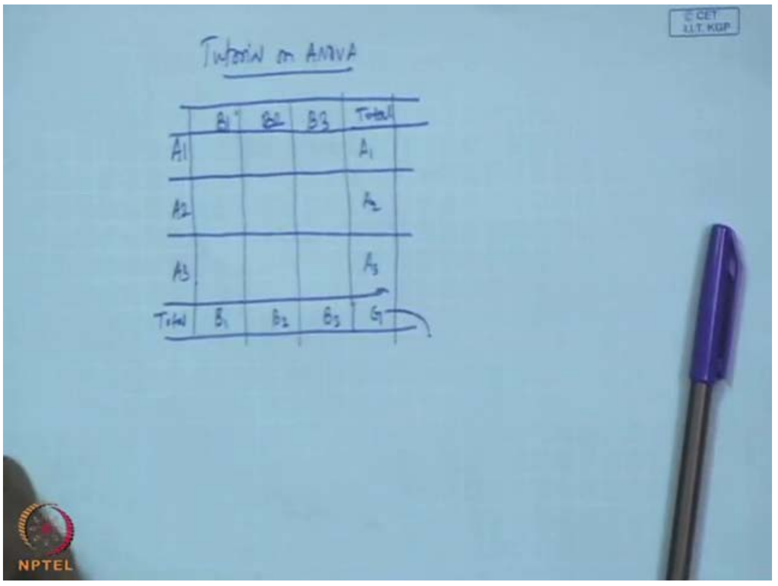
Applied Multivariate Statistical Modeling
Prof. J. Maiti
Department of Industrial Engineering and Management
Indian Institute of Technology, Kharagpur

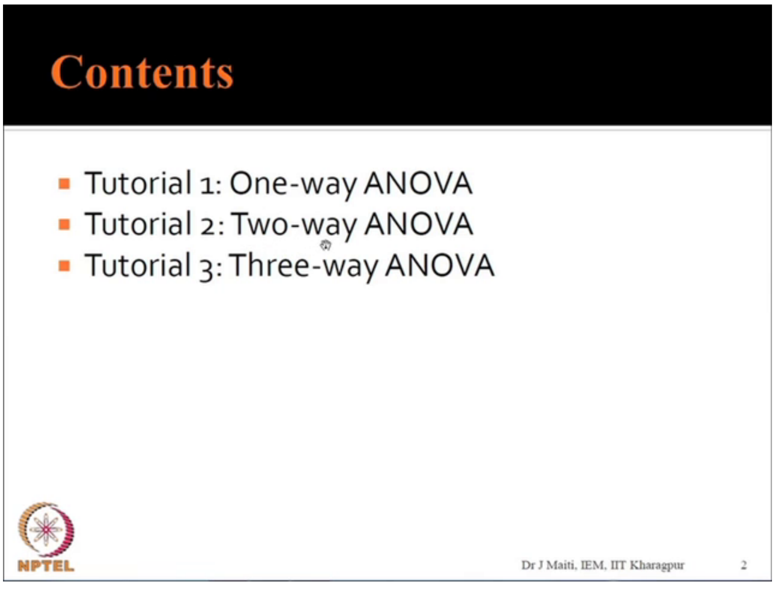
Lecture - 19
Tutorial - ANOVA (Contd.)

(Refer Slide Time 00:24)



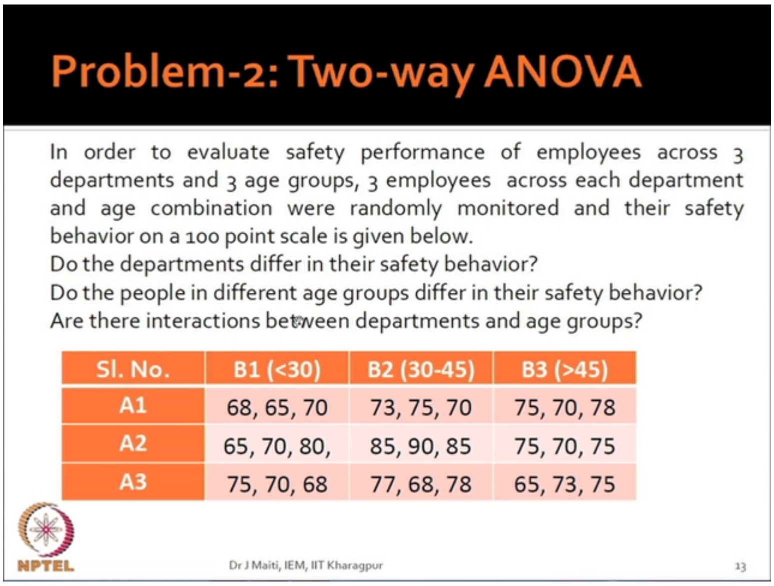
Good afternoon, we will consider tutorial on ANOVA. So, last class what we have seen?

(Refer Slide Time 00:36)



We have seen that we considered two tutorials one way ANOVA, another two way ANOVA and two we have added. Now, one more tutorial on three way ANOVA, we have fully completed one way ANOVA, partially completed two way ANOVA and we have not completed three way ANOVA. Now, I will complete first the two way ANOVA and then we will go for three way ANOVA.

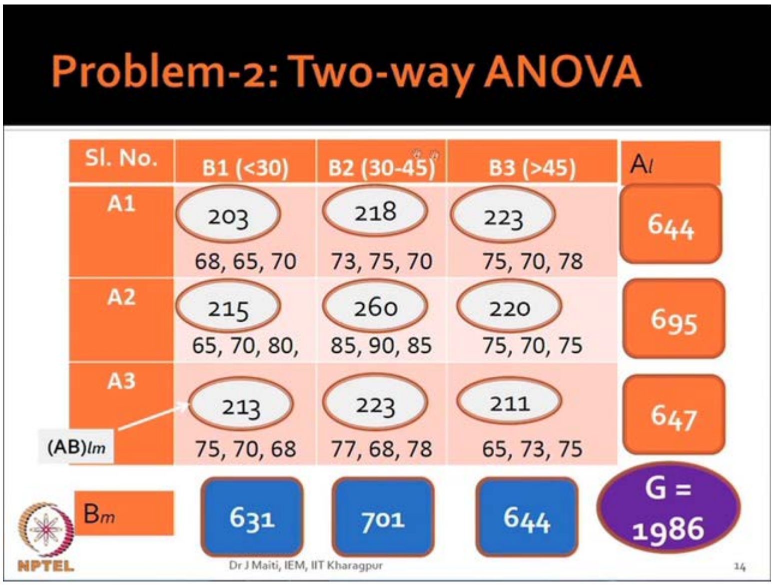
(Refer Slide Time 01:05)



Let us come to the problem for two way ANOVA, as I told you that the issue is safety behavior of a company across different departments and different age groups. We have considered three departments A 1, A 2, A 3; and 3 age groups less than 30 years of age 30 to 45 years of age and greater than 45 years of age. This table shows the different data points collected across departments and groups. For example, in the first cell 68, 65 and 70 that are values, this data points are for the combination A 1 and B 1. Similarly, for A 1 B 2, A 1, B 3 like A 2 B 1, and finally A 3 B 3, so there are 3 samples, so in the sense that the size of the sample for each combination is 3.

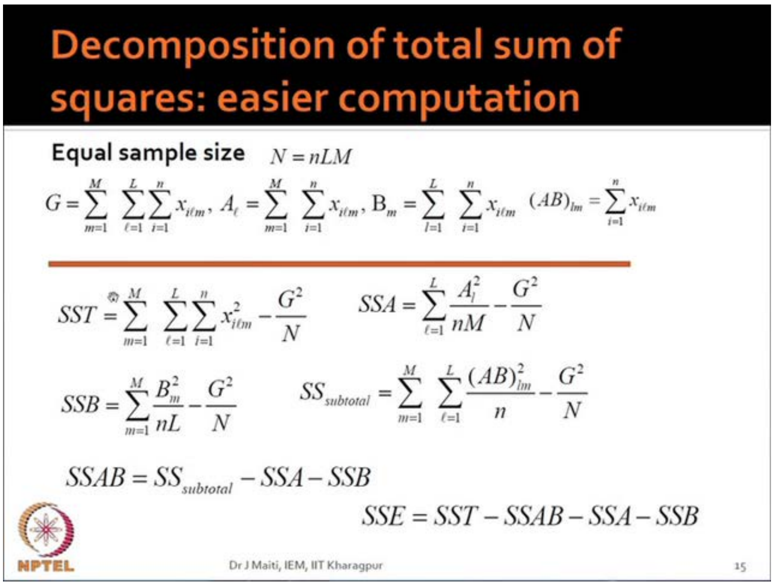
In total, there are 27 observations and across each department there are 9 observations across each age group there are 9 observations. Our objectives are to do the different departments differ in their safety behavior, do the people in different age group differ in their safety behavior. Are there relationships between departments and age groups, so we have discussed this?

