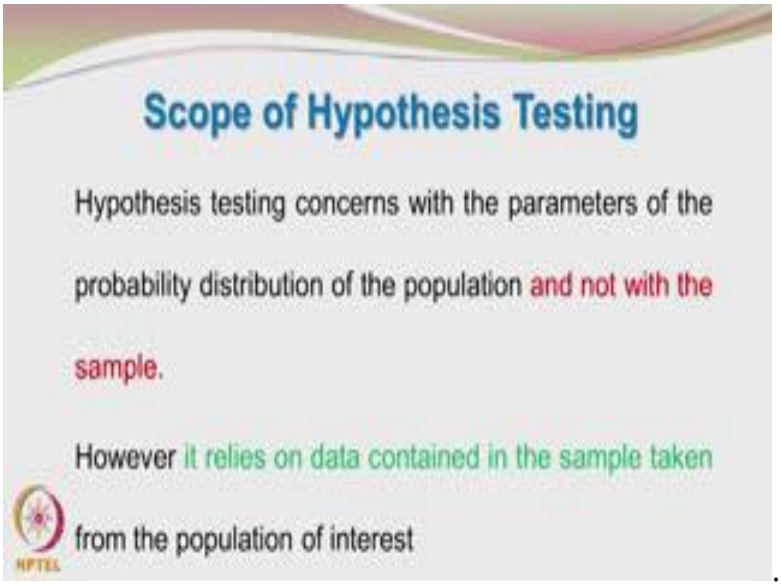
Introduction to Research
Prof. Kannan
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 03
Design of Experiments

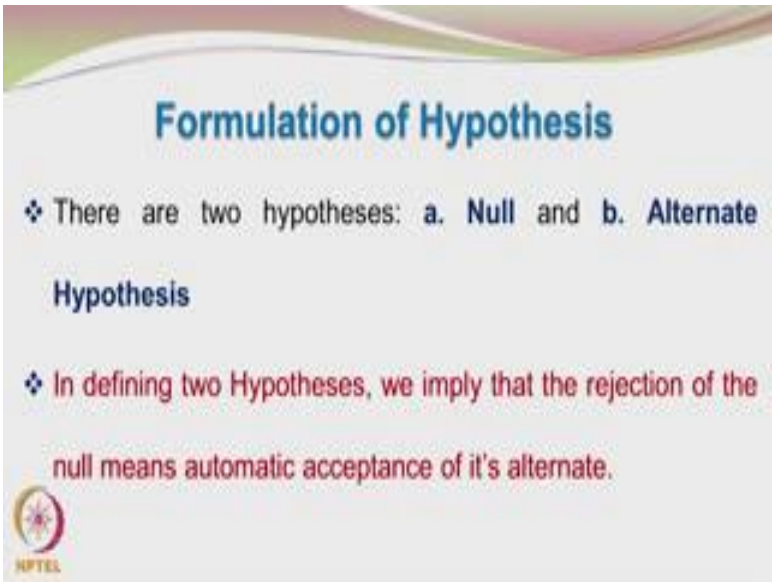
Now, we will be coming to Hypothesis Testing, which is another application involving these statistical concepts.

(Refer Slide Time: 00:15)



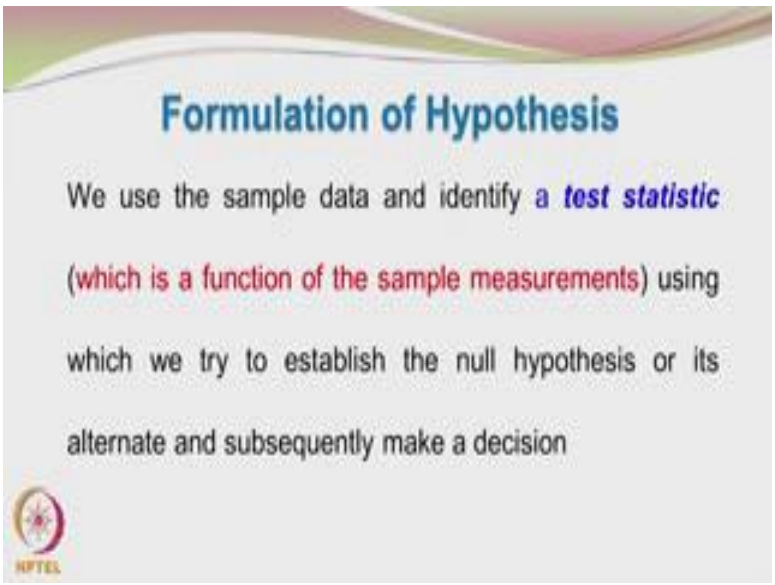
This is very interesting and plays an important role, even if a bit understated in design of experiments. Nobody really looks at design of experiments as a hypothesis testing, but this is a very important part of design of experiments. Here what you do is you specify a hypothesis or you postulate an hypothesis, and use the data contained in the sample to see whether your hypothesis is adequately supported. We are using the hypothesis testing to make decisions on the population; we are not making decisions on the sample.

