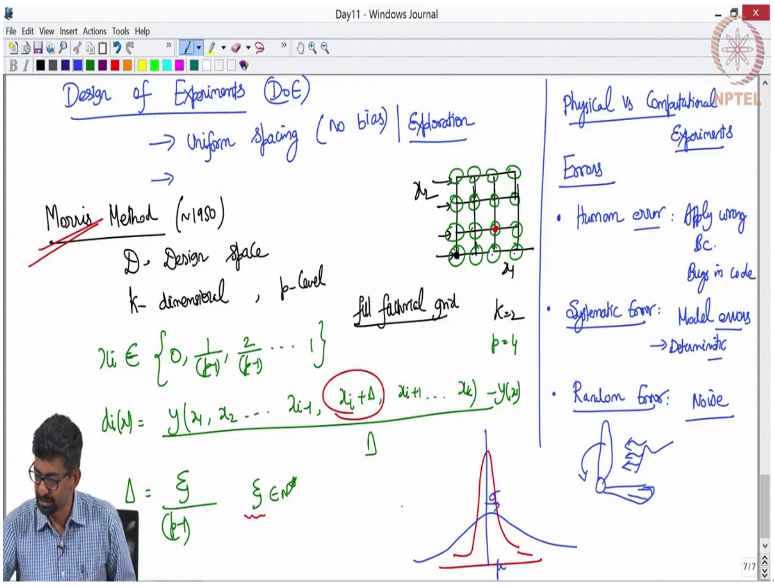
Surrogates and Approximations in Engineering Design
Prof. Palaniappan Ramu
Department of Engineering Design
Indian Institute of Technology, Madras

Lecture – 07
Introduction to DOE – 2

(Refer Slide Time: 00:22)



You recall something that we discussed yesterday in the design of experiments? Nothing? Simply we talked about a uniform property, ok. I want a grid where I want to sample it uniformly, uniform spacing. The good thing about this guy is no bias. I am not making any assumption. I want a uniform property, ok. I want to kind of explore and exploit ok, but we are only discussing about exploration. Exploitation comes later. Similar to oil rigs, first they will explore across the sea and when they find that there is oil at that point, then only they will exploit, ok.

What you do is you first survey your literature. When you find that there is a gap that is when you exploit that one, ok. So, basically for exploration purpose unless you have any information, you are only going to explore. Then, what did we discuss in design of experiments in general; that is it? Sorry.

Student: (Refer Time: 01:47).

Those are just names that we suggested, right. So, we are going to start looking at some statistically designed experiments. We will come to that, but before that I just wanted to tell you something this just to make sure, that is just the terminology physical versus computational experiments, right. So, there are basically from the perspective of errors I am looking at this. Only from the point of errors, what are the different types of errors that can come in, ok?

So, one is the human error. If you look at it, the human error usually introduces only a bias per person intra, ok. So, in a computer sense because let us say that there is a technician who is trying to do a test. He always over tightens the screw. You over tighten the screw and usually it is always because that is what is tight for him whereas, you use another person he kind of does it to the correct extent. So, result of it. The results are going to vary a little bit ok, but this person always over tightens it. So, there is only a bias, there is no variability in the results. The results are always over or underestimated, whereas in a computer experiment, you cannot do anything like that. There is no tightening. It is kind of pretty much automatic right, but there is some kind of an error which is like you apply a wrong boundary condition that could be a human error. It is still not a model error; it is a human error. The boundary condition is a human error per say and of course, your regular bugs in the code. So, these are all only bias. The other type of error is systematic error. These are called the model errors, ok. These are usually deterministic in nature. There is no systematic error in experimental if you look at it usually.