(Refer Slide Time 02:40)



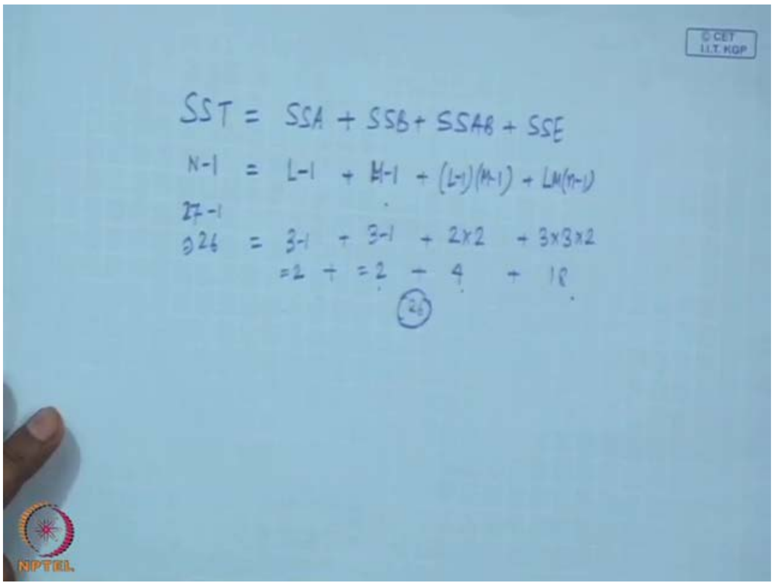
We have seen this table.

(Refer Slide Time 02:43)



Also we have discussed that, what are the different O A, the computational formulas for computations of SST and its subcategories from source variation point of view. So, I think all of you know at this point in time that.

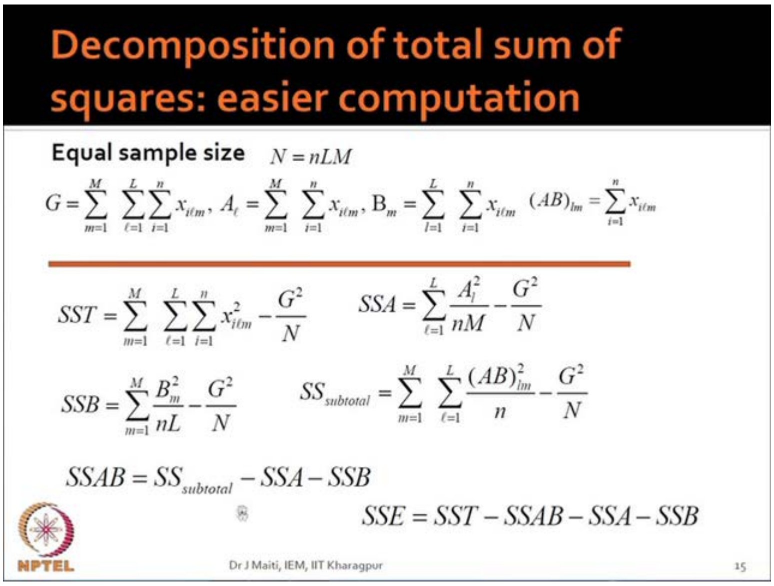
(Refer Slide Time 03:03)



Our total sum square yes sum square total is divided into sum square factor A sum square factor B, then sum square factor A B means interaction A B plus sum square error. So, we have divided in this manner and we have also seen that the degree of freedom is N minus 1 for total sum square, for SSA it is L minus 1, for SSB it is our M minus 1 for SSAB it is L minus 1 M minus 1 and for SSE this is L M into n minus 1.

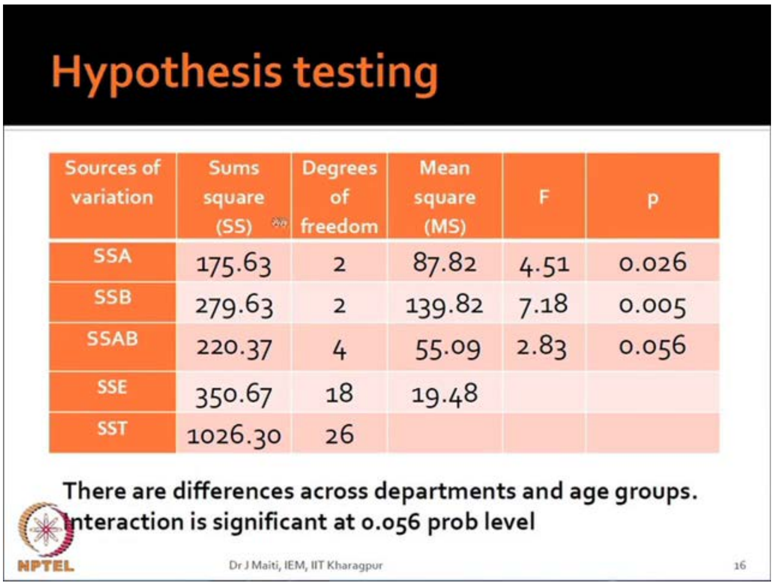
So, that we have already seen, so ultimately for the example given there are total I think 27 observations minus 1, so that is 26 is equal to there are 3 departments. So, 3 minus 1 equal to 2 plus 3 age groups that also 2 plus 2 cross 2 that is 4 and rest will be this L is 3 into 3 into 2 that is 18, so if you add this 2 plus 2 plus 4 plus 18 that will lead to 26. Also you have used the formula computational formula for SST and SSB SSA all those things.

(Refer Slide Time 04:52)



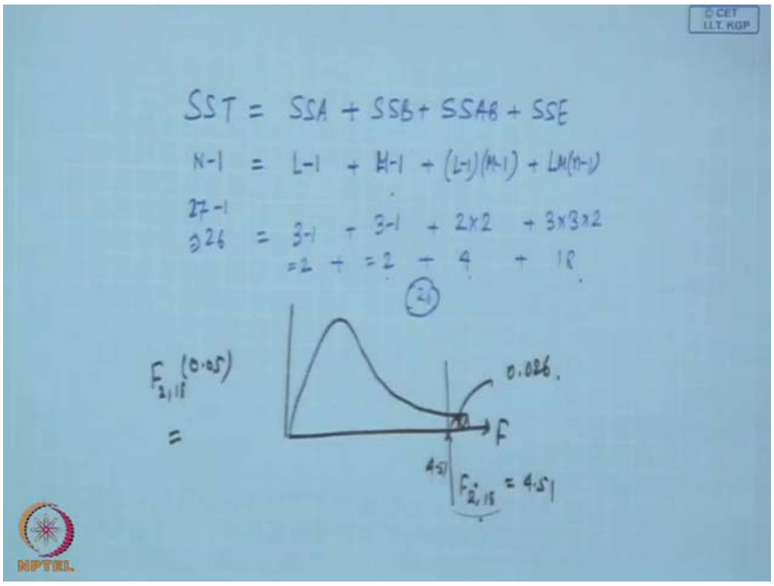
Using this formulation that first you compute the granding, then individual factors level mean factor B levels mean and the combination these are totals. So, first grand total then factor A level totals, B level totals and A B level interaction total, then following this equation you are getting SST SSA SSB subtotal. Then finally, you are getting a SSAB using this equation and this is the formulation final SSE and then what we have done.

