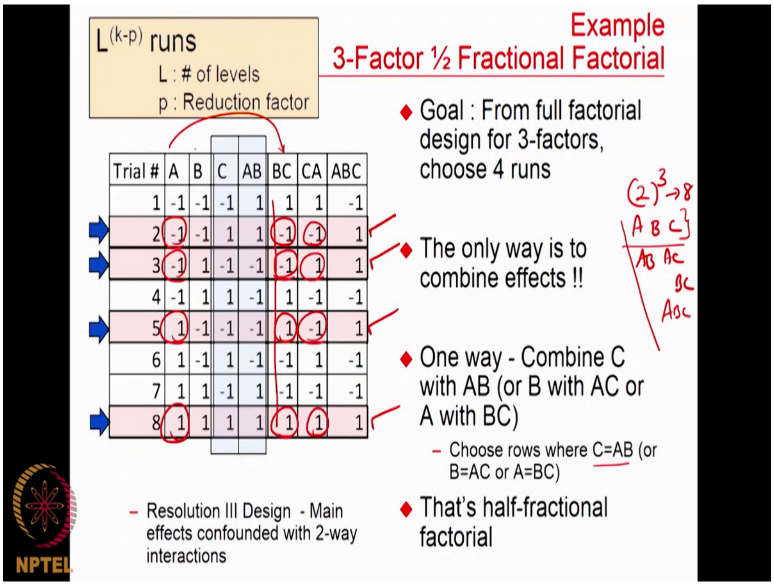
Design For Quality, Manufacturing And Assembly
Prof. Pataniappan Ramu
Department of Engineering Design
Indian Institute of Technology, Madras

Lecture - 16
Confounding Logic & Randomization of Experiments

(Refer Slide Time: 00:16)



One thing that I wanted to discuss is; how do you confound this is we saw in terms of factors right, but what is actually happening in L8, sorry in the orthogonal array, when you go from L16 to L8, how can you recreate? Can you recreate L8 from L16? What is the physics behind you know practically, not physics, but what is the idea behind it? How do you confound? For an instance here is an example of 3 half fractional factorial, right. So, if it is 3 factors, then you are talking about L8. 2 levels means 2 raise to 3 is L8.

How do I get L4? Simple example here you take trial 1 to 8. This is an existing stuff, right. So, I have given, it is not as per what our discussion is. I am just giving you a random assignment here A B C AB BC CA ABC, ok. It is a random assignment and it should be AB A cross B, ok. Now, what I am going to do is this. When I am saying that I am moving from L8 to L4, what am I doing in terms of the factors? Because if it is 2 raise to 3, there are 8 meaning 2 levels of 3 factors. So, if I have A B C, then I have AB AC and BC and of course, I have ABC. So, if I am going to take all these guys into account, it is L8, but I am going to do only L4 which means that I am going to get rid of all these guys. I can only do this guy, correct.

So, what is happening is, I have to confound all these guys. One way that you can do that is you need to combine. For instance, you take A B and C. This is just an example. In this case, you can take anything that you want. What is the meaning is, I will have to build these guys into these guys. See practically what is happening is, you are confounding and you are reducing the size, ok. What you are doing is take this C and AB for now. What I am going to see is this. Where do they share 1 1 they share this particular. So, I marked it here minus 1. Minus 1 meaning the level is the same. In this particular low, the level of 1 and the level of C and the level of AB is the same. In this particular row, the level of C and the level of AB is the same.

