Quality Design and Control
Prof. Pradip Kumar Ray
Department of Industrial And Systems Engineering
Indian Institute of Technology, Kharagpur

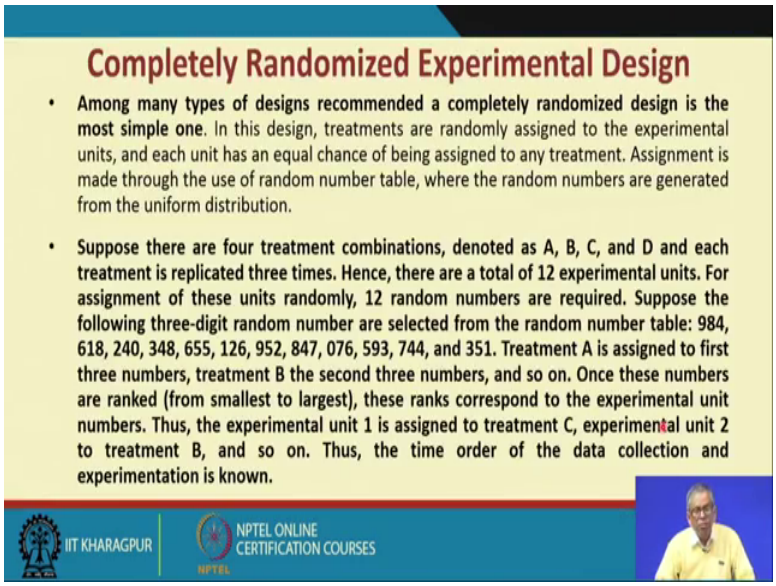
Lecture – 53
Quality by Experimental Design (Contd.)

(Refer Slide Time: 00:19)

In this lecture session on quality by experimental design, I am going to discuss in detail a particular experimental design which is referred to as the completely randomized experimental design. Now completely randomized experimental design is your starting point; that means, what you find that whenever you believe in say the concept of statistical design of experiments. So, your starting point is this completely randomized experimental design only in certain conditions, what you do; that means, you refer to certain variations in randomized experimental designs.

So, the original design you must know and then you also must know that under in a given situation you may have to oft for a variation of completely randomized experimental design.

(Refer Slide Time: 01:21)



So, ah; so, in this lecture session let me explain in detail the completely randomized experimental design among many types of designs recommended a completely randomized design is the most simple one and that is why you say that in any experimentation exercise this is your starting point in this design treatments are randomly assigned to the experimental units and each unit has an equal chance of being assigned to any treatment; that means, this is essentially we say that the randomization principles you have to follow and while you follow these randomization principle you need to follow these principles completely.

Assignment is made through the use of random number table I hope that you are aware of what is a random number table the those who were not aware of random number table please refer to random number table and how to use a random number table, we will explain right now so, but you should refer to the random number table and you also must know that how these random numbers are generated. So, here whenever you use you refer to a random number table what do we assume that these random numbers are generated from the uniform distribution that uniform distribution is referred to as that many a time it is refer to as refer to as the distribution with maximum ignorance.

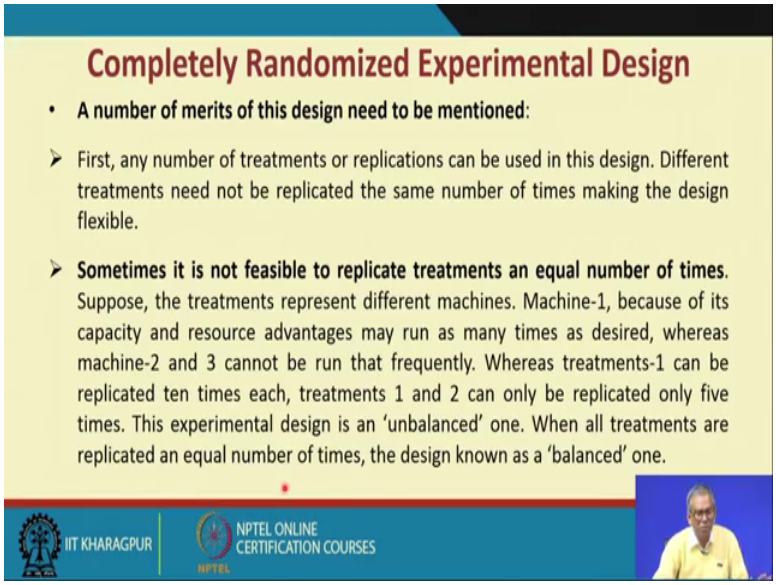
That means even if you do not have say the data the past data as an as an the initial stage in the beginning you may assume that the random variable is uniformly distributed and as you get the data you can the revise your the your assumptions regard related to the

distribution. Now, here is an example how to use the random number table suppose there are four treatment combinations already we have defined what is the treatment combinations denoted as A, B, C and D and each treatment is replicated three times ok.

So, replication principles already we have explained hence there are a total of twelve experimental units for assignment of these units randomly 12 random numbers are required. Suppose, the following 3 digit random numbers are selected from the random number table is it this is the this is available ok. So, random number table; so, you will find in any text books on in statistics you will find that in the appendix. So, random table is given. So, these are the three values 984, 618 and so on 240. So, these 12 values you select, treatment A is assigned to the first three numbers, treatment B the second three numbers and so on.

Once these numbers are ranked from smallest to largest, these ranks correspond to the experimental unit numbers as the experimental unit one is assigned to treatment C experimental unit two to treatment B and so on. So, this will you follow right, there is the time order of the data collection and experimental experimentation is known ok.