(Refer Slide Time 05:33)



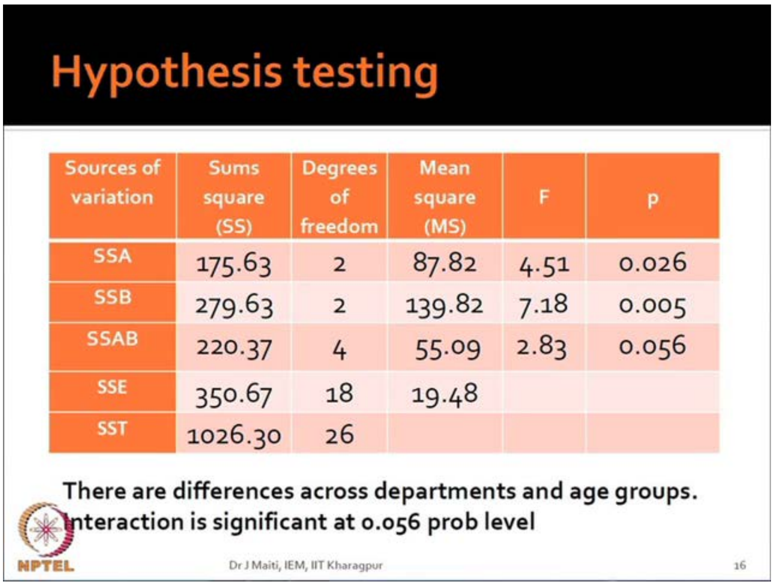
We have also seen that the ANOVA table, this is the ANOVA table sources of variation difference values then degrees of freedom and you found out that 87.82. This is the MS for factor A, for factor B it is 139.82, for the factor the interaction A and B that is 55.09 and SSE 19.48. Then the F statistic values are for factor A 4.51, factor B 7.18, factor A interaction A B that is 2.83 and these are the probability values which we got from the table. Now, what are the degrees of freedom for this, for these degrees of freedom is numerator degrees of freedom is 2 and denominator degrees of freedom is 18. For the second one also, numerator degrees of freedom 2 denominator degrees of freedom 18 and for AB numerator degrees of freedom is 4 denominator degrees of freedom 18.

(Refer Slide Time 06:43)



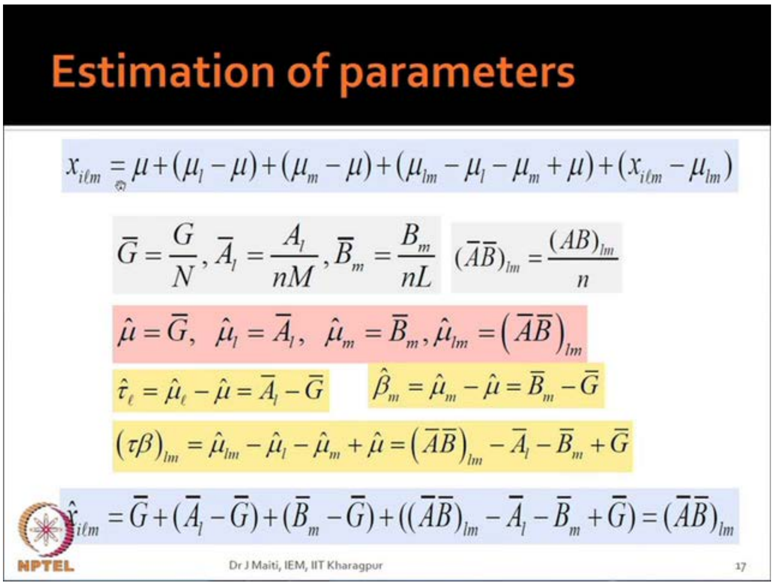
So, all cases this F then we found out F some value you got suppose this 1 is this suppose F 2 18 if this is 4.51 then this probability value is this one is 0.026. So, two way you can do it one is you fix one alpha value and then find out that what will be the F value for that alpha. Suppose if you take F 2, 18 0.05 you will get certain value, definitely that value will be less than this value that is 4.51, so the first factor contribution is there.

(Refer Slide Time 07:31)



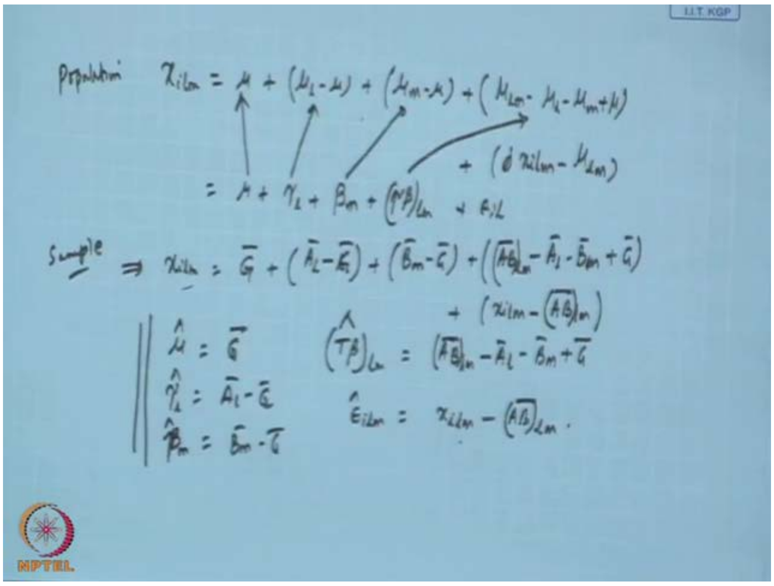
Which is much lower than 0.05 significant level second factor also same the third factor which is the interaction factor it is slightly more than the requirement that is 0.05 that is 5 percent probability level. But, you can consider it significant, so there are differences across age groups, there are differences across departments and interaction is significant at 5.6 percent probability level.

(Refer Slide Time 08:02)



Then I told you, we also discuss this that I told you that how do estimate the parameters. These parameters are written clearly here these are the parameters, so I am writing once more.

(Refer Slide Time 08:18)

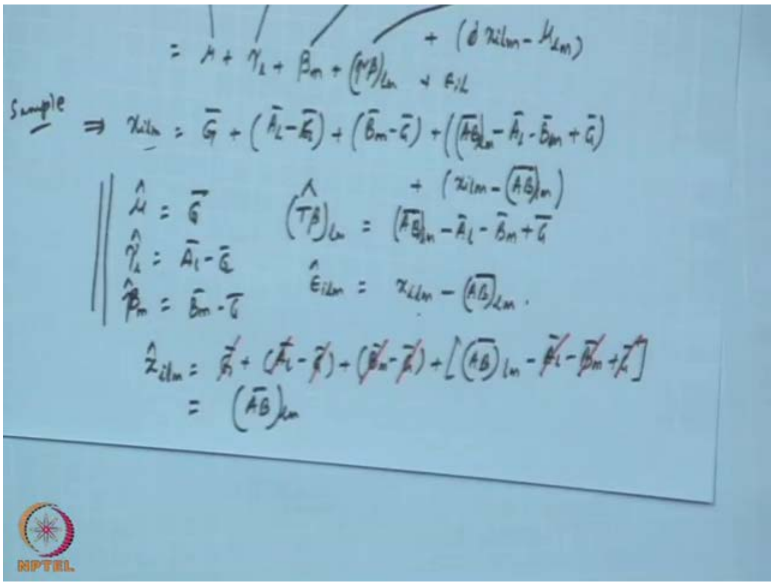


X ilm equal to mu plus mu l minus mu plus mu m minus mu plus mu l m minus mu l minus mu m plus mu then plus X ilm minus mu lm. So, that what does it signify then this is nothing but the grand mean this one is tau l then this one will be beta m. So, grand mean tau l beta m and then tau beta l m this one plus epsilon i l and there sample this is population, from population, so their sample counterpart.

So, if we find out X ilm it will be we have written like this G bar that is the grand mean bar plus we say A l bar minus G bar grand mean bar plus we say B m bar minus G bar plus we say that A B l m bar minus A l bar minus B l bar B m bar not B l A l B m bar minus A l bar minus B m bar plus G bar plus X ilm minus A B l m bar this is the sample domain.