So, whenever we say systematic error or model uncertainties is the word that people are using these days. It is usually deterministic and it is what happens is you have a particular grid, you run the simulation, many times also you will get the same result only for the same loading unless you are going to go and change something else inside. So, if you want to induce randomness, you have to induce randomness in one of your variables otherwise the computer is going to give you the same result every time you do it, but that is not the case in the physical experiment. You take a sample for instance today we call virtual testing. Whatever you are doing in UTM, can also be done in the computer, but in UTM the physical component that you put could have material defects unless you model that in the computer. That guy is going to keep giving you the same stress strain curve. It will always give you only the same stress curve, but in this you will get slightly different meaning like you will not get a curve here and you will get a curve here. No get the variation that I showed you exceed existed will have a band around that, but you will not get the same curve, ok. So, that is called the systematic error, but this is more to do with your computational ok and then, there is a random error. This terminology is important, which you will realize later. This is usually called the noise.

For instance, the same cantilever beam that we spoke about yesterday in a physical sense, you might get different results ok, but in the computer sense if you give one load, you will get the same value every time. In order to induce a randomness what you do is, you model the load to be random ok, but in reality that is called noise. For instance, you design a wind turbine and then, there is a anyway. So, there is a wind turbine and then, there are winds that are coming in this direction and as a result of it, this guy is going to rotate, ok.

So, of course I have a motor connected to this assuming that there is rpm, the wind speed is going to influence my rpm, but do I have control on my rpm meaning do I have control on the wind speed. No because the wind speed is going to vary. Today it will be 8 meters per second because it was windy and tomorrow it could be dry. It could mean only 4 meters per second. I do not have control of that. So, that is called noise. I cannot say this wind turbine will work only for 8 minutes per second. It has to work at least for a range. I understand that you cannot put this somewhere in a place where there is heavy wind. It might not survive, but at least here I cannot do it for only one wind speed, ok. So, that is called noise and you have to account for this noise while you are trying to do. In computer simulation, this noise has to be simulated and it has to be modeled, and there if you have model errors, then it will propagate and the model errors also will propagate.

So, you need to first understand what a model error, what a random error is, ok. You cannot combine all of them to be error and you just say there is an error, but that there error can be decomposed into multiple things. There is problem with your data then you are going to fit the data. So, there could be a fitting error. So, you need to know how much of what is coming from which error ok. So, in order to clarify that what I brought this, ok.

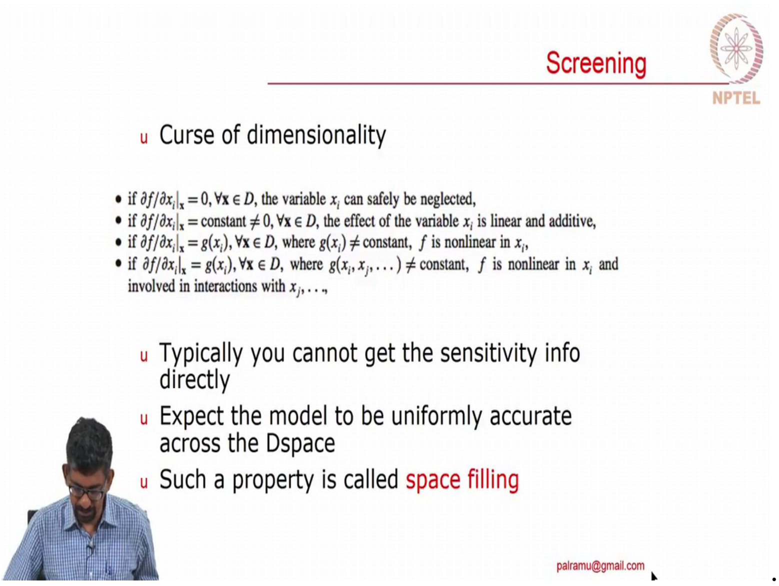
Student: Computer experiments inherently does not have random error

Computer experiments inherently does not have random errors provided you are talking about Analysis models. Analysis models how many other times you do, you will get the same result unless there is a random component into it. Let us say that you do a Monte Carlo into that and obviously, you will get different results.

Student: Getting different result (Refer Time: 08:58).