(Refer Slide Time: 05:12)



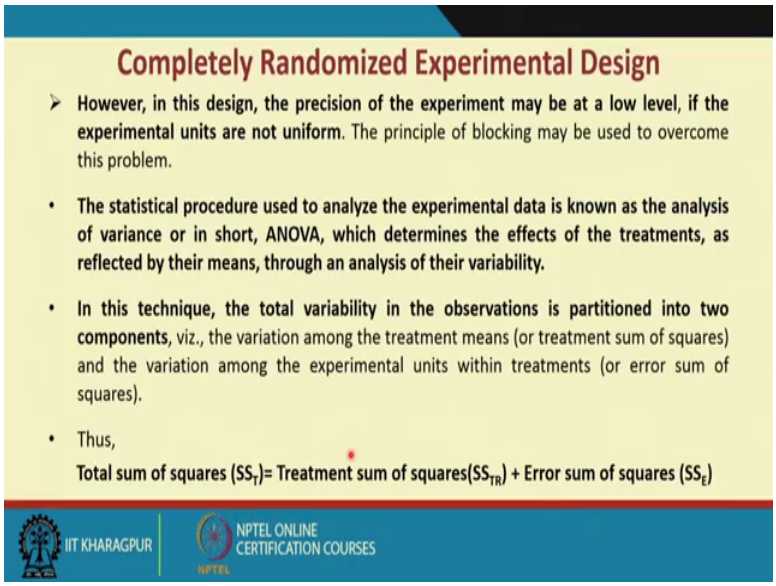
So, this is just a simple example a number of merits of these designs need to be mentioned; that means, why do you opt for this one we have been saying that you know given a condition you must opt for completely randomized experimental design.

So, what are the advantages first any number of treatments or replications can be used in this design, is it any number different treatments need not be replicated the same number of times making the design flexible, is it ok. So, this is point number one what is the next the merit sometimes it is not feasible to replicate treatments and equal number of times you see because of many reasons.

In fact, time could be a constant resource could be a constants or there could be many other reasons suppose the treatments represent different machines machine one because of its capacity and resource advantages may run as many times as desired, is it ok, this is a practical problem we are highlighting.

Whereas machine 2 and 3 cannot be run that frequently ok, whereas, treatments one can be replicated 10 times each treatments 1 and 2 can only be replicated 5 times these experimental design is an unbalanced one, is it ok, when all treatments are replicated an equal number of times the design known as a balanced one though always we prefer a balanced design was the balanced one, but due to some practical problems ah. So, many a time, you may not have the balanced ah. So, experimentation, so, you go for an unbalanced experimental design, is it ok. So, this is unavoidable in majority of the cases.

(Refer Slide Time: 07:11)



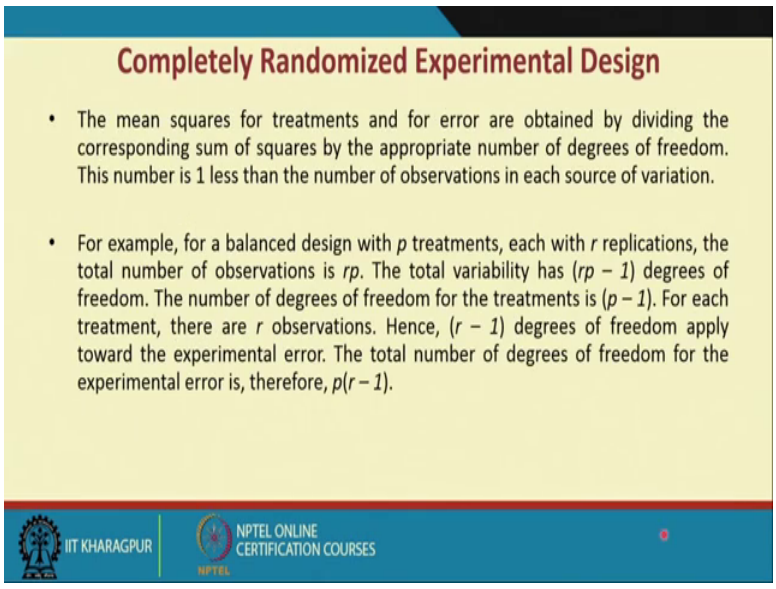
However in a randomized experimental design the precision of the experiment may be at a low level if the experimental units are not uniform. So, this is point to be noted the principle of blocking may be used to work on these problems till now we have we have

not mentioned about say the blocking ah. So, the first you start with randomized completely randomized experimental design and if you find that the precision of the experiment is at a low level then the next best option is that you go for you know the block design ok; so, randomized to block design.

So, the statistical procedure used to analyze the experimental unit this is very important point to be noted, now, what is the statistical procedure this statistical procedure is known as the analysis of variance or in short ANOVA. So, when we take up the numerical problems, we will construct the ANOVA table, is it ok. So, against any such experimentation you carry out. So, there must be a analysis of variance, is it ok. So, what is this ANOVA is this ANOVA determines the effects of the treatments as reflected by their means through an analysis of their variability.

So, in this technique what do you do the total variability in the observations is partitioned into two components variation among the treatment means or the treatment sum of squares this is the term we use treatment sum of squares and the variation among the experimental units within treatments that is error sum of squares, is it ok. So, two terms we use one is the treatment sum of squares and the second one is the error sum of squares. So, the total sum of squares that is SS T is equal to treatment sum of square SS TR plus error sum of square that is SS E is it ok. So, this is relationship.

