Course material

Course:

ENG606 / PHYS 442

Video:

summary

Concepts (extracted from automatically generated subtitles):

Complicated things. Causal model. First time. Main argument. Time components of your explanations. Different tools. Best thing. Lot of bias. Idea of the real theory. Noise etc. Profit of the information. Different levels. Quite very interesting case. Real world. Line of hypothesis.

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

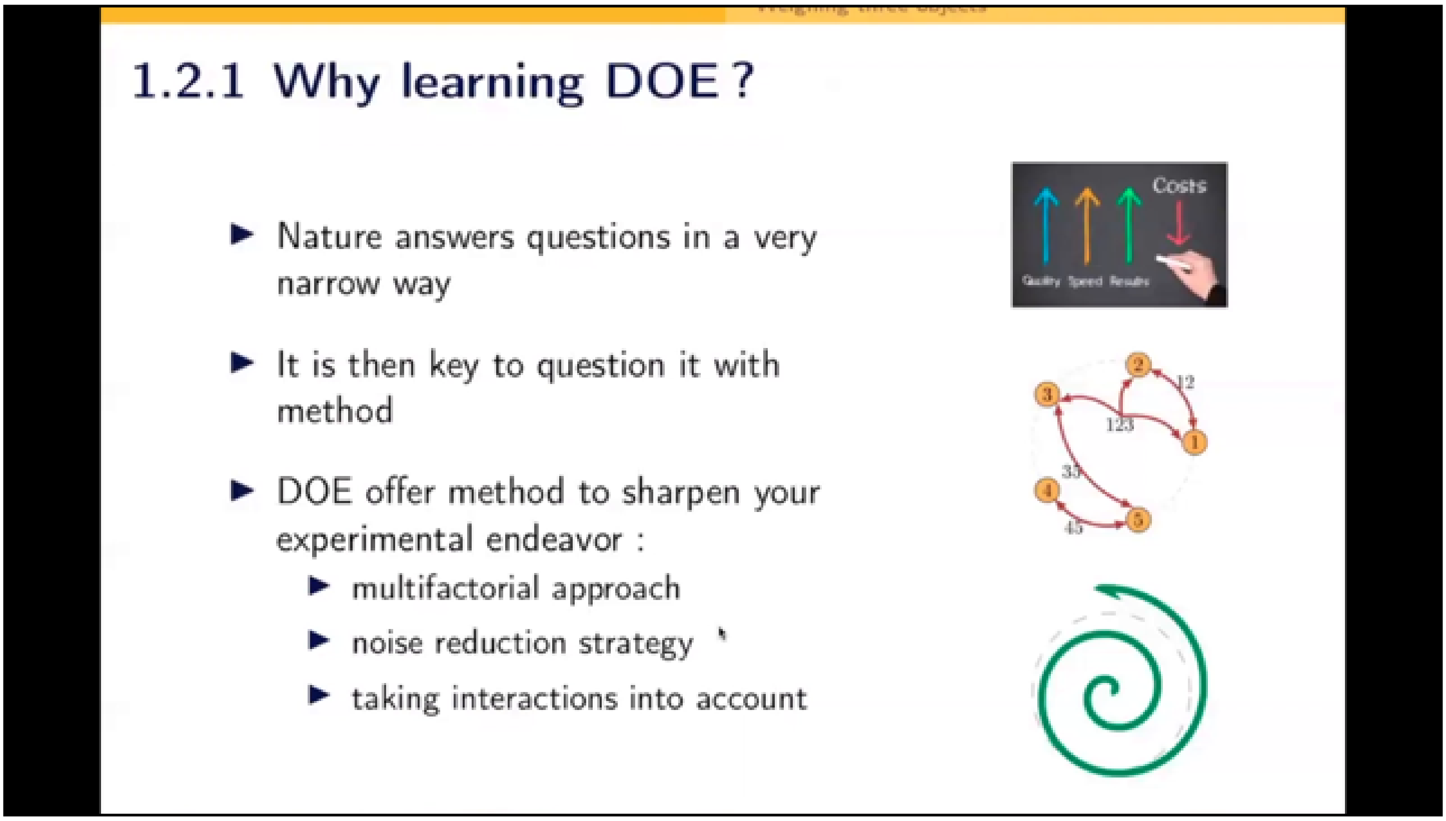
to video

...

notes

summary

0m 0s



These subtitles have been generated automatically So, if you remember, we start with this type of, yeah, what is DOE and why we are learning

notes

summary

0m 1s

1.3.9 Pareidolia

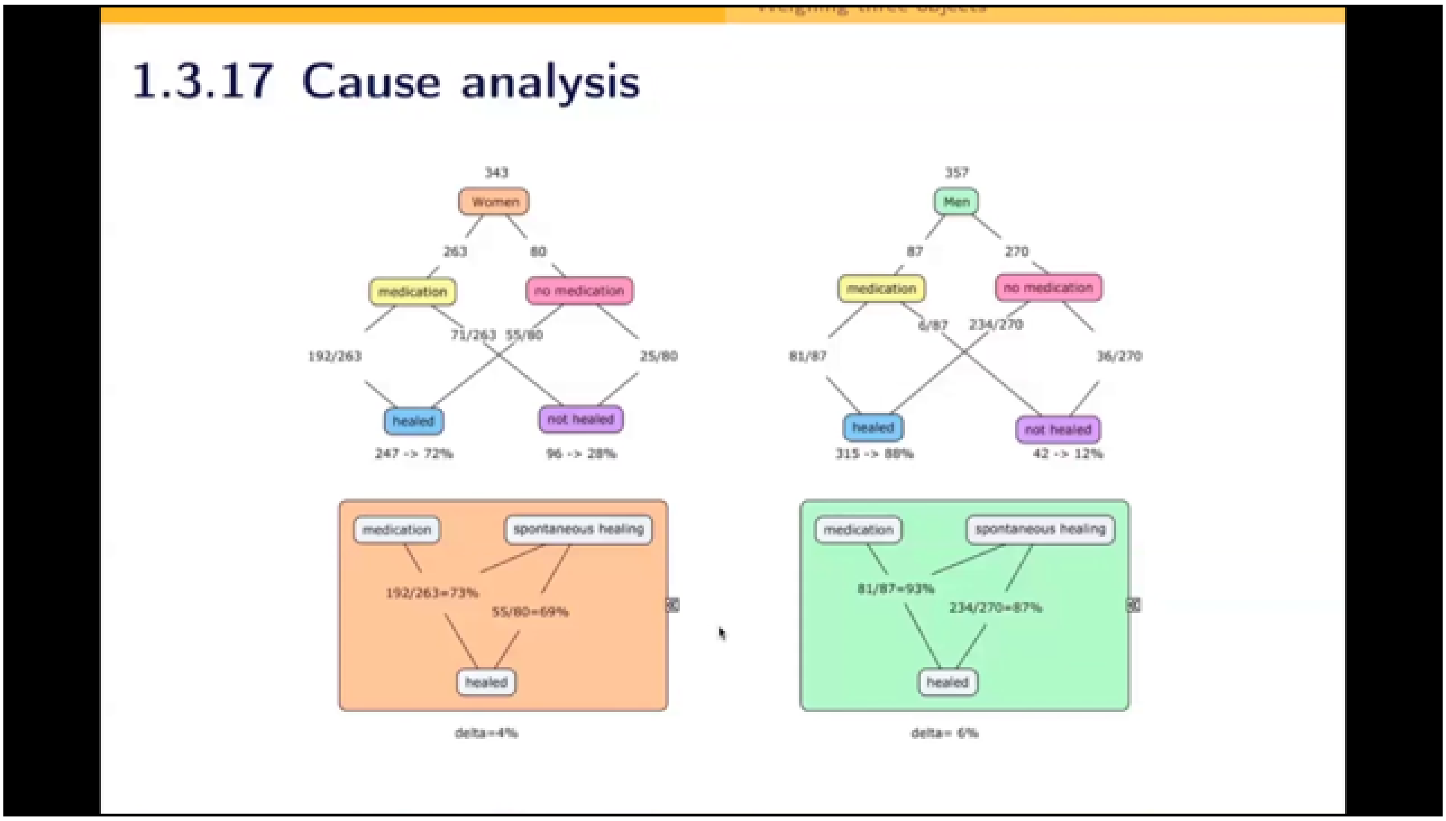
Our brain has the tendency to see motives in random sets as well as a tendency to make a script, which means to tell a story, and give a meaning to an image. Statistics offers tools, like statistical tests, to pass the perception.

DOE. So, the main argument for making DOE is that the world is noisy, multifactorial, with interaction. So, if people ask you why, why making so complicated things? They are not so complicated at the end, but they are not things you have the first time in mind when you think experiments, you say, okay, I will make experiments everywhere. So DOE let you make less experiments that if you don't think, so it's better to think before. It's step by step, but remember, the world is noisy, multifactorial, and you have interaction. That's why you have a very huge interest of having a strategy.

notes

summary

0m 8s

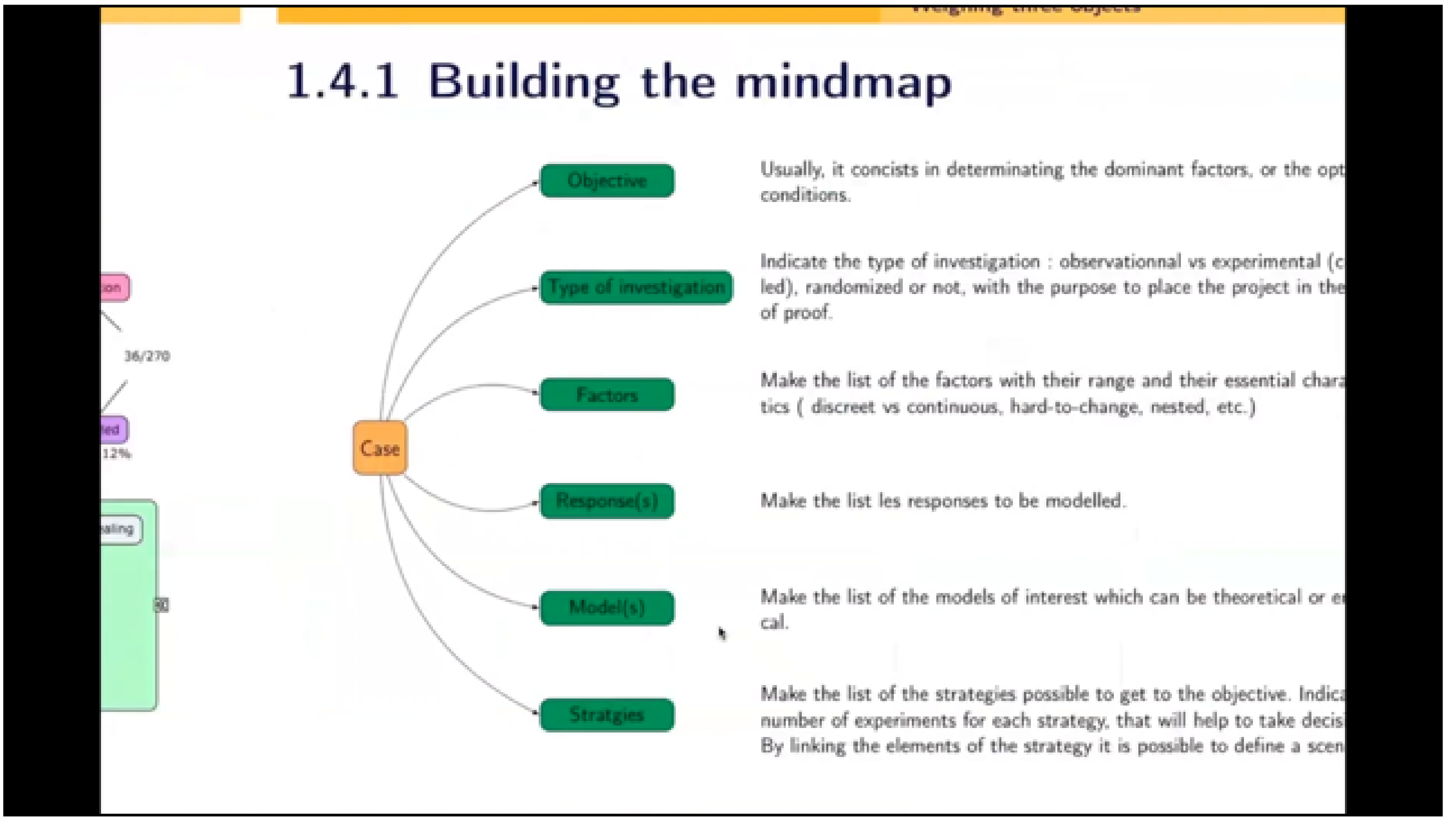


After I try to convince you that when we are analyzing data, we have to be very careful because we have bias. We have a lot of bias and it's important to see things graphically, but not only also to have tests and over. We have different tools because we have all the tendency to see the sun from our door. We said in French, we are midi à sa porte. And it's this tendency of the human, of seeing things, of not accepting the things that are uncertain and trying to interpret even things that cannot really be uncertain. For me, the best thing is paraidolia when you have the impression that your laboratory or your plug is in life, as life is exactly illustrating for me this problem of the bias. Be conscious of the bias, look after them, be careful with this. After I also show you that sometimes the data, depending if you look at it precisely, if you look at it with fraction, be careful with the fraction and be careful. This is also related to the mixture also. There are some time components of your explanations or real component in your recipe, but also component in your reflections that are not apparent and really do this causal model.