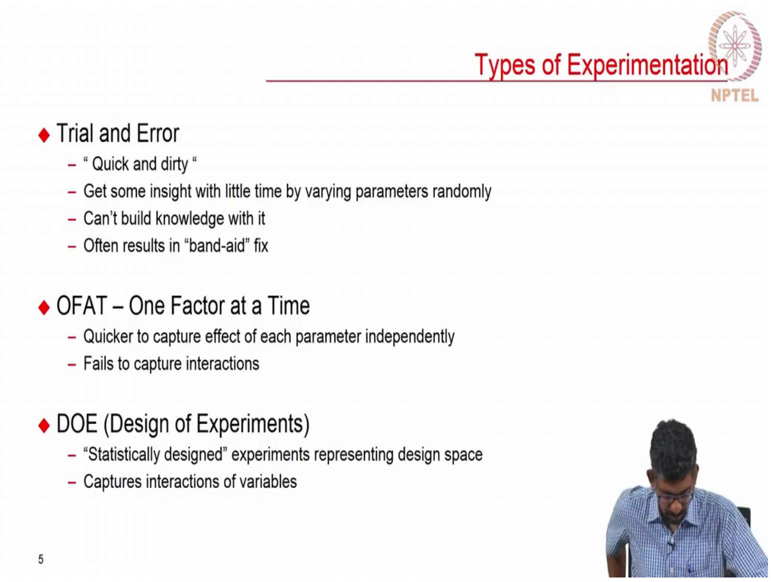
That is what I am saying. What I am trying to tell is let us say that your force in your analysis model you consider your force to be random, then obviously your output will be random. So, that is the randomness, not random error per se. What I am saying is like you always try to compare it with one is gold standard which will be your recall, your cantilever beam or your question is correct. Modeling randomness is not equivalent to random error. So, that is why I said you know like the systematic error is deterministic which is associated to model error, but that is not randomness. So, that is all.

(Refer Slide Time: 09:38)



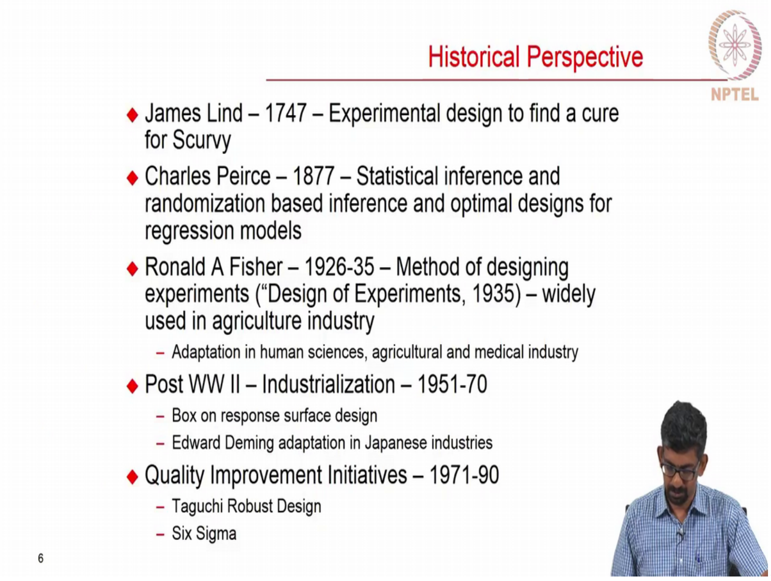
So, yesterday we were somewhere here. We talked about the curse of dimensionality, ok.