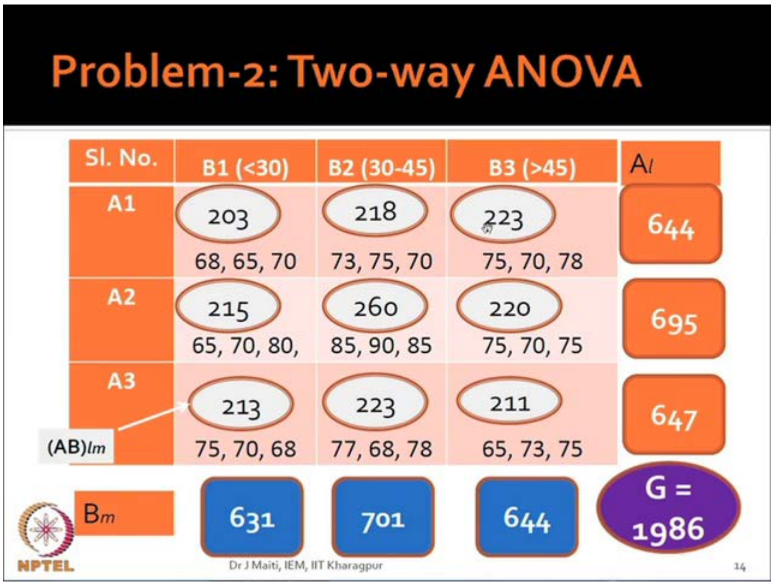
So, that means the estimate of m u is G bar estimate of tau l is we say A l bar minus G bar, estimate of beta m is again B m bar minus G bar, then estimate of tau beta l m that total cap this one is your A B bar that l m. You can give a bracket here A B bar l m minus A l bar minus B m bar plus G bar and your error portion ilm cap which is X ilm minus A B. Again you make a bracket here, A B bar l m, so these are the parameter estimates of parameter, different parameters.

(Refer Slide Time 11:26)



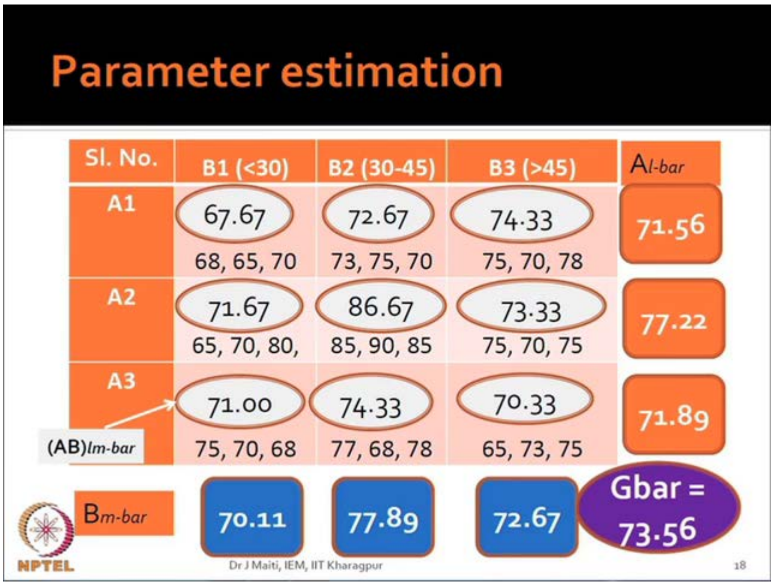
Now, you see then what will be your the predicted value of X ilm ilm, so everything will remain except the error term this one. So, this it will be G bar plus A l bar minus G bar plus B m bar minus G bar plus A B bar l m minus A l bar minus B m bar plus G bar. So, what will happen here you see this few things will be cancelled out G bar, G bar will cancel out. Now, one more G bar is there, G bar and this G bar will be cancelled out A l bar plus Al bar minus will cancel out B m bar plus B m bar minus will cancelled out. So, ultimately what will remain then this is nothing but A B bar l m this is the predicted value correct. So, if I go back now, so what will be the predicted value for each of the cell?

(Refer Slide Time 12:48)



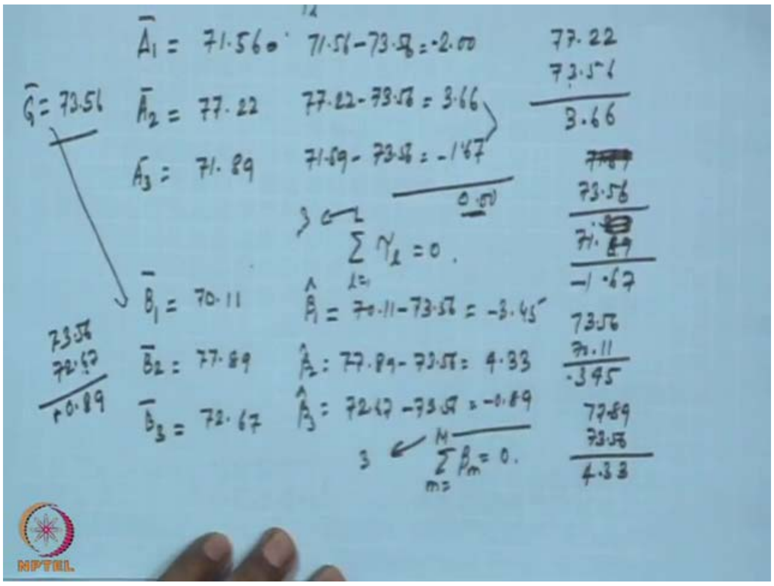
What is the average in this cell that will be predicted value for first three quantities? In the second cell, the average of this will be the predicted value for all three third cell like this. So, every three observations having a same predicted value, so now let us see this final one what we have.

(Refer Slide Time 13:14)



This is what all the values computed, so A l bar is said we will calculate something more with this A 1 bar.

(Refer Slide Time 13:26)



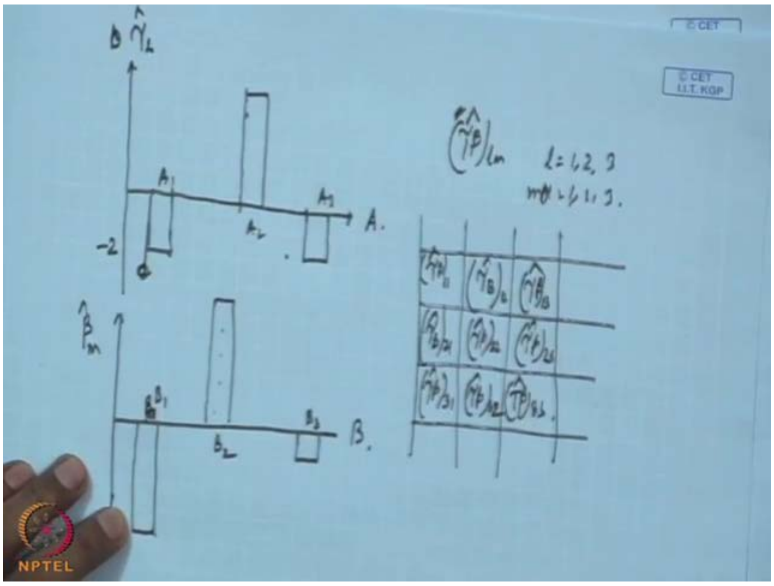
So, A 1 bar equal to 71.56 this is computed, now A 2 bar is 77.22 and A 3 bar is 71.89, what is our G bar, G bar is 73.56 G bar is this. Suppose you want to find out that you are interested to find out tau l bar then what you will do this minus your thing will be 71.56 minus 73.56. What will be the value this will be minus 2, so what does it mean that the A 1 effect department one effect is minus 2 that is with respect to the average we are talking about. Then second one will be 77.22 minus 73.56, what will be the value here 77.22, 73.56, 6 again 6, so this will be 4, so 3.66 and third one will be 71.89 minus 73.56, so 71.89, 73.56, so 3 then another 3.

Then this will be what I think just opposite this is smaller 71.89 then 7, this will be 89, 9 plus 6 15 2, 1 minus 1.67, now you add this this is 0 this 6 and 7 that is some rounding error. Otherwise this quantity will be 0 and I think you can remember that I say the tau l, L equal to 1 to capital L this will become 0, so the same thing will prevail for beta.

So, what will happen what will be your beta 1, so first we will find out B 1, B 2 and B 3. See what is your B 1 bar, B 1 bar let like this B 1 bar is 70.11, what is your B 2 bar 77.89, what is your B 3 bar 72.67, where G is same for this case also. So, then we want to find out beta 1 that is the cap first one 70.11 minus G that is 73.56, what will be the value here 73.56, 70.11. So, it will be 5 then 4 then 3, I think this is minus 3.45, what will be beta 2 value beta 2 value will be 77.89 minus 73.56. So, you calculate 77.89, 73.56 give 3, 3, 3 I think 4 so 4.33.

