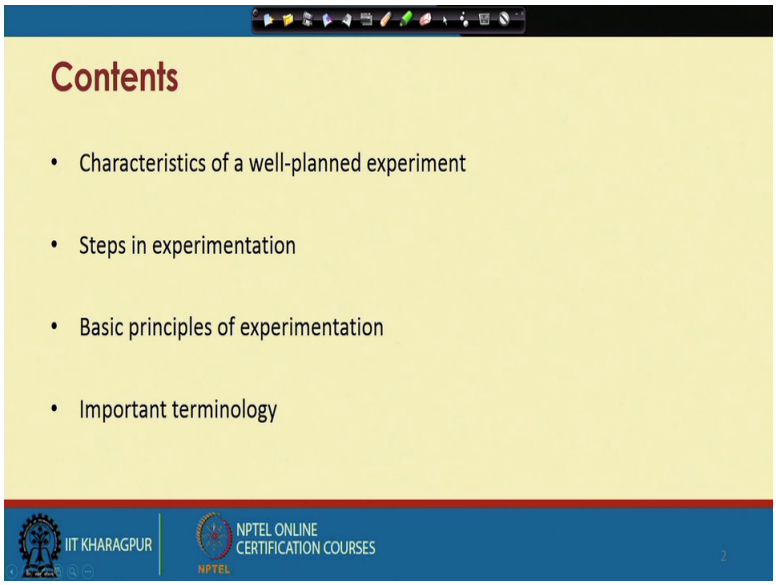
Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture - 02
Principles of Experimental Design

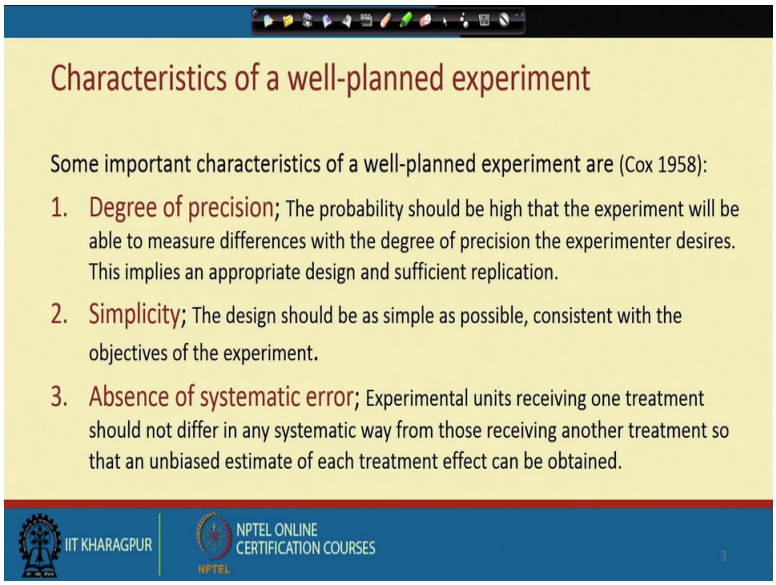
Welcome to the second lecture of Design and Analysis of Experiments. Today we will discuss Principles of Experimental Design.

(Refer Slide Time: 00:27)



The learning objectives for today's lecture are characteristics of well-planned experiment, steps in experimentation, basic principles of experimentation and some other important terminology what we have not discussed in the first lecture.

(Refer Slide Time: 00:49)



As I told you in first lecture that experiment is a purposeful one there are certain objectives that to be satisfied. So, the experiment should be a very good one by a good experiment, we mean there are certain criteria that must be satisfied and some of the criterias can be understood qualitatively, some of the criterias can be understood quantitatively.

So, what are those criteria or what is qualitatively and what is quantitatively, qualitatively means we may not be able to perform certain tests quantitatively means maybe from data or through some statistical test we will be able to tell that yes the experiment edge of that quality. So, quality of experiment design a is a very important one and as a result it must be well planned. Now what are the important characteristics of a well-planned experiment, which is listed by cox 1958 1 is degree of precision. It is something like this suppose you to want to measure the length of a piece of rod, then you must measure it with precision means if you measure in repeated number of times what is it should happen the measurement should be ideally the same.

If it is 10 meter it should be 10 for all the time, but when we talk about a process for which experiment is done there you cannot expect that every experiment or whenever you run a process the output whatever maybe the similarity level, but there will the quality or the response measures in terms of y that we have defined earlier. It will never be the same means it if I measure it a measure by tape a length 10 meter, I will not be

able to measure all the if I measure I will not given to get all the time 10 meter it maybe plus minus 0 0 0 0 1 this side or that side.

So, by precision we mean that what we measure that measurement should be me with minimum error and the deviation in the measurement should not be much. In experiment what is that what you want to measure we have some target something we want to achieve then what we are achieving through experiment by measuring the y or x. So, the difference is the error and that error should be measured very very precisely, the error measurement should not be very very very much spread.

