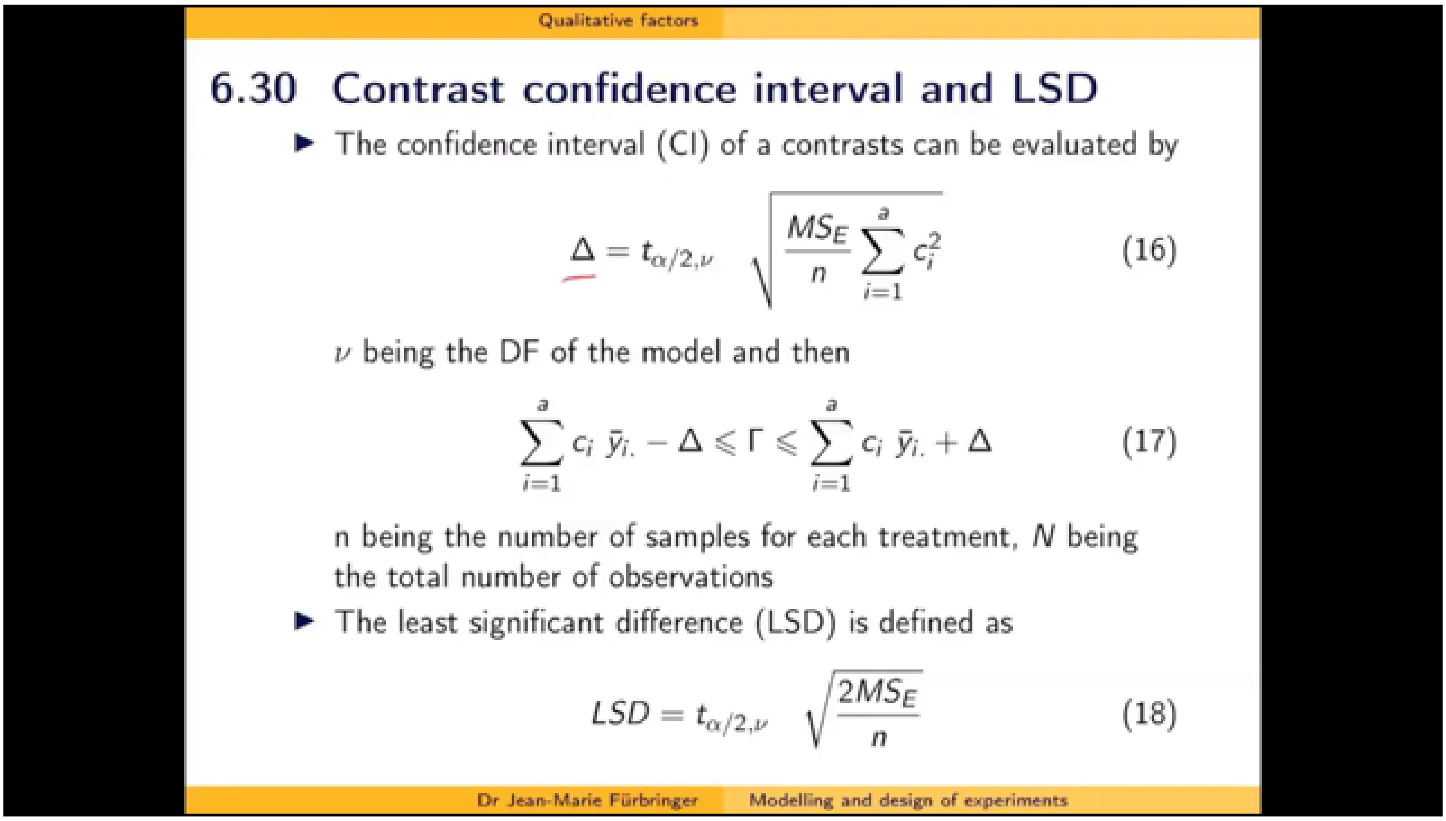
$$LSD = t_{\alpha/2,\nu} \sqrt{\frac{2MS_E}{n}} \qquad (18)$$

the R smaller. So that means also, I introduced it before, that in fact it's not sufficient to make a nanova for really saying, okay, I have an effect of tool. You should be one step further. And this is done with the concept of contrast. So as a standard hypothesis is just to prove that all the effects, they were tau in the other slide, in this slide, they are called mu. That means all average of my tools are around zero. No, what I'm interested is an alternative hypothesis. I would like to know if I really have a difference between one tool and another. If I have an effect of tool, that means that between the tool one, the tool two, the tool three, I should have a difference. So in fact, what I would like to test is better this. That's the difference between two effects is zero. And I would like to reject this hypothesis. So for doing that, we make a contrast, we call it gamma. This is a special sign, is a capital gamma. It's a gamma and it's the linear combinations of my effects. Mu e represents the effects and say is a coefficient. And those say are one minus one or zero. So when I want to make a difference, I make a plus one to one of the tools I want to make a difference, minus one to another. And the rest I pull zero and I like that I can calculate the statistic about this difference. And the statistic that I'm calculating, and this is called T, you see that it will be a student test for proving it. It will be the ratio between this coefficients, multiplying the average of my results divided by a mean square, also multiplying by this type of coefficient, the C coefficient. This formula seems complicated. Most of the time it's the algorithm