(Refer Slide Time: 09:45)



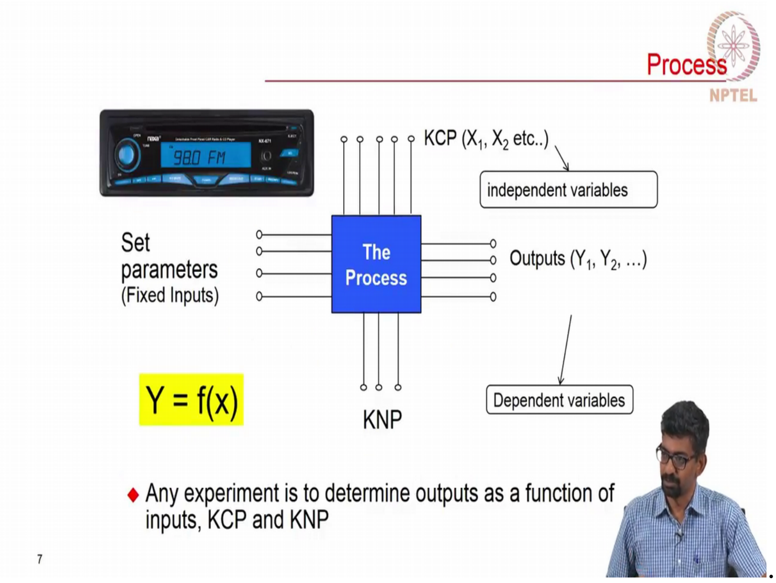
Then, we talked about one factor at a time trial and error approaches. They are only bandit stuff. There is no scientific stuff behind it and design of experiments is what we will focus in this particular stuff.

(Refer Slide Time: 09:57)



We gave you a very quick historical perspective.

(Refer Slide Time: 10:01)

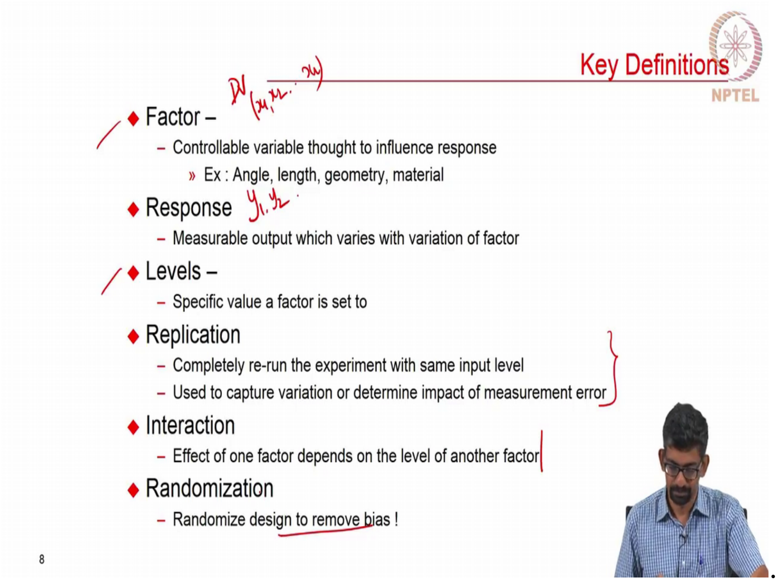


We talked about the P diagram or the where KNP, what does KNP stand for?

Student: Noise parameters.

Noise parameters, now you understand what the noise parameters are. The noise parameters are something that you do not have control on, but you have to design with that ok. So, Xs are your design variables, those are your independent variables and your Y are your responses which are your dependent variable. You are eventually interested in relating Y equals f of X.