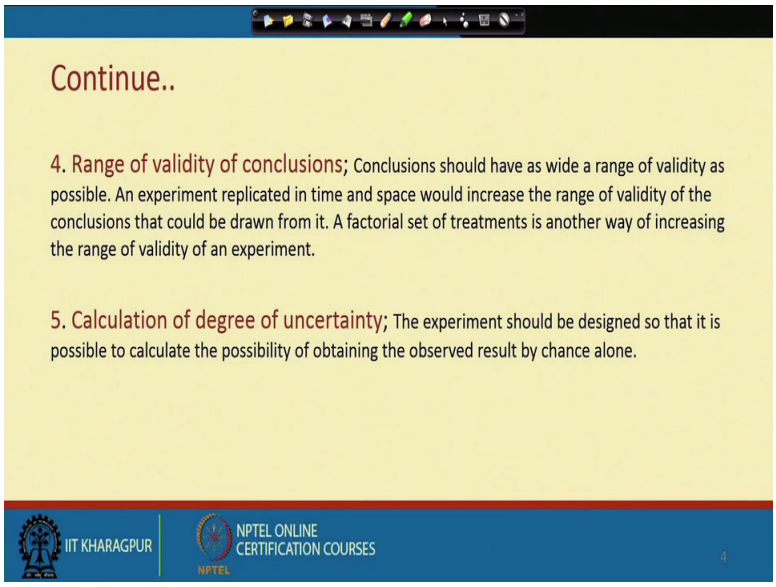
So, by that is why degree of precision is one of the important criteria for well-planned experiment, what is degree of precision the probability should be high that the experiment will be able to measure differences with the degree of precision and the experimenter precision the experimenter desires.

Suppose you want the length your measuring it should be plus minus 0.002 meter it should be within this. This implies and appropriate design and sufficient replication. Replication in the sense how do you know that your precision is more you require many many such runs and then we have many such values then only you will be able to find out whether your precision is as the as the desired level as per the desired level or not. So obviously, the standard deviation is one of the important measures here which later in let us states time you will know that what is standard deviation how it is to be computed. Then simplicity the design should be as simple as possible consistence with the objective of the experiment what does it mean.

Unnecessarily taking many factors many x and concentrating on many z also it is not desirable unless the x purpose or the objective of the experiment desire. So, what you require do to that is why you may consider important factors and variables and that important you make it the simple similarly you will find out later on that in experimental design there are different principles that to be adopted. So, you should adopt the principles you design the material you say select the material design the experiment, run the experiment, in such a manner that it should make it simple. And it should be consistent with the objectives of the experiment, the absence of systematic error experimental unit receiving one treatment should not differ in any systematic way from those receiving another treatment.

So, that the unbiased estimate of each treatment effect can be obtained what do I mean suppose let your measuring one piece, then second time also your measuring, that time you are measuring if you are may you know that the second measurement value is 5 units, then the next measurement value you would think that it will be 5.5 or 5.5 and that kind of systematic bias should not be there. So, it is it is should it should not have any systematic error.

(Refer Slide Time: 07:02)

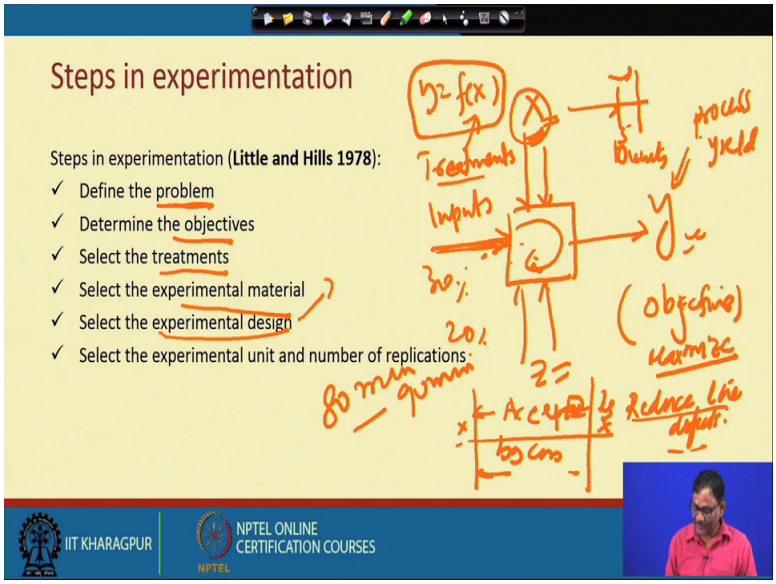


Then range of validity of conclusions the experimental results finally, will be used to aim with to improve the process or design of new product or design new surface, many other things and which is which is not limited to the experimental condition.

So, what you require the results will be generalized for much broader system. Obviously, consider; obviously, that it should not be that that first that it will the factors and other characteristics will not match, but please remember if you do experiment the laboratory it should be it should be reproduced in the field level where actual work is going on, that is what is known as validity. Means it should if it is applicable today it should be applicable tomorrow, if it is applicable in within a particular chemical process, then the in the similar process in others place also it should be applicable. So, that is what is known as the validity range of validity. The factorial set of treatment is another way for increasing the range of validity.