notes

summary

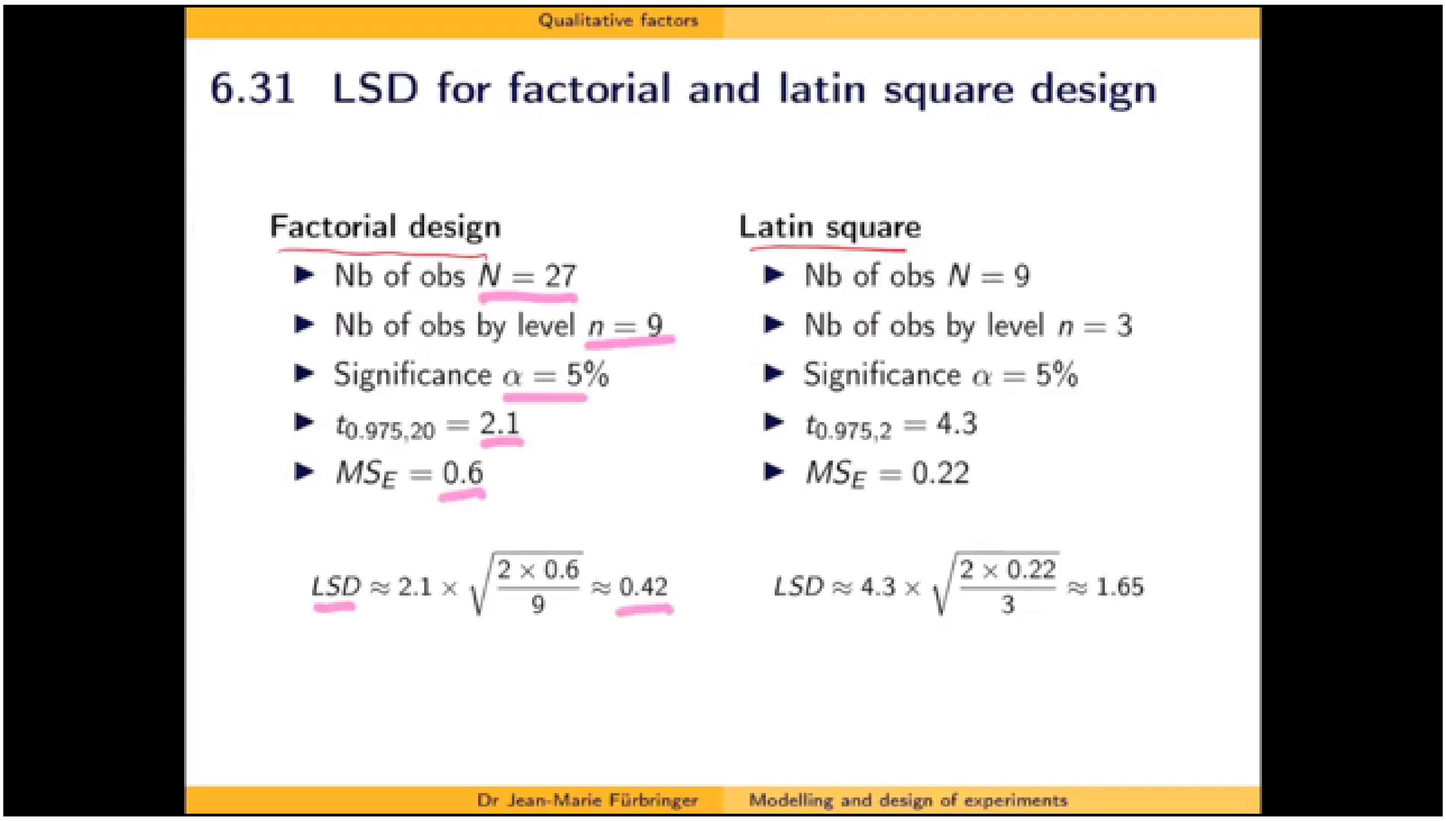
19m 49s



that will calculate it. So I don't spend too much time on and then what could be interesting or calculating those effects and the routines I will show you afterwards of the routines of MATLAB, they are calculating those contrasts. And another thing, so the contrast is when you have a case, but after you can test your experiment and you can calculate what we