(Refer Slide Time: 01:14)



There are two hypotheses: one is the null hypothesis and the other is the alternate hypothesis. As the name implies the alternate hypothesis is contradicting the null hypothesis. If the null hypothesis says that something is working, the alternate hypothesis will say that is not working. If the null hypothesis says the person is innocent the alternate hypothesis says that the person is not innocent. So, when defining the two hypotheses we imply that the rejection of the null hypothesis means automatic acceptance of its alternate.

(Refer Slide Time: 01:49)



Okay so, we use the sample data and identify a test statistic, which is a function of the sample measurements, using which we try to establish the null hypothesis or its alternate and subsequently make a decision okay. So, what we do is, we take the sample, and then extract a sample statistic from it; it may be the variance of the sample or it may be the mean of the sample - these are the two more common ones. And using this estimate, we try to infer about the population parameters. If we are using the sample mean, then we are trying to infer about the population mean; if you are using the sample variance, we are trying to infer about the population variance. So, the decision making is always associated with the errors; nobody can really say that all their decisions have been completely correct. So, we have to see what are the possible errors in decision making.

(Refer Slide Time: 02:58)

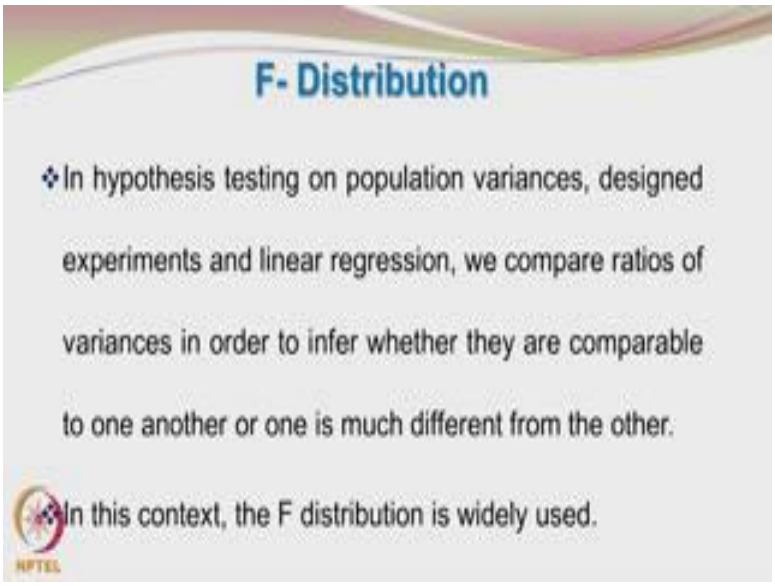
So, let us look at this table. Here the columns are headed as a H naught is true and H naught is false. The null hypothesis is true; null hypothesis is false. And the statistical decision may be either do not reject H naught; we are a bit careful, we don't accept H naught, we instead say that do not reject H naught. And the next one is a bit more unambiguous - reject H naught. Okay so, if H naught is true and that the decision made is do not reject H naught, obviously, you have made a correct decision. But when H naught is false actually, and you have made a decision not to reject H naught, a less serious type II error has been made okay. When you reject H naught, when H naught is actually true, then you have made a more serious type I error. And again, when H naught is false and you have rejected H naught, you have definitely made a correct decision.

Why is a type I error more severe? I can give a very simple example or may be a couple of examples. If the null hypothesis says that person accused of a crime is innocent, and the court rejects that hypothesis H naught, and instead says the person is actually guilty, and convicts him, then a wrong decision has been made, and an innocent person has been punished. So, a type I error is said to be made. On the other hand, if the person is really guilty, but if the court exonerates him, then the guilty person is getting away scot free, and type II error is supposed to be made. Perhaps you may recall that even though many guilty people may escape punishment not a single innocent person should be wrongly punished.

Another example which I can think of, from an industrial point of view, is a company is having a well established process, and a newcomer to the an industry, perhaps a new recruit or an old hand from another industry, comes and says this process is not really that good, I can improve upon the process. The management is skeptical and says, look we have been using this process for the last 20 years so on, and it has been working fine without any problem, and we are making profits. So, when something is working why tinker with it. So, the null hypothesis would be the proposed process is not a really great improvement on the existing process and the alternate hypothesis would be definitely the proposed process is a improvement over the existing process.