notes

summary

1m 1s



Be conscious of what you take in your model and what you forget to take in your model and when you decide to forget something, be sure what you forget it. So it was this situation of this with the medicines that seems to be different for man and woman or when we don't know the gender of people. So be careful with that causal model who could really help you and follow what you are doing

notes

summary

2m 37s

3.2.2 ANOVA for two orthogonal parts

Source	SS	DF	MS	F	p
Partie 1	$SS_{\hat{Y}_1}$	P_1	$\frac{SS_{\hat{Y}_1}}{P_1}$	$x_1 = \frac{MS_{\hat{Y}_1}}{MS_\epsilon}$	$F(x_1, P_1, N-P)$
Partie 2	$SS_{\hat{Y}_2}$	P_2	$\frac{SS_{\hat{Y}_2}}{P_2}$	$x_2 = \frac{MS_{\hat{Y}_2}}{MS_\epsilon}$	$F(x_2, P_2, N-P)$
Résidu	SS_ϵ	$N-P$	$\frac{SS_\epsilon}{N-P}$		
Total	SS_Y	N	–		

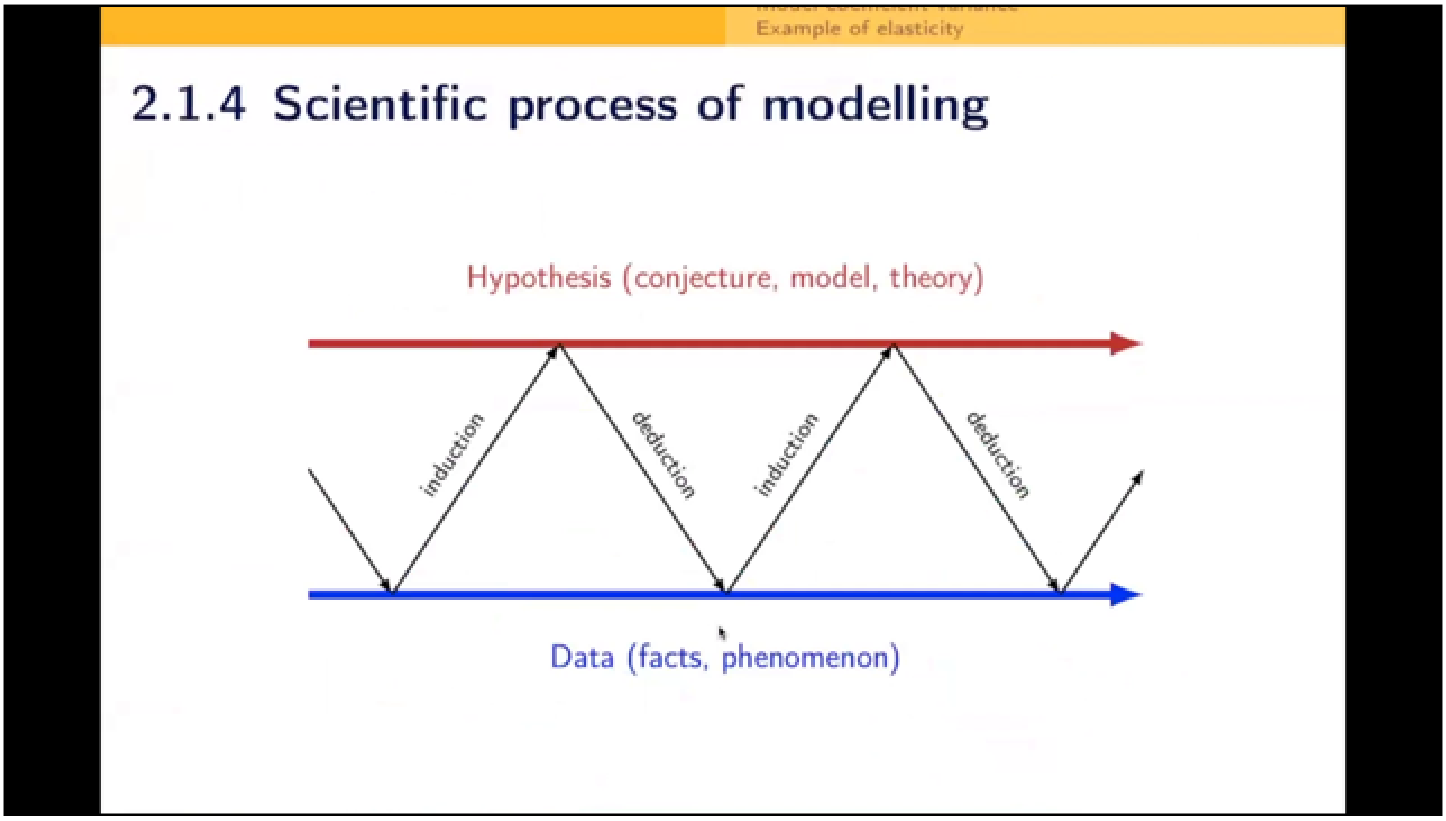
N is the number of runs and P_1 et P_2, the number of coefficients of the parts 1 and 2 respectively, $P = P_1 + P_2$

and etc. I propose you and I show you a lot of examples making this mind map. It's very interesting because you can summarize what you are doing. You can see very rapidly what is your objective, what is your factor, what is your data, what is your model. This is very important because we are playing with these three poles and you know that any

notes

summary

3m 8s

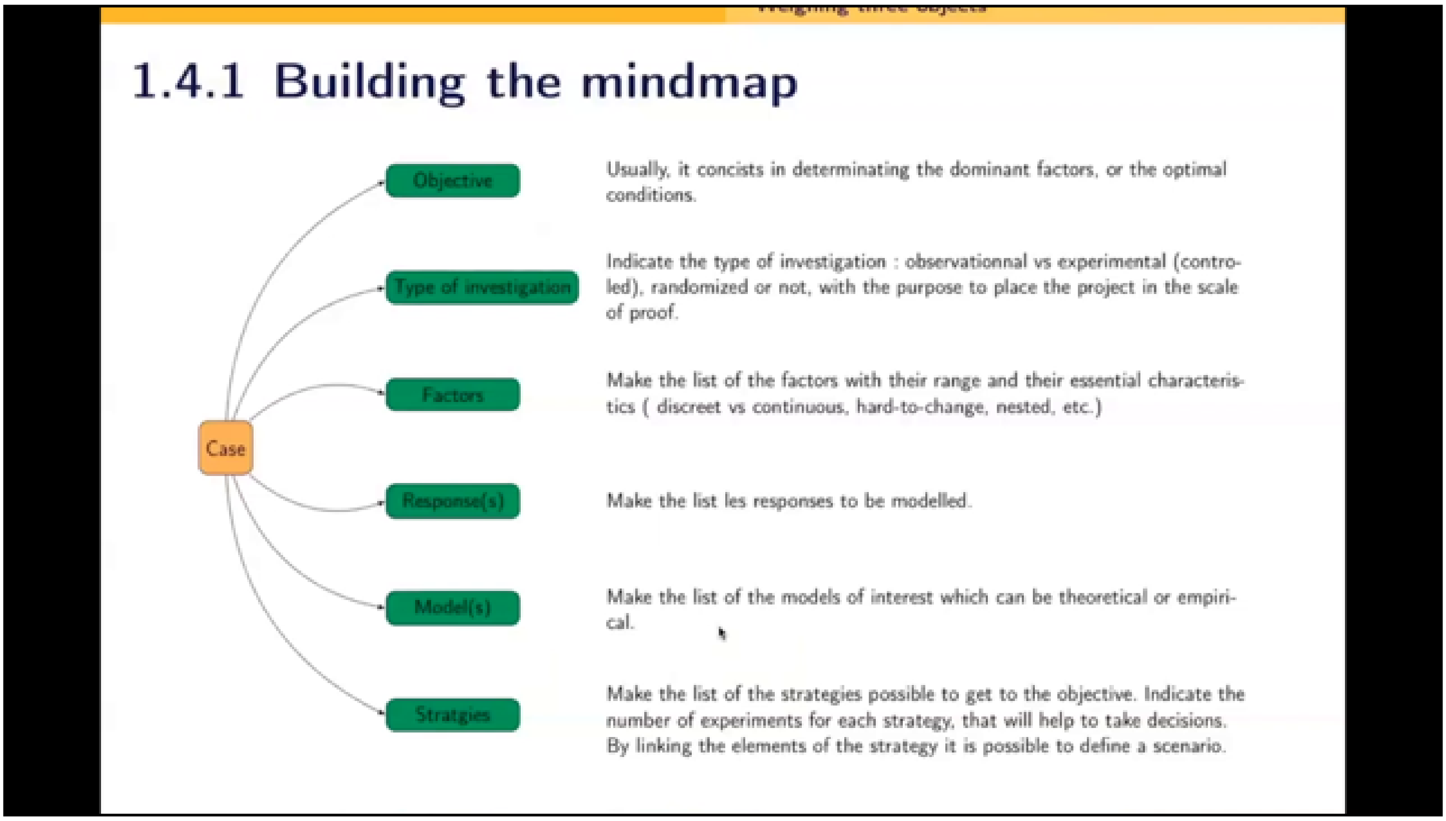


play with three is more complicated than play with two.

notes

summary

3m 37s

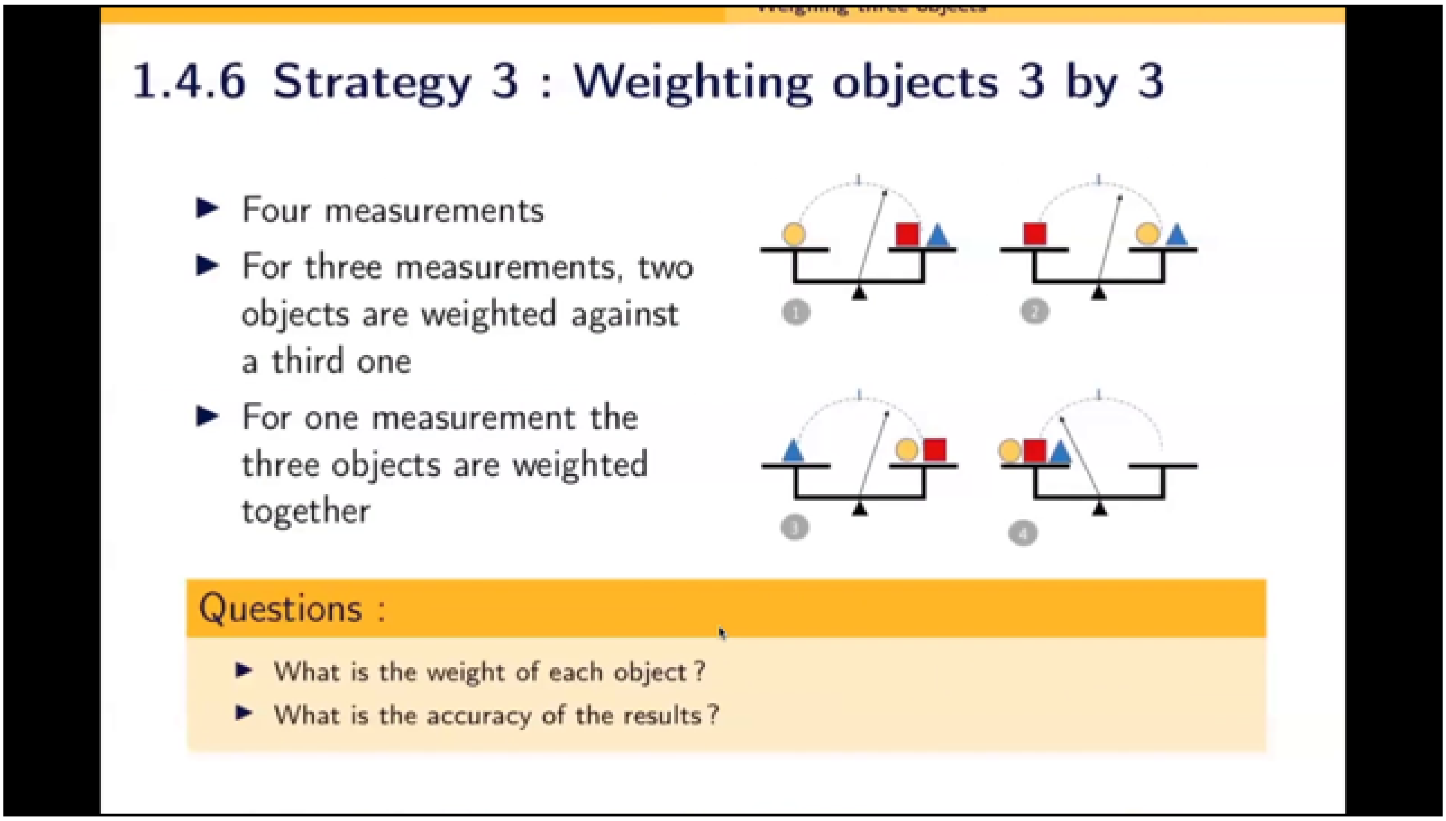


If you remember we start with this type of story. What is DOE and why we are learning DOE? The main argument for making DOE is that the world is noisy, multifactorial, with

