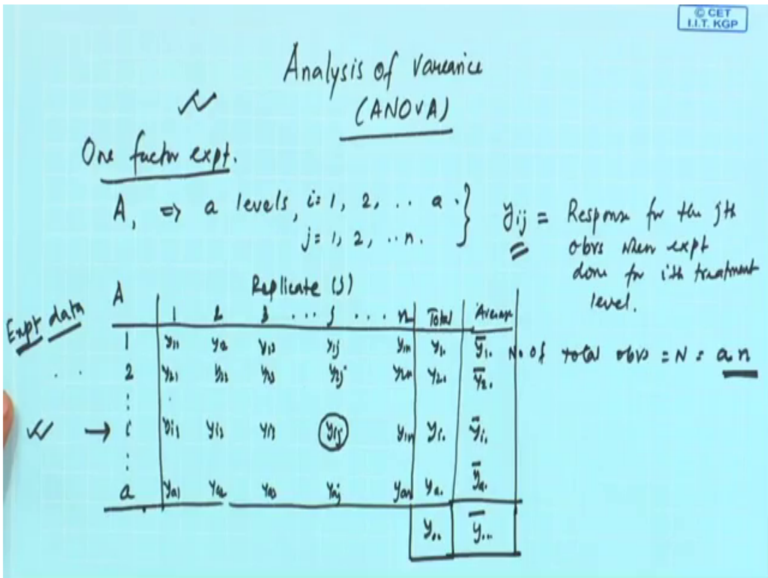
Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture – 13
Analysis of Variance (ANOVA)

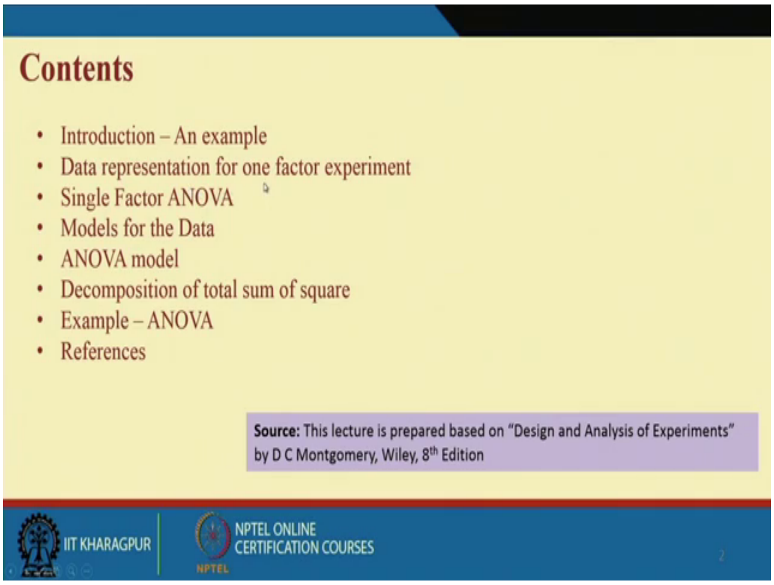
We will discuss certain parts of Analysis of Variance.

(Refer Slide Time: 00:18)



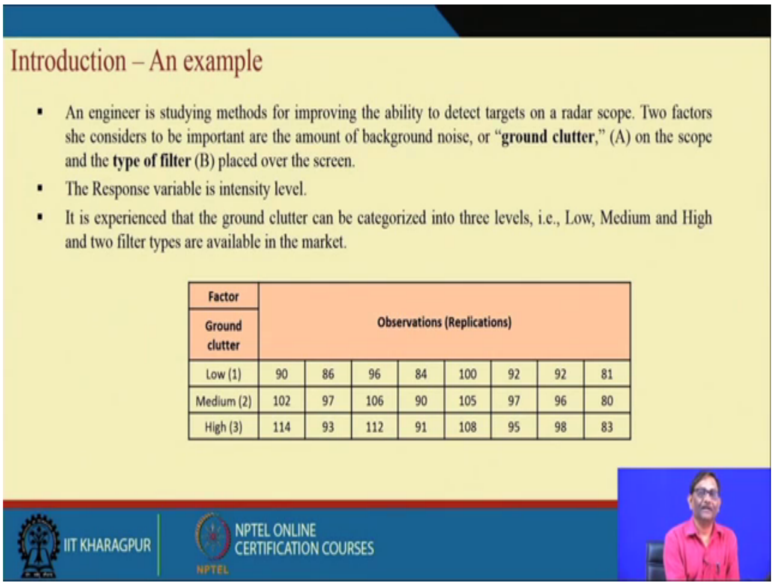
Today’s contents are first we will introduce with an example and then we will see that how one factor experimental data can be represented in tabular form, then we will discuss the partitioning of data, observations for single factor ANOVA and also the models for the data and then how to decompose the sum of squares into treatment effects, error effects and that same thing with an example with later we will explain.

(Refer Slide Time: 00:20)



So, let us start with the same example what we have seen in earlier lectures.

(Refer Slide Time: 01:20)

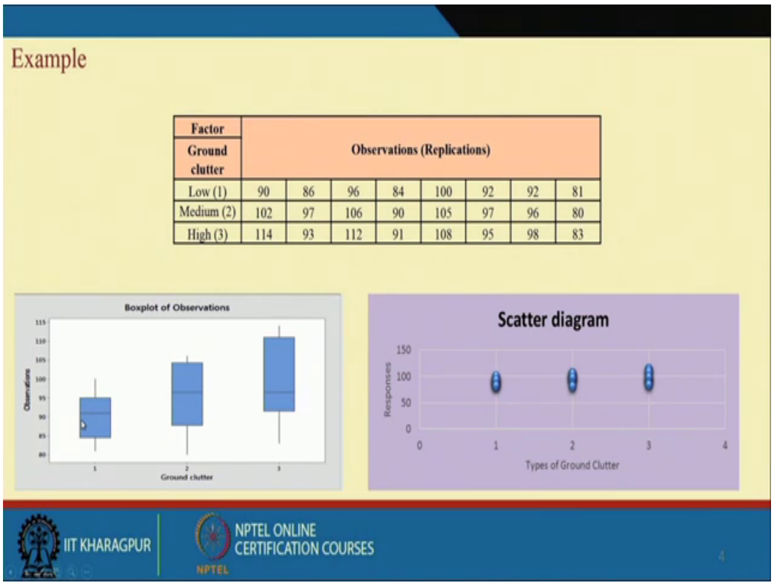


Factor / Ground clutter	Observations (Replications)							
Low (1)	90	86	96	84	100	92	92	81
Medium (2)	102	97	106	90	105	97	96	80
High (3)	114	93	112	91	108	95	98	83