So, if you make a type II error, then even though the original or the existing process is not that good, you are not losing that much; anyway, you are back to where you are previously, and so you are still making profits, and the plant is running. So, a less serious error has been made. On the other hand, if you say that the existing process is bad, it's not good and the new process is better, and that decision is wrong, that means, you are throwing away an actually good process, and then investing lot of time, money, resources, and man power in bringing up a new process, which does not really give any great benefit over the existing process. So, then we say that a type I error has been made. So, this kind of decision making you will employee in design of experiments, for example, you may have several variables influencing your process, upfront you start saying that none of the factors are really significant in the experiment. And then, based on the information provided by the experimental data, you do this hypothesis testing, and then you can conclude if some of the parameters alone are significant and the rest of them are not significant.

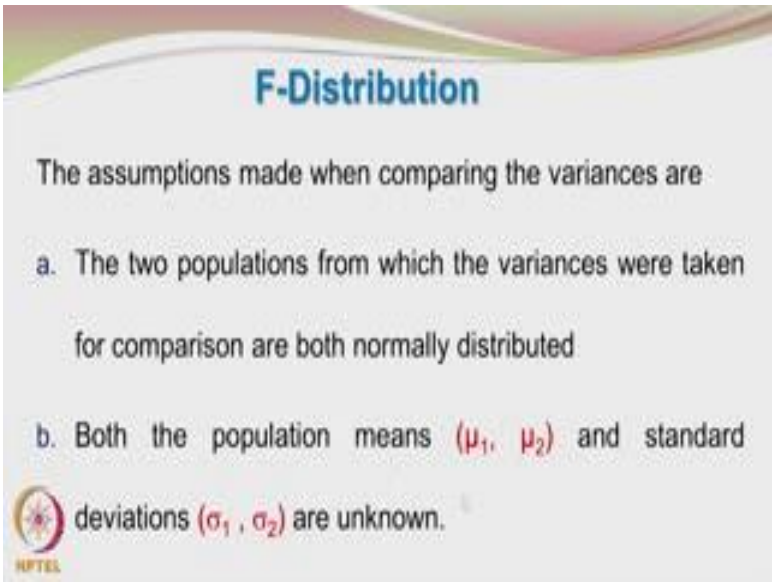
(Refer Slide Time: 07:35)



Now, we come to another important distribution called as the F distribution. The chi square distribution is not really immediately used in design of experiments, but the F distribution finds direct application, and the F distribution is similar to the chi square distribution, and it is based on that. So, when you do hypothesis testing on population variances designed experiments and linear regression, we compare ratios of variances in order to infer whether they are comparable to one another or one is much different from the other. So, when you want look at the variability of the data, the variability can be caused by variation in the factor levels or the variability may be simply caused by random effects on which you have no control of. So, when you claim that your experimental factors which you are controlling are impacting the response of the experiment significantly, then you have to prove that this variability caused by these so-called important factors is much higher than the variability caused by noise or random error.

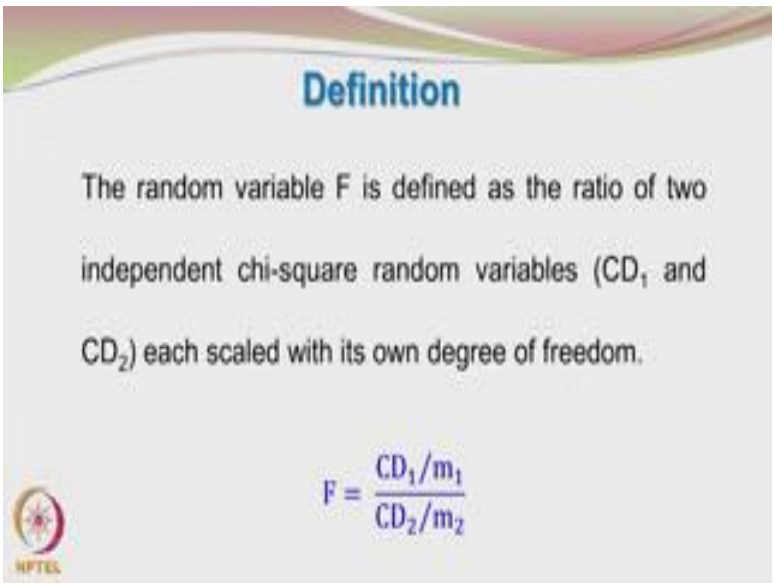
If the variability shown by changing the factors is comparable to the variability due to noise or random errors, then you cannot say that this factor is really making an impact. When you change the factor may be the random effects are causing the experimental response to deviate considerably. So, you have to compare the variability due to the change in factor with the variability because of the prevailing a random error sources, and for this comparison of variability, the F distribution is very useful. So, let us a look at the basics of the F distribution.