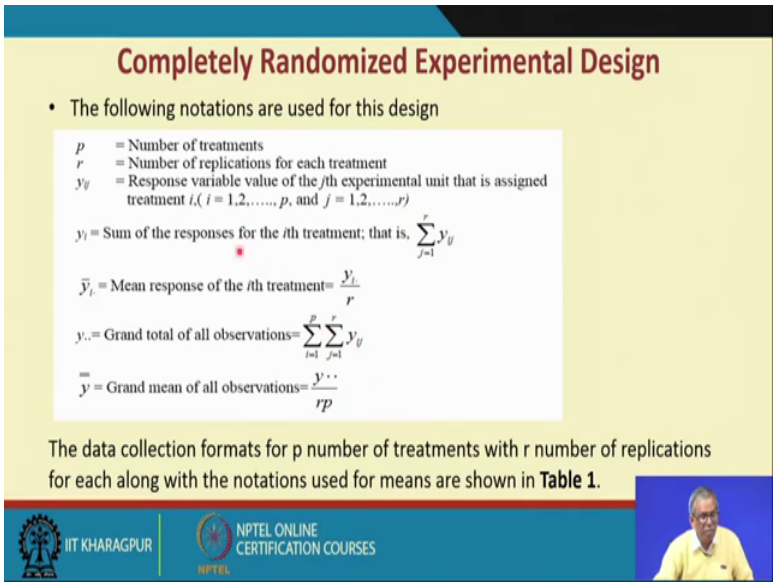
(Refer Slide Time: 09:26)



So, what we try to do that the mean squares for the treatments and for the error are obtained by dividing the corresponding sum of squares by the appropriate number of degrees of freedom this number is one less than the number of observations in each source of variation, is it ok. So, this rule we follow for example, for a balanced design we know what is a balanced design with p treatments each with r replications the total number of observation is r into p its clear the total variability has rp minus 1 degrees of freedom the number of degrees of freedom for the treatments is p minus 1 for each treatment, there are r observations. Hence, r minus 1 degrees of freedom apply toward the experimental error.

The total number of degrees of freedom for the experimental error is therefore, p into r minus 1 is clear. So, this rule we follow.

(Refer Slide Time: 10:30)



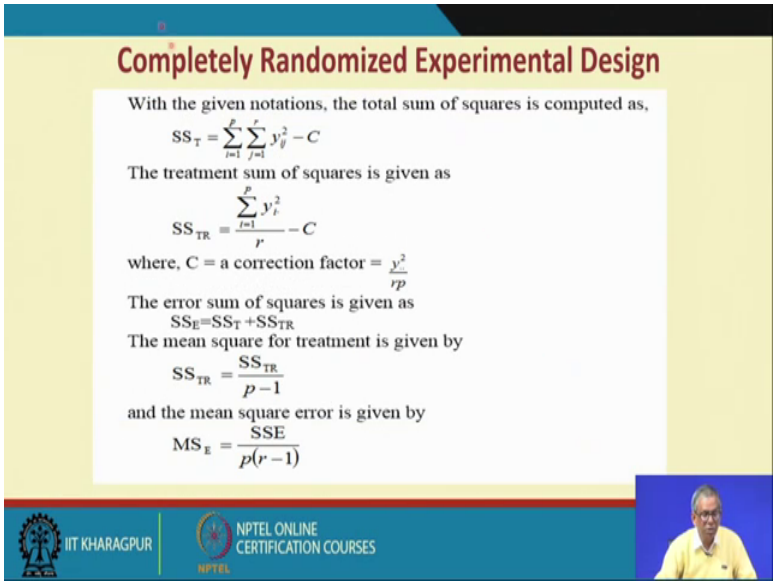
Now, what are the notations we use? So, let me first you know use these notations p is the number of treatments we have already we already know the what is a treatment r is the number of replications for each treatment we know that y are the principle of replication you have to follow yij response variable value of the j th experimental unit that is assigned treatment I is it ok. So, this is the treatment is i; that means, I varies from one to p and j varies from 1 to r is clear.

So, what is yi sum of the responses for the i th treatment that is sigma j equals to one to r yij is clear what is y I dot bar; that means, the mean response for the i th treatment yi dot

divided by r is clear, what is y double dot that is grand total of all observations that is sigma i equals to 1 to p j equals to 1 to r sigma yij and what is y double double double double bar that is the grand mean of all observations; that means, y double dot divided by r into p. So, this is ah. So, this is the, this is the notations you use and you have explained which notation means what.

The data collection formats for p number of treatments with r number of applications for each along with the notations used for means are shown in table.

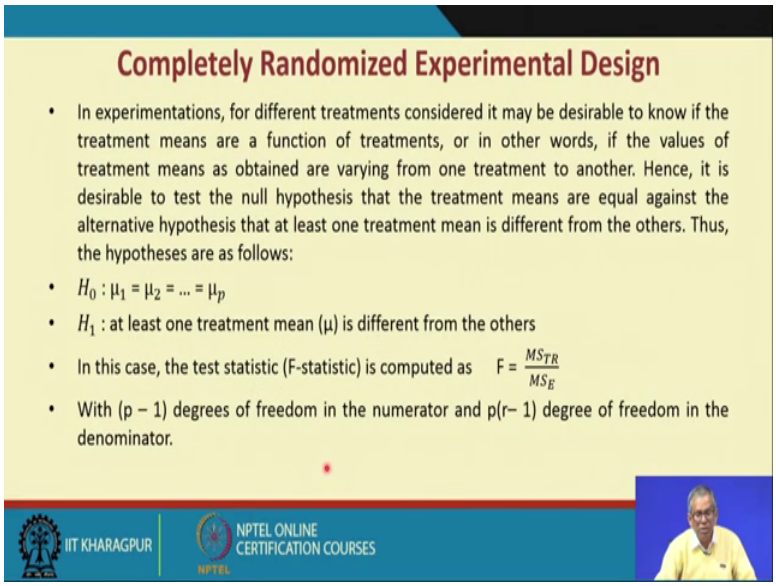
(Refer Slide Time: 12:25)