Suppose the engineer is interested to study the methods of improving the ability to detect targets on a radar scope two factors considered to be important are ground clutter and type of filter response variable is the intensity level at the time of detection that will be that is what is measured. And in this example we are considering only one factor that is ground clutter which is having 3 levels low, medium and high.

For the time being like earlier in earlier lectures we assumed that this data comes from an experiment that is one factor experiment and ground clutter level at low medium and high and there are 8 observations or 8 replicates again against each of the 3 factorial settings or 3 treatment levels.

(Refer Slide Time: 02:25)



So, I also explained in earlier lectures that a use of boxplot. So, from the boxplot we wanted to know whether there is difference in means as and medians when the factor level changes from level 1 to level 2 to level 3. In fact, in this boxplot the median lines are given this horizontal lines middle lines are median. So, there can be somewhere mean here mean here and mean here.

So, from this plot it is visible that there is difference in mean response from when you observe with ground clutter level low to medium to high. But this is a visual one only just and it all it all depends on the scale you are using. So, it is, it can give you go for some kind of other analysis, but it will not clearly give you that whether the difference is statistically significant or not, and in the right hand side scatter diagram is shown. This is just to see that how the data behaves in terms of the relation between this two variables scatter plot region is full of one or to talk about the relation, but from the scatter plot again it all again depends on the response measurement that the scale of the figure. So, here also here from here we are not able to find any significant we are not able to find out much difference between the mean levels.

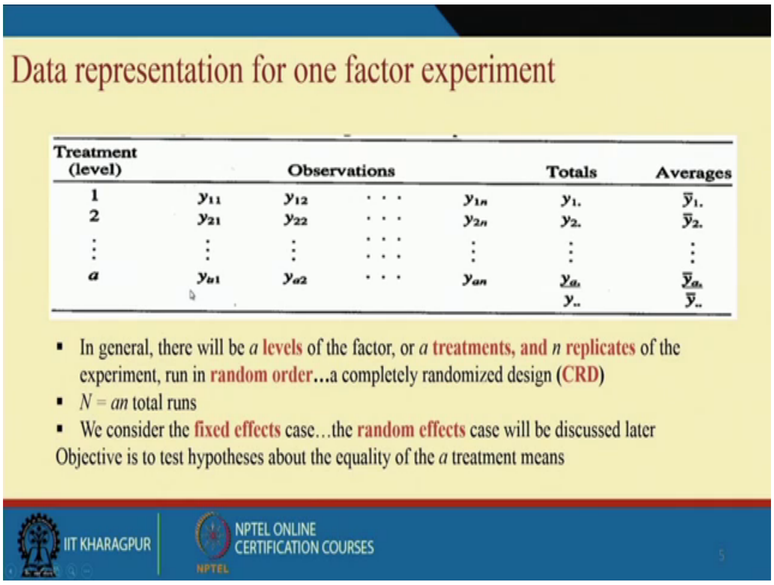
So, essentially what happened then that if there is one factor experiment let the factor is a denoted by a then this factor has a levels, let a mean level 1, level 2 like a levels. So, we say i from 1 to a and then against each level you have n replications j equal to 1 to n and each of the response responses, each of the responses now can be denoted by y ij. What does it mean? This is the response for the jth observations when experiment done for ith treatment level, ith treatment level what I mean to say your factor a is having level 1, level 2 like level i similarly suppose level a.

And there is replica replicate that denoted by j a j equal to 1 2 3 like j like n. So, then the general observation is y ij. So, ith level jth observation that is what I have written here and this is what is the representation of data when you do experiment you do experiment for one factorial experiment again which is having a levels each having n n replications then your data set will be y 11, y 12, y 13.

So, like y 1 j like y 1 n this is y 21 this is y 22 y 23 y 2 j, y 2 n then y i 1, y i 2, y i 3, y ij y in, in the same manner y a 1, y a 2, y a 3, y a j, y a n. So, that means, the number of observations total observations number of total observations if I say equal to a n then this is nothing, but a level times n a n, a level time times n, now one factor experimental data.

So, let us see the slides, you see this, this is data representation for one factor experiment the what I have given in this, this slide this is nothing, but data representation for one factor experiment you may be interested in total here you may be interested in average across ach level here. So, by total what I mean to say suppose the ith total you can write down y i dot. So, then first row total is y 1 dot y 2 dot like here y a dot, if you want to get average then you write y 1 dot average, y 2 dot average y i dot average y a dot average and then someone may be interested to know the total of all the observations y dot dot and grand average y dot dot bar. So, you know how do compute average how to make total all those things. So, this is what is there, but we will proceed further. So, I want to go to the slide.