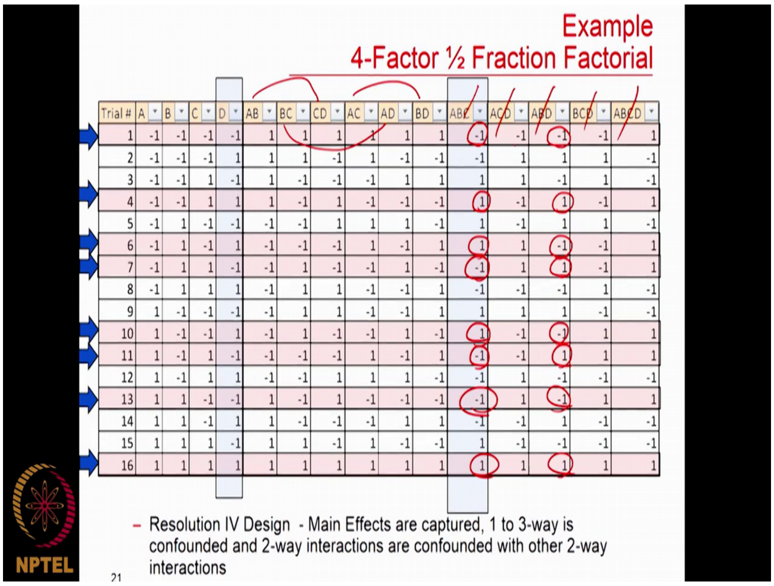
So, I am taking those things. So, you are combining C with AB or B with AC. So, you can do whatever you want. You can do B with, B with AC or you can do A with BC because you are practically trying to get ABC. That is what you are trying to get. So, choose rows where C is equal to AB. That is what we did, ok. So, we are choosing this row, we are choosing this row, we are choosing this row, and we are choosing this row, then you are going to choose only those rows because that AB does not change in that row with respect to C. So, you are just taking that. Similarly, you can take for this guy and it will be the same. If you see let me just draw something for me myself.

So, this cannot be there minus 1 minus 1. So, this one will be there minus 1 minus 1, this one will be there 1 1, this one will be there 1 1, this one will be there. You understand what I am saying minus 1 minus 1 minus 1 minus 1 1 1. The remaining four stuff will be compliment minus 1 1 minus 1 1 1 minus 1 1 minus 1. This cannot happen unless you make sure that you are the balance. You understand that is the reason for using orthogonal array. You can get a subset of L8; you can get it as L4.

So, will it hold good with respect to C A and B? Also, if you see minus 1 minus 1 1 1 minus 1 minus 1 1 1, it is the same 4 rows that will reflect the same stuff. So, what I am going to do is, this I will only take the ones that are marked as the pink one and I will say minus 1 minus 1 1 minus 1 1 minus 1 1 minus 1 minus 1 1 1 1. That will be my L4. I will take only these guys, these 3, this guy and this one. That is all. It will give me my L4.

You get the point. You will not get 7 columns because now you are merging all the columns into the other columns 3. It is only 3.

(Refer Slide Time: 05:45)



Trial #	A	B	C	D	AB	BC	CD	AC	AD	BD	ABC	ACD	ABD	BCD	ABCD
1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1
2	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1	1	-1
3	-1	-1	1	-1	1	-1	-1	-1	1	1	1	1	-1	1	-1
4	-1	-1	1	1	1	-1	1	-1	-1	-1	1	-1	1	-1	1
5	-1	1	-1	-1	-1	-1	1	1	1	-1	1	-1	1	1	-1
6	-1	1	-1	1	-1	-1	-1	1	-1	1	1	1	-1	-1	1
7	-1	1	1	-1	-1	1	-1	-1	1	-1	-1	1	1	-1	1
8	-1	1	1	1	-1	1	1	-1	-1	1	-1	-1	-1	1	-1
9	1	-1	-1	-1	-1	1	1	-1	-1	1	1	1	1	-1	-1
10	1	-1	-1	1	-1	1	-1	-1	1	-1	1	-1	-1	1	1
11	1	-1	1	-1	-1	-1	-1	1	-1	1	-1	-1	1	1	1
12	1	-1	1	1	-1	-1	1	1	1	-1	-1	1	-1	-1	-1
13	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	-1	1	1
14	1	1	-1	1	1	-1	-1	-1	1	1	-1	-1	1	-1	-1
15	1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