So, what is this table? So, we will show you the table. In fact, so, with the given notations the total sum of squares is computed is this one SS T that is the total sum of squares, is it and the treatment sum of square is given by this right sigma equals to q to p yi dot square divided by r minus C where C is a correction factor. So, that correction factor is y double dot square divided by r into p the. So, the error sum of squares is given as SS E equals two SS T plus SS TR and the mean square for treatment is given by this one, is it ok.

So, SS TR divided by p minus 1 and the mean square error is given by MS E to SS E divided by the degrees of freedom that is p into r minus 1 ok.

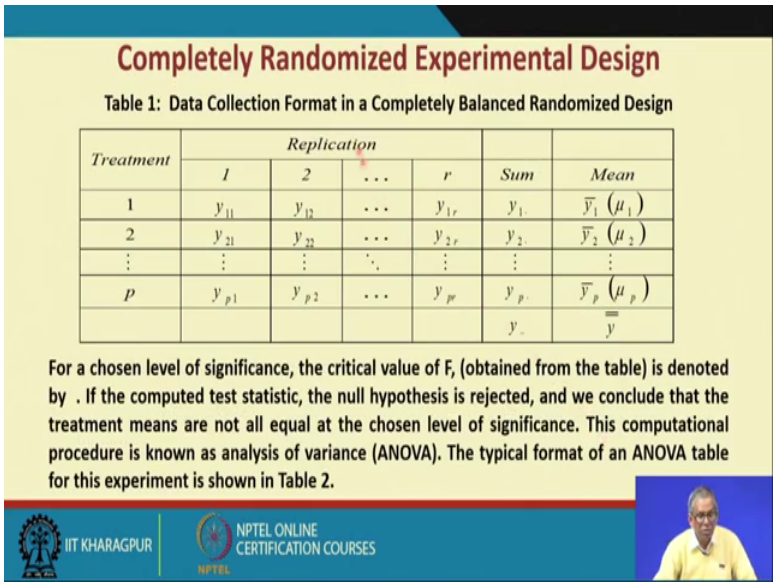
(Refer Slide Time: 13:21)



So, this is a standard formulations we have. So, in experimentations what you do for different treatments consider it may be desirable to know if the treatment means or a function of treatments or in other words if the values of the treatment means as obtained are varying from one treatment to another, is it ok. So, the variations are to be traced essentially, hence, it is desirable to test the hypothesis that the treatment means are equal against alternative hypothesis that at least one treatment mean is different from the others, is it ok.

So, the treatment mean is actually your result and that you are trying to test, is it ok. So, whether in the treatment values the mean value the differs are not thus the hypotheses are mu 1 equals 2 mu 2 is equals to mu p and what is alternative hypothesis H 1, there is at least one treatment mean mu is different from the others, is it ok. So, in this case the test statistics f statistic is computed as f equals to mean square treatment divided by mean square error, is it with p minus 1 degrees of freedom, in the numerator and p into r minus 1 degrees of freedom in the denominator in the denominator, is it ok.

(Refer Slide Time: 14:45)



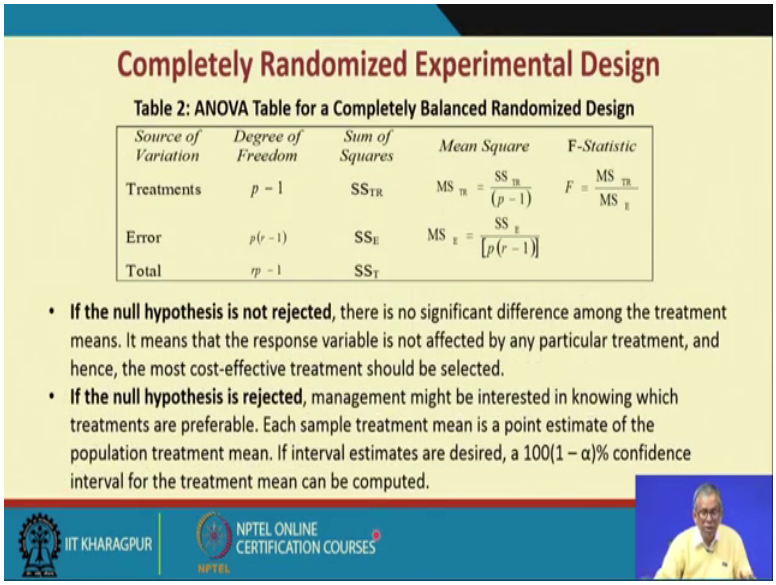
So, so, this is the table, you have you have the treatment 1 and 1 2 up to p treatments you have each treatment is replicated r number of times is it and then this is the way you collect the data y 1 1, y 1 2, y 1 r, this is y 2 1, y 2 2, y 2 r, this is you look at this table, I think all the notations are made very very clear the notations which you have used already. So, y p 1 is it ok, it is a p a throw and this is y p 1 and similarly ypr. So, you get the sum; that means, the sum against the first row is it whatever may be the value of r that is a variable that is why it is dot.