Now the range of validity is very very important you should consider the factors in such a manner that they gives you the general things. Then calculation of degree of uncertainty so the experimenter should be designed so that the experiment should be designed show that it is possible to calculate the possibility obtaining the observed result by chance alone. Now it should it should it should it has a it should be when you are in a position to know the what is the uncertainty level is there. So, you will be able to calculate that that uncertainty and at the same time. Basically, what happened the uncertainty will be random only it should not be the systematic uncertainty when a systematic way of inducing such things.

(Refer Slide Time: 09:21)



Now, for the time being so, what I let us concentrate on a process it has inputs, it has outputs, it has controllable factors, it has uncontrollable factors. So, inputs, x outputs, z controllable factors, y is your response variable that is.

So, objectives will be said in terms of the problem; problem means suppose y suppose if y is the is related with the process yield process yield, then what you want to what will be your objective you want to maximize the process yield now object will be maximize process yield. Suppose your objective is that quality; quality of the output say quality then maybe you want to reduce the defect.

So, you must know where lies the problem? So, define the problem is it a process which is constantly producing out of the desired range, suppose this is the range which is

accepted by which is accepted by customer and this is the range which is rejected is it producing here is it producing here or it may. So, happened that the amount what is produced this is basically somewhere you want this is the target 10 to the power 5 units.

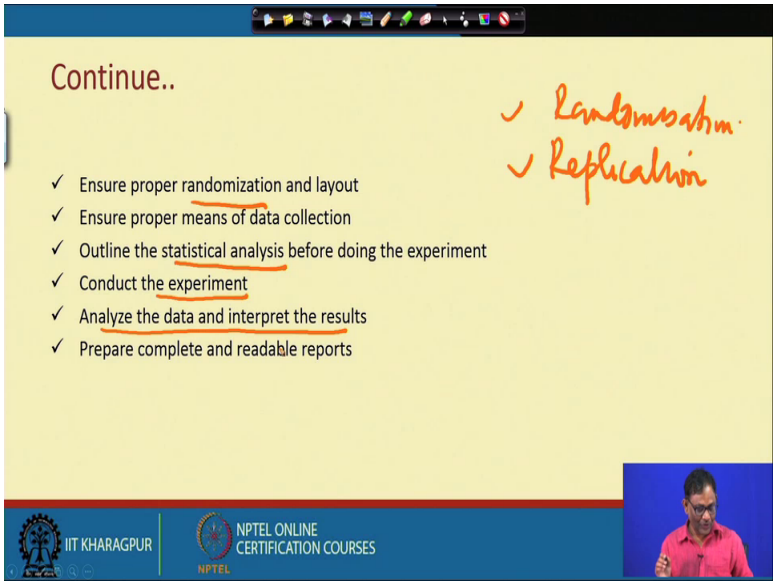
But you may be producing here. So, that you want to increase this. So, you must know what is the problem, by saying these it is not the problem is not coming now you have to know that depending on this what are the x and what are the z and what kind of inputs are there and what is this process all those in detailing you have to do. Now I as I told you then what is the first is the problem; problem mean that it is a quality problem, process is not performing as per the desired specifications. Then what will be your objective improve the quality of the output. So, in this case reduce the defect or if it is a maybe if it is a dimensional variable you produce within this range something like this. Then comes the select the treatments.

Now if I as I told you in first class that y is a function of x. So, this x is factor if these are nothing but also known as treatments. Because you can treat x in different ways, what I mean to say better way better way we have to be say that you can treat the process by manipulating the x so that the output will be of desired quality. So, select the treatment minimize is you select the factors x controllable factors, which are important for this functionality then selects the experimental material actually you are giving certain inputs here it may be raw material it may be a component parts it may be a fashion so; that means, when you want to design experiment this inputs very important. Suppose if patient with certain edge group is important instead of these you are even maybe choosing patient of all edge group then the purpose will not be solved.

Similarly it may. So, happened that when we are producing a particular quality of molten metal for steel making then the coke what you are is your input this would have person certain percentage of s or otherwise ways and the certain percentage of quality, but instead of that you have a different percentage of as; that means, it will not be correct one for explain. In the products regular production as content in maybe 30 percent, but when we are doing experiment you have taken call of 20 percent as content then this is not the correct material you have chosen. So, experimental material it is one example from raw material point of view there will be another components also for all those things you have chose accordingly.

Then select the experimental design now experiment by experimental design we means that how you will manipulate the process during experiment there are different kinds of designs available. For a further time being I can say that there will be one factor at a time that can be on type of design, all multiple factors together. Now when if you take multiple factor together you may go for all levels of each of the factors you may go for a some level of demand factors for example, the reaction time in the earlier example we are giving it is wearing from.

(Refer Slide Time: 14:45)



It is it is varying from 80 minute to 90 minute in the pros chemical reaction case so, but what I mean to say here that so experimental design. So, what I mean to say that 80 minute (Refer Time: 15:04), but somebody someone may choose 3 levels, for 3 levels for someone may chose 3 levels for this one that factors that particular factors.