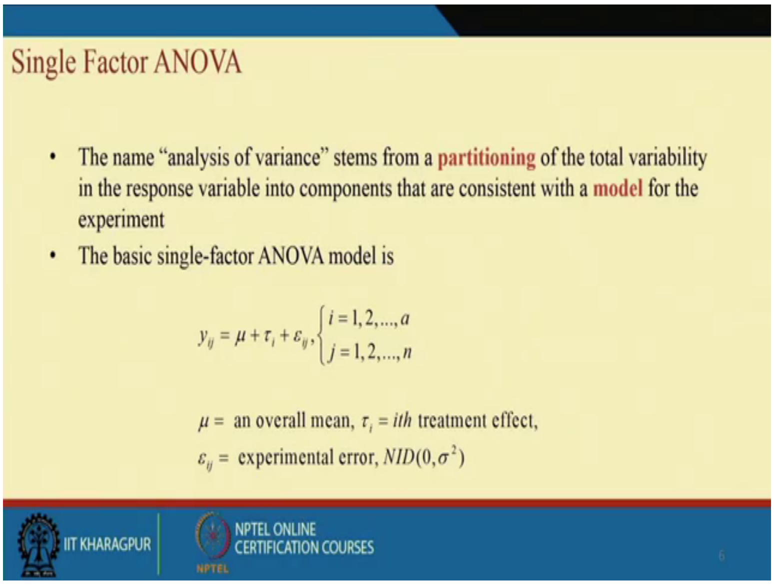
(Refer Slide Time: 09:31)



So, you see n replicates and you all know this is random, random experiment. So, how to make randomization that also we have discussed in first few lectures, this is a complete randomized design case.

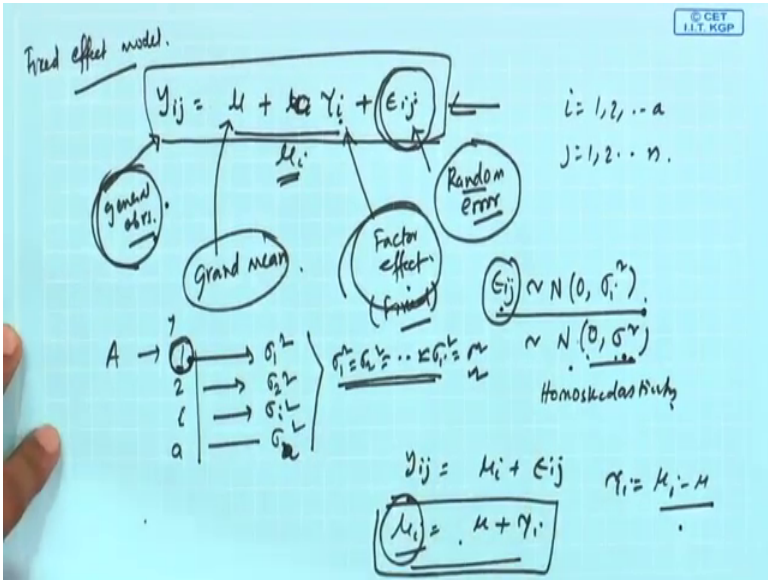
I told you n total a n total runs, and here we can consider two kinds of modules here a fixed effect it mean the treatment fixed treatment otherwise random treatment random effects. So, we will discuss later. So, why we are we have done this experiment, we have done this or we will be doing this experiment to know that whether the means across different level levels differ or not, that is what we want to test it will be a hypothesis testing case where you have more than two levels or two treatments and you want to test whether the mean response across the different treatments are equal or not. So, equality of treatment means across various treatment levels.

(Refer Slide Time: 11:01)



So, with this we will now formally define the single factor model. So, single factor model will be defined like this y ij equal to mu plus mu i mu plus of i will write like this tau i plus epsilon ij this is more general 1.

(Refer Slide Time: 11:12)



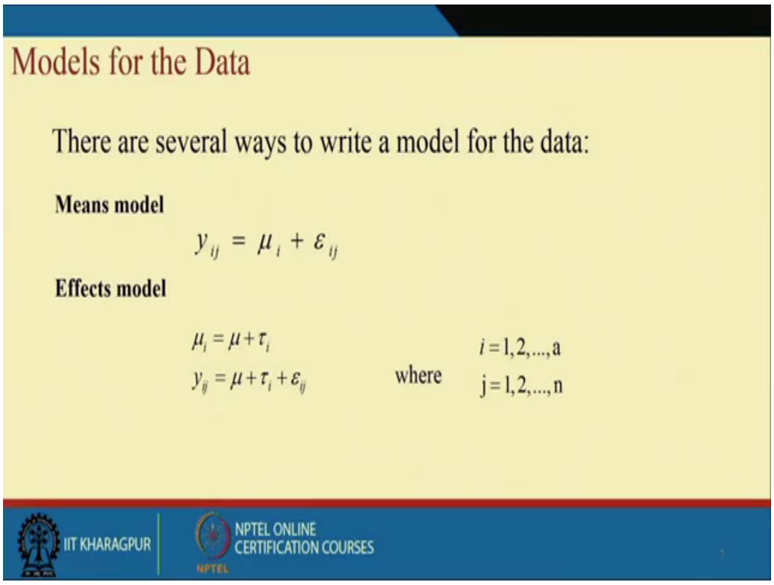
So, mu plus tau y is basically mu i that also we can write, but we will start with this mu plus tau i. So, what is the y ij? This is a general observation. What does it mean by general observation? It means the response value for the jth observation and when you expand ith experiment level 1 or on the experiment of ith level of factor a. Then what is this mu? This is known as grand mean. What is this tau i? Tau i is because you have one factor the factor effect other way we can see the treatment effects what is this portion this is the this is the random error path. So, if tau i is fixed it has a particular effect that is constant then this model will be known as fixed effect model fixed effect model.

Now, this in this diagram show what is i, i equal to 1 to a, what is j, j equal to 1 to n this is our, this is our model that is one factor ANOVA, single factor ANOVA model. So, here very important concept is this, this is the random error this random error means it has certain distribution. So, epsilon ij will be normally distributed with 0 mean and sigma i square variability, if the variance across the different levels i we have a with 1 2 like a levels if the variance of y when experiment done is done at the factor level a at level 1 then the variance is suppose sigma 1 square, then this will be sigma 2 square and this will be sigma a square or ith case this will be sigma i square then we will write like this.