So, this is y 2 dot and this is y p dot, is it ok. So, and you add them; that means, it becomes y double dot, it is clear. So, sum and then you get the mean. So, you get the mean that is y 1 bar, this is y 2 bar, y p bar. So, this is referred to as the mu 1, mu 2, mu p, is it ok. Now you are comparing whether they are equal or they are unequal, is it ok. So, if you say equal, then you say it is the null hypothesis, if you say that at least one is different from all others; that means, this is the alternative hypothesis H 1, is it ok. So, for a chosen level of significance say alpha say 0.01 or 0.05.

The critical value of f obtained from the table is denoted by this. So, if the if the computed test statistics the null hypothesis is rejected and we conclude that the treatment means are not all equal at the chosen level of significance, is it ok, this computational procedure is known as analysis of variance the typical format of an ANOVA table for this

experiment is shown. So, I have already mentioned that what is ultimately now you need to analyze the variance.

(Refer Slide Time: 16:58)



So, ah, so, the analysis of variance is very very important. So, you create the ANOVA table. So, what is the typical format of an ANOVA table for a completely balanced randomized say the design.

So, the; so, the source of variations basically the treatments or the error and the total degrees of freedom is p minus 1 error is p into r minus 1. So, the total degrees of freedom is r p minus 1, is it ok. So, if you add this you will get r p minus 1 sum of squares is SS TR, SS E, this is SS T, fine, then the mean square is SS TR divided by the degrees of freedom.

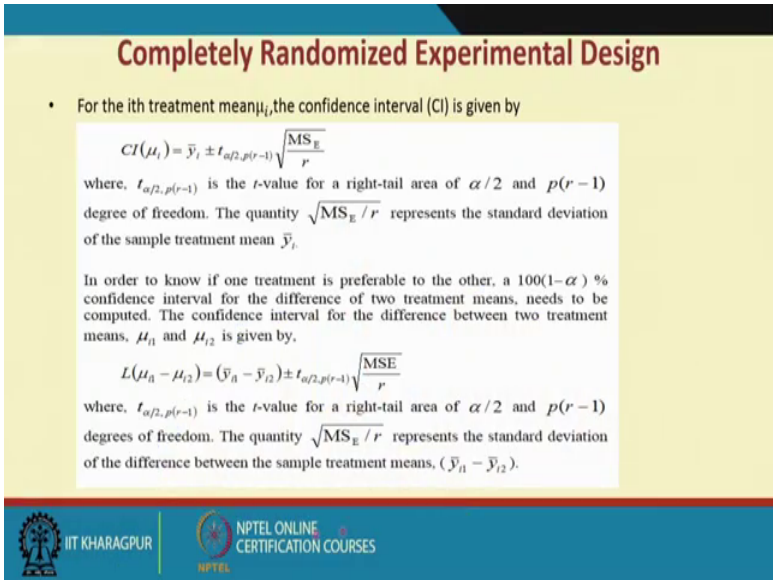
Similarly, mean square error is a sum of squares of error divided by the degrees of freedom and then F statistics is this is MS TR, ok, the treatments divided by MS E, is it ok. So, this is the F statistics, you compute if the null hypothesis is not rejected, there is no significant difference among the treatment means if the means, it means that the response variable is not affected by any particular treatment and hence the most cost effective treatment should be selected. So, that is your conclusion, is it ok.

So, what do you try to do; that means, when you look at the ANOVA table, is it ok, as well as the results from your hypothesis testing. So, if you conclude like this; that means,

what do you conclude; that means, that the most cost effective treatment, I need to select you have already considered that the p number of treatments, is it ok. So, which one is the most cost effective if the null hypothesis is rejected management might be interested in knowing which treatments are preferable among p lambda of say the treatments each sample treatment mean is a point estimate of the population treatment mean fine.

If the interval estimates are desired a hundred into 1 minus alpha percentage confidence interval for the treatment mean can be computed I have we have been telling all the time we have been preferring you know the interval estimate to say the point estimate.

(Refer Slide Time: 19:30)



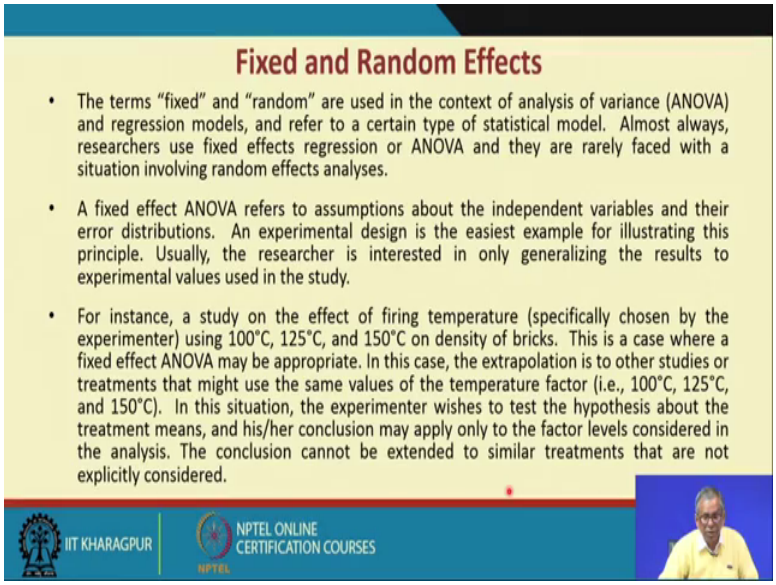
So, here how do you get this the interval estimates as a confidence interval for mu i that is already y i bar you have computed and this is the, I know the assuming you know the t statistic. So, the t alpha by 2 that is alpha by 2 and this is the degrees of freedom p into r minus 1 into root over say ms of error mean square error and this is r where t alpha by 2 pr minus 1 is the t value for a right tail area alpha by 2 and p minus 1 degrees of freedom, is it ok.