So, similarly temperature it you may chose a 170 to 180 you may chose that 150 170 5 different levels you may those things will be discussed later on, but further time being you play in understand that the treatment or the a controllable factors x that how do you choose, how do select the different levels for x by level we mean that, if it is a temperature it can be continuous variable any value possible, but most of the time it is not possible to consider all the values all combination you may chose a choose a low or medium and accordingly there may be your factorial design. Now if you choose the factorial design one can kind of result data you will get if you use factorial design with

some other additional constant like central points or you may go for mixture design depending on the depending on the requirements you have to choose the design experimental design will be talked to you in subsequent classes.

Select experimental unit and number of replication very important experimental unit and number of replication. So, I will tell you when the principles of a design doe that what do you what do you mean by replication, then insure proper randomization and layout. So, please keep in mind this was one is I told you replication, another one now you are saying randomization. So, I will I will explain what is randomization and replication, with some example and then you ensure proper means of data collection. So, if it is experimental it is a manual data collection or it is a offline, online or it is a sensor based what kind of the your data when, which one is based you that you do then once you have data out the you required to do statistical analysis.

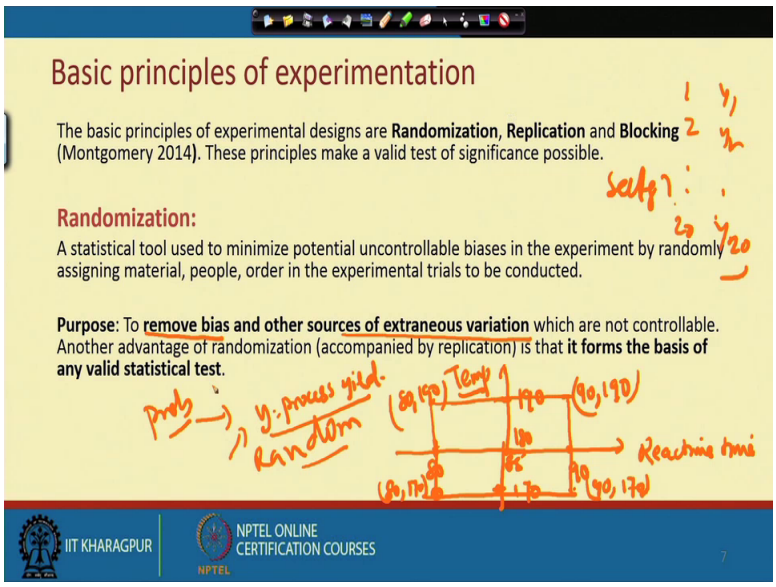
So, as you required to do statistical analysis of the data that will be obtained from the experiment. So, you should note wait till the data is available what it is said that before the experiment you assume know all the other things you must be we must be while equip that, what kind of data you are going together out of the experiment and what kind of analysis will fit to that data. So, some outline you must have, then you will do the experiment, then after the data will that data whatever data you get that data you analyze and interpret the result prepare and complete readable reports.

So, I am repeating this yes. So, let me repeat when you do an experiment you must be well prepared for that experiment and they are the certain steps first problem must be very much specific you have every corner; you know the every aspects of the problem. You have clear cut objectives you know; what are the controllable factors and you are in a position to select particular appropriate made experimental material and you know variety of designs and from that variety of experimental design you will be you will be choosing the approp correct one. And also you know that what, but what are the experimental unit you are going to use as well as you know that what will be the number of times you will be conducting a particular state of the experiments and then the experiment should not be biased or systematic, it will be properly randomized and layout mean in the what are of experiment also should be randomized, then you must see that when you go for experiment the data collect how you will collect the data. So, the data collection process must be known before hand. And also once you know that what are

the data that are the when to come then you must be in a position to know what kind of statistical analysis to be conducted. And then what will happen when you conduct the experiment after that you will choose the appropriate statistical techniques for the experiment.

For example if you are interested to compare to the performance at 2 different treatment levels these statistics is sufficient, but if it is more than 2 3 or more you required to use annova. So, that mean you must have that much understanding that what kind of data are going to be generated through the experiment, then you will analyze the data and analysis things will be told to you later on and then. Obviously, the results must be interpret it and you please keep in mind that it is a statistical issue. So, you are thinking is statistical you are interpretation must be statistical and from the statistical thinking to practical problem solving that much that type of mind set you must have then prepare complete and it able reports the entire experimental reports.

(Refer Slide Time: 20:43)



So, as I told you that the fine the steps are known, but at the same time you please understand that the there are certain principles which are to be fulfilled, otherwise your experiment will not be considered a correct one or good one when we always will be subjected to error which are not acceptable-unacceptable errors. So, the 3 most important principles of experimental designs are randomization, replication, and blocking. I will first read out what is the randomization then I will try to explain with an example.

Randomization a statistical tool used to minimize the potential uncontrollable bias is in the experiment by randomly assigning material people order in the experimental trials to be conducted.

What is the purpose to remove bias and other sources of extraneous variation which are not controllable? So, another advantage of randomization is that it forms the basis for any valid statistical test. Let us consider this example that reaction time the chemical process the reaction time these are based temperature and; obviously, we know that the output is y output is process yield. For the time being let us assume that you are considering 80 degree 80 minutes and suppose this is 85 minutes and this is 90 minutes, this range for time and this 1 100 suppose 70 and let it be 190 then in between it will be 180. What are it mean the range suppose the operating range, for the process we are considering 80 to 90 for reaction time for temperature 170 degree 500 190 degree 500.