Suppose we do one kind of assumptions here that that sigma 1 square equal to sigma 2 square equal to sigma i square equal to sigma square then we can write that this epsilon ij is 0 sigma square; that means, the response values y values across different levels. So, level 1 n observation level 2 another n observation it will if you have substantially large number of observe experimental data you will find that the from grater the variability across this values across this values, across this values, across this values will remains same or similar all most equal then we can assume that variability across different level of y across different level is equal that is sigma square.

Then this error term this error term follows normal distribution with mean 0 and sigma square variance this is a very very important assumption. Later on I when I talk about assumption we will come back to this, this is known as Homoskedercity as assumption skadersty city. Those things we will discuss later on.

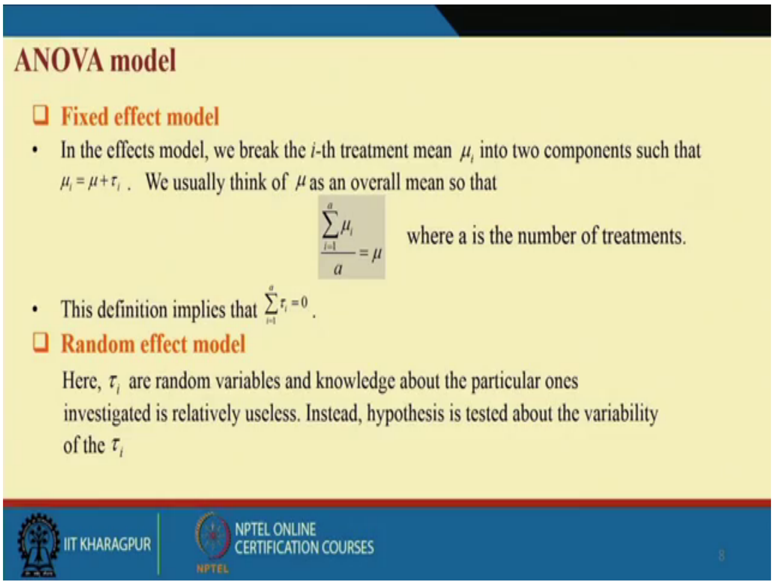
(Refer Slide Time: 15:27)



So, as I told you that mu plus mu i. So, I you can also write that y ij equal to mu y plus epsilon ij. So, then what is mu ij, mu ij equal to mu plus tau i. So, this one is known as partitioning also you see that the total observation a particular experimental run y of values what you got you partition into grand mean into fixed factor effect plus random error.

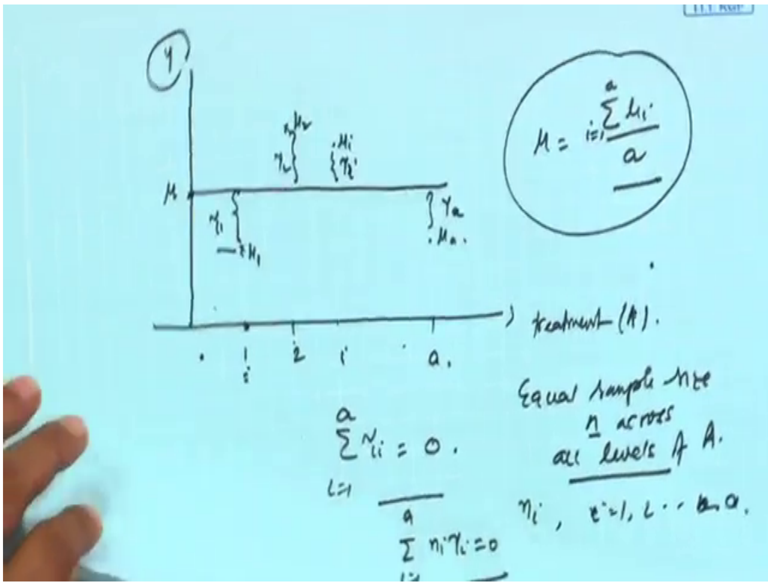
So, it is partition into 3 components overall mean, factor effect and error. So, overall mean plus factor effect is giving you the mean at each level. So, mean at this level is mu 1, mean at second level is mu 2 mean at ith level is mu i. So, then what we are saying that these tau i equal to mu tau i equal to mu i minus mu you see from here tau i equal to mu. So, mu i is mean at is of y at each level ith level mu i, if i equal to 1 this is a first level. So, this is what is our model data representation and model.

(Refer Slide Time: 16:58)



And here interestingly what is happening. Suppose I plot the data at this side y, suppose this is my mean, suppose you are talking about these are the treatments treatment our treatment is a.

(Refer Slide Time: 17:02)



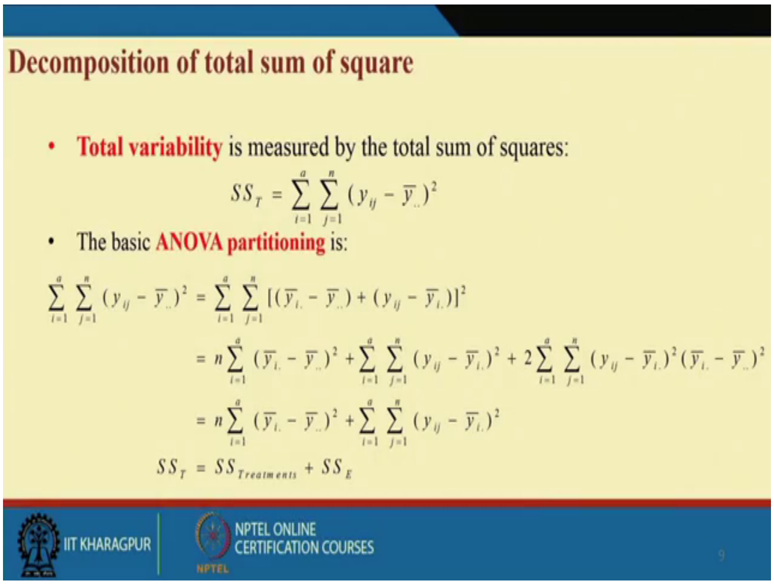
So, at different level first level second level like you it is ith level this may be a th level. So, everywhere there will be a mu. So, I first level mean maybe here would you which is mu 1, second level mean maybe here which is mu 2, ith level meal may be here which is mu i, ath level mean may be here which is mu a. So, these are the mean at of y of y at different level and then average or mean of the means is the grand mean.