(Refer Slide Time: 09:40)



So, again, we start off with the set of assumptions: the two populations from which the variances were taken for comparison are both normally distributed. We are comparing two variances and we assume that both of them are coming from populations, which are normally distributed. And we also do not know the population parameters namely the means and variances.

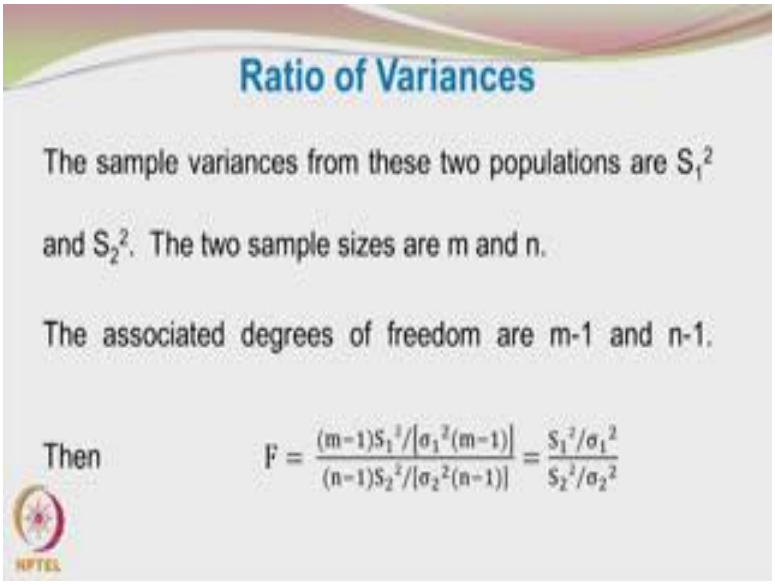
(Refer Slide Time: 10:02)



So, how do you define the F random variable? The F random variable is defined in terms of the chi square distributions. It is simply defined as the ratio of the independent chi

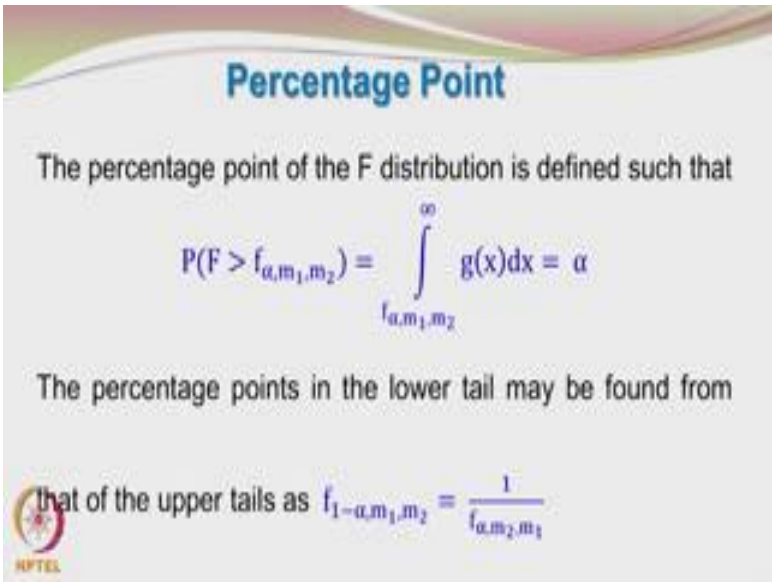
square random variables. Let's call them C D 1 and C D 2. Each of which is scaled by its own degree of freedom okay. So, we take the first chi square distribution, and then we divide it by m 1; m 1 representing the degrees of freedom for the first chi square distribution Then, we take the second chi square distribution and then divide it by m 2, and m 2 representing the degrees of freedom for the second chi square distribution.