So, the quantity is this one that is root over MS E by r represents the standard deviation of the sample treatment means y i dot bar in order to know if one treatment is preferable this is the standard procedure I am explaining for all you know design of experimental ah. So, the methods we use or we employ in order to know if one treatment is preferable to the other a 100 into 1 minus alpha percentage confidence interval or the difference of

the 2 treatment means needs to be computed the confidence interval for the difference between two treatment means is this one mu i 1 and mu i 2 is it ok. So, it is given by this one, is it ok.

So, these expressions you have, right, you just take the difference on where this value is the t value for a right tail area of alpha by 2 into p into 1 minus degrees of freedom and this quantity that is root over MS E by r represents the standard deviation of the difference between the sample treatment means is this one, is it ok.

(Refer Slide Time: 21:29)



So, this you need to consider there will be fixed and random effects as we have been telling you that the fixed and random effects are used in the context of analysis of variance and the regression models essentially you assume a linear regression model initially and this fixed and random effects referred to certain types of statistical model almost always researchers use fixed effects regression or ANOVA and they are rarely faced with the situation involving random effects analysis.

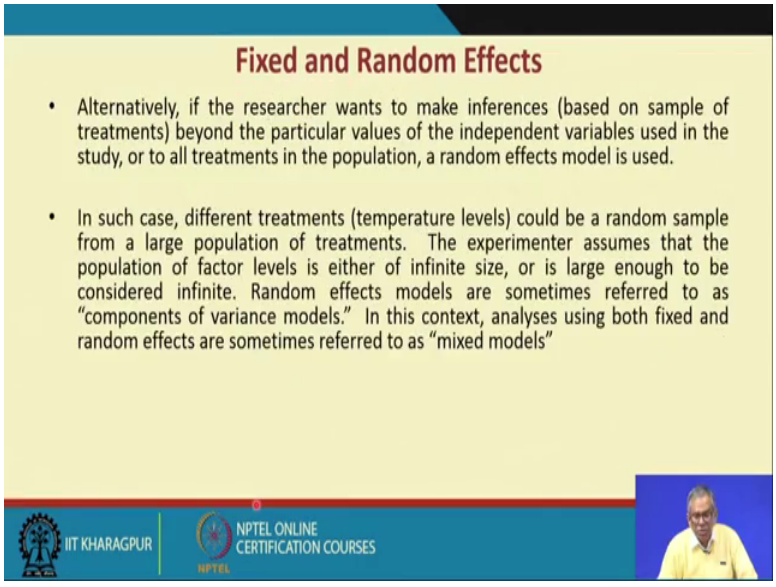
But in certain cases, you may have to opt for in a fixed effect ANOVA refers to it refers to assumptions about the independent variables and their error distributions an experimental design is the easiest example for illustrating these principles usually the researcher is interested in only generalizing the results to experimental values used in the study. So, generalization is your goal for instance a study on the effect of firing temperatures specifically chosen by the experimenter here in this case, firing temperature

using 100 degree Celsius, 125 degree Celsius and 150 degree Celsius on density of bricks, this is the case where a fixed effect ANOVA may be appropriate.

In this case the extrapolation is to other studies or treatments that might use the same values of the temperature factor; that means, when the extrapolation is required maybe you may go for say the random effects and analysis. So, the 100 degree Celsius, 125 degrees Celsius and 150 degree Celsius these three temperature levels, you consider . Now in this case, the experimenter wishes to test the hypothesis about the treatment means and his or her conclusion may apply only to the factor levels considered in the analysis, is it ok; that means, you are saying that this system will continue only these three levels I need to consider.

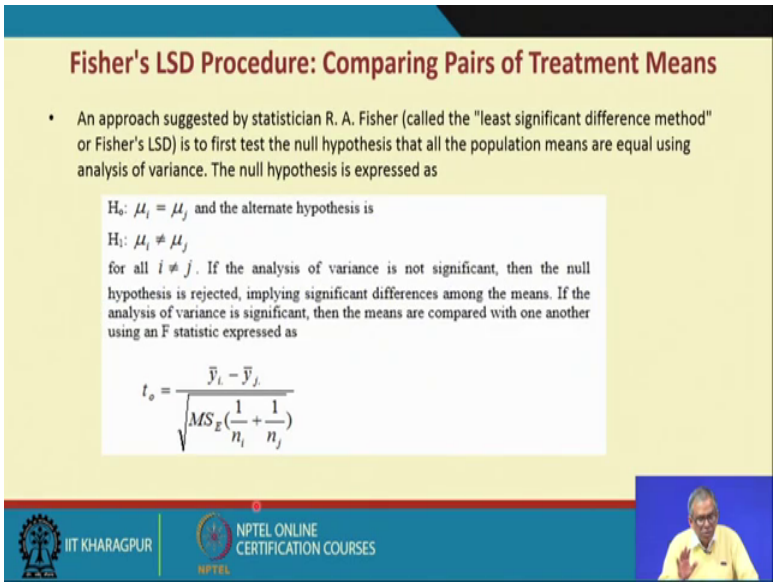
The conclusions cannot be extended to similar treatments that are not explicitly considered, is it ok. So, then in that case you have to go for the random effects model, is it ok.