Now, let us suppose I say the treatment, treatment mean how I will conduct the treat the process the reaction process. So, that I will be able to have, but I will have the yield values. So, my treatment can be the reaction time at 80 degree 80 minutes and 90 minutes, as well as temperature can be 170 degree and 190 degree. So, I can create a combination like this. So, this is a combination which talks about 80 minute reaction time 170 500 temperature this is my one setting, this setting is your 90 minutes 90 minutes and again this is 170 second setting, third setting it is 90 190, 4 setting here may be your 80 190.

So, what I what I mean to say here now see that mean the treatments reaction time and temperature there to treatment variables that basically your factors controllable factors a treatment level is 80 and 170, another one is 90 and 170, 90 190 80 180, these are known as treatment levels. So, what we say that now how do the experiment you will you will you will keep reaction time at 80 a minute and temperature 170 you can do experiment. So, question is which one you will choose first this one this one this one this one who is settings. So, settings which one first or which one second will there be any what are you cannot chose selectively you have to randomize it you will choose in with using certain random numbers. So, that what happened first time you made do experiment here second time here third time.

He where when in the second time may be here like this. So, there will be random order in choosing the settings at the same time maybe in a particular setting to may do 4 5 experiment a 4 5 runs, when 4 5 times 80 minutes 170 temperature 80 minutes you will run the run the process. Similarly here similarly here you can repeat this one for 5 times, but you will not repeat at a time mainly in one go first experiment is over first second run is over like the in the same setting 5 times you cannot do, what you required to do you again you have to suppose 5 5 5 5 20 runs will be there those 20 runs will be randomized.

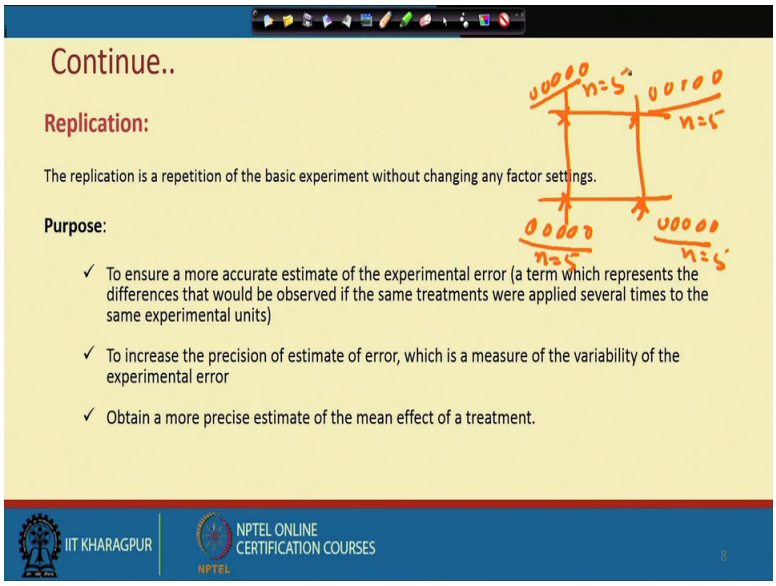
So, that is what is your known as random randomization, why we are talking about material it may. So, happen that the same amount of material the input will not work for everything. So, similar material will come and the which material you will chose that also to be randomized, suppose it cannot be completed within a 8 a what period you require 4 5 people to do in 4 5 6. So, that time also that also to be randomized.

Student: (Refer Time: 26:44).

So, it is a huge issue because unless you randomize you what will happen there will be systematic biases. So, you want to remove this bias and randomization also helps in this process randomization also helps you to eliminate out the extraneous variations. And as you know if you do randomization then the y the whole process yield, in this case this will be a random variable also because the process is randomized the experiments are is randomized all the runs are based on randomized experiment, a outputs are based on randomized experiment. So, in that case what will happen whatever output you get suppose in total if there are 20 20 such runs then you will be I think 20 such y values? So, all are random. So, that mean this will follow a certain kind of probability distribution.

Most of the times normal distribution we will find. So, randomization is a very important to 1.

(Refer Slide Time: 28:01)



Now, Replication; in the previous example you have found out 4 different settings ninety 170 80 170 90 170 then 80 180 90 180 4 different settings. So, you can at least you can conduct one experimental run at each settings, if you do one set a time that mean you are not doing any replication, you are doing single getting single results for y. That will note help you to find out the errors because unique your way there are many later on we will know that there are many vary many vary variables many levels together. So, many factors together so all those things what happened they will their effects once you estimate with single observation you may not be able to estimate all of them.