(Refer Slide Time: 10:46)



So, when you expand the chi square distribution, you see that C D 1 is given by m minus 1 S 1 squared by sigma 1 squared that is the definition for the chi square distribution one and then when you divide that by m minus 1, the m minus 1, the numerator and the m minus 1 the denominator gets canceled, and then you get S 1 squared by sigma 1 squared. Similarly, when you do the same thing for the second chi square distribution, you will get S 2 squared by sigma 2 squared. So the F distribution in fact becomes S one squared by sigma one squared divided by S 2 squared sigma 2 squared. S 1 square and S 2 squares are the sample variances of the first and second samples respectively, and sigma 1 and sigma 2 are the standard deviations of the first and second populations respectively.

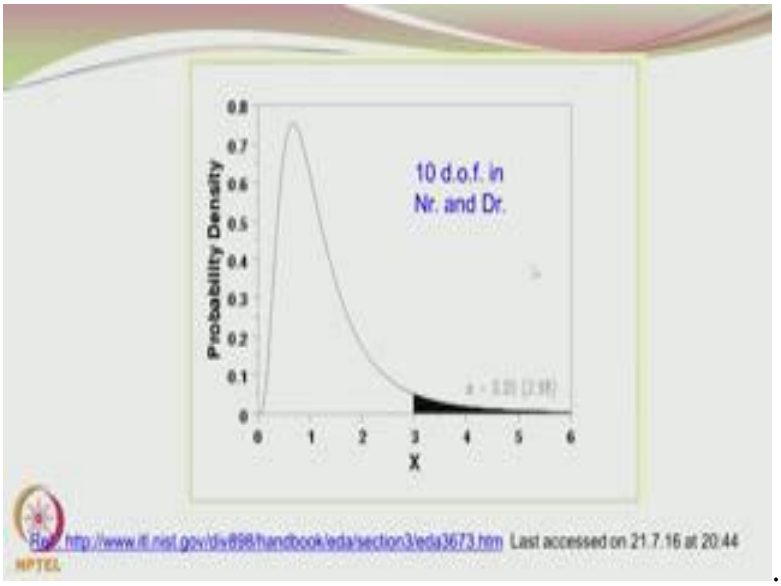
(Refer Slide Time: 11:34)



So, we get back to our percentage point. We saw the percentage point cropping up in a normal distribution and in chi square distribution; it's not a surprise that it is now coming in the F distribution. The F distribution is slightly different from the other two distributions we saw previously, because it has two degrees of freedom. One called as the numerator degrees of freedom, so m minus 1 is the numerator degrees of freedom, and n minus one is the denominator degrees of freedom. It is convenient to represent the numerator degrees of freedom first followed by the denominator degrees of freedom. So, then, we have probability of F greater than alpha m 1 m 2, and f is the random variable, f random variable and f alpha m 1 m 2 is a particular value. So, we have to identify f alpha, m 1, m 2 such that we get a probability of alpha.

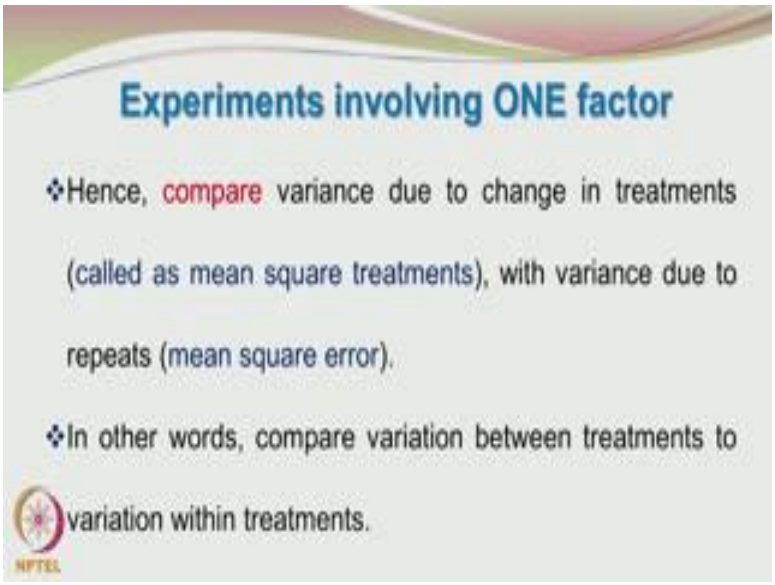
So, going back to the mathematics g of x represents the mathematical form of the F distribution, we will not get into it right now. And so, we have the mathematical form here; we integrate it between the limits f alpha m 1 comma m 2 to infinity, and we have to identify what is the value of f alpha, m 1, m 2 such that the area under the curve is equal to alpha - the required probability. So, there is another interesting mathematical manipulation here which I will leave you to go through.