Then what is the effect? Effect is the difference between those two. This is tau 1, this one is tau 2, this is tau 3, tau i and this one will be tau of a. So, it is it is nothing, but difference between individual level mean and the grand mean, again grand mean can be computed from the individual level means like grand mean is some total of mu i by a i equal to 1 to a. So, these a restriction or constraint this leads to some of tau i equal to 0, i equal to 1 to a means this plus this plus this plus this if you do you will get it 0.

Obviously we are considering equal sample size equal sample size that is n across all levels of, levels of a when you are doing experiment (Refer Time: 19:20) otherwise what will happen if the sample size varies sample size is n i i equal to 1 to 1 to a then it will be i equal to 1 to a n i tau i this will become 0. Now, this is the, this is the restriction and this help us in doing many things. Now come back to the slide.

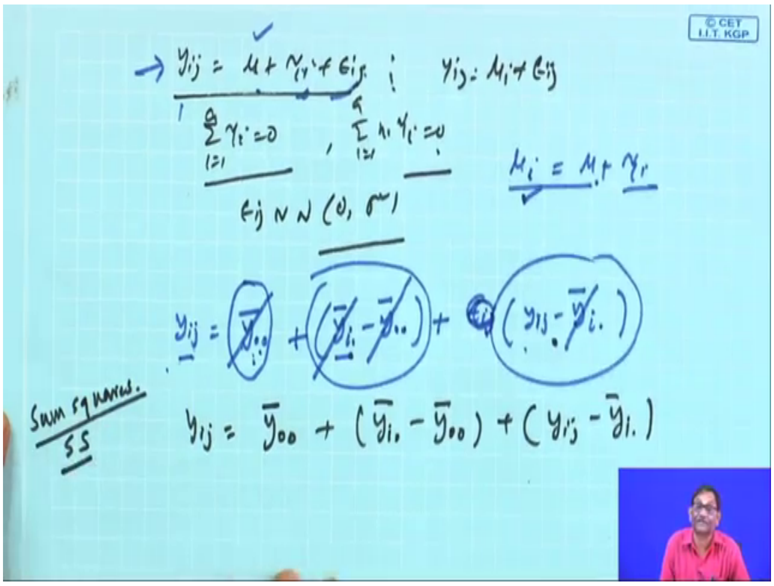
So, in fixed effect model we consider that tau i is fixed you just see that within mu as and then in the random effect model you see here tau i are random variables and knowledge about particular once investigated is relatively used less instead hypothesis tested over the variability of tau i. So, you do not have fixed value for tau i it is a random one, it is a random variable. So, instead of mean you will we will be interested in the variability of tau i, but ANOVA is analysis of variance even though it is analysis it is named as analysis of variance, but it is actually tests the equality of population means.

(Refer Slide Time: 20:38)



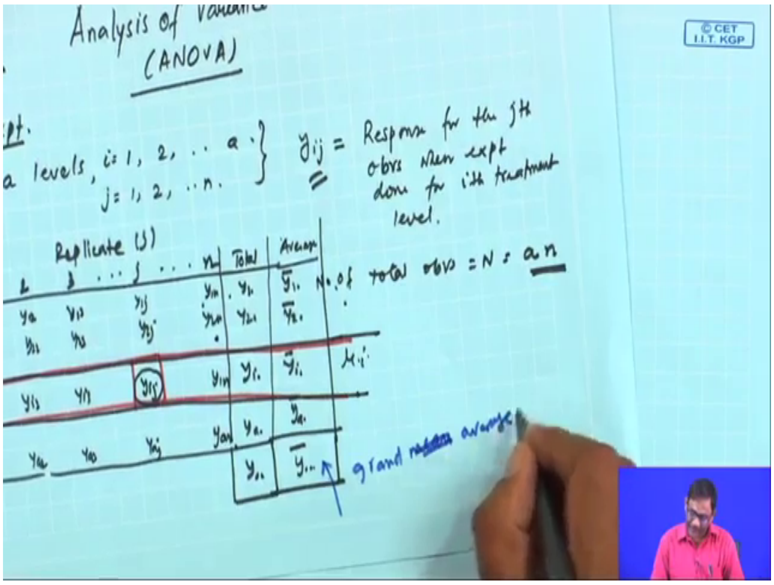
So, that mean when you say y ANOVA model you have to say y ij equal to mu plus tau i plus epsilon ij and then definitely you will be saying epsilon tau i equal to 0 i equal to 1 to a or i equal to 1 to a n i tau i equal to 0. This you can also write like this y ij equal to mu i plus epsilon ij and epsilon ij is normally distributed 0 and sigma square sigma square that is that is what is what we have learned so far.

(Refer Slide Time: 20:45)



So, now, come to the point that you have done the experiment once you have done the experiment you will get this kind of data. So, well the row of interest for us now is this row, this ith row to do some kind of and the very important observation y ij which is basically of j a. So, if you say, if you say that y ij is this from the population point of view when experiment is not done this is what is the model. So, wants experiments it done you will get some data y ij then what do you require you require the mean value that will be y dot dot bar I have given you the grand mean, this is my grand mean and these are individual means or averages. So, instead of mean you can write grand average also and individual level average.