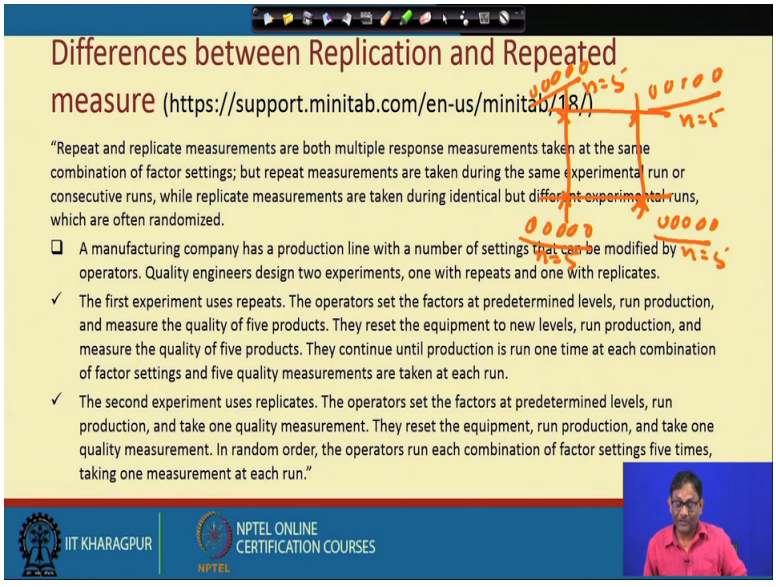
If you want to estimate the all the effects you will not be able to estimate the error. So, what you require you require in each setting you require more number of experiment to be conducted. As I told you I have 4 set 4 different settings this one this, this and this here instead of one you conduct maybe 4 or 5 runs, here you conduct 5 runs here you conduct 5 runs here you conduct 5 runs.

So, now, will you conduct at a time 5 runs here at a time 5 runs here no you do you will not do experiment like this here you have to this 20 experimental runs will be randomized. Now whatever may be thing after completion of the experiments you will find out 5 observations for this identical setting, another 5 observation for this identical setting, another 5 observation for this identical setting, another 5 observation for this identical setting?

This is known as replications n equal to 5. So, what you write the replication is repetition of the basic experiment without changing any factor settings, this settings you are not changing here you are repeating this again and replication with randomization we will be doing. What is the purpose to ensure a more accurate estimate of the experimental error, to increase the precision of the estimate error and to obtain more precise estimate of the effect of the treatment. Though these this continuity basis all are quantity been nature once we complete the basic statistics part that time I will again come back and I will tell you what how to compute the accuracy, how to compute the compute the calculated the precision of the estimated and how do we say that it is basically effect is the you are getting precise effect for the main effect precise estimate for the main effect that will be told to you.

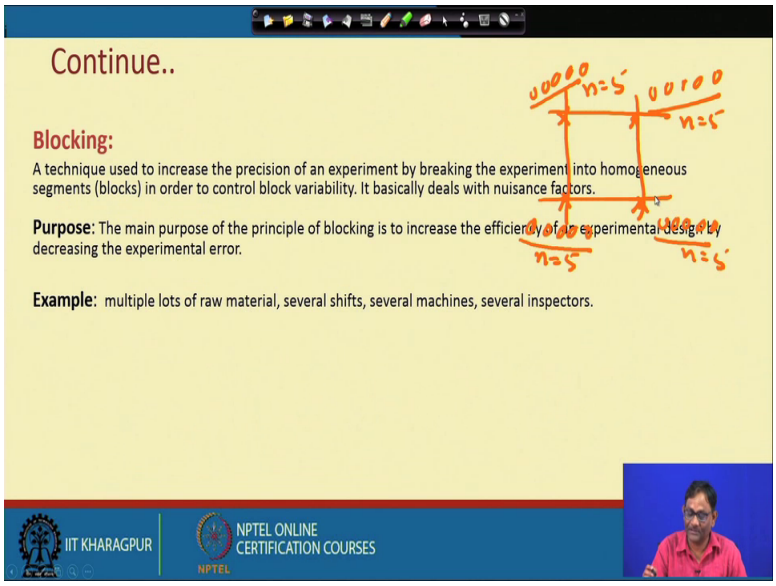
For the time being you understand this replication means that in every setting there will be more than one experimental runs, and by randomization all the setting this settings will be chosen randomly based on certain random; randomization scheme and also each of the runs at each of the setting that will also not be in or a at a time or in sequence it will be it will be also randomized random what you is.

(Refer Slide Time: 31:55)



So, there is a important concept called he repeated measure and replication. So, I just read out what is repeat and replicate measurements are both multiple response measurements taken at the same combination of factors setting like here.

(Refer Slide Time: 32:18)



These 2 factors setting this is first combination, this is the second combination, third combination, fourth combination. So, you are taking multiple measurements.

But repeat measurements are taken during the same experimental run or conjugative runs, if you do if you repeat here suppose 5 times you are taking here, then you are going 5 times here, then this repeat measurements while what is replicate while replicate measurements are taken during identical, but different experimental runs which are often randomized.

So, if you randomize that what will happen first you run will be taken here, second run may be taken here maybe 30th one more run 20th run again will be taken here. So, if you choose a setting and then a repeat the experimental run for n times that is repeat measurements or a, but if you know that there will be 20 experimental runs at 4 different experimental settings and each run is randomized in such a manner that, you will you when whenever you choose this setting of this setting this setting this settings for the experiment it will be it will be based on the randomized scheme only then this is replication.