notes

summary

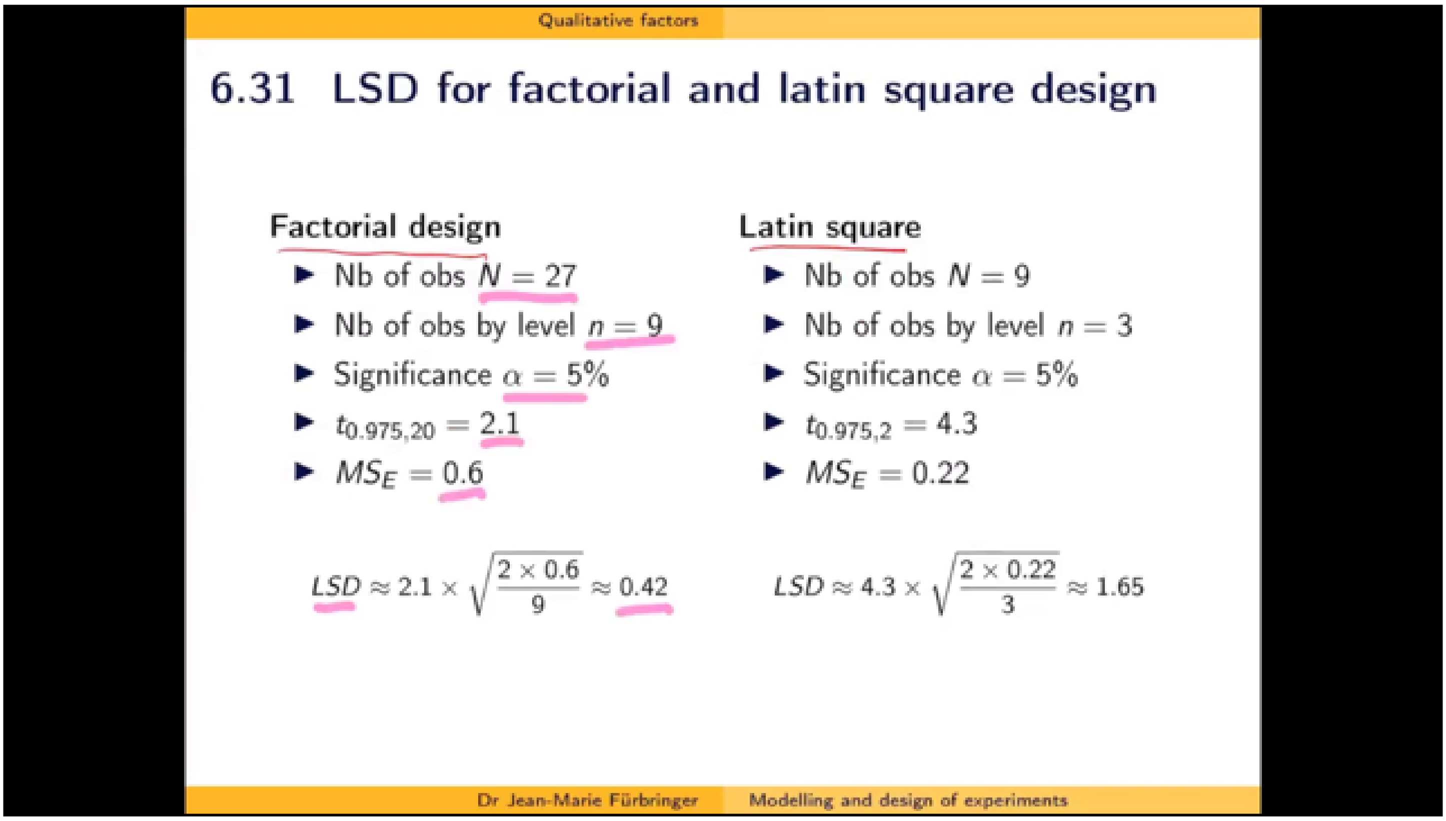


called the confidence interval on the contrast. And this you do it by quite making the same type of calculation, you have your mean square error, you divide it by the numbers of motherment and you make the summation of your coefficients of plus one and minus one when they are square. So it's the sum of C square will just be two when you are comparing two coefficients. If you want to make something more complicated, it will be different, but most of the time it will be two. And you multiply by this T student distribution, we already see it when we calculated confidence interval for coefficients parametric situation. So like that you are able when you are calculating a contrast to have a plus minus so you can see what is the contrast and with this, with the contrast when you this difference is bigger than zero. So you consider that you have an effect in the can. The same thing as you calculate confidence interval in other situations. After there are another thing that you can calculate which is qualifying your experiment is the LSD, the least significant difference. You understand that when you are preparing a test, so you are applying a factorial design or you are applying Greco-Latin square, what is important is what you can detect. If you have very small effects, it would be more complicated and you will probably need more measurements for detecting an effect. If your effect is very clear, probably with a few experiments only you can detect it. So this LSD, it's calculated as two times the mean square error that you have in your experiment divided by the numbers of experiments that you have. You take the root square and you multiply it by the student coefficient and this will give you a value and this is let's say the minimum type of effects that you