notes

summary

3m 43s

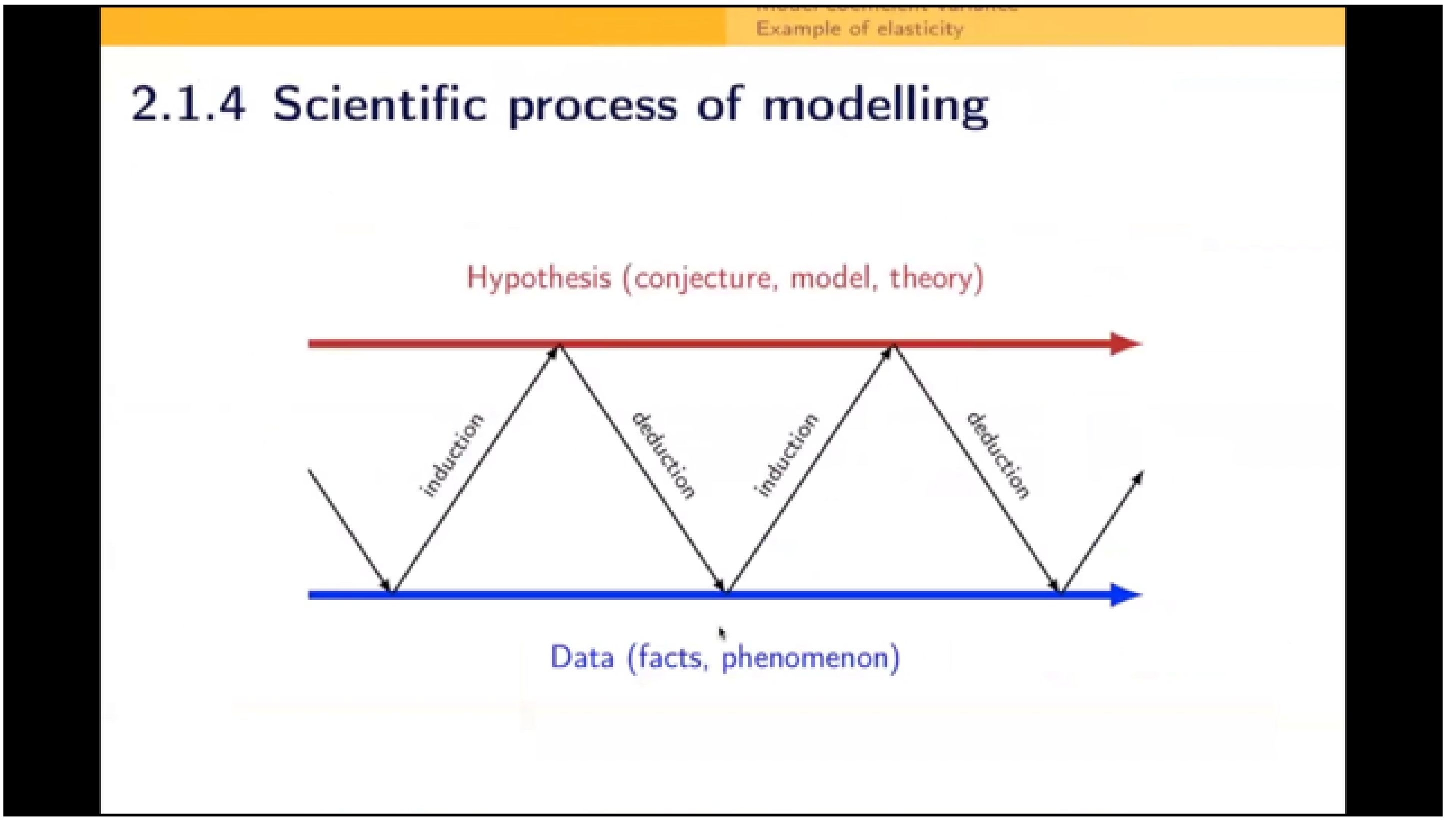


So for me this type of mind map helps you. You can change them, you can adapt them to what you want but it's very important. It's important to follow during your work because a few things will change so follow your mind map. And something which is key, remember that the strategy. So to understand you have strategies step by step and draw them, explain them especially the PhD students when you are making your project, not just list all the possible strategies. Understand when to use them so I will start to do that depending on what I will obtain. I will do that and do that after step by step. More you have a step by step strategy, better will be your strategy because at the start you don't know a lot. At the end you know a lot but it's too late so try to have steps because it helps you to take more profit of the information it's coming step by step when you are making experiments. So make model when you increase the degree of the model or you increase the numbers of factor you are considering could be one of those strategies. After I present here this type of situation in the scale it's very easy. Typically if you want to present them to somebody else which is very new to DOE it's

notes

summary

4m 3s

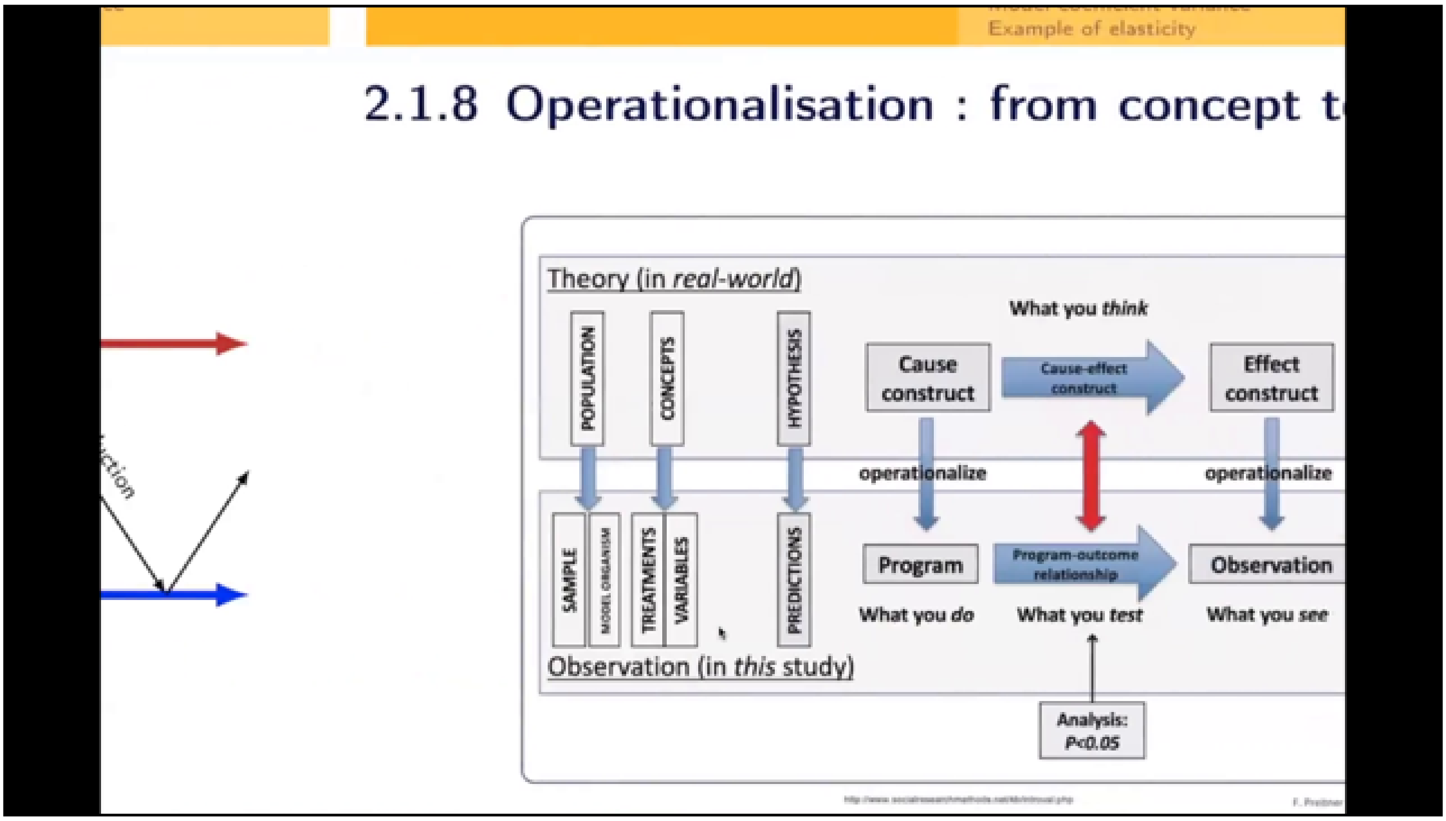


a good example because you see that in a very simple situation you can go in a situation when you have two sigma square as a variance to a situation where you have one divided by four sigma square you improve your quality of your data eight times with a very simple manipulation with the same cost. It's a quite very interesting case for showing and explaining to others and perhaps for you also to understand why it's why we're improving what we have an advantage of doing that what we have a very high payback when we are doing DOE is because of the variance because the law of the variance when we are making difference the variance is adding and so we have interest of multiplying the occurrence in which we can make comparison in these letters thanks to the student distribution to diminish the variance.

notes

summary

5m 25s

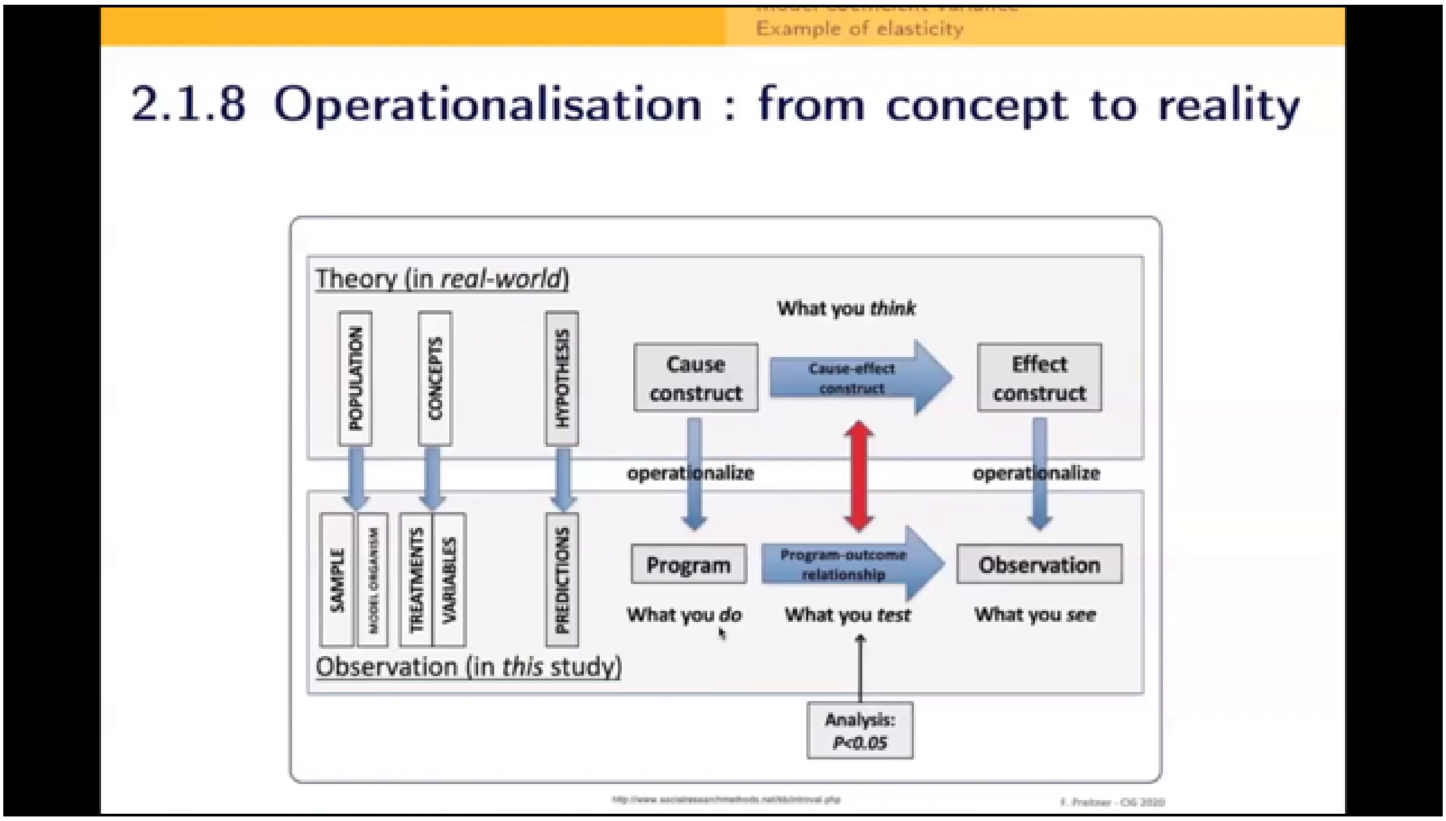


After I also present you how we are advancing this is also a global science philosophy principle that's also personally we are going step by step and we are following all the time and thesis is exactly that a master project could be also very close to that we have one line of hypothesis are the ideas a theory and we have one line of fact and data and we are all the time going from one to the other and design of experiment will help you to do this but remember that the tricky words here is induction which is not so easy that we think sometimes it's just a role of the same importance and the same dimension but in reality induction is more complicated than deduction because deduction you just apply rules of logic when induction you have rule of logic but you have to have quite a nose for choosing the right data avoiding noise etc which is not the case in deduction.