Here one example a manufacturing company has a production line with number of production line with number of settings that can be modified by the operator, quality engineer design 2 experiments one repeat and another one with replications. What is the first one repeat one the operator said the factors at predetermined levels like here 80 90

and 170 180 predetermined levels, run production and measure the quality of 5 products, 5 the products at a particular set settings. They reset the equipment to a new level like here and then again run production and suppose take 5 products of an quality value.

They continue until the production is run one time at each combination or 4 are completed that is repeat measurement, what is the replication the operator said the factors at predetermine levels run production and take one quality measurement, they reset the equipment run production and take another quality measurement, in random order the operators run each combination of the factor settings 5 tens taking one measurement at each time, what I explain that is and I have taken this one from the min tab support.

Student: (Refer Time: 35:08).

Another important concept is blocking you have you have heard that the noisiness are noise way uncontrollable variables. Suppose you consider any productions of where raw material are supplied by different venders what is the guarantee that all venders are giving you the equal amount of quality. It may not be in that case of raw material create in machines variable or uncontrollable variable, sometime suppose there are different operators of different training and training exposure different work experience, different call education level, but they are doing working in the same machine or to same productions of.

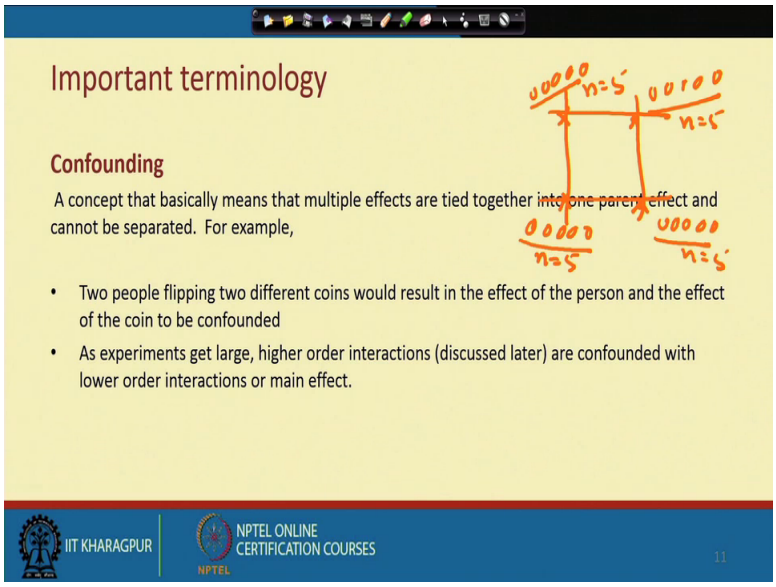
So, that also create some kind of uncontrollable or controllability or noise, suppose you want to you want to do experiment and in that case if you do not block the raw material coming from different vendors or the operators of different quality then the then what will happen the whatever at the error term will increase while term will increase in the sense the error value will increase. So, what you what you require it is always advisable that you do the experiment with a set of homogenous, homogenous material or homogenous people when experimental a material or experimental image should be homogenous. If you cannot do this then then what will happen the heterogeneous effect will be there and error will be more.

So, blocking is a very interesting technique here is a experimental technique what it does it basically help in in conducting experiment a experiment for the homogenous said together in the sense that by breaking the experiment into homogenous segments.

Support a particular set of raw material will be used for all experimental all treatments settings, then another experiment raw material will come, then again all those experimental settings with that raw material the experimental run will be conducted in that manner. So, then we can we can have data for all the factor settings for and for each of the raw material group that complete set of runs are available. In that case your analysis when you do analysis you will be in a position to find out the block effect or you will get the factor effect by controlling the block effect.

So, the main purpose of this principle is blocking is to increase the efficiency of the of the experiment experimental design by decreasing the experimental error, here example multiple lot multiple lots of raw materials several shifts several machines several inspectors.

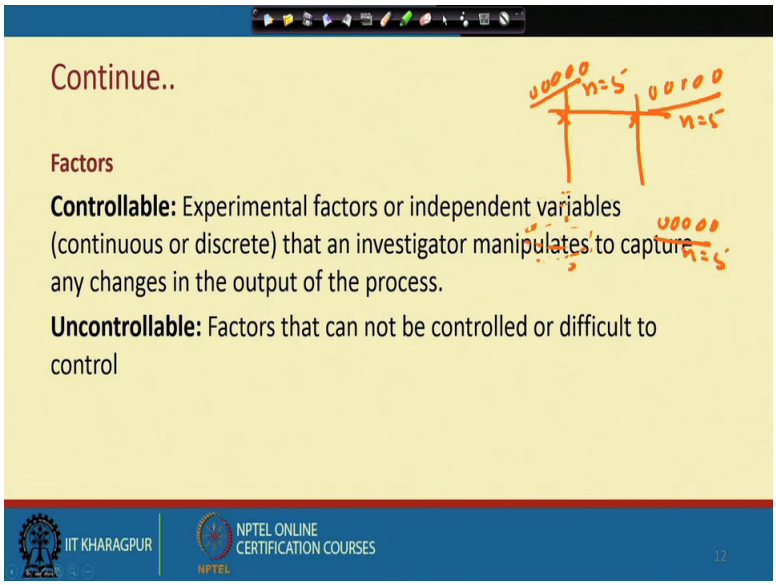
(Refer Slide Time: 38:14)