notes

summary

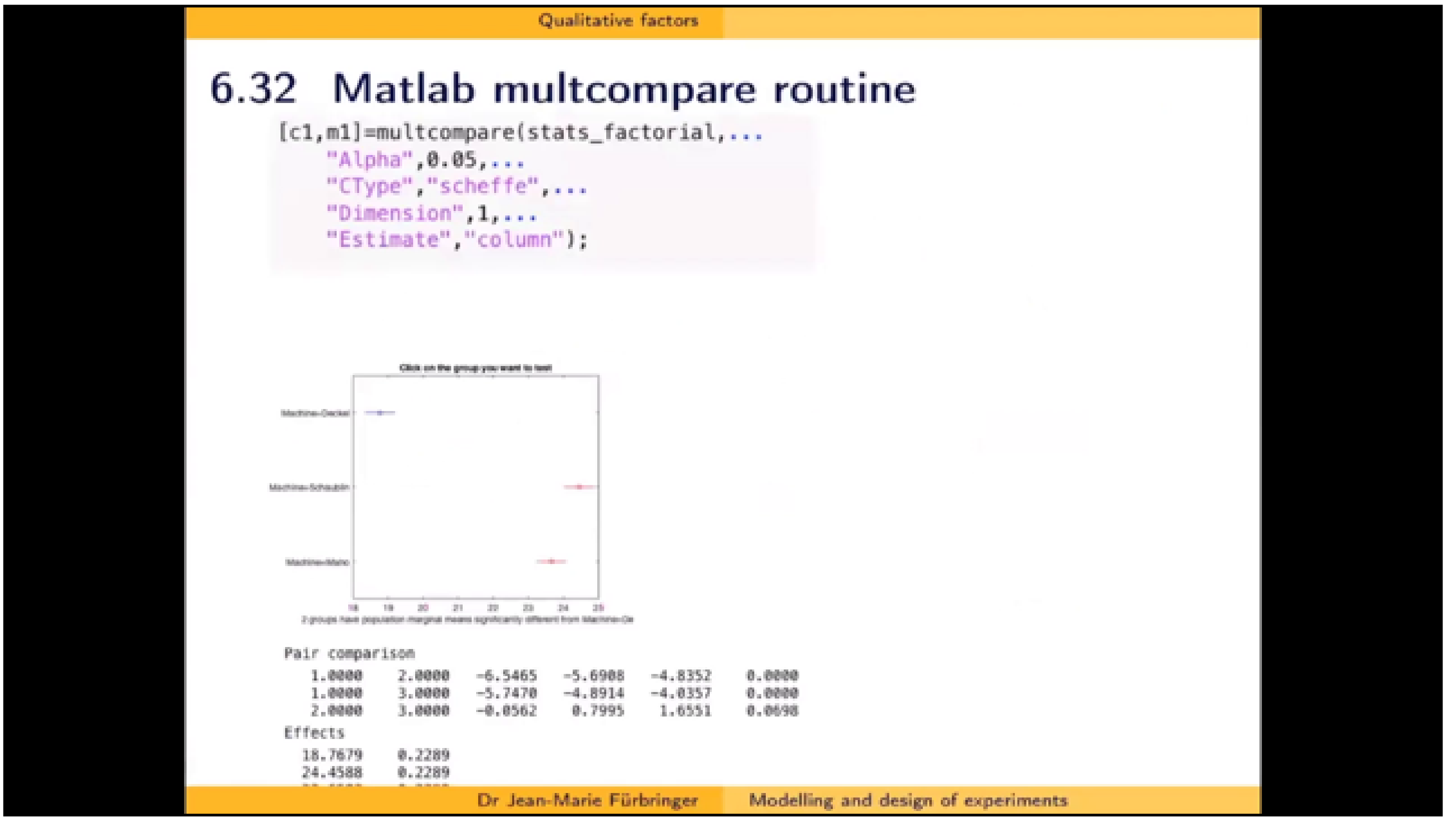
22m 49s



can detect. It's all the time in statistics you are swimming in uncertainty and the risk is all the question is all time to manage the risk, control the risk and understand if those risks are so you can verify if you know more or less the size of an effect. You can verify if your experiment would be able to detect it or at the reverse you have made an experiment and you say okay, if the effect is smaller than that I would not be able to detect it but perhaps it exists. It's all the time the case. It's not because you don't know something that these things doesn't exist is this philosophy of Kant with the phenomenon and the new man. We have access to the reality only through phenomena but the reality is more complex and there are sometimes things we cannot detect our instruments, our strategy is not sufficient. So I have calculated the situation for the two cases that were illustrating this chapter and we had if you remember a factorial design where we have made 27 experiments. We have made all the possibilities with my operator, with my tool, with some machines and I had another situation I called Latin square where I just made nine observations. So you are here only the results of the calculation so with the factorial design we have 27 measurements is the cost of your experiment. By each level we have observed nine, we have nine measurements for each level of each of my different categories. I have put a significance level at five so I have calculated the student coefficient and it was something else too. That means that when you have standard deviation you have to multiply them by two for having confidence interval. So I have a mean square of my experiments that was of 0.6

notes

summary



and then I can calculate an LSD of 0.42. That means that I would be able to detect effects that are smaller than if you remember we were talking in microns so it was something that if I have defects if I have a problem provoking defects in average bigger than 0.4 microns I would be able to detect them what is the cause of of my defects. In the Latin square we only have nine so it's three times cheaper. Each level it tested only three times when it was tested three more times or nine times in the other situation. I have the same level of confidence as reference five percent and in this case because I have less degrees of freedom I only have two degrees of freedom you see here the two before I have 20 degrees of freedom in the previous case than my student coefficient is the double as the other 4.3 and when I'm calculating the mean square I have less experiments than my mean square is even smaller but so is a student's distribution which is putting against the equilibrium in a situation and when I'm calculating my LSD I get 1.6 meaning that I would be able to detect causes that provoke errors bigger than 1.6 microns. So you see as I so happy to repeat it each time there are no free meals in statistics. Latin squares are very good but they are less precise it was exactly the same thing when I present you for example the Dolert design and the other the composite design for the second degree coefficients the Dolert design is very good because you can get results with less experiments but that means that the confidence interval will be bigger. Eventually it's simpler to make nine different experiments and you make replicates and perhaps it's cheaper to make three times my latin square and it