notes

summary

6m 22s

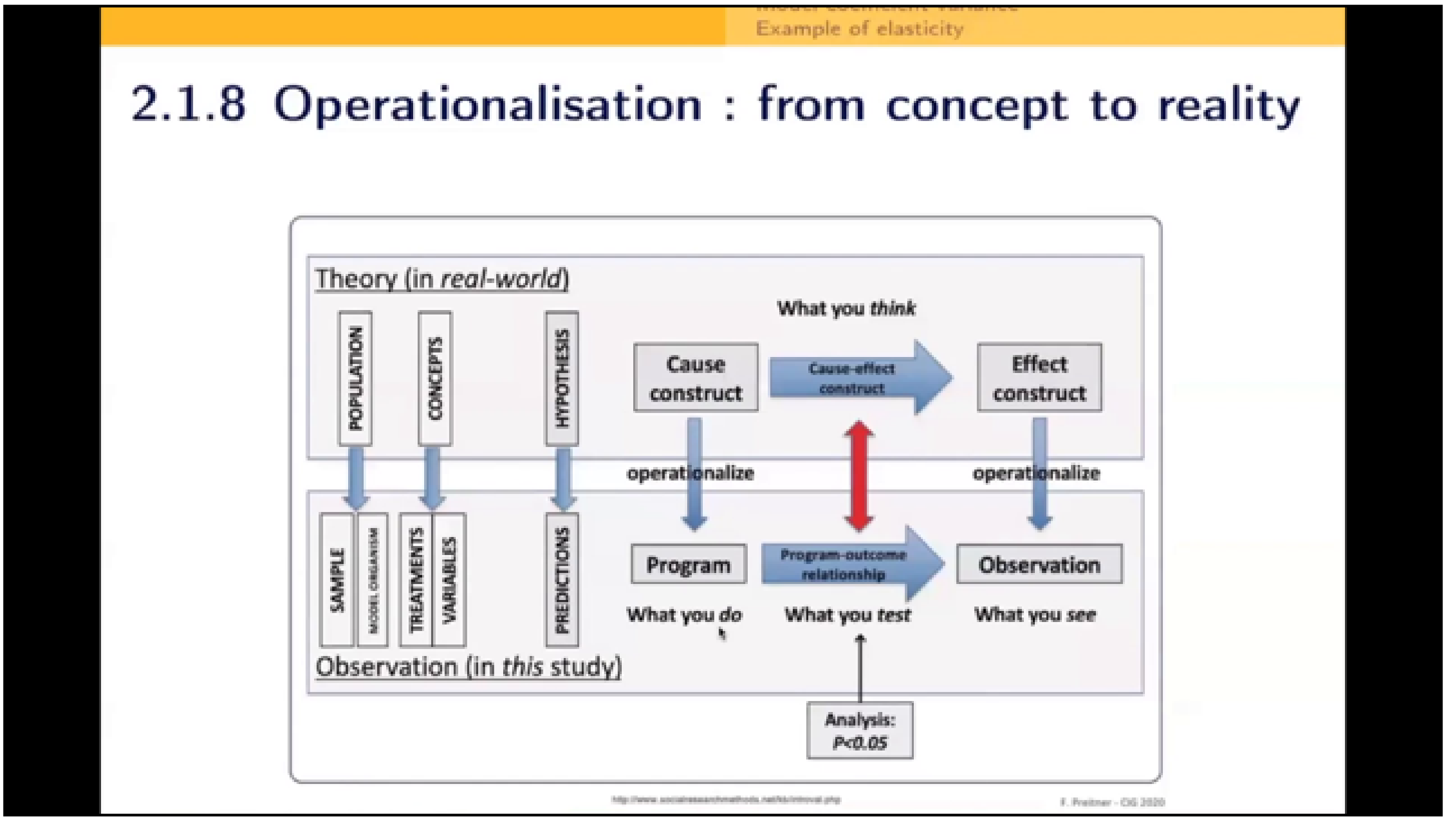


After I also explain you in this two lines approach you have the idea of the real theory which should apply to the real world and what you are doing in the lab you remember also my comics when the guy was trying to jump the canyon and goes through a device and finish as a stone so okay there are a few elements that you need to do I didn't spend a lot of

notes

summary

7m 31s

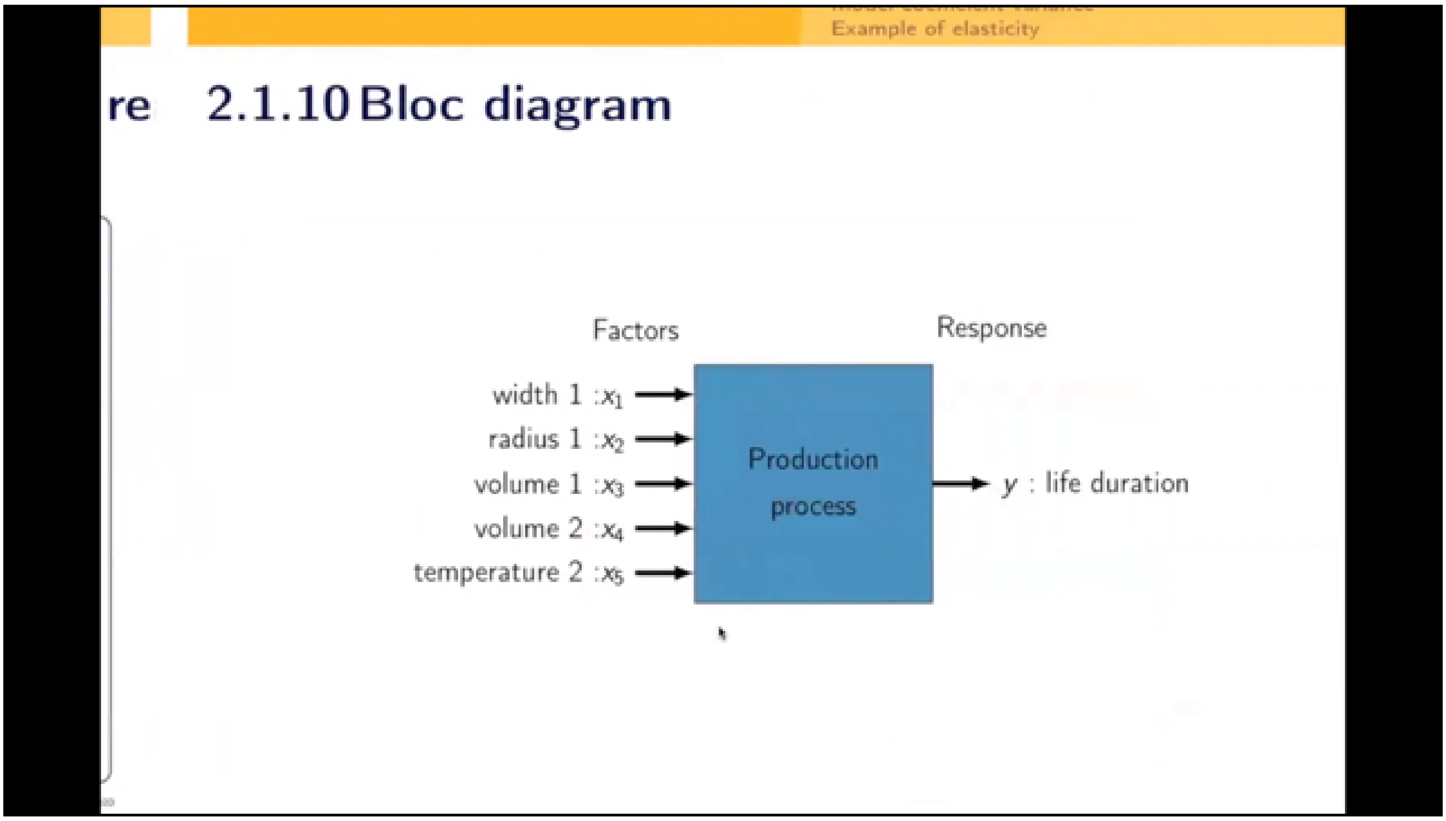


time it wasn't the large view of the course but you all the time to be sure that what you do in the lab is applied to what you would like to apply in reality when you are outside of the laboratory because the objective of an experiment is not to stay in laboratories to go outside of the laboratory and to say something on the real world so we have a few concepts because this idea of construction so you have to be coherent in your construction and when you are in the conclusion you really need to prove that you do things correctly scientifically in the lab but you also have to prove that what you did in the lab is coherent to what you have also outside of the lab the lab and so you need in your conclusion to go step by step and concluding on those elements I do my job statistically correctly I do my

notes

summary

8m 1s

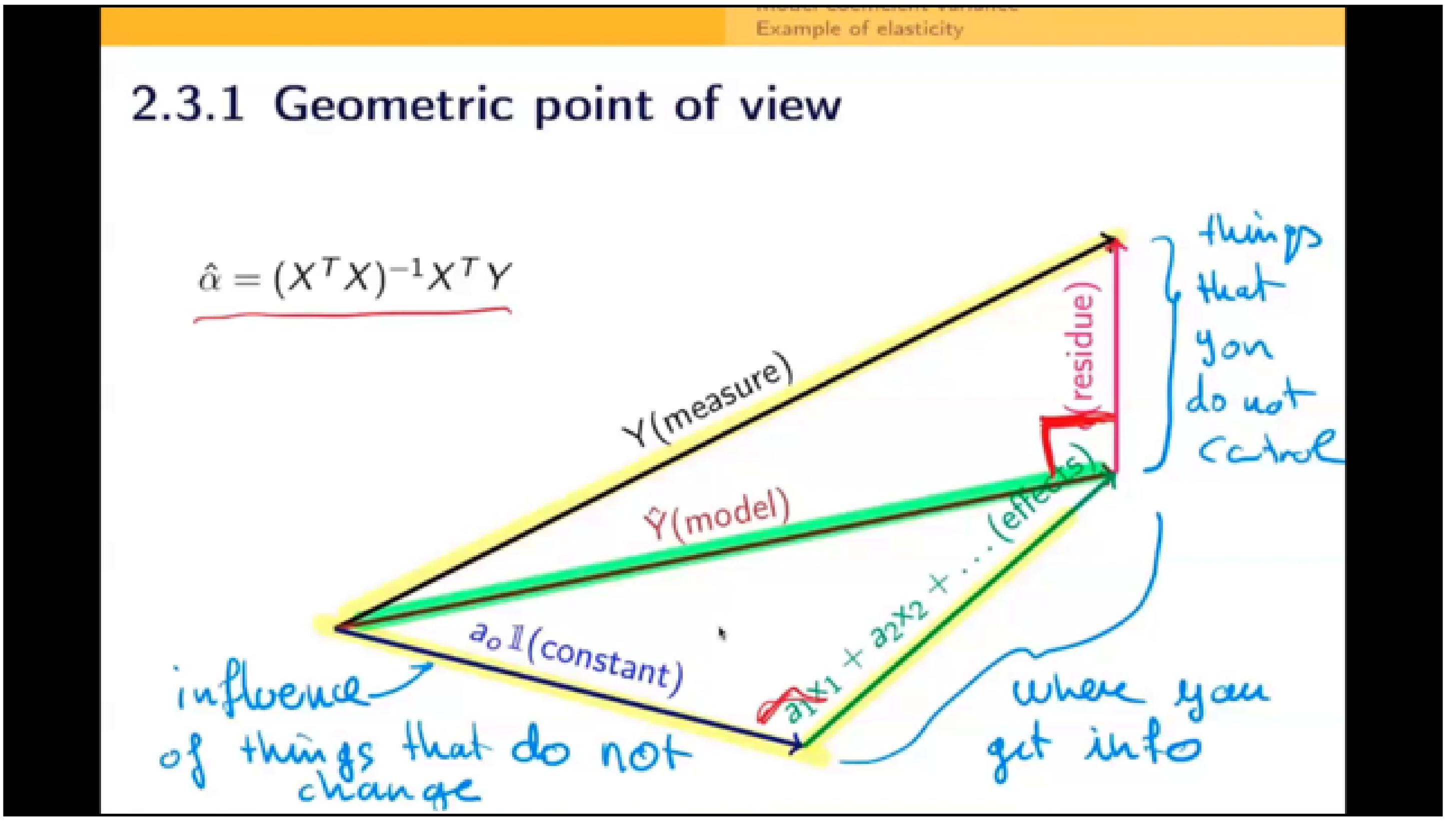


physics correctly I do my statistics correctly and I also do my adaptation from the lab my samples my my selection of what I wanted just to be sure that it really represents