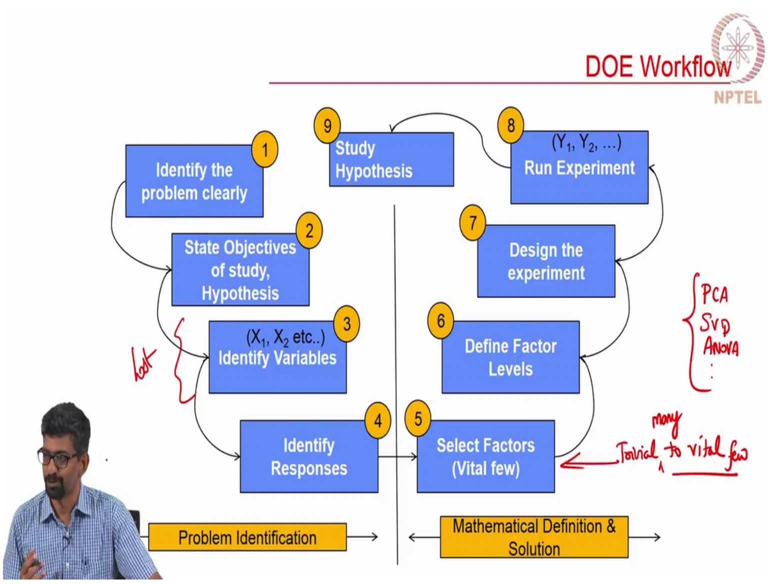
(Refer Slide Time: 10:33)



We did quickly look at these things yesterday this factor and levels are nothing, but your design variable. The response is nothing, but your Y. I mean whatever we saw y 1 y 2 this is nothing, but your x 1 x 2 x n, ok.

So, the levels are nothing, but what are the different values that your x i can take. This is the reason I spoke about the errors, ok. For instance, replication in a computer model how many ever times you run, you will get the same value ok, but if you go to the experiment and you do the same tensile testing, you might get slightly different stress strain curves. That is the reason, we also have talked for interaction that today we will see a little bit in detail and it is usually advised to do randomness because in reality your situations are random and you want to test your prototype or whatever it is or even your designs for random. So, randomize your designs to remove any kind of bias.

(Refer Slide Time: 11:38)



We very quickly looked at or taking user inputs from there. The important point that you might want to remember here is this is lot of stuff ok. What we call here is trivial to few vital few they call it trivial, many to vital few.

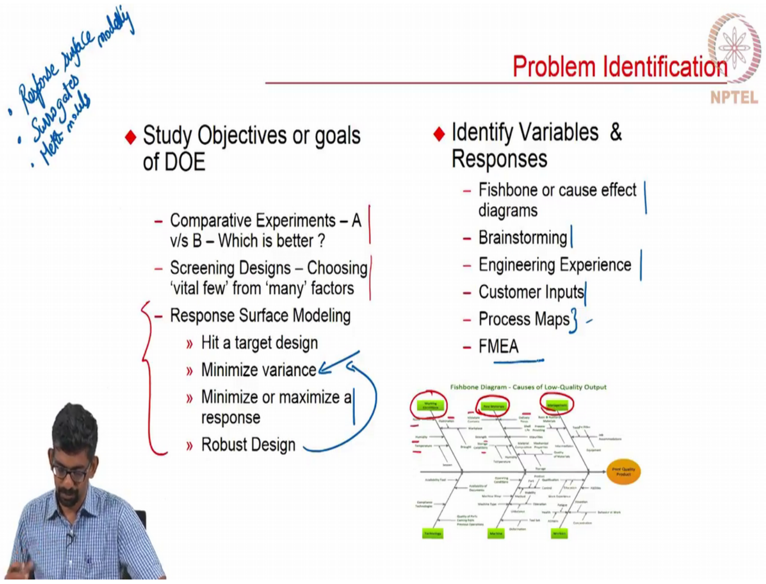
So, basically this is called a factor analysis. Do any of you know of any factor analysis approaches? It is a math technique principle component analysis singular value decomposition no problem ANOVA. ANOVA you should have heard mechanical engineers analysis of?

Student: Variance

Variance ok. So, like that there are multiple techniques, ok. So, nothing as the name suggests it gives you the principal components. It tells you which factors are more important compared to the other factors. Of course, you will not get 100 percent representation, but you will get about 95 to 97 98 percent representation and you will have to describe what is the level of intensity or to what extent you want to have the variations explained, ok. So, you can use any of these methods as well.