Similarly, beta 3 cap will be 72.67 minus 73.56, so it will be 73.56, 72.67, so 9 6 plus 1 7, 7 plus 8 15, I think minus 0 this will be this then ultimately this will be minus 0.89. Again this sum will be beta m, m equal to 1 to capital M this will be 0 here our capital M is 3 here capital L is also 3. If this is the case, then I think in one of the example I show you that.

(Refer Slide Time 18:14)

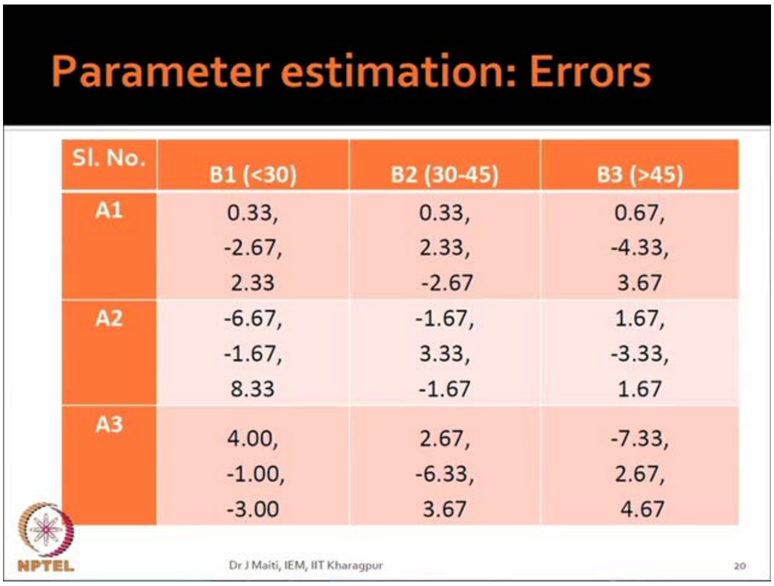


Suppose the age effect I want to plot, age effect this died suppose this is for A 1 this is A 2 and A 3 then what is your A 1 value, A 1 value is minus 2. So, A 1 is if I go like this this is minus 2, minus 2 and a 2 is 3.66, so 1, 2, 3.66 somewhere here and minus 1.67, 1.67 this will be somewhere here. Similarly, these are all, this is what is tau L estimate and these are this is A so far A 1, A 2, A 3. Similarly, you require to find out, also this type of diagram is better to give for easy understanding, so what is our beta 1 minus 3.45.

So, 1, 2, 3 like this, so minus 3.45 will be somewhere here that is beta 1, here it is for B 1, so write B 1 then second one is 4.33, 1, 2, 3, 4 this one for B 2. Similarly, other one is minus 0.89, so you can say this side is beta n and this is B capital B. Now what will be the tau beta l m, tau beta l m how do compute, so how many l equal to 1, 2, 3 m small m equal to 1, 2, 3, so your combination will be 9 combinations.

So, ultimately you will be getting a table like this where 1, 2, 3 then 1, 2 and 3, so what is required what is this tau beta cap 1 1, tau beta cap this is 1 2, tau beta cap 1 3 is coming yes. So, similarly here tau beta cap 2 1, tau beta cap 2 2, tau beta cap 2 3, then tau beta cap 3 1, tau beta cap 3 2, tau beta cap 3 3, you require to find out all those values.

(Refer Slide Time 21:28)

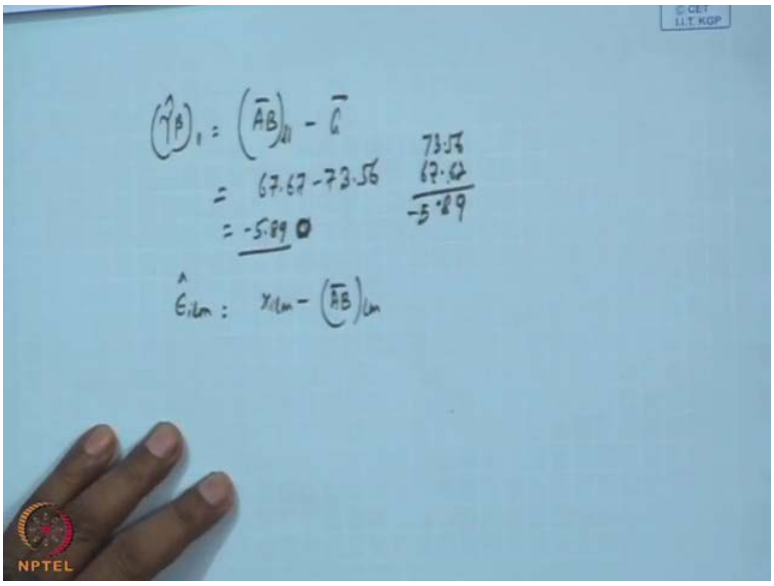


Parameter estimation: Errors

Sl. No.	B1 (<30)	B2 (30-45)	B3 (>45)
A1	0.33, -2.67, 2.33	0.33, 2.33, -2.67	0.67, -4.33, 3.67
A2	-6.67, -1.67, 8.33	-1.67, 3.33, -1.67	1.67, -3.33, 1.67
A3	4.00, -1.00, -3.00	2.67, -6.33, 3.67	-7.33, 2.67, 4.67

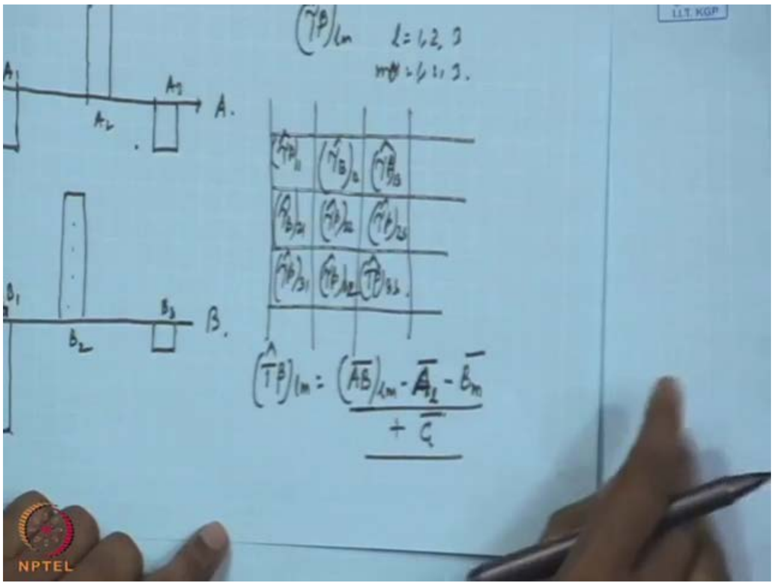
Now, once you calculate these, all those error beta another things then what is required to know. This is what is the error component, how do you calculate the error part then error part is epsilon i l.

(Refer Slide Time 21:49)



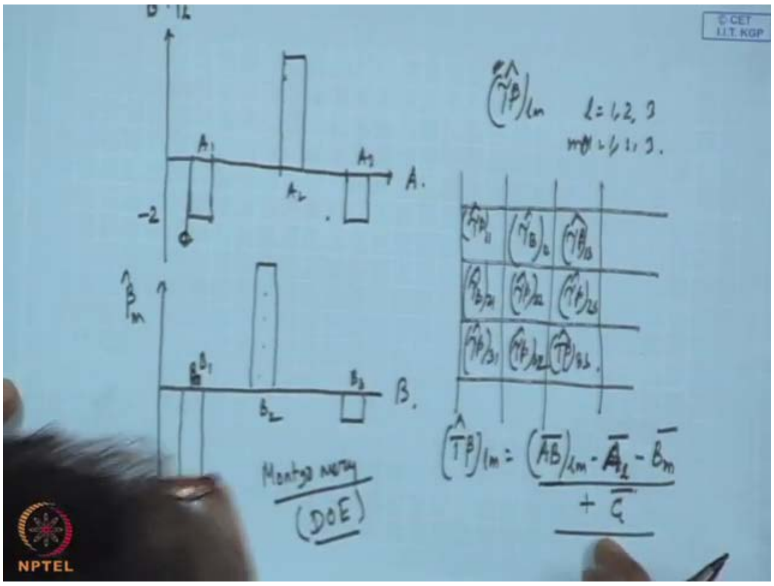
Ilm, what will be this one X ilm correct this minus A B bar l m.