notes

summary

8m 50s



the big picture not only the small picture you have perhaps too much of this type of graphic is really my vision you come with your problems and your technology and for me everything summarized to a thing like that I have input and output and I'm trying to make a model so when you work there are one moment when you are the engineer really doing the job thinking to your technological problem to your specific theoretical problem and there are another moment when you need to be just a statistician managing information understanding from where is coming the information and where goes information from where is coming the noise and how do I do that my output is not influenced too much by the noise and for me is this slide that represent that after I have insist quite a lot on this graphic which is the geometrical interpretation of a model when we are making a model in fact we are making projections the equivalence of geometrical projection and you need to understand the different levels the data the model the different effects understanding that with the when we are doing that we are separating the world or the signals from the world in three elements signal which is not so interested because it's constant things we are not analyzing we are not changing during the experiment the parts which is where we get new info we have analyzed we have moved factors we have analyzed the effect of some factors and a part

notes

summary

9m 6s

3.2.2 ANOVA for two orthogonal parts

Source	SS	DF	MS	F	p
Partie 1	$SS_{\hat{Y}_1}$	P_1	$\frac{SS_{\hat{Y}_1}}{P_1}$	$x_1 = \frac{MS_{\hat{Y}_1}}{MS_\epsilon}$	$F(x_1, P_1, N-P)$
Partie 2	$SS_{\hat{Y}_2}$	P_2	$\frac{SS_{\hat{Y}_2}}{P_2}$	$x_2 = \frac{MS_{\hat{Y}_2}}{MS_\epsilon}$	$F(x_2, P_2, N-P)$
Résidu	SS_ϵ	$N-P$	$\frac{SS_\epsilon}{N-P}$		
Total	SS_Y	N	$-$		

N is the number of runs and P_1 et P_2, the number of coefficients of the parts 1 and 2 respectively, $P = P_1 + P_2$

which is which have moved during my experiments and is represented so noise it represents effect of something I'm not able to manage I'm not able to manage because too costly or because it's inherent to my process I have a noise I cannot change I have noise in my measurement I have noise in my my process and so understand that in fact we are make separating our information our sum of squares coming from the experiment in these three parts so we usually let the constant the part because we cannot sing so much we say a lot of things on my effect because this is a product of my job and we use to compare this with what I'm not controlling because what I want is that what I'm controlling my effects must be significantly greaser than my noise if my noise is too high so it exists other type of techniques like lock in etc I'm not entering at all in this in my course but you know that it exists a few situation but in the standard situation you want effect at all let's say at least three times bigger than the noise for starting to consider them we do that we finish that making a nanova so you know I've seen the structure of the anova we see a variance variation in my data and I'm trying to attribute the origin of this variance and so I'm calculating sum of squares I'm calculating degrees of freedom because if a random variable is more degrees of freedom a theory she can have more sum of squares so we have to adapt things to be fair between the different source and we after we compare the mean square with the mean square of the residue and we have then the possibility to have fissure ratio and p values but if you are working with simulation deterministic

notes

summary

10m 49s

3.2.2 ANOVA for two orthogonal parts

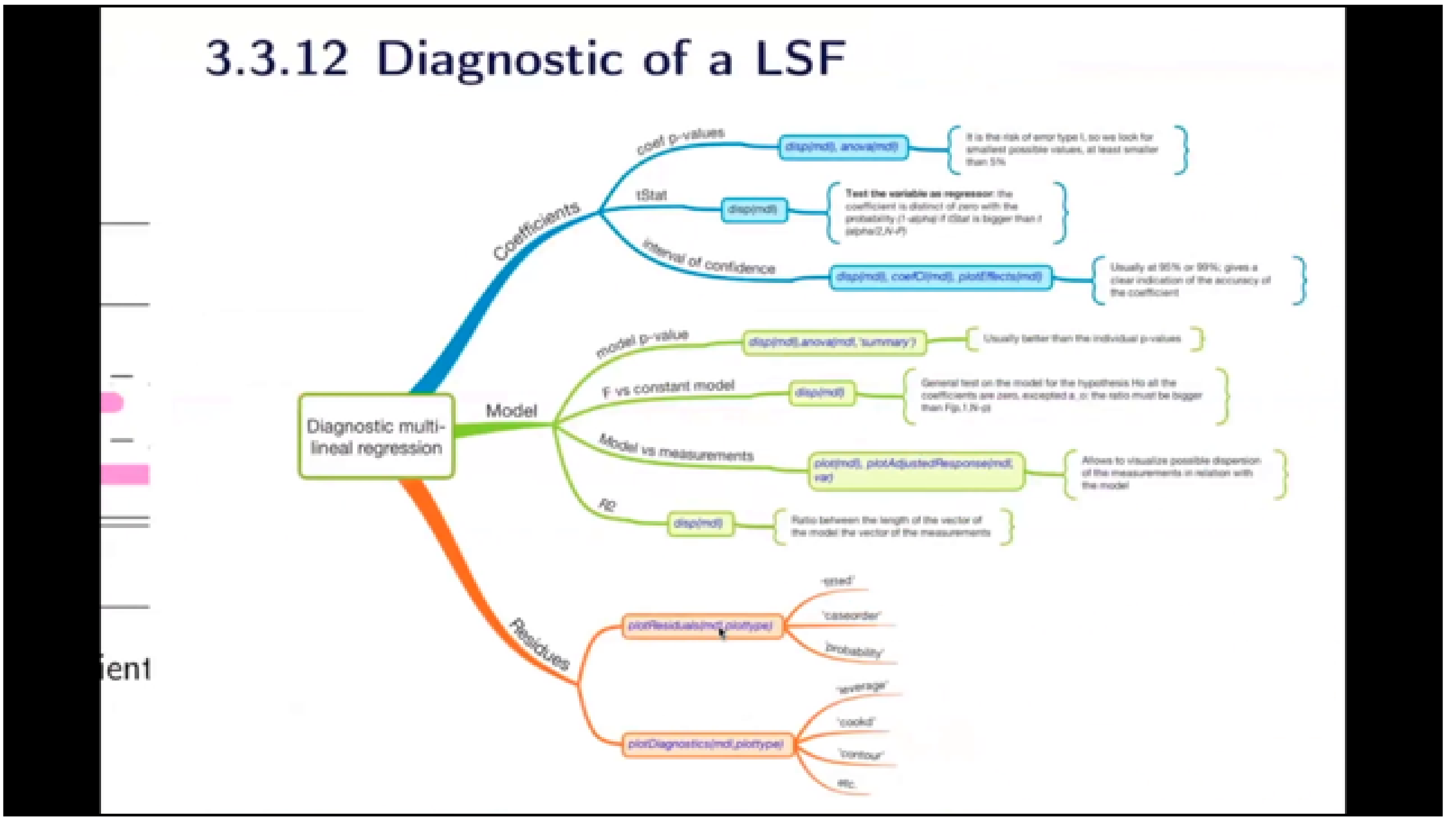
Source	SS	DF	MS	F	p
Partie 1	$SS_{\hat{Y}_1}$	P_1	$\frac{SS_{\hat{Y}_1}}{P_1}$	$x_1 = \frac{MS_{\hat{Y}_1}}{MS_\epsilon}$	$F(x_1, P_1, N-P)$
Partie 2	$SS_{\hat{Y}_2}$	P_2	$\frac{SS_{\hat{Y}_2}}{P_2}$	$x_2 = \frac{MS_{\hat{Y}_2}}{MS_\epsilon}$	$F(x_2, P_2, N-P)$
Résidu	SS_ϵ	$N-P$	$\frac{SS_\epsilon}{N-P}$		
Total	SS_Y	N	–		

N is the number of runs and P_1 et P_2, the number of coefficients of the parts 1 and 2 respectively, $P = P_1 + P_2$

simulation you do not have origin do not have noise so please don't present me an ova with a situation when you have deterministic simulation the p value is a nonsense the f value of the mean square yes f sense is the sense that you say what I'm not explaining in my model could be related to the size of what I'm explaining but please don't have p values it's not nonsense because if you are not playing with random variables