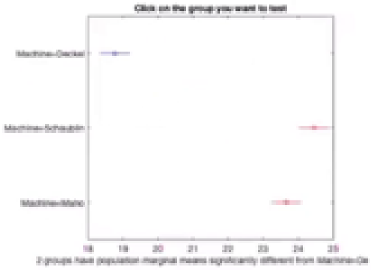
notes

summary

27m 25s

6.32 Matlab multcompare routine

```
[c1,m1]=multcompare(stats_factorial,...
    "Alpha",0.05,...
    "CType","scheffe",...
    "Dimension",1,...
    "Estimate","column");
```

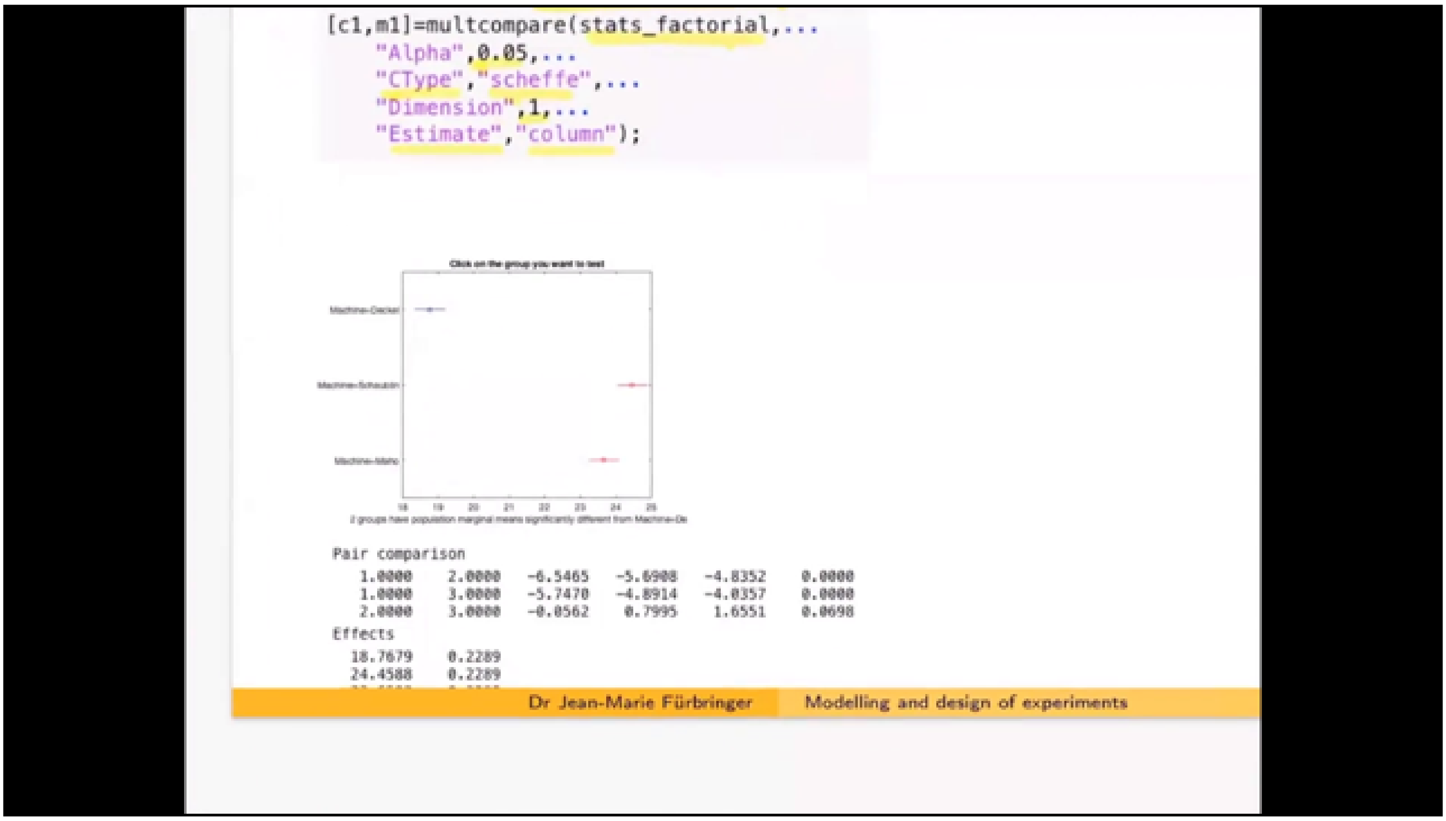



```
Pair comparison
    1.0000    2.0000   -6.5465   -5.6908   -4.8352    0.0000
    1.0000    3.0000   -5.7470   -4.8914   -4.0357    0.0000
    2.0000    3.0000   -0.0562    0.7995    1.6551    0.0698
Effects
   18.7679    0.2289
   24.4588    0.2289
```