(Refer Slide Time 22:13)



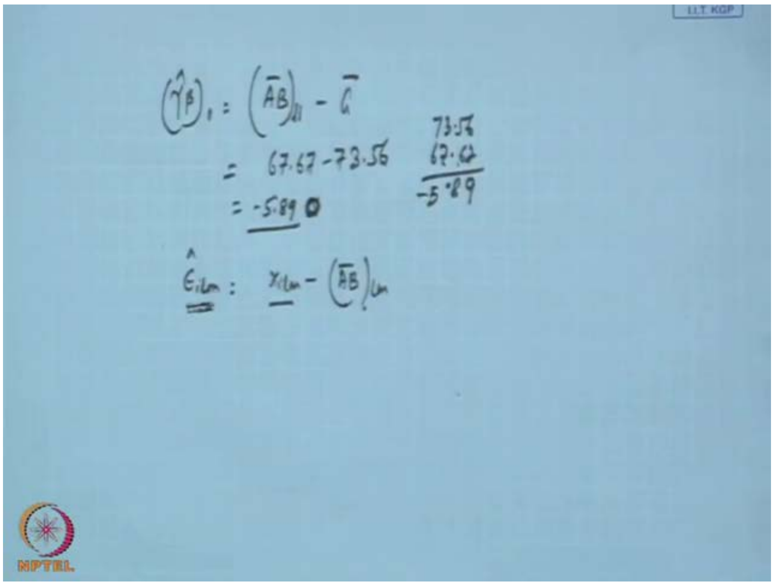
In this case this tau beta lm formula and in this minus B m cap plus G this SI the formula you have used, this is the formula we have used, so this formula we will be using.

(Refer Slide Time 22:40)



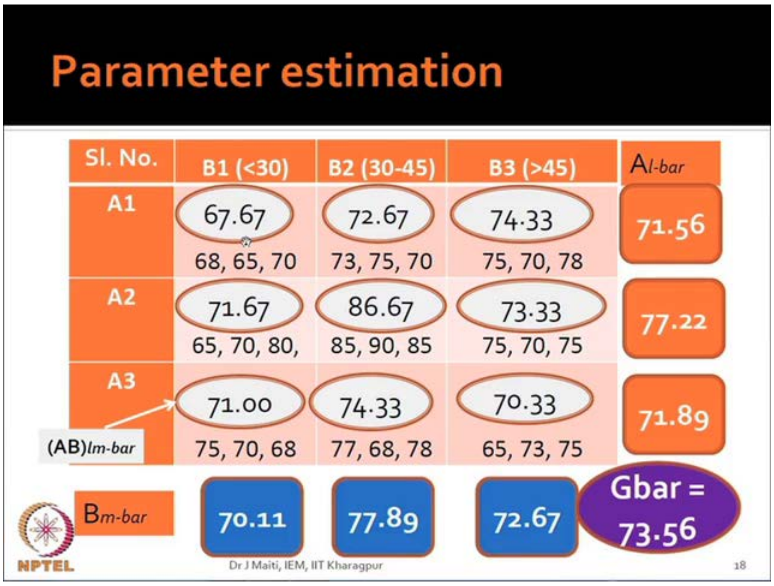
In Montgomery it is given clearly, you go through Montgomery, this one design and analysis of experiments design, analysis of experiment theory. This A B l bar minus A L bar minus B m bar plus G bar that is the formula we have used.

(Refer Slide Time 23:04)



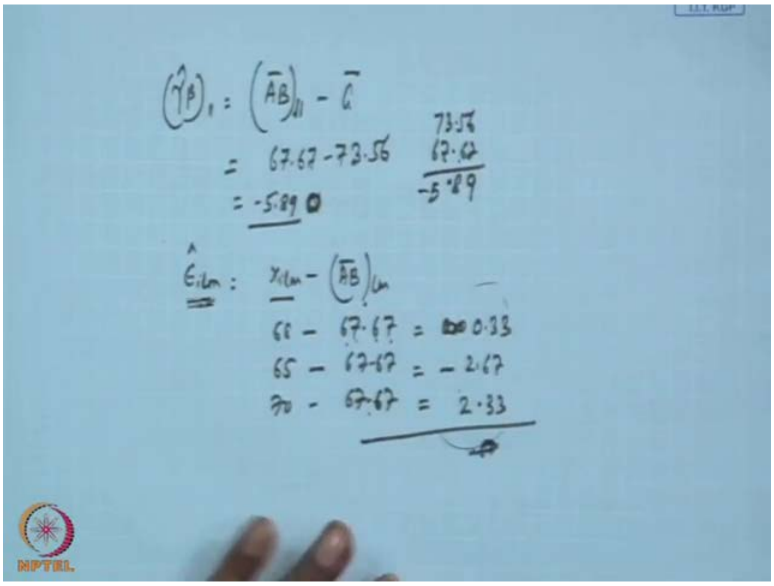
So, by this manner as now you will be able to compute that errors also, errors is X ilm minus A B bar L m, so A B bar L m you have already computed.

(Refer Slide Time 23:16)



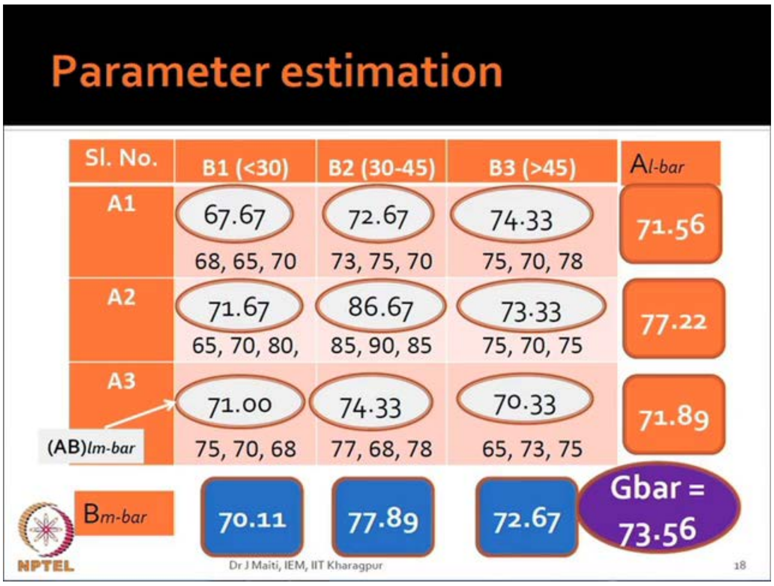
This is the A B bar L m, so 67.67, now what is xil x 111 68, so 68 minus this similarly, 65 minus this, 70 minus this, in this manner you will calculate.

(Refer Slide Time 23:37)



So, that means the first three observations 68 minus A B L wise 67.67 then 60 minus 67.67 then 70 minus 67.67. So, this one is I think 0.0 no 0.33 3 3 8, correct this one will be 65 minus 2.67 and this is 3 3 8, 2.33 again you see that sum is 0 the sum is 0.

(Refer Slide Time 24:26)



So, in this combination the parameter X ilm bar L m bar, all those things return.