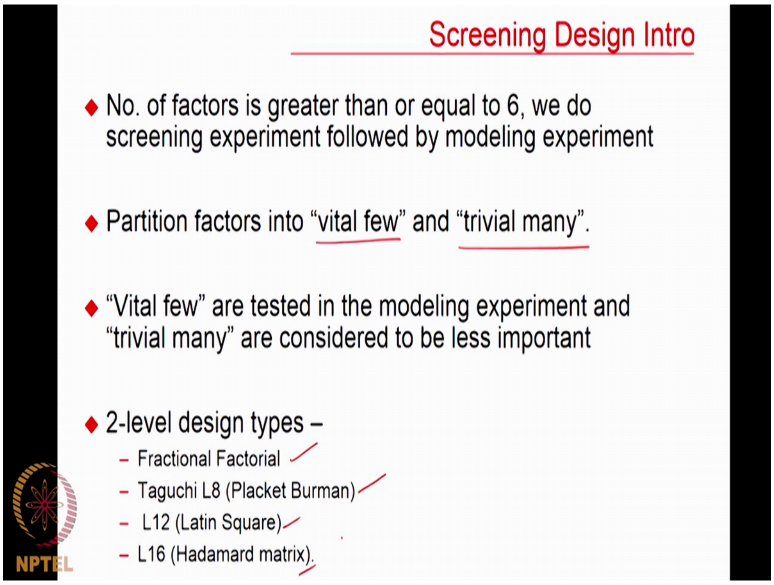
So, your result, same can happen for other stuff also. So, if you look into this, this is larger like L16. If you want to bring it down to L8, what they are doing is they are taking D and they will factor it with ABC, because you want to first delete ABCD. So, I will take minus 1 minus 1, then I will take 1 minus 1 minus 1 1 1 1 1 minus 1 minus 1 1 1 anyway. So, if you see the pink is already marked only to say that, and then, you take this guy. So, these things you are collapsing column wise and you are collapsing row wise also.

So, finally it will boil down to be A. You just need to do only this. The remaining is automatically taken care of. Let us see for C, for instance minus 1 minus 1 ABD, right. Minus 1 1 cannot go, 1 minus 1 cannot go, 1 1 minus 1 1 cannot go, minus 1 minus 1 1 1 1 minus 1 minus 1 1 minus 1 minus 1 1 1 1 minus 1 minus 1 minus 1, minus 1 minus 1. I should say minus 1 1 1 minus 1 1 1, it will automatically be there. You just have to do only two guys and the remaining will automatically fall in place. You could have done, we could have taken C and ABD also, ok.

Now, what will happen, this guy, this guy, this guy, this guy, this guy is gone. So, then I can only 1 2 3 4 5 6 7, I am bringing it down to 8 columns, right. So, resolution 4 main effects are captured one. So, this is 4 factors half-fractional factorial experiment, ok.

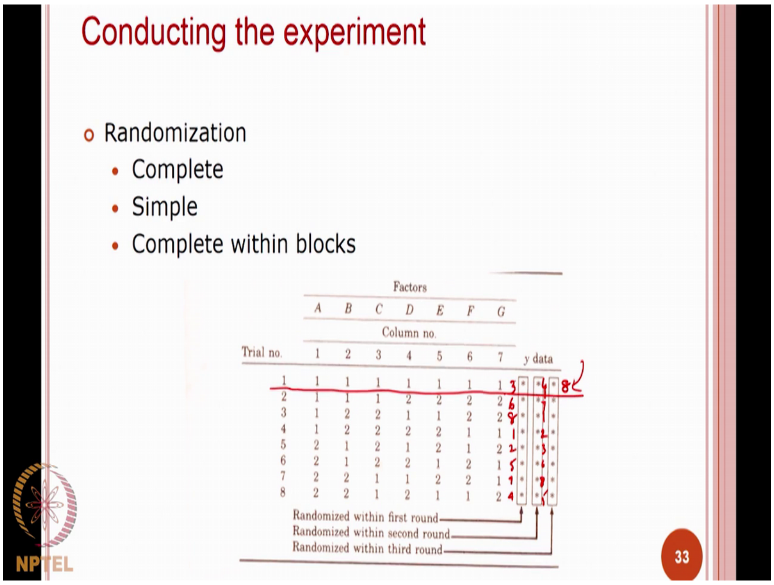
Then, you might also want to A B C D, ok. This is a 4 factor experiment that we are doing and then, you will have to do confounding to bring the sky down back to help because if you just take only these four, it will give you a larger 1 2 3 4 5 6 7 8 9 10, but you need to bring it down to 7 because it is an L8 array and then, you will have to do confounding to bring those guys. So, the confounding will be between, AB CD BC AD CD. The other one that is left you understand what I am saying. We did now this for 4 factors; I mean the 4 factors interaction. The next one you might want to do is this between ABCD, then BCAD, then ACBD. If you do that level of confounding, then it will bring it down to three more levels down. That is all. So, it will be 7 factors then.