could be interesting to see what's happened probably would be very very very comparable in this aspect if I would have make three times my latin square and perhaps it was experimentally simple. So here is what you

notes

summary

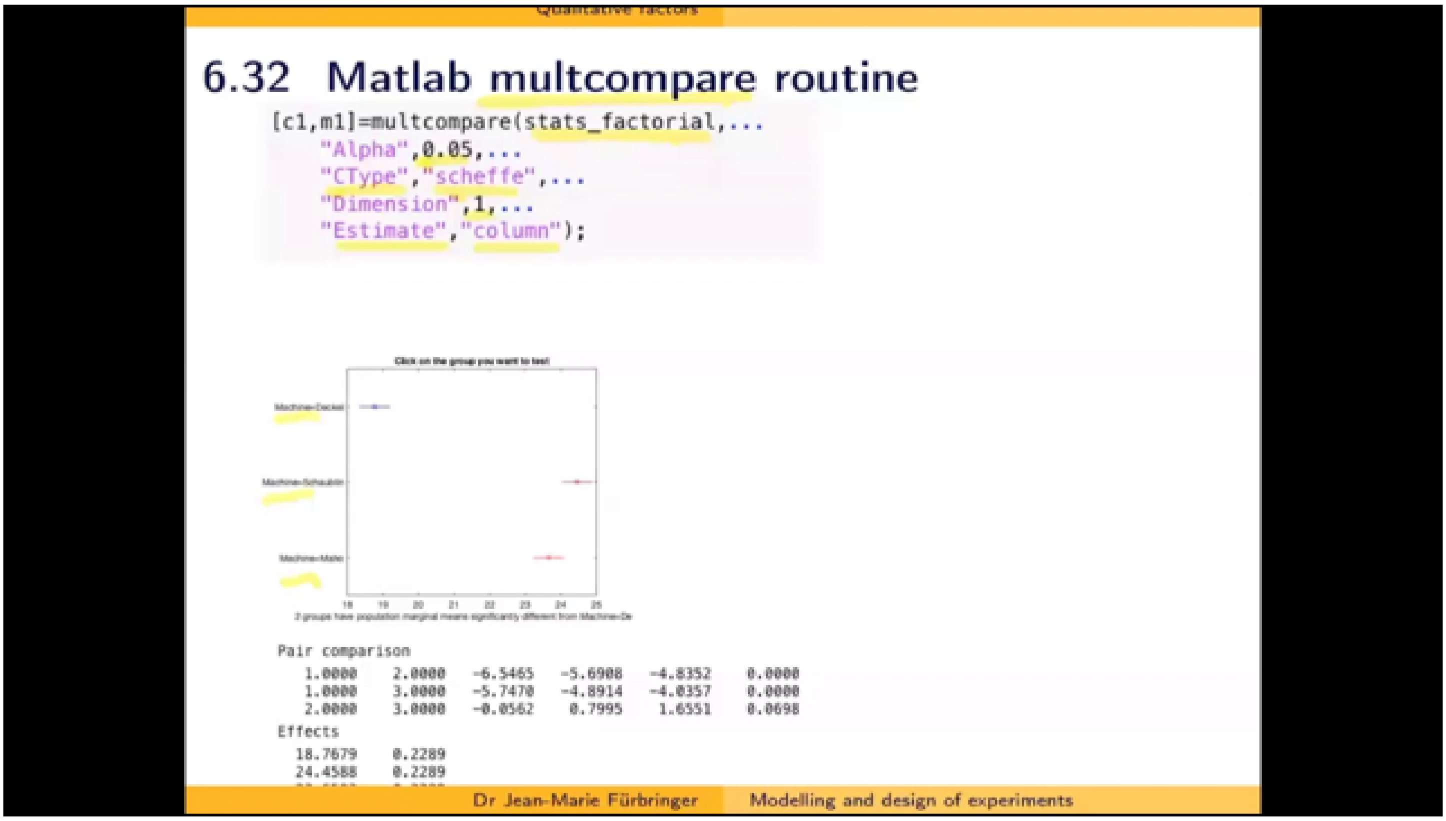


can do for getting those numbers with MATLAB so there are one routine which is very nice which MULT compare and this routine will do all the calculation for you. You have to have made this analysis that we see last last week so the ANOVAN routine the ANOVAN routine would have provide you with the statistics of your data the model that you have and you have to tell the type of the level of confidence you have to give the type of the coefficients that we are using at this C values and so the name for those are Sheffield and I don't know the other I don't know them it's the Sheffield. You have to tell the dimension in which you are working in this case was one because we are testing the coefficient of one factor and we have to tell I don't remember why we have estimate and column privilege where in your data you have to find things I don't remember