So, they are just saying the workflow this is how it goes. So, as you can see the left side of this graph I mean the graph in the sensor picture, it is called the problem identification, ok. Oftentimes in academia we only do the 2nd part of it which is the mathematical definition and solution, ok. So, this is usually done in the industry set up most likely in the concept design stage, ok. You do not know what are the factors, you do not know you need to find out or identify the factors that needs to be tuned. Then, they come to us and say we have about 35 parameters. How do you reduce the number of parameters there are?

(Refer Slide Time: 13:37)



So, this is one way of looking at it. I am not going to get into the details. How many of you know anything on the fishbone diagram. You know a fish of course you know how it bones looked like most likely like this ok. So, what they say is this they kind of take different I do not know whether this is clear, ok. I will just read out for instance this is the working conditions, they call this raw material, they call this management. Like that they have multiple barks in which there are small branches for instance working condition; what they are saying is noise is one case, illumination is one case, humidity temperature is one case. Just to give you an idea in raw material, they are saying moisture content, delivery times strength storage conditions like that under that in illumination, they are saying brought in temperature, they are saying it is season dependent. So, you can keep on building this fishbone diagram, ok.

Sometimes this is also used in terms of FMEA ok, but that is at a slightly later point. Failure Mode Effects Analysis you need a different similar kind of a graph, where you will say what are all the different components and inside the components what are all the sub components, what can go wrong if the sub component goes wrong, what will happen what is the type of failure. So, that is called failure modes. Each one is a failure mode, right. So, failure mode and the corresponding f x so, these kind of stuff will be helpful for you to build such kind of diagrams, ok. So, this is a fishbone diagram kind of stuff, but your overall idea at this point in time is to you know what is the goals of your design of experiment is it to compare in a cardinal sense or in a ordinal sense. What I mean is cardinal means 21.2 and 23 which is bigger you know, or should I just tell you A is better than B ok. Those are the kind of stuff.

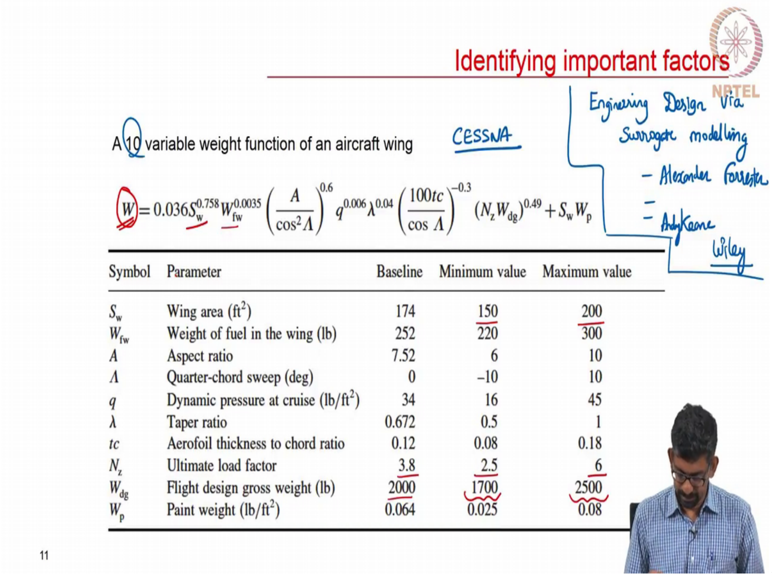
So, comparative experiments A versus B which is better, then there is also screening designs which let you choose the vital few from the many factors, ok. This is where you probably use your PCA kind of stuff. This just tells you like this gives you a tiring the top 6 components they belong, but they do not give you a cardinal value, ok. Then, this is something that people are focusing on because it kind of gives you both a bit of both doubles. It is called the respond surface modeling are the Meta models are the surrogate. So, please note this throughout the course I might just use the word response surface modeling or response surface, surrogates, metamodels in an exchange fashion meaning if I say any of this, it means the same, ok. It does not have, it has nothing to do with the modeling technique that is used, ok. I could call it response surface modeling. I could call it surrogate or I could call it Meta models; they all mean the same.