(Refer Slide Time 24:38)

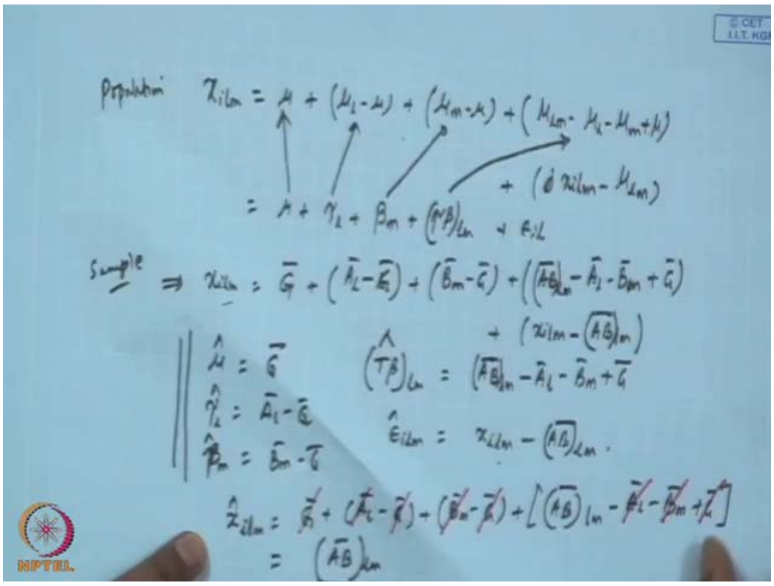
Parameter estimation: Errors

Sl. No.	B1 (<30)	B2 (30-45)	B3 (>45)
A1	0.33, -2.67, 2.33	0.33, 2.33, -2.67	0.67, -4.33, 3.67
A2	-6.67, -1.67, 8.33	-1.67, 3.33, -1.67	1.67, -3.33, 1.67
A3	4.00, -1.00, -3.00	2.67, -6.33, 3.67	-7.33, 2.67, 4.67

Dr J Maiti, IEM, IIT Kharagpur 20

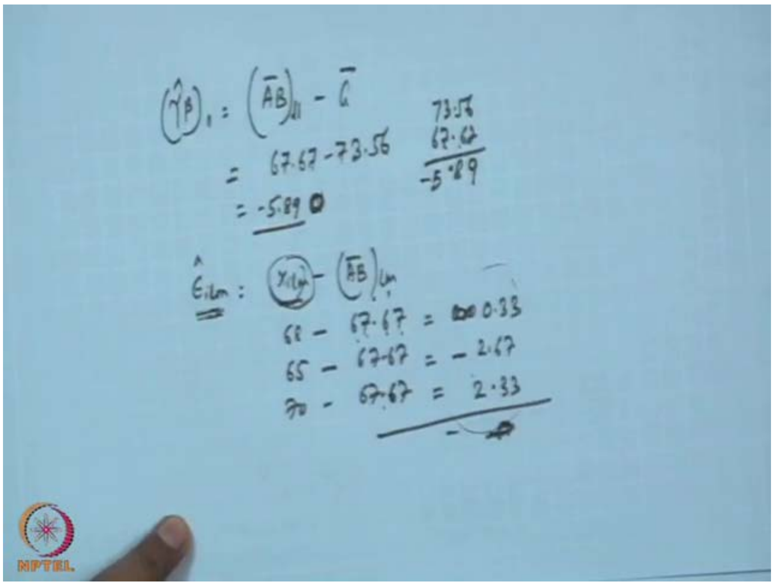
You go to this 0.33 minus 2.67 plus 2.33. What do you do with this errors, these are all errors because you have stabilized earlier I think I am correct that your predicted value is A B bar L m.

(Refer Slide Time 25:07)



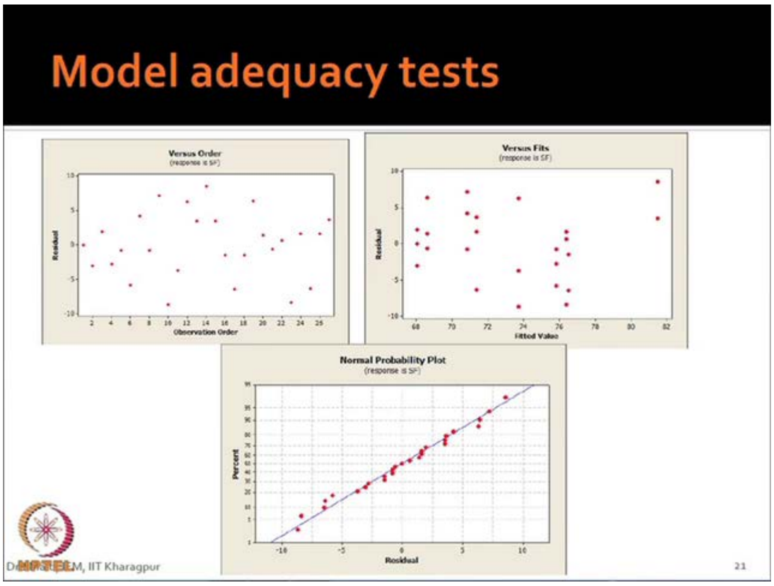
Predicted value is A B bar L m.

(Refer Slide Time 25:09)



Actual value is X ilm, so the error will be the actual value minus predicted value, so in that sense it is correctly calculated. So, now what you will do with this error you have to for model adequacy test.

(Refer Slide Time 25:26)

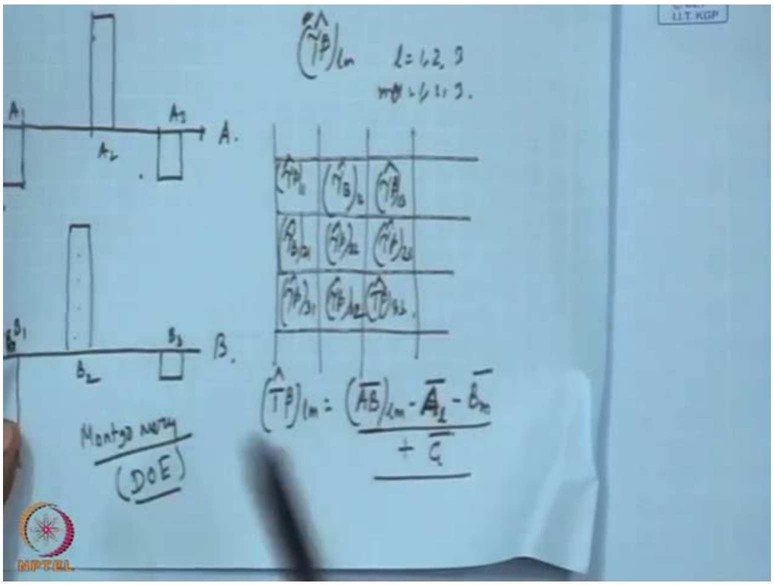


What are those model adequacy tests, first one is you have to find out the normal probability plot I am giving you the formula how to go about in first class. So, then see it is almost in a straight line, so you can say this is normal, second one is you are seeing the independency of the observations. So, in order if you see ultimately residual versus observation order, there is no relationship it is a hapazard, random, third one is you want to check the variability across different that groups this level.

So, here also you see we are finding out much that the pattern almost it is basically not saying departure from equal variability, so the adequacy part is also tested. Now, this is our two way ANOVA part, so should I repeat again or it is. So, two way ANOVA part is you have four sources of error variation, one is factor A, factor B their interaction and error term random error this four will make up the total sum square.

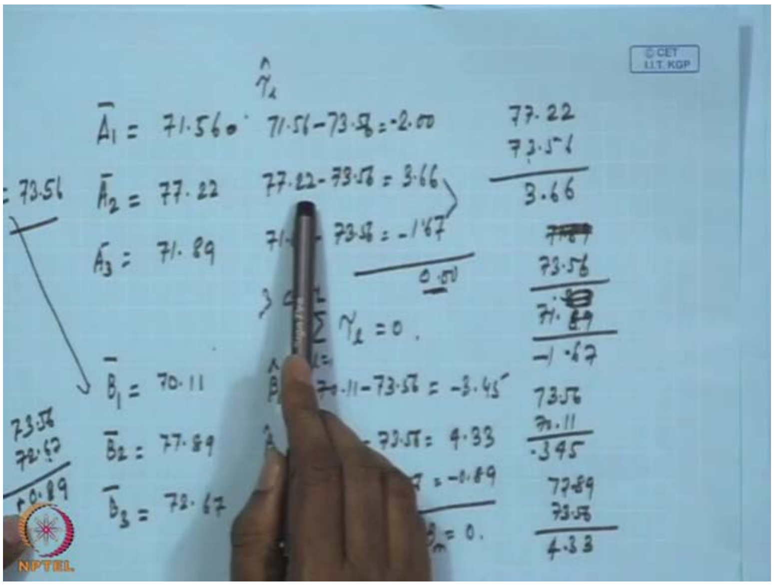
You have to find the degrees of freedom test it, once your test is result by hypothesis then you go for parameter estimation. Otherwise, no need of estimating the parameters, so you use the parameter estimation following the rules given to you except the other only one case that the tau beta.