notes

summary

30m 3s

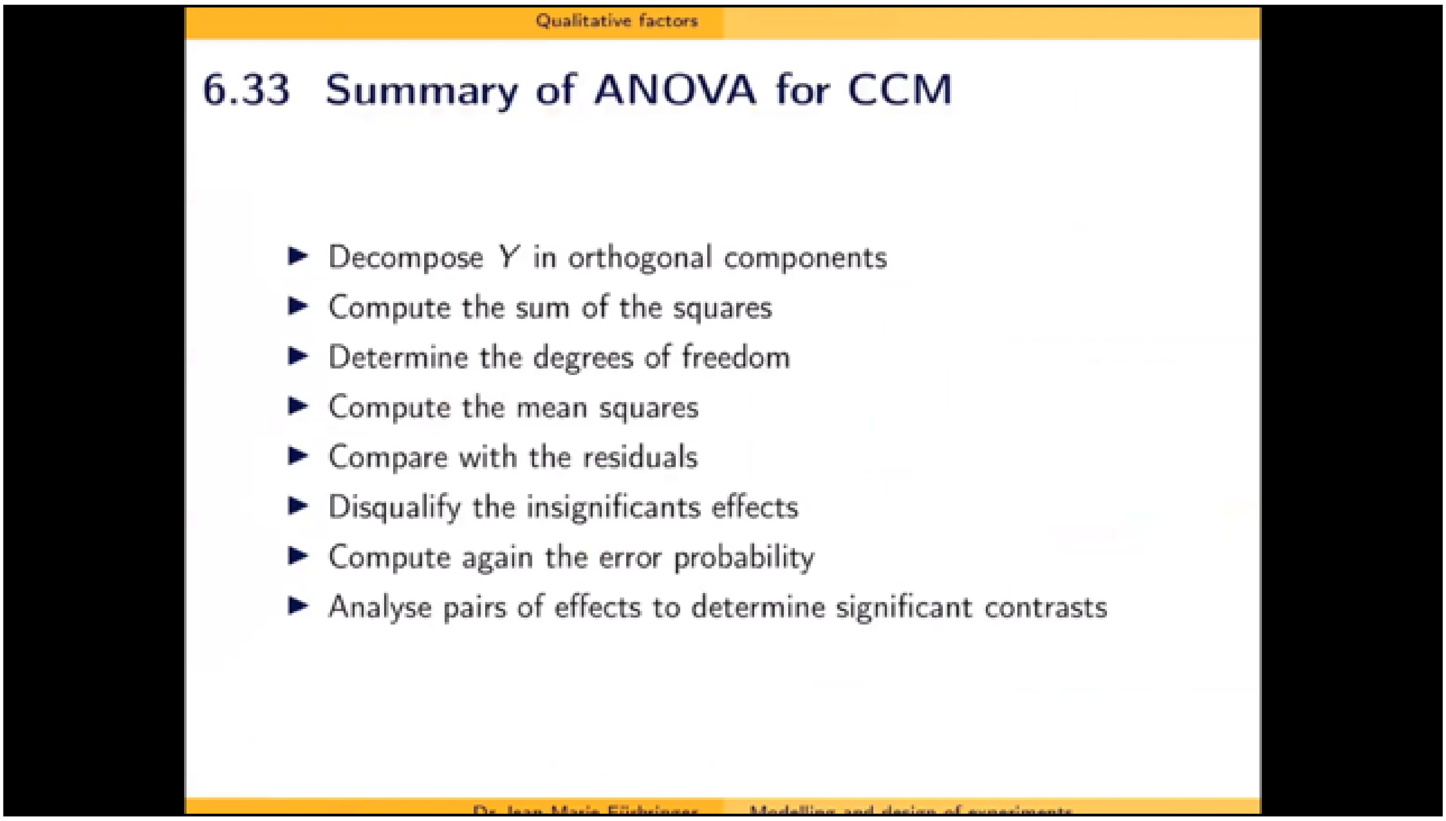


this and what you get is this result you will get a sort of dot plot kind of where you have the three different effects for your machine yeah it's the machine it's written on the left side so you see the three effects so in my dot plot they were all at the same level so here they put them at different level and the one is in blue the two other is in red because we are testing it if you realize this routine you will see that it's a sensitive graphic you can you can click on the