(Refer Slide Time: 08:54)



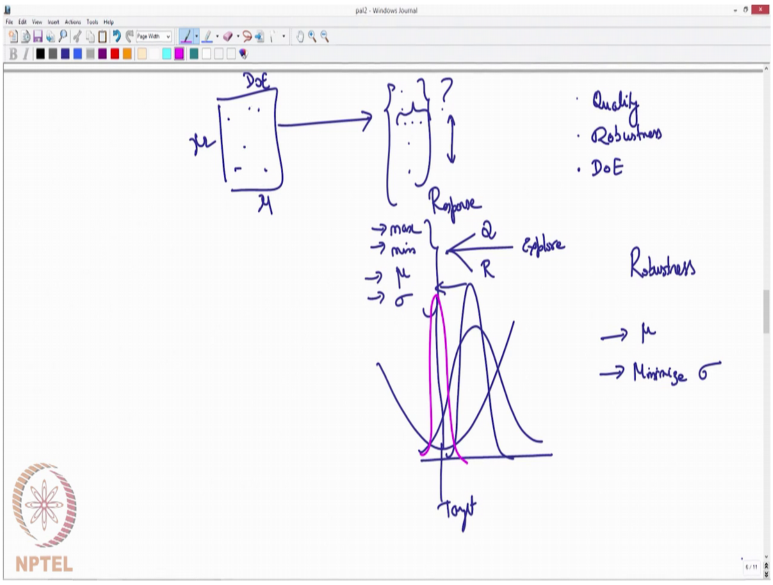
So, the type of screening experiments that people usually do is, this is primarily, it comes from the industry requirements. If the number of factors is greater than 6, what they usually do is, they do a screening experiment first. You do not want to do 2 raised to 6 and stuff like that. Even fractional factorial they take the vital few from the trivial meaning this is something that we have discussed, vital few are tested in the modeling experiment and not to be less important there are different types of designs. You can take a fraction factorial, you can take Taguchi, there is this Latin hypercube is something that we will talk about, and there is also Hadamard matrix that you can talk.

(Refer Slide Time: 09:34)



The next one is about conducting the experiments itself. Let me just give a brief.

(Refer Slide Time: 09:45)



So, where did we start? It goes back. So, there were 3 things that we wanted to do. One is define this is the stuff that we wanted to do, right. So, we wanted to understand quality from a definition perspective and we need to be able to measure quality metric, your quality loss function. The other one, the second one is robustness. You need to understand the definition of robustness which we understood as minimizing the variation and you should be able to quantify robustness. How? Using a signal to noise ratio, then we wanted to understand the relationship between quality and robustness which is also we have done.

Now, from there we did this guy. We did this, but in order to get these data and all that we wanted to do something called the Design of Experiments because you need to have a bunch of data whatever. You call this is as your response; I should have some entries here. How do I get these entries? I need to conduct some experiments somewhere and that will feed into these responses. This is my variable x 1 and x 2. Under these combinations, I do the experiments. And, then only I will get this because otherwise I will have to infinite experiments here to get these values.

So, I need to do. So, this is what your DOE is and that is something that we checked here. You cannot do all the combinations, but we will be able to do a fraction of the combinations. This is what we have done, but please understand what we are trying to do here in this case is, yes you can get a max, you can get a min, you can get an average, you can get a standard deviation of all this and then, you might want to do a quality or a robustness study or you might just want to explore your design space. That is one thing.