Now, if you want to minimize variance of your response usually, that is called your robust design. I want to do that. I want to minimize or maximize my response. We saw examples yesterday, ok. Robust design means this guy there is also a reliability based design. Then, I want to hit a target design for instance it could be cost or it could be a specific design curvature or something, ok. For any of these things which is very specific response, surface modeling is a good way to go about, ok.

So, identifying the variables and responses and then, identifying the vital guys from the trivial many, ok. So, you can use a fishbone diagram, you can brainstorm, you can bring in legacy talking with people. It is also called expert opinion, engineering experience, legacy of the company, customer inputs. You talk to the customer and understand what needs to be done, process maps FMEA as I pointed out. So, there are different ways of looking at it under FMEA you can look at fast, there are different algorithms. That is not the scheme, that is not the focus of the course, ok.

So, basically there is a bunch of variables and then, you get a subset of those variables to run your optimization.

(Refer Slide Time: 18:09)



This is a just to give you an idea of how do you go about doing such a thing, ok. So, this one book that I am referring for this entire Metamodels section, design of experiments in Metamodels section says you know.

Student: (Refer Time: 18:28).

Now, what is the title of the book?

Student: (Refer Time: 18:33).

Engineering Design or Optimization.

Student: (Refer Time: 18:47).

There is a book. It is a Wiley publication by Alexander Forrester. I forgot who is the 2nd guy is. The 3rd guy is Andy Keane. It is a Wiley publication. It is a very nice book. I am not sure there is a one edition in 2010. I think there is another edition in 2014, ok. This is a book that you might want to use. So, I have just lifted this example directly from the book. So, this is actually representative of CESSNA vehicle in aircraft, CESSNA aircraft. So, it is a 10 variable problem and it is a weight function that we are talking about, ok.

So, what are the 10 variables are given here. For instance, this S w means wing area, then F w, W fw is weight of the fuel in the wing, and this particular expression is taken out of another paper or a textbook. There they have this explicit equation what we are going to assume here is you do not have this equation. You pretend that you do not have this equation, but we are going to generate the data out of this equation. If this equation did not exist, but you had the data let us say how do you make conclusions, ok.

So, what they are assuming in this particular case, these are the baseline values ok. For the wing feet 174 is the baseline meaning the nominal value. The minimum value could be 150 or maximum value could be 200 like that they are giving. What is the ultimate load factor? The minimum value should be 2.5 and the maximum value should be 6 whereas, your baseline value is 3.8. This is not necessarily uniform if you look at it 2000, 1700 and 2500 that tells you, ok. So, 2000 maximum is 2500. So, it is only 2500 off on this side, where 1700 is only 300 off on this side. So, there is a minimum and maximum value and there is a nominal value or the baseline value which means this is what my current designer is.

Now, the question is there are so many variables here. So, if I want to do full factorial experiment which is we will discuss that in detail in the next. So, we want to understand which of these guys effect the most, the weight. So, of course you can bring in expert opinion, you can bring the legacy ok, you can have common sense. You can say oh for sure the wing area is going to affect the weight ok, but they have listed all these factors after a brainstorming session let us say and you want to understand not just is it going to be, but you also want to rate by how much or which one has more effect, ok. One way of doing that is what they call the contour plot, ok.

Before doing this maybe I will just I will just switch. We just need to introduce a method because there is something called the Morris method. So, I guess somewhere in 1950s or 70s I am not sure, came up with this idea on sampling basically how to understand, but you will see how this is related, ok. So, the point does assume that there is D is your design space and then, it is K dimensional and then, in each dimension I have p levels.

So, for whatever problems that we discussed yesterday, it was two dimensional; x 1 and x 2. So, k in that case would be two and p level. So, what I am saying, I am going to have 1 2 3 4 levels, ok. So, this is one level, this is one level, this is one level, this is one level.