notes

summary

31m 22s

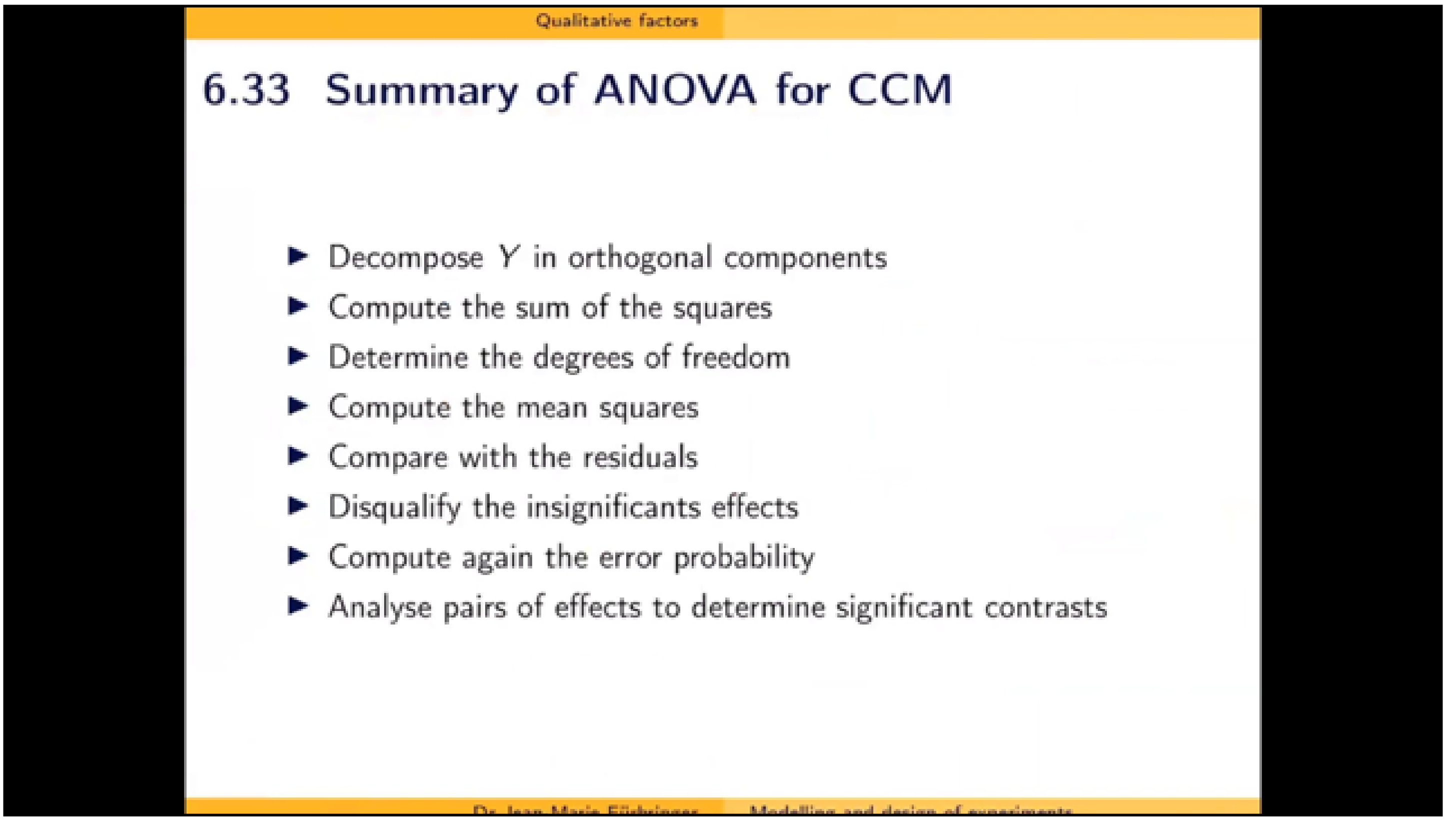


different effects and so now I have clicked on the blue one and it's telling me it's comparing me to the two other so this blue one is really different than the two other you remember there are two machines they were really close and when they change they place when we make a latin square or when we make the factorial and after it makes a pair comparison so it was between the one the first one and the second one but you have to check which one is the second one probably the first one is the blue one and the chauvelin is a second one and it giving you the contrasts 5.6 and it's giving you the confidence interval for the minimum and the maximum they could have so the difference between these two tools is 6.5 microns and 4.4 microns so this first tool the tool decal is in comparison to the the machine chauvelin is provoking in average 5.6 microns of error less and it's in between 6.5 and 4.8 and it's giving you the probability of this difference so it's a p-value the last one is a p-value so it's giving you a p-value which is smaller than thousands of a percent it's like zero zero zero it's because it's very very small smaller than your your numbers your digital numbers of your digits after you see the first one with the third one so it would be this one compared to this one you see that the difference is a little bit smaller you see also that this effect the red one is an effect which is closer to the blue one again you have in average 4.9 microns that is going between 5.7 and 4 it's a minus sign so it's less for the first one so it's a diminution of the error and a probability which is also very good and

notes

summary

32m 1s



after I'm comparing the two red between them the second with the third you see that the difference of them is crossing the zero so it's between minus 0.05 and 1.6 so there are a small difference but not really significant and you see that the probability is 7 percent smaller than the 5 percent that you use usually as the the analysis so in this case I would say okay I clearly have an effect between the machine decal and the other but between the shirblin and the Mao I'm not sure I really have a difference and after it gives you sorry for my slides which is cutting but it gives you the values of of the effect so as a summary of this before just having a pause we have made ANOVA in the constant coefficient model CCM is constant coefficient model so we are decomposing my data in orthogonal composite it was a sweeping you you you remember this operation where we are and each of those elements are orthogonal one to the other we compute the sum of square that means usually we don't do because we don't do things in excel but okay the sum of square of each of those rectangles we determine the degree of freedom remember in this type of model is the number of level minus one for the effect is the product of the degree of freedom of the main effect for the interaction and it's one for the constant so we can after that compute the mean square we can look at the residual disqualify insignificant effect and after we can compute the error probability mainly the alpha more seldomly the beta and we can analyze pair of effects for determining the concept what really matters

notes

summary