notes

summary

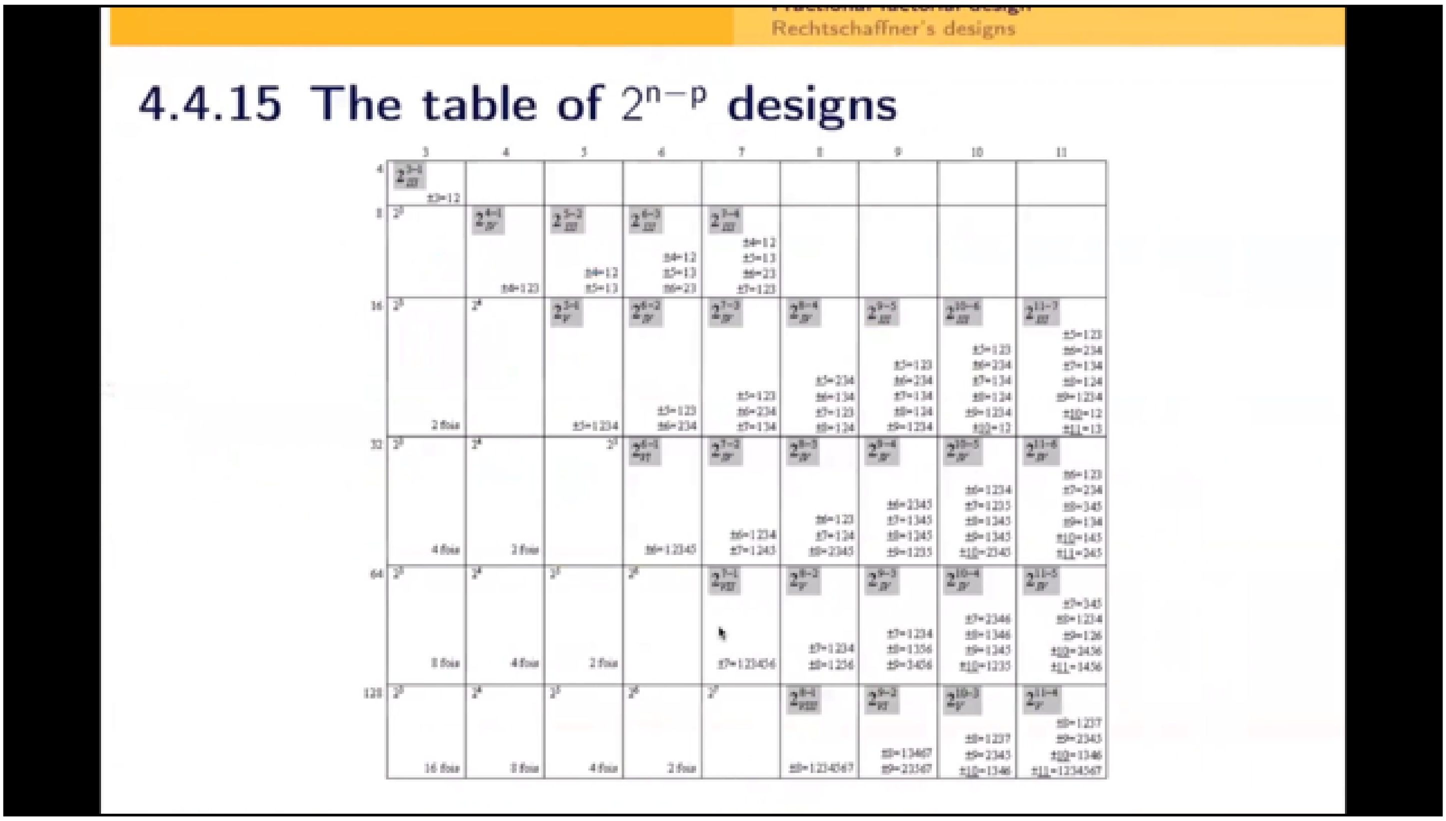


you are playing with projection with vectors but not whose random variables

notes

summary

13m 25s

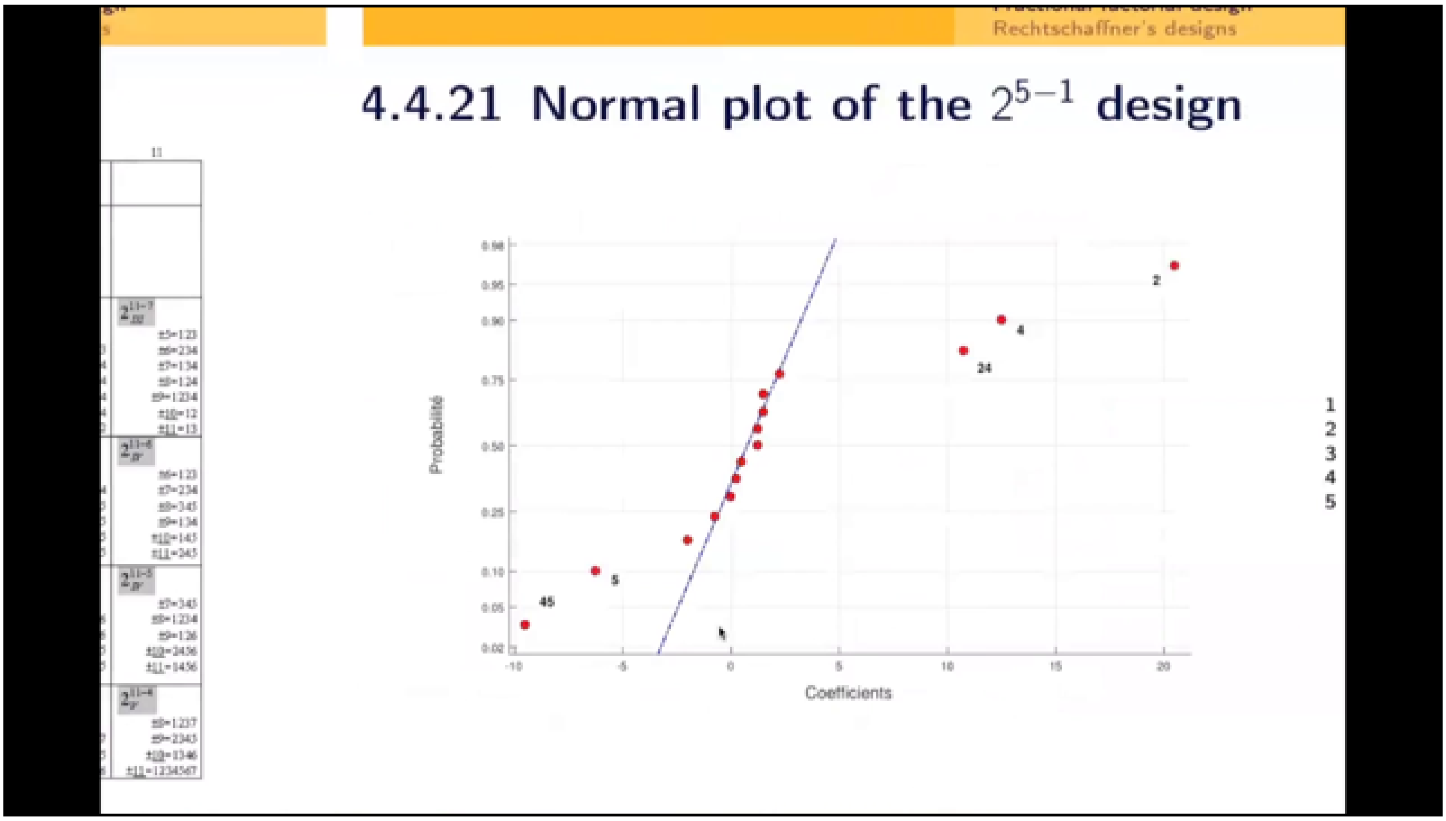


after I present you how to make least square feet was I didn't make a big explanation of the square feet I just use them matricially and I present you a few tools a few algorithms that you can use afterwards for checking so making a fit is not sufficient you after you need to come after that and and try to prove that's what you have done is wrong and if you don't succeed in proving that it's wrong it's okay you can accept it but so it's again in your role so first you are an engineer a physicist after you have been like a statistician and after you are somebody like try to criticize what you have been doing and it's also a very important role just to be sure that what

notes

summary

13m 30s

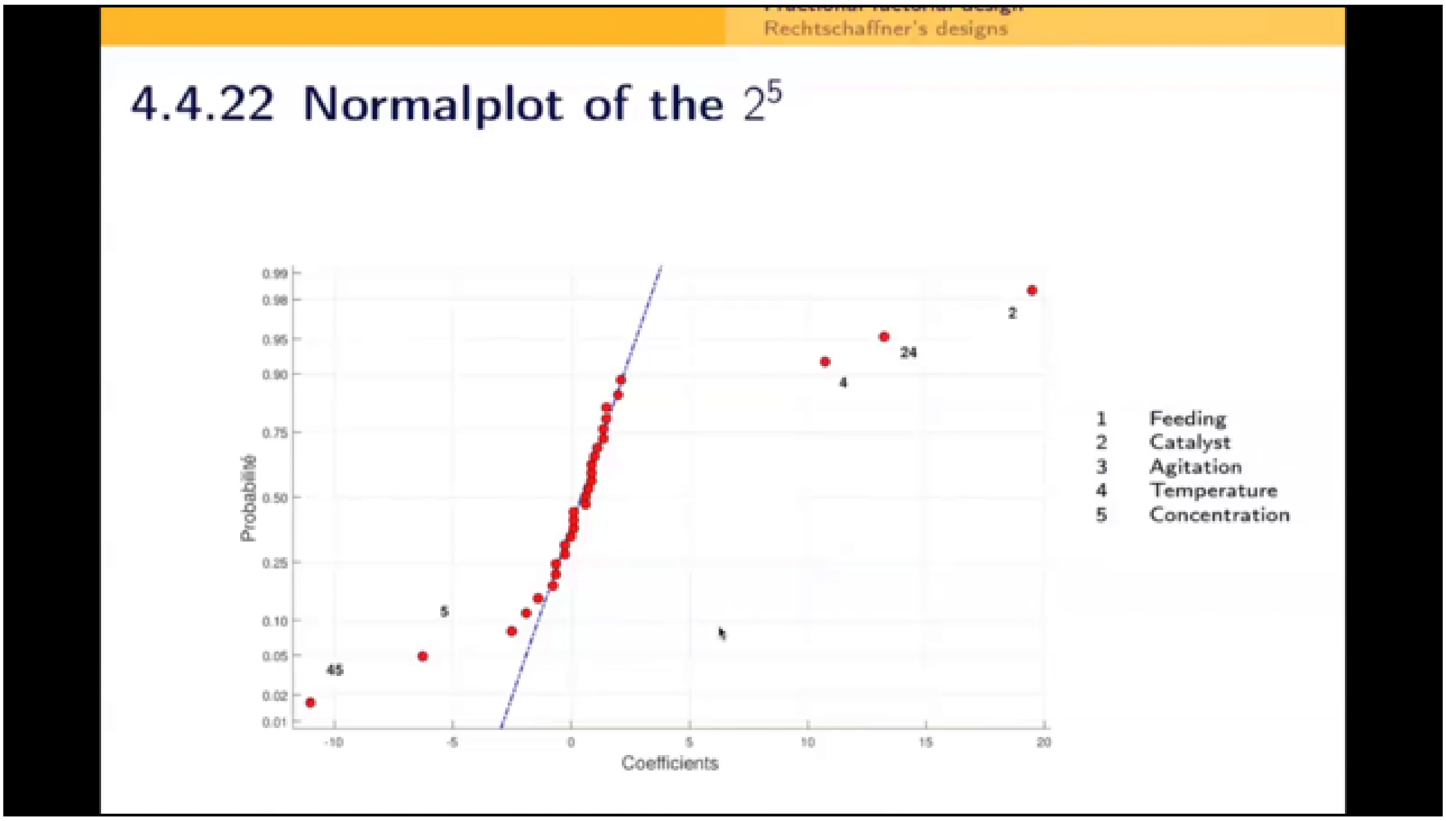


you have doing is is robust after I present you some design I don't want to go back through all the design I present you the um uh placate and brumann design but something that could be quite to be remembered is this future of fractional factorial design and these slides coming from box explaining what you can do and prof take profit of these slides just for explaining what you are able to do depending on the numbers of factors and depending on the numbers of experiments that you are allowed to to perform them the money is the time that you have and also step by step going from low resolution to higher resolution for discovering step by step what what is what what

notes

summary

14m 21s

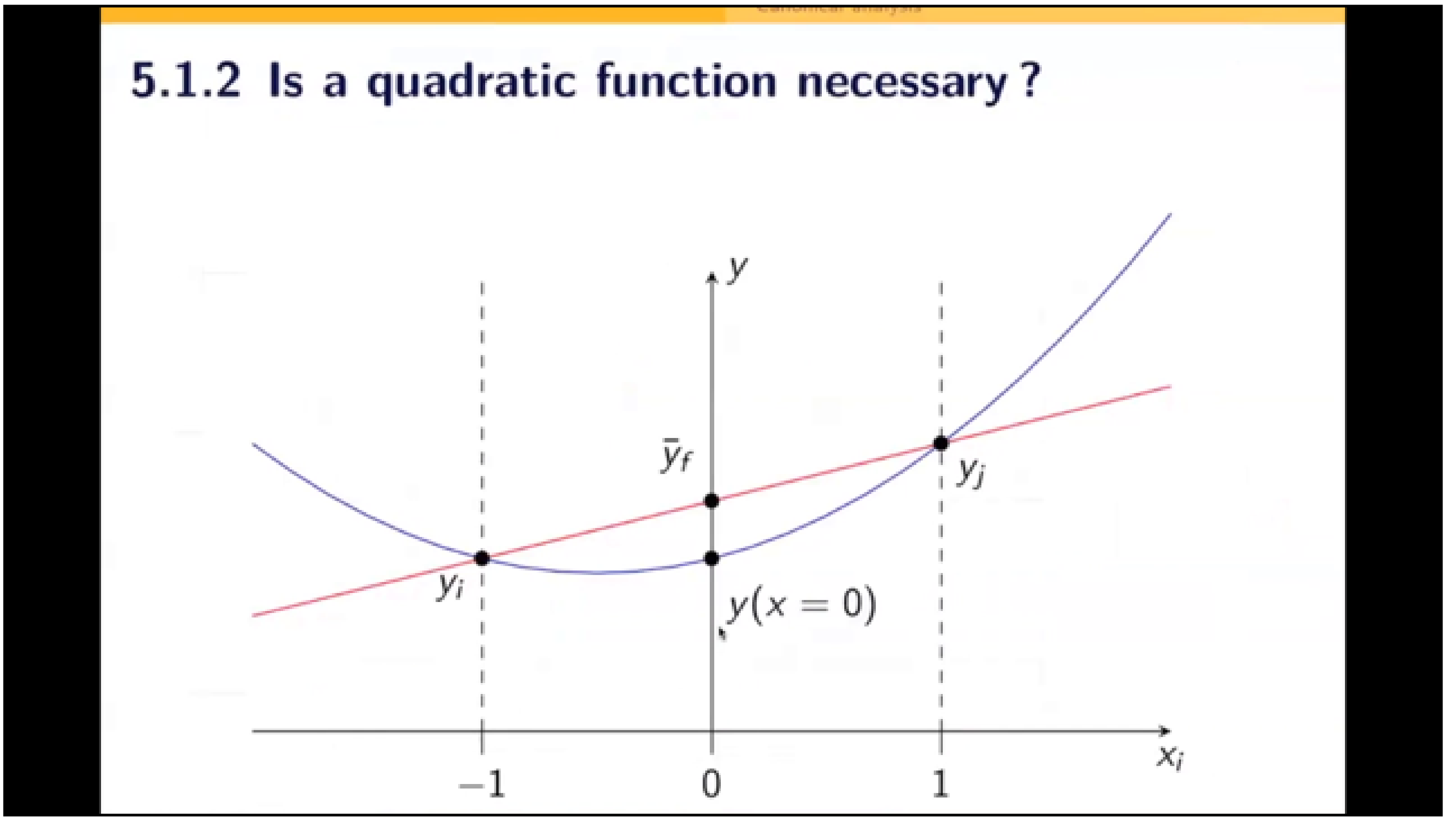


is important I show you a few elements the normal plot could be very interesting in two situation words for checking that you have a lot of of factors but not all are effective effective having an effect so not all your factor effective and I an interesting result is the difference between these two slides so one was obtained by the fractional factorial design the other was obtained with a full factorial design so you see that in both situation you obtain the same result at the end in one with a lot more information and your noise which perhaps not what is most important so this is also the idea of using fraction of what you do going step by step try to don't draw a very big design from the start and that you need three three years or ten years for finishing you better go step by step learn things step by step