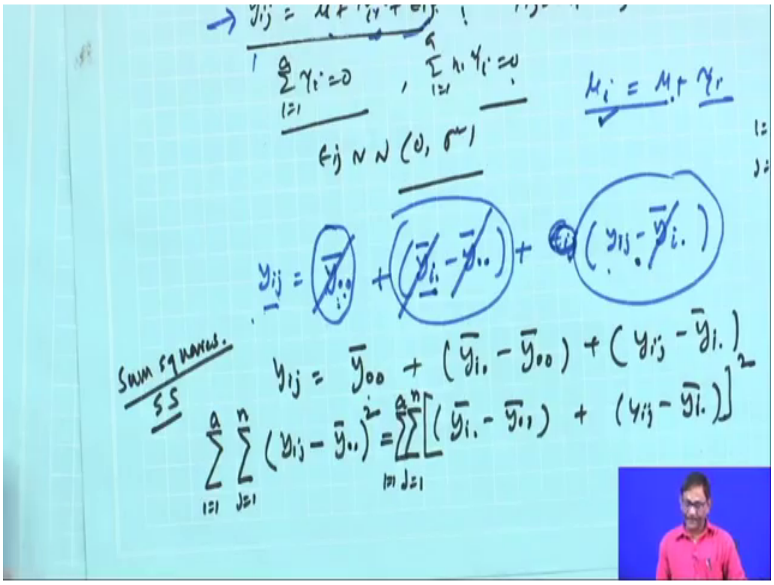
(Refer Slide Time: 22:24)



So, this one will be estimated by this and this tau i, tau i you will be basically y i dot bar minus y dot dot why because we know mu i equal to mu plus tau i and tau i equal to this. So, mu i mu i estimate is mu i dot bar mu s estimate is this, this, so plus this epsilon ij. So, y ij epsilon ij, now you may write this, this one as epsilon ij you may write y ij minus y i dot bar. First mu corresponding to this, tau i corresponding to this, epsilon ij corresponding to this each right hand both side left hand side right hand side equal y dot dot bar these will be cancelled this will be cancelled y ij equal to y ij. So, this is known as portioning the observation.

So, what we have done here then we have written y ij equal to y dot dot bar plus y i dot bar minus dot dot bar plus y ij minus y i dot bar. So, one observation sing single general observation is partitioning into grand average, treatment effect, error. So, from here we will derive a concept called sum square, that is known as SS.

(Refer Slide Time: 24:44)



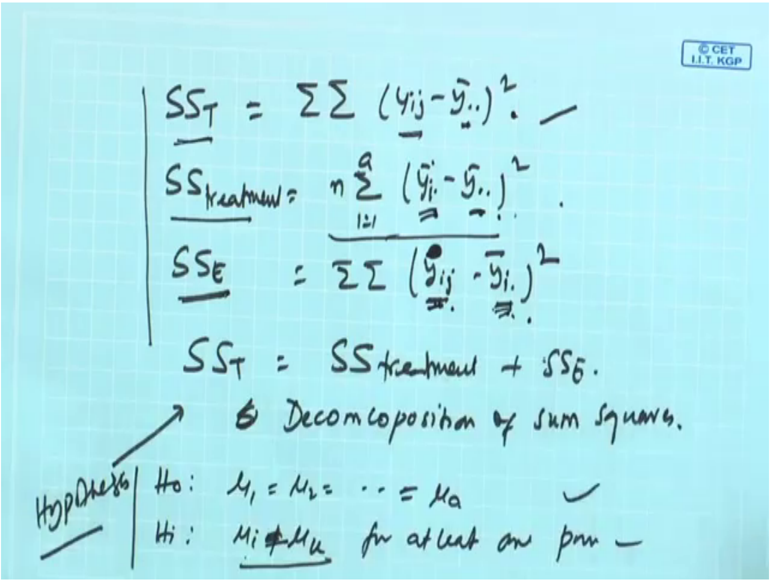
So, what you do you just write down y ij minus y dot dot bar left hand side then right hand side will be y i dot bar minus y, y dot dot bar plus y ij minus y i dot bar.If you take square then you have to square it and then take summation then you sum here. So, s i and j both are there. So, i stands from change from 1 to a j stands from 1 to n. Suppose I will write j equal to 1 to n here you write j equal to 1 to n here.

Similarly another summing will be there not only across not suppose if I put 1 j equal to 1 to n that mean you are summing up y 1 plus i 1 plus y i 2 plus y i 3 like this. So, this you are subtracting by y bar and then making square and summing up. So, if you make then there is another portion is 1 to a. So, that also you sum up. So, i equal to 1 to a. Then another sum i equal to 1 to a.

What will happen if you expand this? Now, see the slides see what I have written here i equal to 1 to a j equal to this square now this one when you take double sum then you see you are doing this and finally, these portion left hand side is becoming n a n times i equal to 1 to a y i dot minus y double dot bar double dot bar square plus this here double sum is there. So, this portion when each of the observations are subtracted by their grand mean and taken square and makes sum up this is the sum square total, this is sum square total it is total. Now, here what is happening each of the treatment means are subtracted from the grand means or grand average each of the treatment average is subtracted from the grand average and then you have taken their sum S square and sum across all the levels and then this one is nothing, but some square treatment and rest is some square error.

So, from here to here is a very simple one this quantity remains valid this quantity remains valid and this quantity becomes 0 because j equal to 1 to n if you sum up here you will find out that this quantity will become 0. So, one quantity this y i this y ij, one of these two will be this one will be I think this one, this one will become 0 and ultimately this is simple one it is not a big one big issue because if you write j equal to 1 to n then this will be n y bar square will be there and this also will be n y bar square then this n y bar minus n y bar that will become vanish, (Refer Time: 28:02) vanish and you will be having total sum square equal to treatments sum square equal to individual sum square. So, what you are getting then?