The other one that you also need to appreciate that I pointed out in the quality, the tensile strength. Example is there could be variation within an experimental set and that also you need to quantify. It is clear that there will be variation between the experimental setup for obvious reasons, but within an experimental setup also there is variation. That is why car engine number 1 and car engine number 7 are different though they are coming out from the same. After the class is over, all of you are going to step out. The instructor was the same, instructor was not different for Rakshith. It was not different for Vamshid, the instructor is the same. You think the information that you got from me is the same when you get out of this room.

The information that the guy, that guy number 1 and the guy number 7 is different. You all went through the same process. I am the same guy, it is the same lecture that I am giving, it is the same recording that is going on, but what you have picked from me is different. There are several factors that had affected that in a similar fashion, but you are all humans and there is a complex dynamics that relates me and you, but in a robotic setup also, it might not be that kind of a variation, but still there are smaller variations which is influenced by different factors. When you eliminate humans out of the loop and you put the machine into the system, it brings down the variation, but still there will be variations.

So, there could be an intravariation which you need to account for. So, it is not enough. You do a sigma here, but you also need to find a sigma here and often that becomes the interest because I want to find the best SN ratio along this because it is not enough. Please understand what I did, what have I told you about robustness, what have we discussed about robustness minimize what?

Student: Variation in the remains, the effect of cost without actually controlling.

The problem is you are still in class 1 after we have run 2 classes in between your answer is right, but it comes from class 1. Minimize sigma; you are minimizing the variance variation that is correct, ok. What he is told is also correct. You are kind of minimizing the effect without touching the cost itself. That is correct, but mathematically as an engineer when you are looking at it, I am kind of trying to minimize the variance. Is it all?

Student: We should find the maximize.

I need to take care of my mean. It is not always maximized unless you make it a cost function, ok. You need to deal with your mu. It is a two stage process that is something that you need to appreciate. You remember this graph that I do for you. You make your process strict here. You are still losing out because your target is here. There is a mean shift. So, you will have to push your model here. Unless your model is here, you are not gaining what you should gain. So, there is a mean shift that you need to worry about and then, you have to worry about your variation. There are two things.

So, that is something that we will talk about, ok. This is the basic funda that I wanted to give you, so that we can start talking about this. The whole idea if you recall that we brought this quality robustness, all that is because of variability, it is because there is randomness in most of the things that we try to do which is sometimes beyond your control, sometimes it is under your control, sometimes it is beyond our control. That is why we classified the uncertainties into two as aleatory and epistemic uncertainty. Aleatoric is something that you know is inherent variability. You cannot do anything about it; you have to live with that.

Epistemic uncertainty you build a better model and then, you might be able to take advantage of that. So, in a similar fashion we want to understand or rather characterize the uncertainties which are an output of a random process. So, I should be very clear about the fact that I do not introduce bias when I am trying to recreate the random nature, otherwise if there is no randomness, you do not get this correct. You will get only one dot either across the column, sorry either across the column or across the row. So, there is randomness, there is variability, hence you are getting different values for the same experimental setup point 1.

Point 2 is you are creating DOE, you are running the simulation, you are running the experiment and then, you are doing this. What does it mean is, you are trying to recreate the variability. I am giving you the car number 1, I am giving him car number 7, and then I am monitoring, but I cannot build the bad product and give it to you and monitor and understand what the design changes I will be able to do in the design stage. Again I cannot build and give it to my employees and follow them for 2 years to understand what their characteristics are. So, what happens is, I will have to simulate this situation.

So, I have to simulate because computers will do you what you ask it to do. So, you say 2 plus 2 is 4. You program it that way, it will always do whether you operate whether I operate; it will say 2 plus 2 equal to 4. Now, I have to introduce, I have to simulate the variability. When you do, it will give 2 plus 2 is 4.2 When I do, it will give 2 plus 2 is 3.8. When he does, this is just an example 2 plus 2 is only 4. What I am saying is when you are putting process A plus B, let us say that it should give you 4, but there is variability in process A and B. So, when you do that, when I am saying when you do that, it could be just me, but when I do the 1st time, it will give 4, the 2nd time it might give 3.8, the 3rd time it might give 4.1, the 5th time it might give 4 and that 22nd time it might give 4.8, ok.