(Refer Slide Time: 23:59)



So, this is; so, here we are referring to fixed effects model, is it.

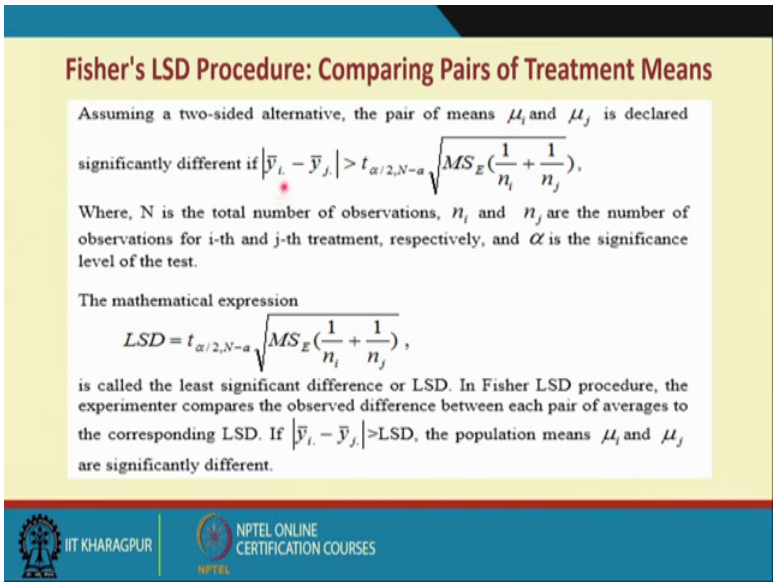
(Refer Slide Time: 24:05)



So, just ah; so, the one particular case we are referring to that is the Fisher's LSD procedure comparing pairs of treatment means an approach suggested by statistician ra Fisher called the least significant difference method or LSD least significant difference method other it is also referred to as the Fisher's LSD is to first test the null hypothesis that all the population means are equal using analysis of variance.

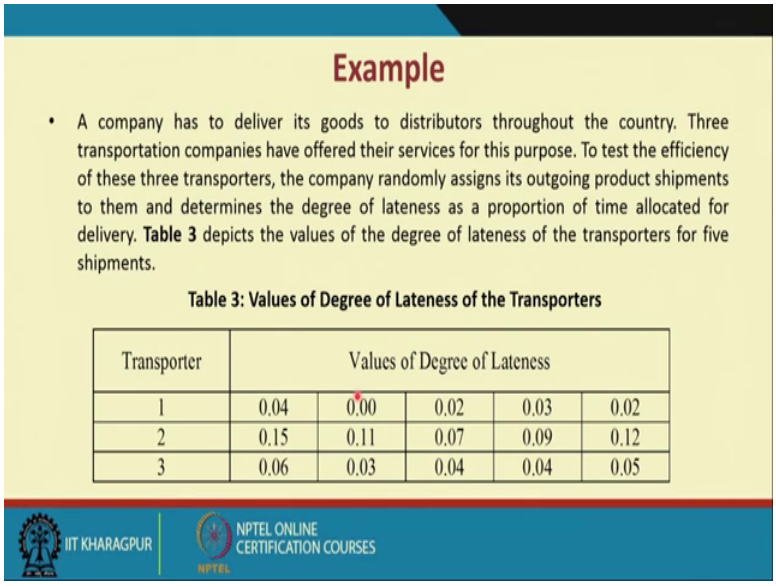
We have already referred to ANOVA the analysis of variance table. So, the null hypothesis is expressed as this one; mu i equals to mu j and the alternative hypothesis that is H 1 that is mu i is not equals to mu j for all I not equals to z, if the analysis of variance is not significant then the null hypothesis is rejected implying significant differences among the means if the analysis of variance is significant ok, then the means are compared with one another with an F statistics expressed as this one. So, this is your F statistics and that statistics is given by this one y i dot bar minus y j dot bar divided by root over a mean square error into 1 upon ni plus 1 upon nj, is it ok.

(Refer Slide Time: 25:35)



So, this is the procedures; that means, two sided alternative. So, ah; so, just refer to this particular case, right. So, you may use this particular notations and n is the total number of observations ni and nj are the number of observations for i th and j th treatment later on when we refer to a particular example. So, we made very very clear and this is the mathematical expression for LSD, is it ok. So, this is called the least significant difference or LSD right. So, many a time we do that.

(Refer Slide Time: 26:11)



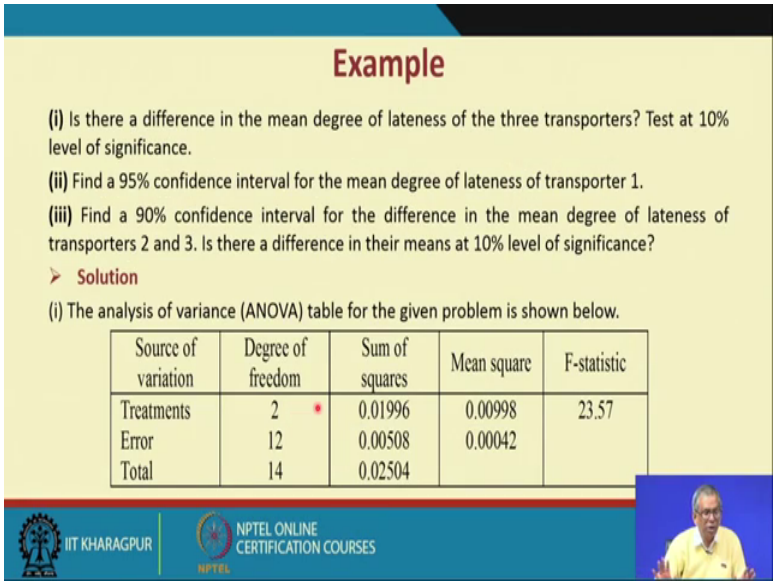
Now, before I conclude these sessions let me explain one particular example, is it related to design of experiment a company has to deliver its goods to distributors throughout the country.