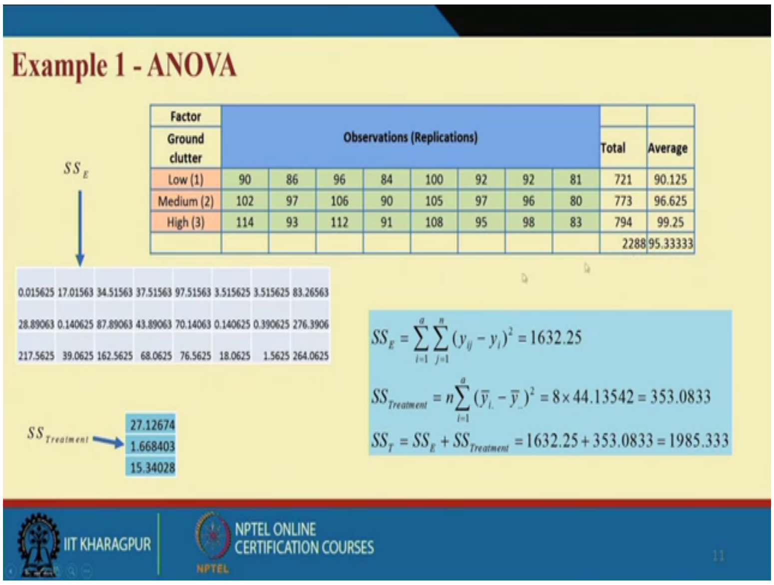
(Refer Slide Time: 28:12)



You will get SS T equal to double sum y ij minus y dot dot bar square. You will get treatment equal to n, 1 sum i equal to 1 to a y i dot bar minus y dot dot bar square and there will be SS error again double sum then you this is nothing, but y ij minus y dot bar this square. So, interested in what you have achieved SS T equal to SS treatment plus SS error this is known as sum square decomposition of sum squares, decomposition of sum squares.

So, we will see the one example here and ultimately in ANOVA as I told you at ANOVA we test null hypothesis that mu 1 equal to mu 2 equal to mu a and alternate hypothesis is mu i not equal to mu k for at least one pair mu 1, mu 2, mu 3 like this for at least one pair there is a there is a difference this. So, this one, we take this is the hypothesis what hypothesis we test what are the hypothesis null hypothesis is all means are equal alternate hypothesis at least one pair is not equal. So, that is H1, H 0 and H 1 here.

(Refer Slide Time: 30:04)

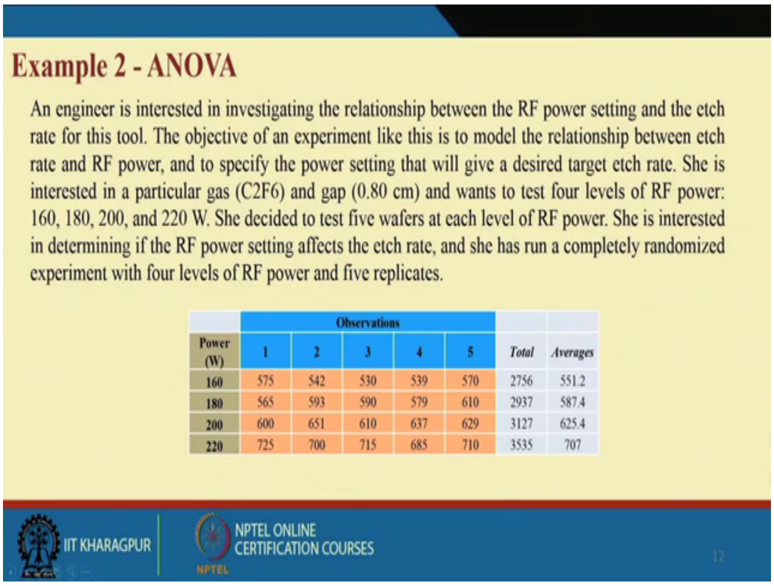


So, we will see how the sum square decomposition is taken place you just see this slide here we have H number of observations and you find out the total and their average second average, like this average this is grand average. We use this formula and we got that SS, SS E equal to 1632.25, SS treatment is coming 353.08 this is nothing, but using this formula only.

So, what is y ij? Suppose 90 minus grand average 95.33, this square is this, this square 90 minus 95 point like this then 86 minus this. So, make it and square it and this one this square. So, every observations will be subtracted by their respective by the grand min grand average and then square it and then when you make the total you will get this value, you will get this value.

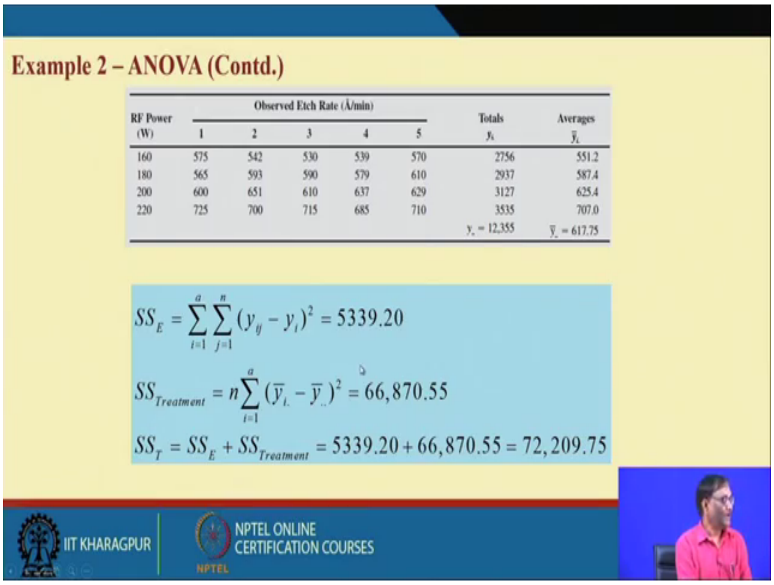
Now, for the treatment how many treatments are there, there are 1 2 3 treatments. So, what is these 3 treatment method treatment average minus grand average, its square then second treatment average and grand average square and third treatment average square then take the sum you will get SS treatment level like this. So, essentially what we basically calculate, we calculate that SS treatment and SS total and then SS error will be the SS treatment minus SS total and accordingly you will get the values.