So, I have to simulate the variability now. So, I have to bring in some kind of a random parameter into my formulation and then, simulate this variability. That is what we are trying to do in general in the experiments. So, how can you do this randomization? There is something called Complete Randomization. So, simple randomization is incomplete within the block. So, we will see what it is and very important from an experiment perspective is you should not bias certain things; when you trying to do an experiment, you should not bias certain things.

You remember the example of our basketball players. You want to measure the height of the students within IIT. So, you just walked, get out of the class and then, you walk across and you find a basketball court and then, you went and measured all the players, then you said sir, I measured about 40 people which is a good representation and good sample. Not good representation; it is a good sample, but can you go with that data as IIT students height you cannot because it is likely that tall people play. There is a conditional sense that only tall people play basketball. So, it is a biased sample.

To begin with, in a similar sense, you do not want to bias your experiments. The other day we discussed what might be an ideal way to go about it. So, you choose different years, different departments. You select representatives from each of that and then, you measure people across the cross-section in a similar fashion. When you are trying to do the experiments also, you should not, you should make sure that you do not bias experiments, ok. We will see what that bias is.

Complete randomization is meaning let us say that you are doing L8 experiment, whatever let it be the example. There are 8 experiments, but you are also replicating each of these experiments, you are repeating these 8 experiments. So, imagine that you are trying to characterize the variability of a material property of a particular material that you are interested in, let it be steel. So, what you do? You do a nice dog bone shape and then, you put in the tensile testing machine and then, you do this pull out test and it will give you a stress strain curve for a multiple things.

If you can do with that in one of the factors let us say is the specimen size or from which sigma manufacturers you got that specimen, it could be either of that. So, if I keep that specimen and or the variability of the person, so basically what I do is, I keep the setup the same and I only change the specimen. First I do the people from company A. So, this takes my entire 4. I did like 5 specimens come from company A, then I go for lunch, I come back and then, I do for company B in the afternoon, 5 specimens, then I go for tea, I come back and in the evening shift I am doing for 5 specimens from company C, fine. I asked you to repeat 5 times. You have repeated for the three company specimens. Is that fine or if you think there might be some kind of a bias in this.

Student: The temperature conditions are not same.

There is, yes. The temperature conditions need not be the same between morning and evening because you fix your AC at 24 degrees, but then during the day, the temperature went up. So, the inside temperature could be different and because of that it could be, ok. Still it is supposed to maintain 24, but in the morning between 9 and 11 you had only 2 people visit. You and they just open the door only twice, but between 12 and 3, there were 5 people came and someone kept the door open for minus minutes and was chatting with you. So, you lost enough temperature out that necessarily does affect the material property little bit. There could be factors like this.

So, now what is happening is if you take a particular batch and test it with the same situation, then you are biasing it. So, you do not want to create such a bias. This happens for an instance in sometimes you go for an interview, people are getting interviewed. So, in the morning batch usually the interviewers are also fresh, ok. So, they expect something, but over that over the course of the interview, we figure out that we have not matched the number of people that needs to be selected and we also kind of inbuilt now we understand over expectations are more or our expectations are less. So, we get adapted to the interviewee.

So, sometimes the way they get drilled at the end is not as much as they got drilled in the morning. This happens even in our department VIVA class and all that stuff. See that is why we try to keep it a small subset. So, that it is across the course. When you take the 1st year courses, I do not know whether all of you are in the same section or it could be different sections also. So, then what they do is, there is an instructor A might offer it in a different way, instructor B, instructor C. And then finally, you try to normalize it because there is variability within an investigator and there is variability between the investigators also, but I need to be fair. So, what we do is, finally we normalize the grades and then, we release it together.

In a similar fashion, the idea here is that I am trying to simulate the variability and I need to make sure that there is no bias. So, the complete randomization what it says is you randomize the experiments completely. Here is an example. L8 is repeating 3 times, so totally 24 experiments. So, what happens is you randomly run these 24 experiments. You do not say I will fix the experiment in the experimental setup 1 and run three specimens. You do not do that because then what happen you are biasing. That is not what you want to do. If you go back to the example that I gave my first experiment, what I will do is, I will take supplier 3, his 3rd specimen I will take and I will run it.