Three transportation companies have offered their services for this purpose to test the efficiency of these three transporters the company randomly assigns its outgoing product shipments to them and determines the degree of left let lateness as a proportion of time allocated for delivery; that means, this is this is this becomes your the response variable, table 3 depicts the values of the degree of lateness of the transporters for five shipments. So, how many transporters you have? 1, 2, 3, is it and this is basically 1, 2, 3, 4, 5, ok. So, depicts the values of these for 5 shipments; that means, 5 replications.

So, the corresponding data you have; that means, these are the values of y; that means, response variable and what is the response variable over here that is the lateness degree of lateness. So, first whenever you say this is the response variable. So, you have to define it and you must specify how to measure them; that means, here we are assuming that there is a measure and you know whose data are required to calculate this value and this is suppose 0.04 and then this value is 0.15 which one is preferable whether 0.15 or 0.04.

So, those details you must know right. So, how to compute the value of y that also must be explicitly specified.

(Refer Slide Time: 28:15)

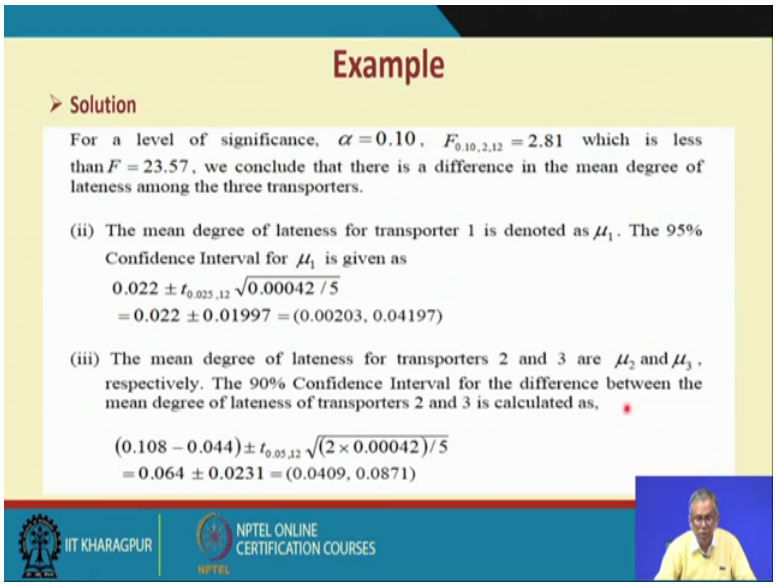


So, this is the value now the question is there a difference in the mean degree of lateness of the 3 transporters, is it ok; that means, three treatment combinations test at 10 percent level of significance find a 95 percent confidence interval for the mean degree of lateness of transporter 1 when the transporter 1, you have collected data five times, is it ok. So, you have a mean lateness right.

Now, how to determine the 95 percent confidence interval for the mean degree of lateness for transporter 1, is it ok, treatment one find a 90 percent confidence interval for the difference in the mean degree of lateness of transporters 2 and 3. So, I think is this part, already we have discussed and the formulations are already given. Now when we explain the problem; that means, you can relate this problem with those the formulations if there is there a difference in their means at 10 percent level of significance; that means, first what you are trying to do you access the performance of transporter 1 and then you next time what you try to do; that means, is there any difference in the performance between transporter 2 and transporter 3 ok.

So, you create these analysis of variance like the treatments the degrees of freedom will be p minus 1 that is two error is 12 and the total is fourteen sum of squares you apply the formula ok, you get all these values and the mean square is this one right when you divide this value by the corresponding the degrees of freedom and ultimately F statistics you get, is it ok.

(Refer Slide Time: 30:04)



So, you are applying this formula and alpha is 0.10 corresponding a value you get like this 2.81 which is less than F equals to 23.57, we conclude 5 7, we conclude that there is a difference in the mean degrees of lateness among the three transporters. So, that is your conclusions.

So, the mean degree of lateness for transporter 1 is denoted by mu 1 the 95 percent confidence interval you calculate by applying the formula. So, this is the confidence interval the mean degree of lateness for transporters 2 and 3 are mu 2 and mu 3, these notations we use the 90 percent confidence interval, again you compute for the difference between the mean degree of lateness of transporters transporter 2 and transporter 3. So, how do you calculate this we apply the formula. So, this is the confidence interval that is 0.0409 and points 2.0871 ok.

(Refer Slide Time: 31:05)

So, this is the typical example please go through it and I hope that you have understood that how important these the design of experiment is and in subsequent a lecture sessions as a part of assignments, I will give you several examples on the on the design of experimental or design of experiment methods, is it ok, there are many methods only one particular method. Till now, we have discussed and please go through these numerical problems already discussed ok. So, you refer to this particular textbook, there are many other examples we will be dealing with in subsequent lecture sessions.