(Refer Slide Time: 31:51)



Second one is a like this, another object here what happened this is the first example what I have given you in, the first from the first class onwards we have given you. This is the second example, second example is as engineer is interested in investigating the relationship between the RF power setting and etch rate for this tool and here basically response variable is the etch rate and you are using different power, power 160 watt, 180 watt, 200 watt and 220; that means, 4 levels. So, there are 5 replications and this is what are the experimental data coming and then these are the total and average. Now, find out SS error, SS total and treatment and SS total.

(Refer Slide Time: 32:35)



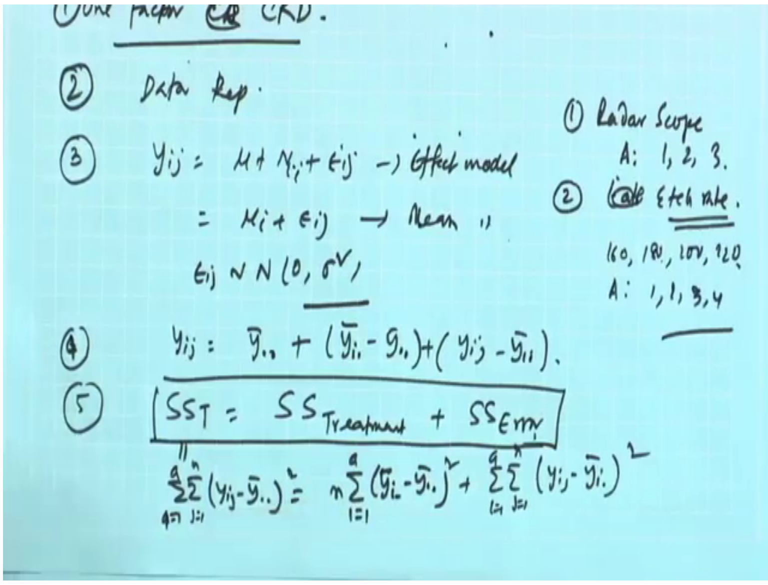
So, here let me give you tell you that there is some typographical mistake. So, please follow this when you are talking about SS T that mean each of the observation is subtracted by the grand average and then that difference is taking some taken squared and then overall sum. When you are talking about SS treatment? Every level average is subtracted from the grand average then there that difference is squared then using this you are getting this and when you are talking about SS error that mean your every average value y ij sorry y ij, not y ij bar y ij every observation is subtracted from their respective from their, respective level average and then this one is summed up.

What does it mean? Here 575 minus 551.2 then 542 minus 551.2 so then the level one all these observations are subtracted by the respective average and then that one is squared and sum and summed up. So, there is similarly they are this one that this 4 different levels 4 into 5, 20 observations will be summed up using this formula will be summed up and then you will be getting SS error.

So, what did the mistake here typographical mistake here is, there is a bar here y i dot bar. So, if you here it is y i dot bar minus this the 3 levels, here one bar is missing please put that bar y i dot bar. So, I hope the calculations are correct it is done in excel. If there is a mistake please mention in the in the discussion for to check.

So, this is what is decomposition of sum squares. So, I hope that you understand. So, what we have discussed?

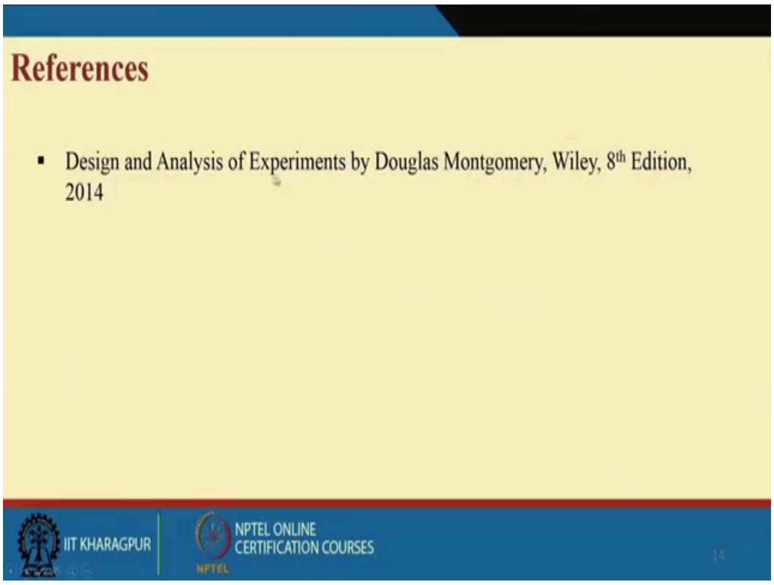
(Refer Slide Time: 35:14)



We have discussed one factor one factors CRD complete randomized design. Then we have we have seen that data representation in tabular form, then what we have seen that the ANOVA model, then we have seen that is the effect model this is what is effect model or you can write this one as mu i plus epsilon ij this is mean model and then here also we have said that the error is normally distributed with sigma square 0 sigma square. And then we have seen that a general observation how it will be partitioned to how it will be partitioned to different component, different component. From there we have seen that how the sum square total is partitioned into sum square treatment plus sum square error and we have given a general formula for calculation of these, these equal to double sum y ij minus y dot dot bar square equal to n i equal to 1 to a y i dot bar minus i double dot square plus here it will be double sum y ij minus y i dot bar square and here it is j equal to 1 to n, a equal, i equal to 1 to a, i equal to 1 to a, j equal to 1 to n.

And then I have given you two example, one is the that that is radarscope example where ground clutter level A has 3 levels and another one is the each etch rate e t c h etch rate example where basically some power is used 160, 180, 200 and 220 so that means, the factor a has 4 levels 1 2 3 and 4, 4 levels are there. And then how to compute the different things SS T, SS (Refer Time: 37:43) is given to you. Please practice it.

(Refer Slide Time: 37:52)



Thank you very much and I have taken all these lecture, those things a from the book Design Analysis of Experiment by DC Montgomery, Wiley, 8th Edition.

Thank you very much. We will go to next, we will see next lecture now.