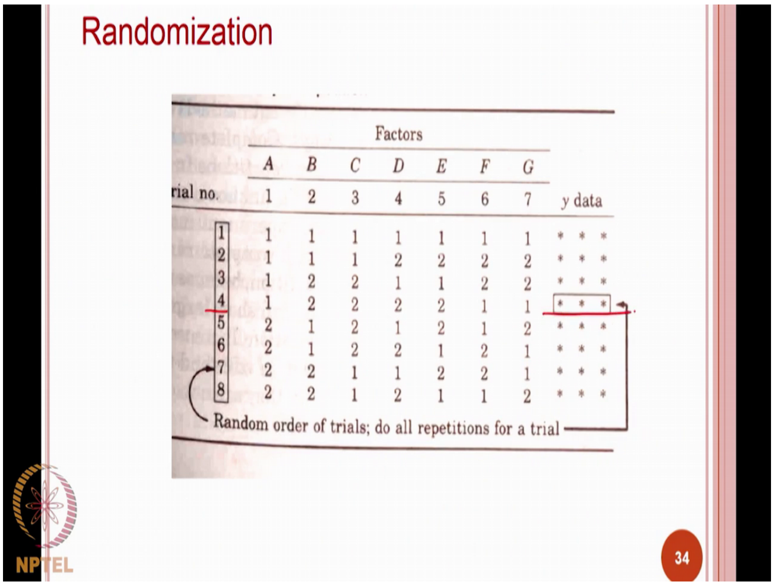
Then, the 2nd experiment that I would run is, I will go to supplier 2 and I will pick his 5th specimen and run it. That is complete randomization. That is exactly what is happening here. So, just one second. So, what they are saying is I am running between 1 and 8 in a random fashion, but then I run I make sure that I run at least one experimental setup, once in the first 8 times. That is what this is. Then, I go to the 2nd step and then, I run it in a random fashion. The 1st time I might have run, I am just writing something, ok. This is let us say I am writing here the times, in, not the times the order in which I did, ok.

I probably run this guy the 1st time, this guy the 3rd, 2nd time this guy, the 3rd time this guy, 4th this guy, the 5th, 6th, 7th and 8th. This is the order in which I conducted the experiment. The first one was experimental setup 4, the second time was experimental number 5 and then the third one was experimental setup number 1. This is random. I just write a random number. Do not make a copy of that. This is a random. If you ask me to write it again, I might not be able to reproduce it. This is a random class. In a similar fashion, I went and wrote this guy I did this guy also. So, this guy is 1 2 3 4 5 6 7 8. So, it is a random order.

So, what I am trying to tell is, let us say if it was a same type of experiment that we are talking about, this one was done in 4 noon, this batch was done in you know late afternoon, but there is variability between this and there is variability between this. Let us say this was done in 8 in the 3rd experiment. If I am doing 80th, this guy was the last experiment that I did for the day and you can see this guy was the 3rd and this guy was the 4th specimen that I did in the afternoon.

So, it is spread across for this is for experiment number 1 meaning experimental setup number 1 and I did in different times of the day. So, I am able to capture the variability without any bias. This is complete randomization, but this is not possible always because sometimes let us imagine that I need to control humidity for one of my experiments, ok. It is a large even our lab you can take, but let us say that you have a nice precision controlled situation. I want to control the humidity, but every time you change the humidity control, it is going to cost you some money and then, what do you want to do is, you want to fix that experimental setup, that humidity level and change all other variables, finish the experiment, then go to the next humidity level, finish all the combination, then go to the 3rd level. That is called kind of a simple randomization. So, what you do is, this you fix particular experiments level in this case 4 which means one of the conditions is fixed and then, you repeat those experiments 3 times.