Another in interesting terminology what I am not discussed in the first lecture that is known as confounding. So, confounding is a important one it is a concept that basically means that multiple effects are tied together, into one parent and what will happen that you will not be able to separate them. We will not be able to separate them for example, 2 people flipping to different coins would result in the effect of the person and the effect of the coin to be 2 people flipping 2 2 different coins. And you want this experiment you will not you will do you take the outcome into consideration and do some kind of analysis. And finally, take some of the next sense.

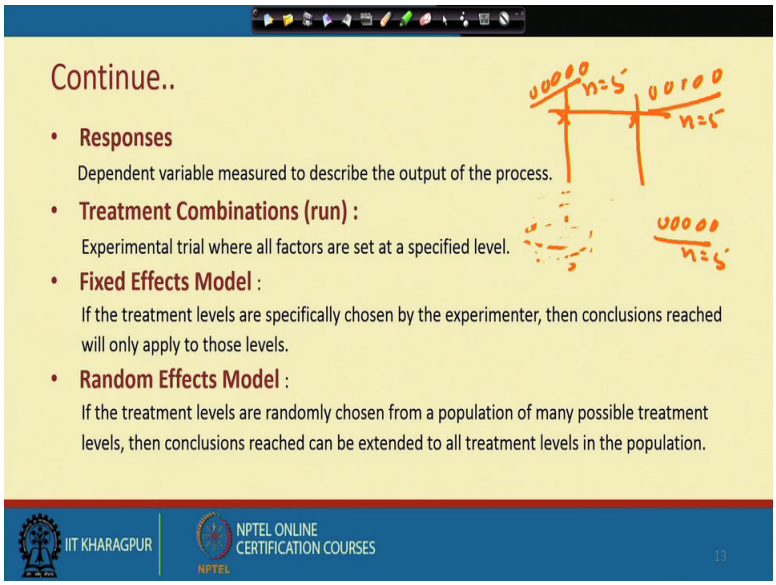
In that case what happen the people and the coin this is confounded you cannot separate out, this situation is known as confounding. Later on when you go for higher order factorial design I will let you know that how the different interaction effects are confounded with other interaction effects as well as with main effects.

(Refer Slide Time: 39:40)



So, some more like factors controllable uncontrollable already told you response I have given you treatment combination given you.

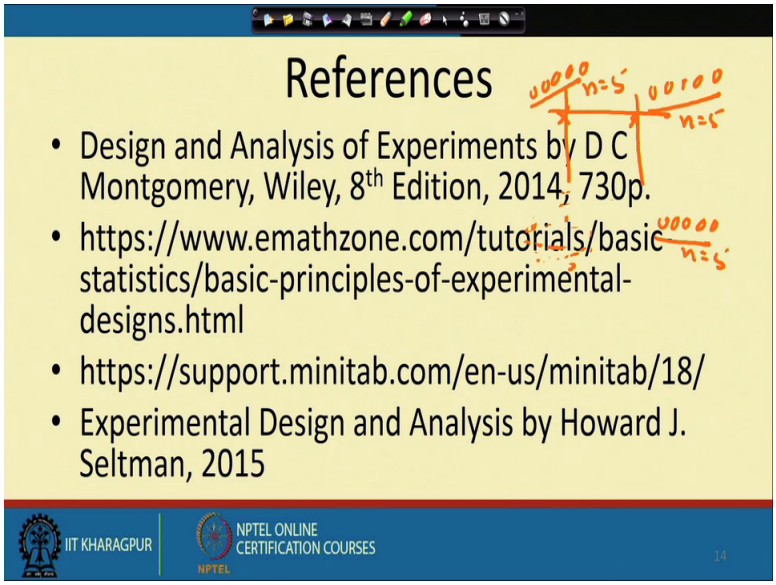
(Refer Slide Time: 39:47)



Now, fixed at that model and the random effect model as I told you that suppose the in the example reaction time I said eighty minute and ninety minute if you fix it and do the experiment get the result analyze it this is fixed effect model, but if you allow the reaction time to be also probabilistic in nature random you do not know when what is the settings, but if the settings will also be probabilistic in nature. So, then what will have that then in that case the model what you will develop later on we will see that will be random effect model.

So, it all relate to the fact fixed effect and random effect model will relate to the controllable factors, if the controllable factors are fixed well in advance before the experiment and then the experimental fixed experimental settings are there and on each of the fixed experimentals you will do the experiment get the results where y is when response variable is a random, but the controllable variable values are not random that is fixed effect model in random effect model the controllable variables values also randomly selected it will be randomly selected not fixed beforehand then that will call random effect model.

(Refer Slide Time: 41:11)



So thank you, and I have use several references like first Montgomery then to that websites and also another book by H J Seltman, where written in 2015 Experimental Design and Analysis.

Thank you very much.