(Refer Slide Time: 13:04)



Now, let us look at the F distribution. For this distribution, we have to specify the numerator and denominator degrees of freedom; in this particular case, it is actually 10. Alpha we set it 2.05; that is usually very popular; for example, you know the 95 percent confidence intervals are based upon a value of alpha of 0.05. So, the f alpha m 1 m 2 is identified here to be 2.98. So, when you look at the area under the curve beyond the F the value of 2.98, you will find area under the curve to be 0.05.

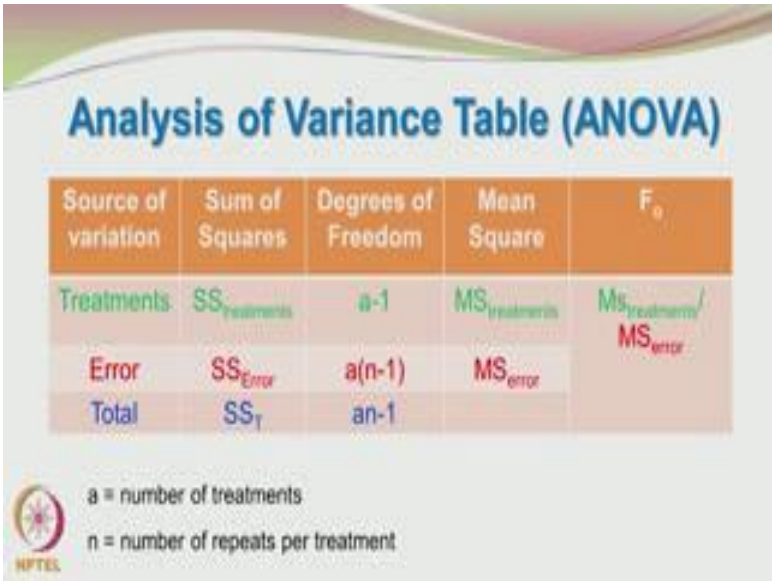
(Refer Slide Time: 13:44)



So, now let us get into simple experiments. These are very scarce, but this is a very good basis for illustrating some key concepts. So, what we do here is we are looking at experiments involving only one factor. So, what we look at here the comparison of variance due to change in treatments with the variance due to repeats. So, when you do repeats, you find variability in the data. If all the factors are kept at the same values, then the variability can be due to only random factors. We assume that all the equipment, instruments are all working fine, there are no systematic errors. So, the error is only from the random components and to identify the error from the random components, you have to do the repeats of the experiments. So, the more number of repeats you do, the more or better idea you have about the experimental error.

You may be asking - what is meant by this treatment whereas this treatment is it a medical treatment or what kind of treatment it is? It is just a classical term, which has survived over the years. So, we just call it as the levels of the factors. So, what we have to do is remember that we are considering only one factor and they may be having different levels or different treatments. For example, it can be fertilizer A can be one treatment to the land, and fertilizer B can be the second treatment to the land, and so on. So, essentially, as I said earlier we are comparing the variation between treatments to variation within the treatments. Good !

(Refer Slide Time: 15:49)



Now, we will be coming to a very popular and very important table called as the ANOVA table. ANOVA stands for analysis of variance. So, you can see that finally, everything is boiling down to the analysis of the variability. So, this table has several columns and the first column is the source of variation. So, it can be treatments, and then, it can be the error. As I said earlier, variability from treatments and the variability from error, and this column contains a sum of squares of treatments and this is sum of squares of error; I am not getting into all the details. So, the variability due to treatments is represented in terms of the sum of squares of the treatments.

Normally, we don't take the absolute variability or the actual variability because some variability with the reference to the defined average value can be positive and some variability can be negative. So, instead of taking the variability as it is, we take the square of them. Similarly, we do the sum of squares of the error. So, I am not getting into how these sums of squares were actually evaluated, but I am just indicating that these are measures of variability in a squared form. A is the number of levels for the particular factor or the so-called number of treatments.