(Refer Slide Time: 29:18)

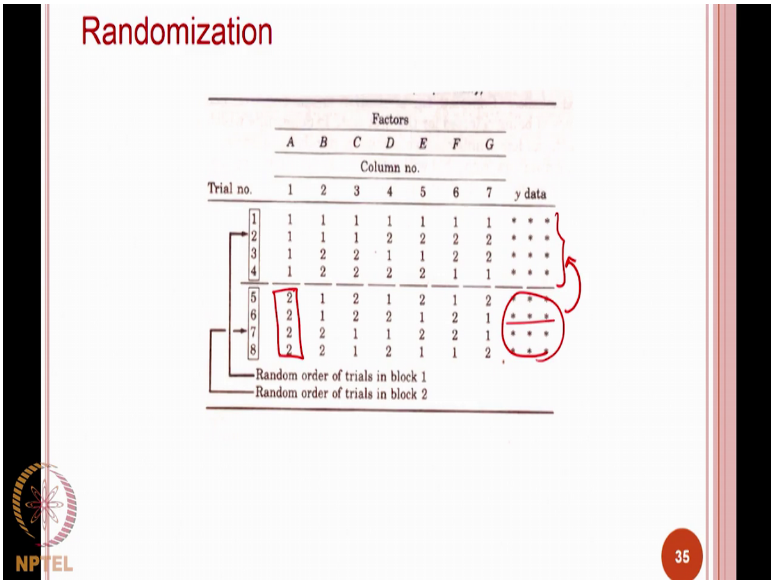


	Factors							
	A	B	C	D	E	F	G	
rial no.	1	2	3	4	5	6	7	y data
1	1	1	1	1	1	1	1	* * *
2	1	1	1	2	2	2	2	* * *
3	1	2	2	1	1	2	2	* * *
4	1	2	2	2	2	1	1	* * *
5	2	1	2	1	2	1	2	* * *
6	2	1	2	2	1	2	1	* * *
7	2	2	1	1	2	2	1	* * *
8	2	2	1	2	1	1	2	* * *

Imagine that I am trying to do different types of studies on different types of grips in the tensile test that I am talking about. Putting the grips itself to the holder is a big problem. It takes considerable amount of time. Even your test gets over before, I mean your tests will take a few minutes, but getting this guy making sure they are aligned, all that takes a while.

So, if I am dealing with 3 different grips that I am trying to factor in, then what I might want to do is this. Put grip number 1, do 3 specimens, then remove them, put another grip, do 3 specimens and then, go in a step instead grip 1 specimen, number 3. Remove grip 1, grip 2, and specimen number 2, remove grip 2, again grip 1. It is going to be troublesome because every time you would remove a grip and put, it will take anywhere between 1 and half to 2 hours. It is time consuming and you might not want to do that kind of variability. Then, this is what I will do. I will fix one variable you understand, right. When I put the block, that block is repeated, the columns are not ok. I am fixing that particular stuff.

(Refer Slide Time: 03:30)



	Factors							
	A	B	C	D	E	F	G	
	Column no.							
Trial no.	1	2	3	4	5	6	7	y data
1	1	1	1	1	1	1	1	* * *
2	1	1	1	2	2	2	2	* * *
3	1	2	2	1	1	2	2	* * *
4	1	2	2	2	2	1	1	* * *
5	2	1	2	1	2	1	2	* * *
6	2	1	2	2	1	2	1	* * *
7	2	2	1	1	2	2	1	* * *
8	2	2	1	2	1	1	2	* * *

The last one is a block 1 level. This again the same humidity condition will help. So, let us say that you have 2 levels of the humidity. What I will do is, this I will fix this particular level and I will do all the experiment that corresponds to this 3 times and then, I will do. So, earlier you fixed only one of the experiments. If you remember I blocked only this guy, but now what I will do is, this I will go with the levels. So, I will do this and then, I will do all these levels first and then, I will do this level next. So, that is this block randomization. So, you are blocking this guy and then, you are blocking this guy.

There are people who will go and do variability between the blocks also. That is when you want to do variability analysis. ANOVA stands for Analysis of Variance or you want to do a correlation study, you want to do an error analysis and all that you can do that. But, that is all that is out of scope for us.