The kind of it gets over here 1 2. So, this is one level, this is one level, this is one level, this is one level. So, there are 5 levels in this, whereas 1 2 3 4 levels in this, ok. So, let us get rid of the top level. So, 1 2 3 so, it is just for the sake of ease of explanation.

So, then this is called a full factorial grid. It is only a grid you have not taken points, but provided you take points here like let us say let us call a full factorial design, ok. So, it is called a full factorial grid and in this particular case what I am going to do is I am going to say x i is 0. This is nothing, but the grid here, 1 over p minus 1 2 over p minus 1. It runs to n over r 1 basically ok.

So, what it means is if this unit was 1, then this would be 25th, this would be 50th. Sorry 25th 50th 75th 100. That is what it would mean, meaning 0 1 over 4 2 over 4, no no sorry 1 1 over 3 2 over 3 3 over 3 which is 1, ok. That is how it goes because the levels I would have as three levels in this particular case.

Student: 4 level for p minus 1.

P minus 1, sorry I have 4 levels and it is p minus 1. So, p is 4 and that is what we took here, right p equal to 4, sorry. So, 1 over 3 2 over 3 and then, 3 over 3 will be 1. So, that is how you do it, and then, d i of x. So, for a given baseline value let us say that d i of x is y of x 1 x 2. You keep going like this x i minus 1 x i plus delta. This is an important point. This is how I get that graph minus y of x, divided by delta. What I trying to do is this and yeah the delta is a small perturbation. Let us say it is basically you can say some it is in the numerical space that is all. Do not worry about that equation. What I am trying to do is, this I am taking a particular value here let us say the red one and I am perturbing it by delta. That is what this guy means it is easy. This is for generic purpose. I have just written this equation, but if you relate to this particular case, it is only x 1 x 2.

So, you can just for discussion you can have x 3 also. Let us say then you are x 2 plus delta ok, but remember that you are going to do this for each case. So, I am going to move this guy delta in each dimension. That is what this i refers to, because it is k equal to 2 and p equal to 4. At each point I will have two part of this guy a little bit to understand what the variability is and what the delta is. It is just a small epsilon divided by the corresponding p minus 1 that you are talking about that is all. So, if you do not worry about this definition, you are just taking this point and you are perturbing this a little bit in this space. If the point was here just perturbing the point a little bit to understand what your responses, but please understand you will have to run a simulation or experiment to do that and that is where the whole catches ok.

So, this was proposed by this guy called Morris, under what assumption he said that I want to understand the elementary effects. Elementary effects means is x 1 affecting y more or x 5 is affecting more or x n is affecting y more. He said this is called the elementary effects, not the interactive effect, ok. Interactions came, but he said there are two outcomes that I can have. One is it should have an elementary effect and if that is not having, then it means it has an interactive effects, but he also takes a probabilistic perspective to that what he says is if I assume that the effects come out of a probabilistic distribution, because this is going to be random, right.

So, the effects are going to follow a probabilistic distribution and if a particular variable has a very high sensitivity meaning y has very good sensitivity for a particular variable, then the mean of that distribution will be large for that particular variable whereas, if the variance is large which means your distribution is flat, then it means that it plays a role in the interaction. If x 1 has what we call like heavy effect on y, then the mean when I perturb it, the mean of that particular x 1 should be large in case if the variance is large meaning it should take a high value. You understand what I am talking about, right. So, this is a distribution that I am talking about. It is a normal distribution. First that is the idea here and then, this is the mean value mu and this is a standard deviation.

So, what is this is the mean value is more, then it is the elementary effect. If the standard deviation is going to be large, it means that the equation is one of become I mean the graph is going to become like this, ok. That means it has a role in the interaction ethics, ok. That is all he says,. So, large spread means the variable is involved in interactions. If there is a large central tendency which means your mu, then it means that it has the basic are the elementary effects, the single factor interaction and the single factory importance.