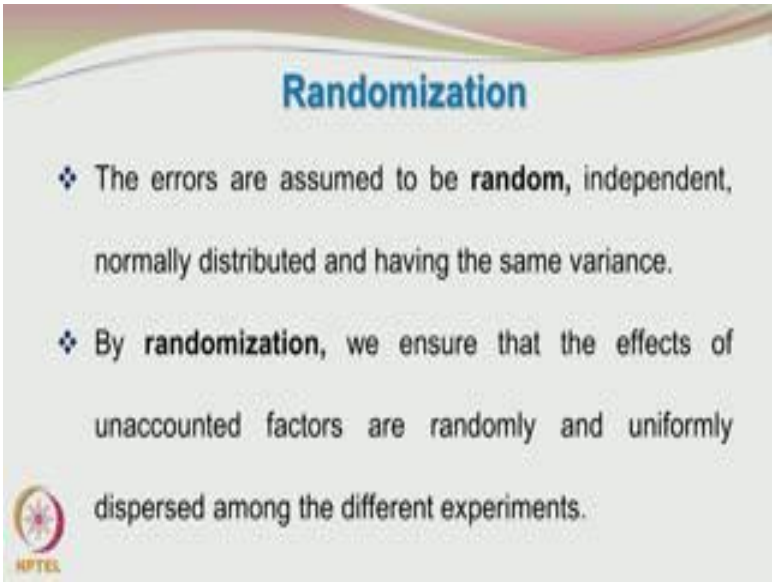
So, what you have to remember is the degrees of a freedom you have to keep in mind, so that rather than comparing the sum of squares of treatment with the sum of squares of error, because the sample size for the treatments and the sample size for the error may be considerably different. So, to put them on a same basis, you scale each of the sum of squares by the appropriate degrees of freedom. So, you divide by a minus 1, where a is number of treatments, and for error - to get the scaled error squared, we divide sum of square of error by a into n minus 1; a into n minus 1 is the degrees of freedom for the error component. So, when you divide the sum of squares by the respective degrees of freedom we get the mean square treatment and the mean square error.

And what we are doing here is comparing the two mean squares, mean square treatments by mean square error. Let us see if I can correct this small typo here. There should be capital S, okay that takes cares of it and so here we go.

If you look at the mean square treatment it is similar to the variance. We know the variance is the sum of square of the deviations. Here also when you calculate the sum of squares of the treatments, you define a suitable average, and then, you find the deviation of the response with the respected average, and then square them, and each deviation is

squared, and then all the deviations are added to get the sum of squares of the treatments. And when you divide with the degrees of freedom we get the mean square treatment. So, this is a kind of variance okay. And then, similarly, you also have the error term, where you have the mean square error, where the sums of squares of errors are divided by the degrees of freedom. So, you compare mean square treatments by mean square error to get the F value and based on the probability values associated with this F value, you can make the adequate conclusions.

(Refer Slide Time: 19:54)



There is another important thing you have to remember in mind or keep in mind when you do experiments is the randomization of your experimental runs. Normal tendency, in any experimenter - a new experimenter, let us put it like that is to systematically vary the factor settings. He may or she may go from a low factor setting to a high factor setting for one variable, and then take the next variable, and the again do in a very systematic manner that shows the organized mind of the experimenter, but in reality, it is better to do the experiments in a randomized fashion.

What I mean by randomization is to mix up the order of the runs - why should we do it? Suppose we are looking at a photocopier, and we are looking at three different toners used in the photocopier, let us say that you put for each toner you do three trials, and look at the picture quality, and then decide which toner is better, and the which toner is not so good, and so on. So, in this very illustrative example, what you can do is do toner

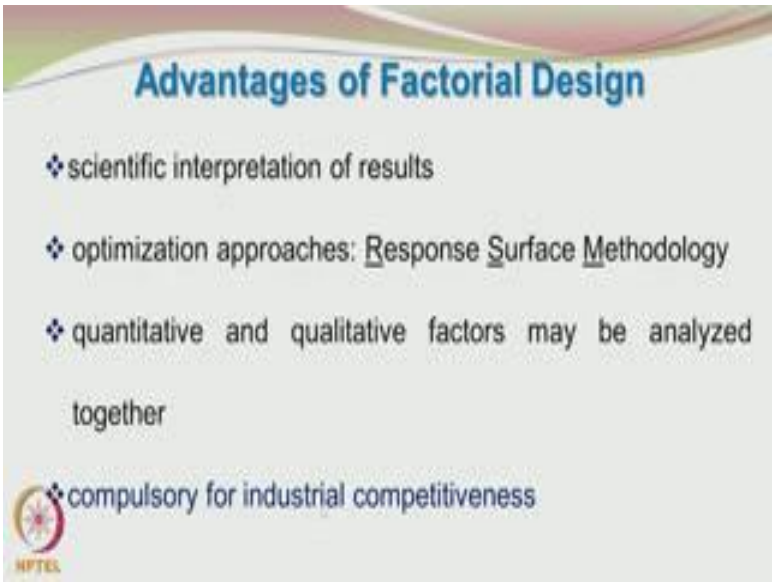
A for the first 10 minutes - 3 copies with the first toner, and then 3 copies with the second toner, and 3 copies with the third toner. When you do that, the photocopier may be getting heated up. So, by the time you come to the third toner, the external heating of the photocopier may lead to poor quality copies for the third toner okay. So, this will lead probably to the wrong conclusion that toner C is not good. What you should have actually done is the heating of the photocopier is inevitable and probably not controllable, and you have to finish all these evaluations within 30 minutes.