notes

summary

15m 6s

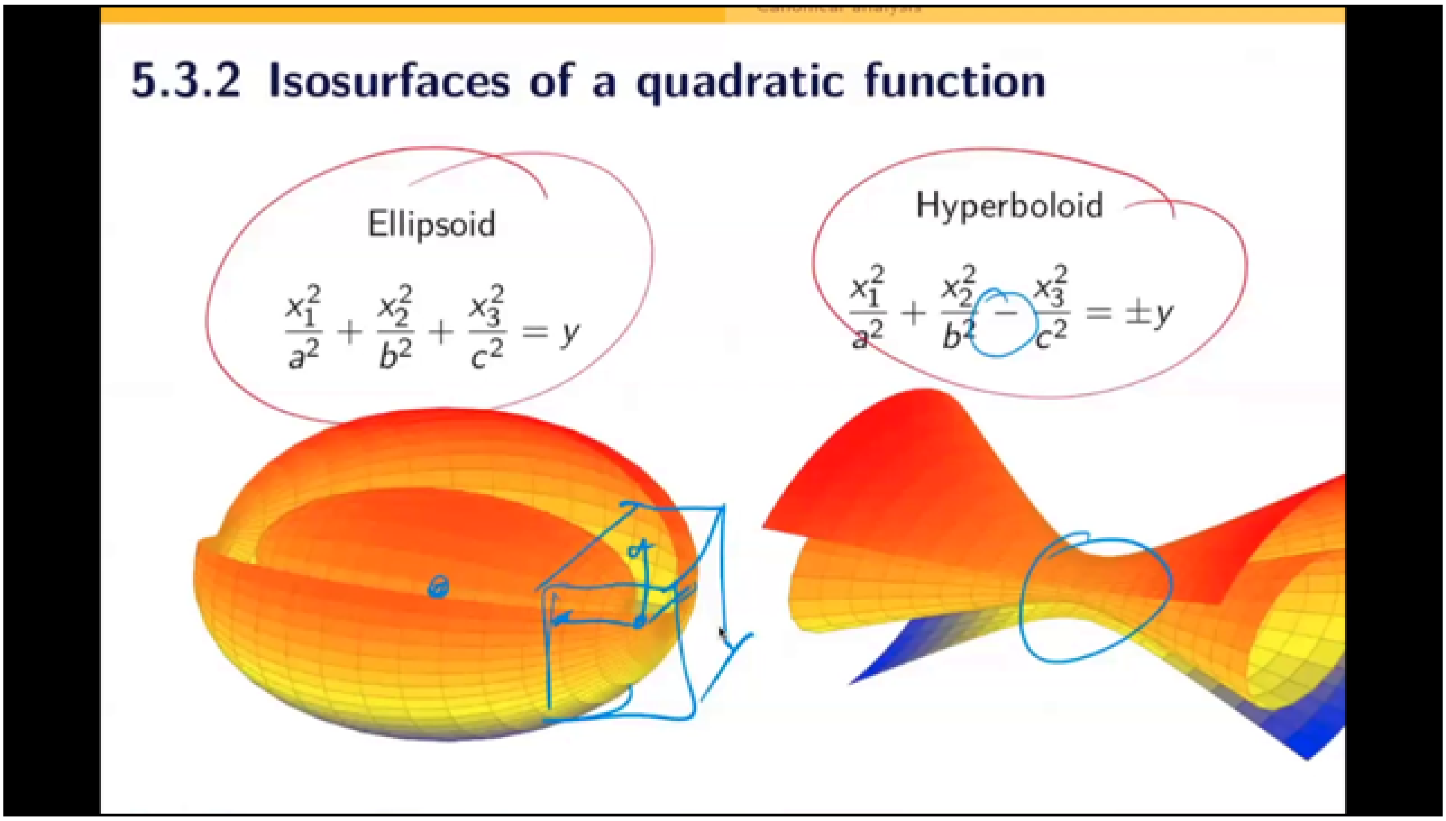


probably you will be you will finish with more interesting information unevenly less work perhaps you will not economize work sometimes yes sometimes not but at least

notes

summary

16m 13s

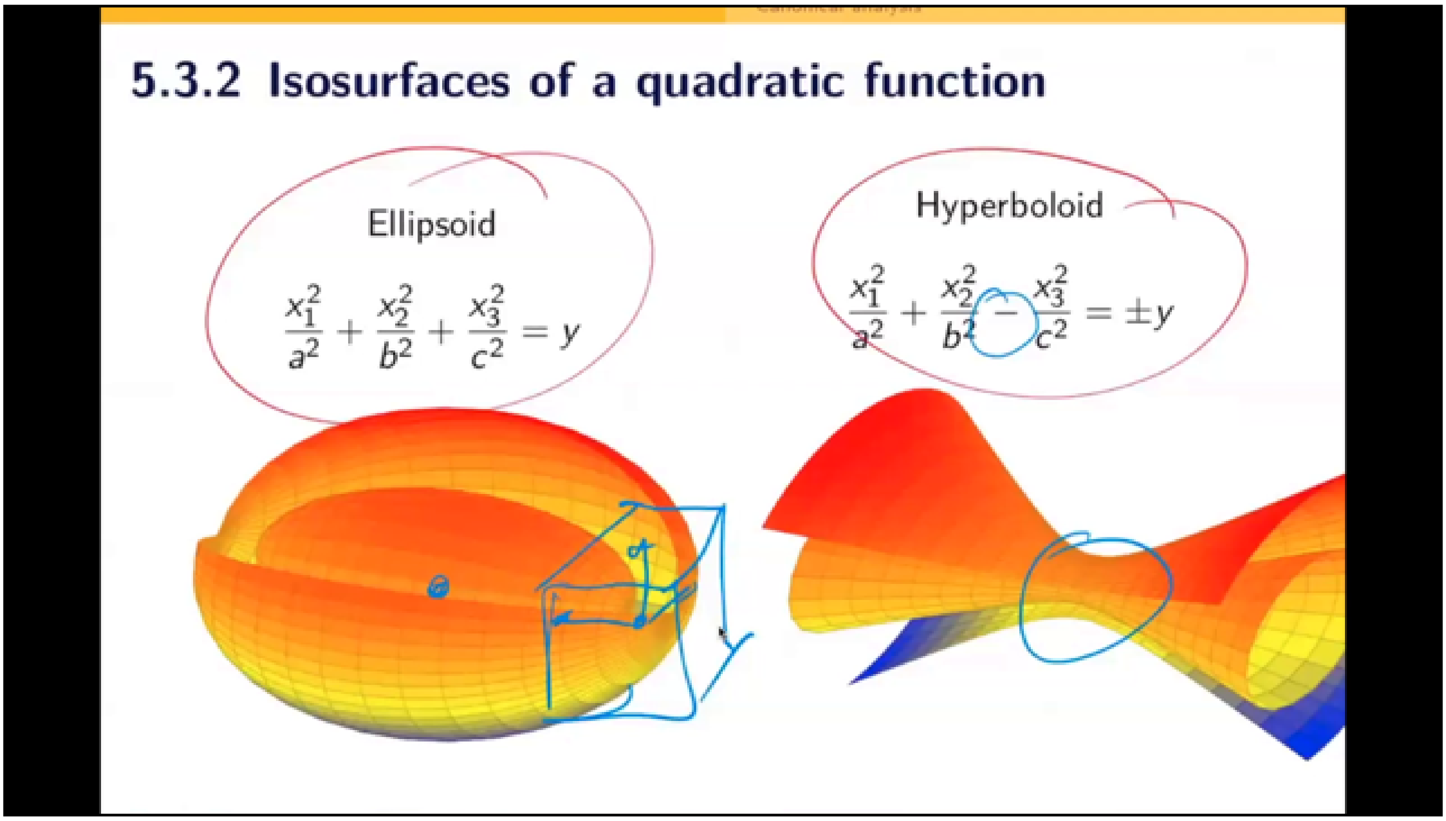


you learn things step by step after we'll see this concept of lack of fit usually try to work with the situation with what we call porcimony makes the model as simple as possible not great again but as simple as possible and so this is a concept of lack of fit and I've show you different situation in which we can calculate it the algorithm of fit if they can usually calculate them but you have to remember for having lack of fit you need to think you you need more data points at the minimum if you want to check if something is just a straight line or eventually a curve you need at least more points than just making a straight line and you need repetition because you would like to separate two types of errors that are not from scratch separated in ANOVA are the lack of fits or the fact that your model is not adequate and the problem of repetition of your measurements that you have some noise some random noise so the lack of fit is for separating these two source of error after we have seen something I call canonical analysis for the second degree probably this view of going first the first degree and after secondary is a bit strange for physicists usually we make the model as as complicated as possible from the start and after we simplify we have tried to see you the interest of going from the simplest and only going to the most complicated so linear but sometime we arrive to quadratic situation and the quadratic situation could appear very complicated in your coefficient a lot of coefficient but you can simplify it looking it as quadratic functions and quadratic functions of two times very simple ellipsoid or hyperboloid or you have your evolution which is the same when you are not same speed but

notes

summary

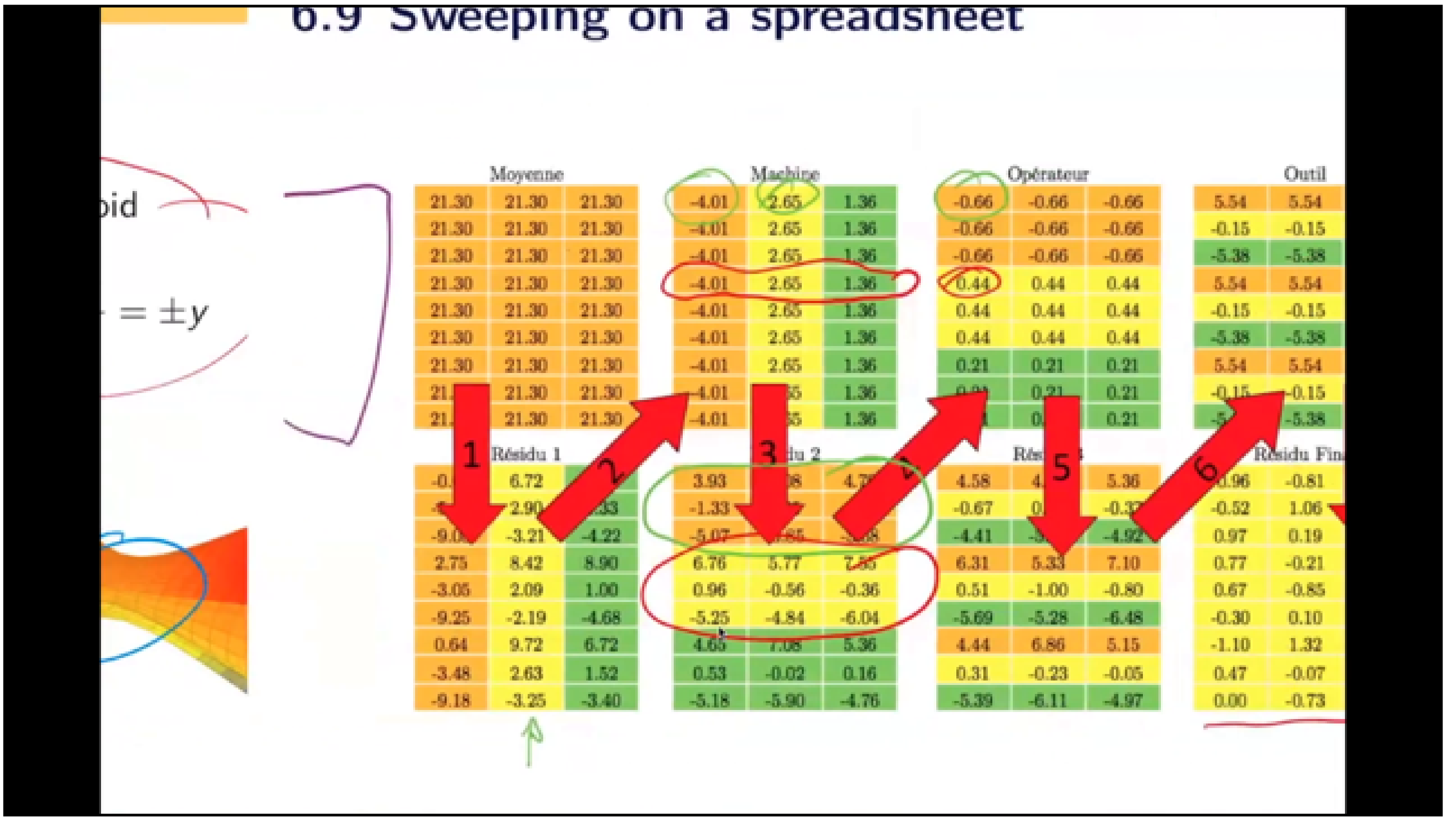
16m 25s



the same evolution when you are going outside of a center is your center the center of your domain or not it's another question what is the orientation of this main direction what we call the eigenvectors but