So, what you can do is you can mix up the order of the runs, you can put toner A or toner 1 first, then followed by toner 2, then by toner 3, and then 2, 3, 1, 1, 3, 2 like that, if you do it in randomized fashion, then the higher temperature effect is also reflected or indirectly influencing the performances of all the three toners than a specific toner. So, by randomization any unaccepted effects is evenly spread or distributed cross all the factor settings. So, randomization is a good idea and I request that when you are doing experiments, you please randomize your order of the runs.

(Refer Slide Time: 22:52)

So, randomization is implemented by running the designed experiments in a random fashion and the allocation of the experimental material to the different runs is also done in a random fashion.

(Refer Slide Time: 23:01)

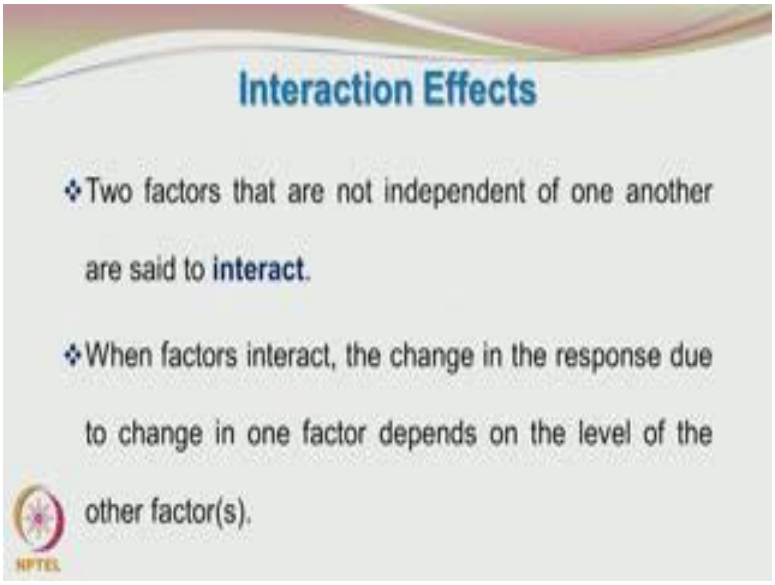


Okay now, we are coming to the design of experiments slowly and one of the important experimental designs is the factorial design; you might have heard about the 2 power n factorial design, where 2 is the number of levels and n is the number of factors. So, for example, if you are having three factors a, b, and c – temperature, pressure, and concentration, each factor you investigate at two levels - a higher level and lower level; that is why we have the number 2 okay. So, we end up actually doing 8 experiments; 2 power 3 which is equal to 8; 8 experiments are being done; and each factor is investigated at only two levels a higher level and the lower level.

So, there are many advantages of factorial design. You may think okay why should I do only at two levels, it looks better if I do it at multiple levels; you are right, but even with the two levels of many factors, you can economize your experimental program and still get valuable information. And even 2 power n becomes quite a large number when the number of factors become too large. So, the advantages of factorial design are: you can scientifically interpret your results; it enables optimization approaches like the responder phase methodology. And one another beauty of this factorial design is both qualitative and quantitative factors may be analyzed together. For example, if you are having three different types of catalysts catalyst 1, catalyst 2, catalyst 3, you don't really a consider them as any continuous entities, because you are not representing them in terms of the b e t surface area, and poor volumes, and things like that.

So, these are discrete entities. On other hand, you may also want to investigate the effect of temperature or pressure which can vary over a continuous range. So, the design of experiments involving the factorial design, you can have both the mix of discrete and continuous factors or continuously varying factors. And now, it has been realized that for industrial competitiveness factorial design is compulsory.

(Refer Slide Time: 25:17)



And again, you can do limited number of experiments to recover maximum information. The normal way of doing the experiments is to vary one factor at a time; this is not a very good strategy, because it leads to more number of runs and also it has not pick up the variability cost by interaction between the factors okay. One factor can influence the other factor; the value taken by the first factor can determine the response when you are changing the second factor. When you are doing experiments involving, let us say, two factors A and B, the interaction between A and B may also influence the outcome of the response. When you are doing experiments from one level of B and going to the other level of B, the response may depend upon the value of the first factor. Okay when you are changing B, if the value of A determines the response considerably, then the two factors are said to interact.