notes

summary

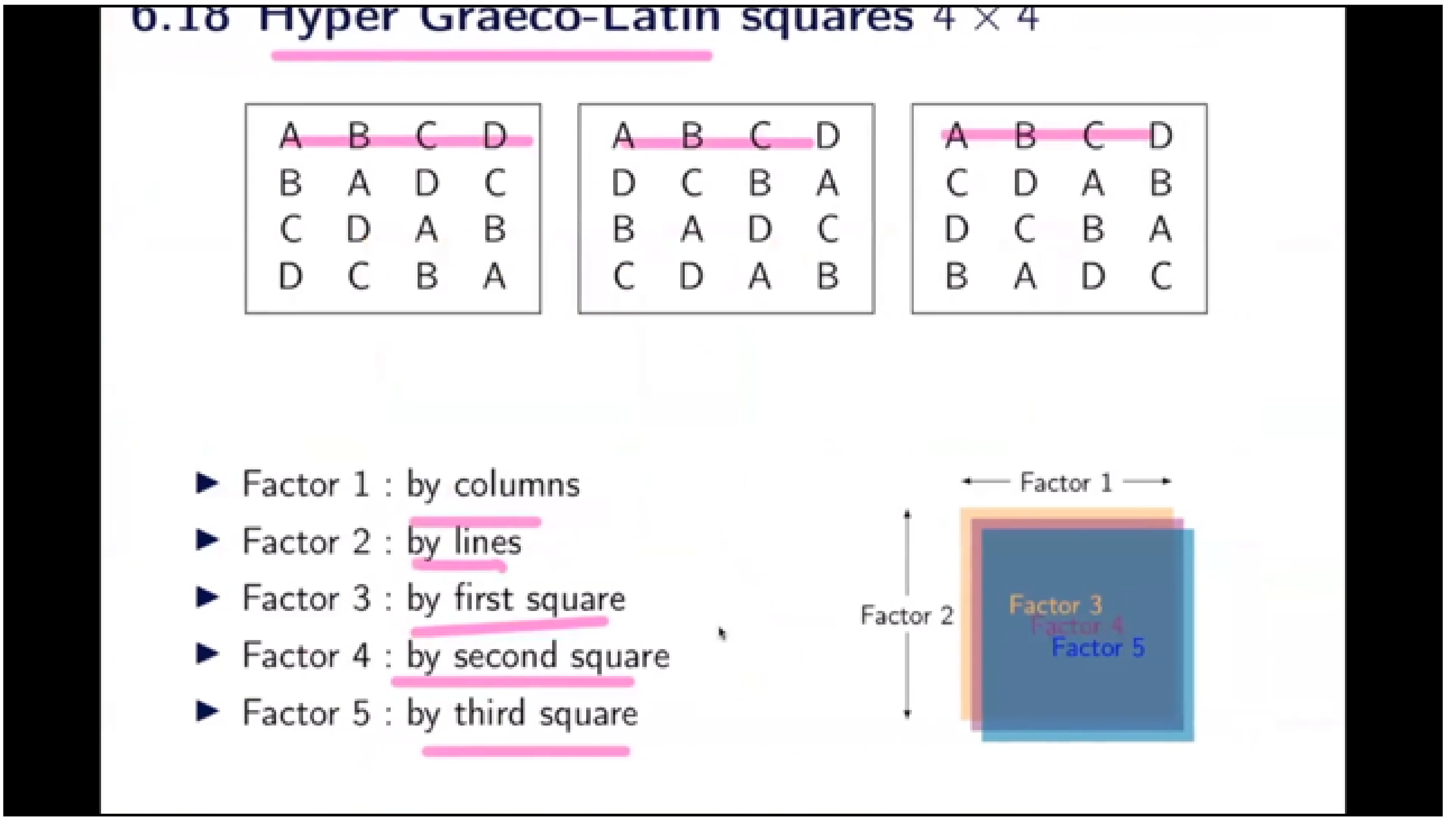


we have an evolution which is the same when we are going outside of our center or we have different dimension and some dimension where you have your property increasing in some direction when you have your property decreasing and you have different type of situation

notes

summary

18m 49s



after i show you the sweeping it was an interesting way of analyzing data remember this image that it's a statue you can observe the junction you can observe the muscles but it's not working it's not a robot it's just a picture of something and you can guess a few elements but it's not a mechanistic model it's really a statistical model of things interesting for summarizing analyzing data probably in in many situations not sufficient for making projection and what could be good or bad to do and to to improve and understand how it works so it lets you guess how it works but it doesn't let you explain how it works

notes

summary

19m 4s

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)

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

- The t-statistics is then

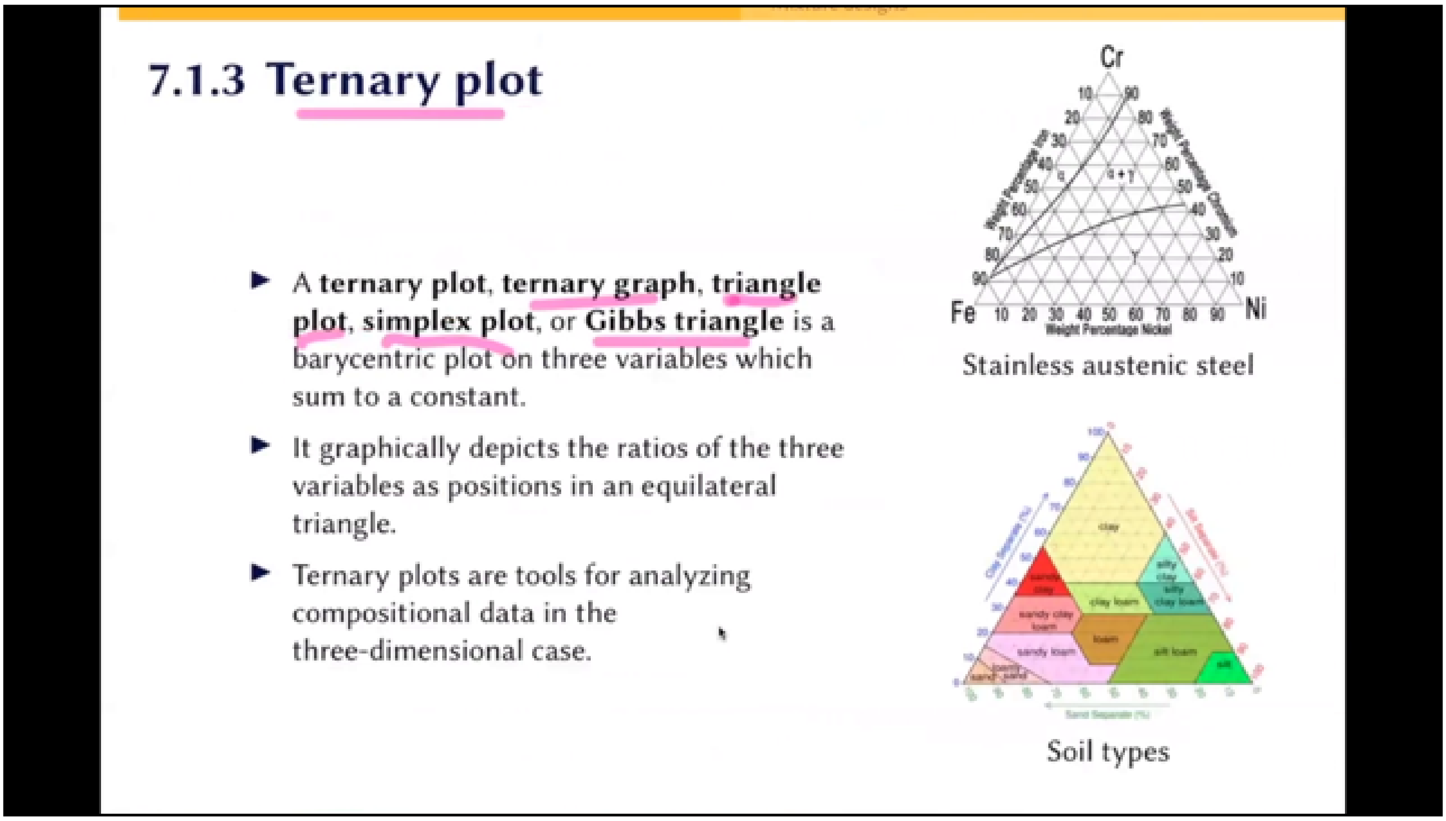
$$t_o = \frac{\sum_{i=1}^{a} c_i \, \bar{y}_{i.}}{\sqrt{\frac{MS_E}{n} \sum_{i=1}^{a} c_i^2}} \quad (15)$$

and i show you a few optimization and the Greco-Latin square was an example of an optimization of this type of situation but remember i hope nothing my my sentence there are no free mail in statistics in mass so each time you you win in the direction you are losing in another just choose to win what's interesting for you to win and just to lose in direction so you are where you are not so much interested to to win something so if your noise is not so well developed perhaps in the first approximation it could be a very good decision

notes

summary

19m 53s

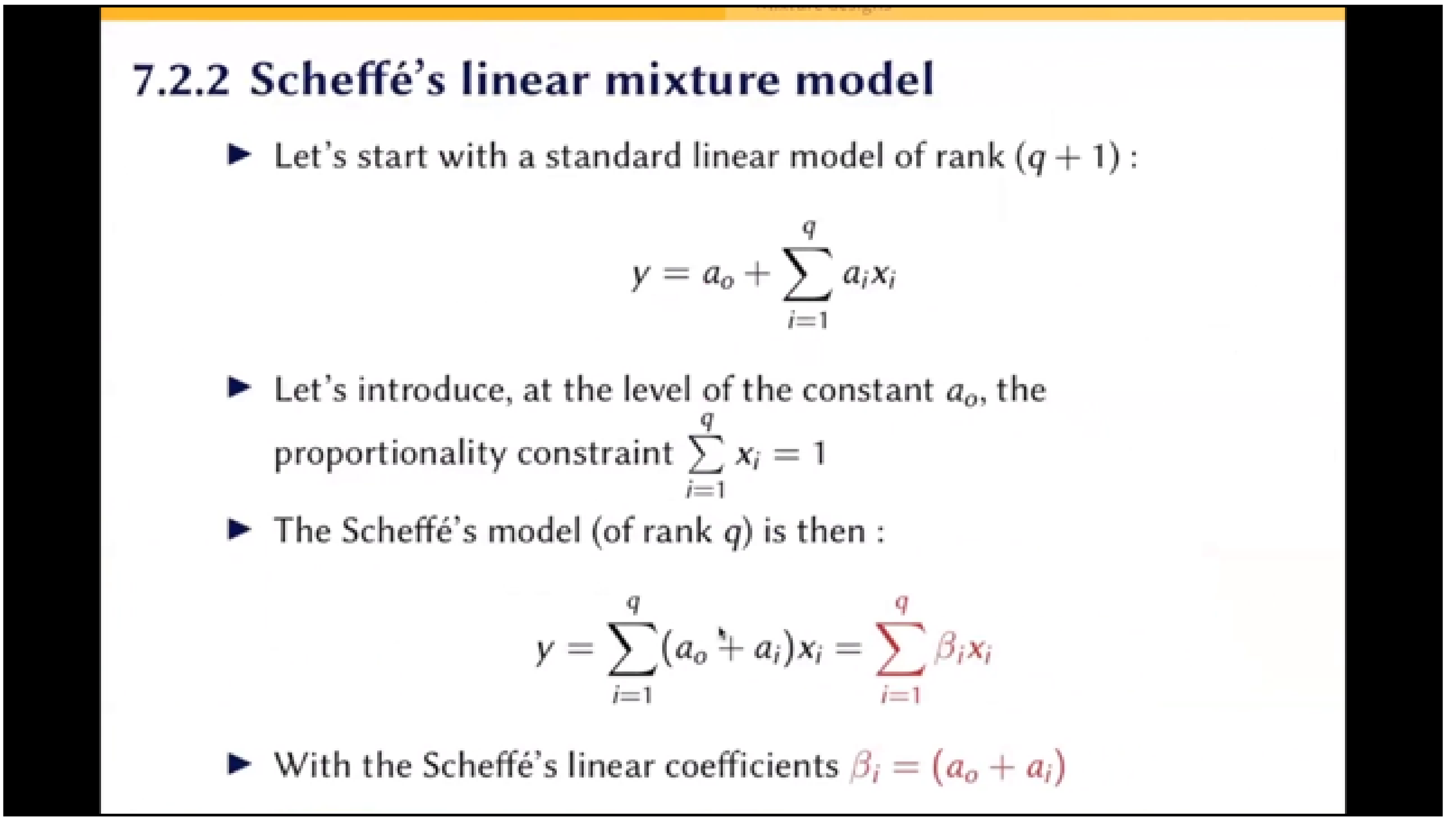


after i also explain you with this type of thing you cannot just make a nanova a lot of people just make a nanova of effect you need to differentiate the different effect and so the concept of contrast is also very important in this and you you need to go a

notes

summary

20m 34s



little bit further now you have the capacity to be to go a little bit further ternary plots ways of presenting mixture and conscious that this mixture chapter it's a little bit missing a few a few elements so it was just for giving you an idea of how to advance in that and in the future year i will probably develop this this chapter a little bit a little bit more and today i present to you this model the sheffet model it's also a very interesting concept for working with mixture so this was a summary of what i wanted to explain you

notes

summary

20m 58s