(Refer Slide Time 27:15)



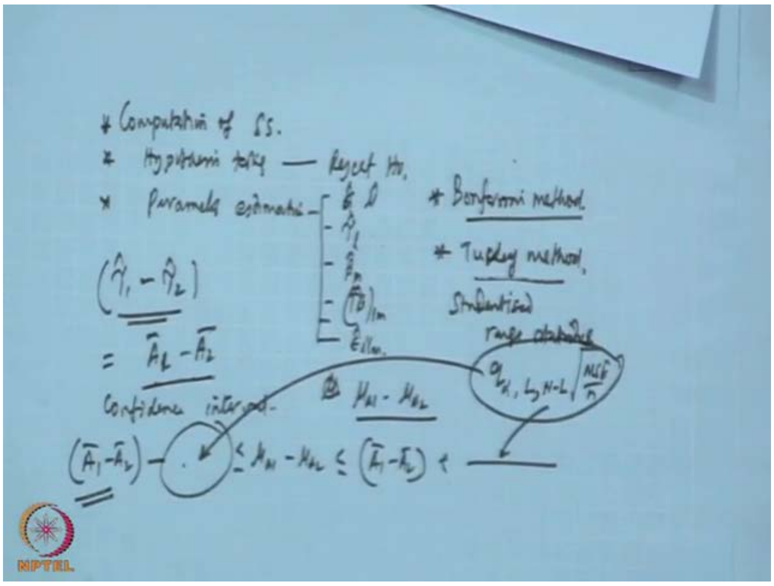
This one, this is the formula we have computed almost all other parameters here and you will be able to use this. Now, you will know that what the values of different parameters are, so, are you not interested to find out.

(Refer Slide Time 27:32)



Whether A the tau 1 and tau 2 are different or not.

(Refer Slide Time 27:42)



Yes, so first is your computation of S S, then hypothesis testing correct then followed by what you will do? You will find either you will reject hypothesis h 0, once h 0 is rejected then what you will do, you go for parameter estimations. By parameter estimation, what are the things you want to estimate, here that is grand mean we want the tau l, we want the tau l. So, grand means we want the tau l to estimated beta m to be estimated then tau beta lm to be estimated and also we want the error to be estimated.

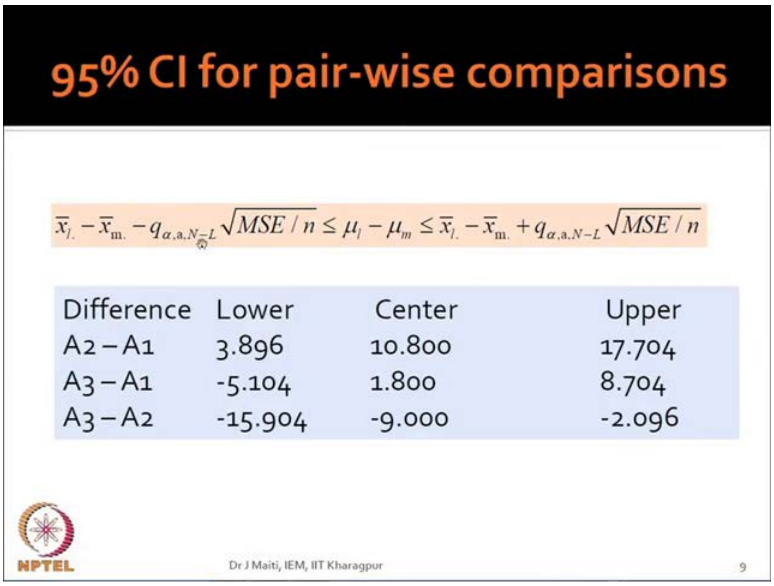
Now, in earlier class I have said that there is tau 1, tau minus tau 2. So, you must be interested to compare which of the population or the factor effects are different and this one is nothing but A L bar, A 1 bar minus A 2 bar. Can you remember this, the two level mean differences what test you will use for confidence interval.

Student: Sir, minus and find out the expectation values.

Yes, so expected value is what you are looking for that your mu, I think I can say. If I say A 1 minus mu A 2 for the factor and two different levels, then what you require. You also require to know that what will be the that A 1 bar minus A 2 bar, that is the sample value minus, then A 1 bar minus A 2 bar then plus some values or you got from this values ok.

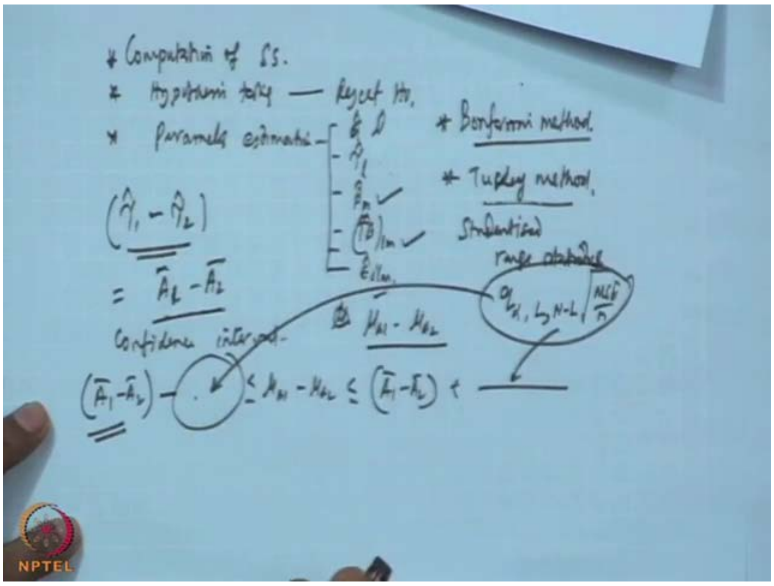
So, you can go for Bon Ferroni method, you can go for morning I say the Tukey method. If you use Bon Ferroni method, ultimately you will be using t distribution and number of comparisons are there, that the alpha value has to be adjusted or weighted using the number of comparisons. If you go for Tukey method, you will be using the studentised range statistics, studentised range statistics that is basically q alpha number that for what this is for factor A. So, number that is the number of labels in factor A, then what is the error degrees of freedom this into MSE by n, so here this one will come. Here, this will come here we can see that what we get in earlier in one factor case

(Refer Slide Time 31:24)



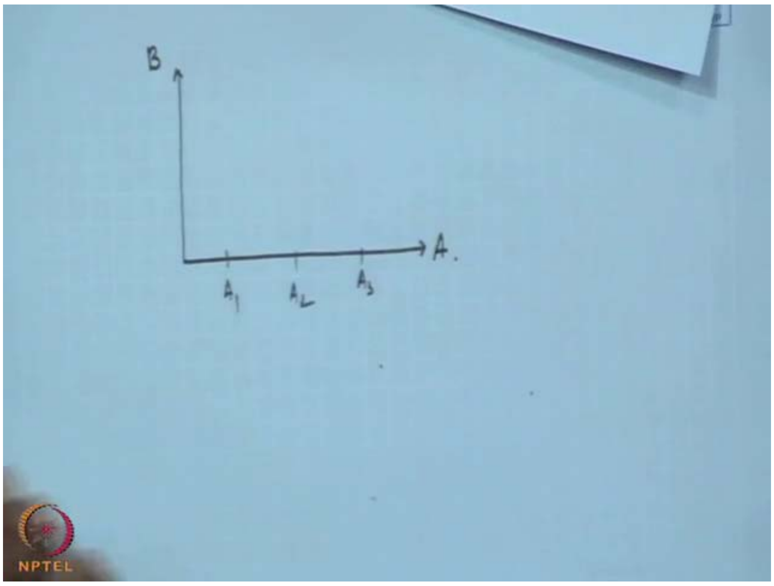
You see that q alpha A l MSE by n, so we will follow this formulation.

(Refer Slide Time 31:39)



Similarly, you will go for beta m, you can go for tau beta L m and all those things. Then once these things are completed, you will be able to find out which are the factor levels are different for factor A point of view for factor B point of view. You will be more interested to know this one the interaction effect.

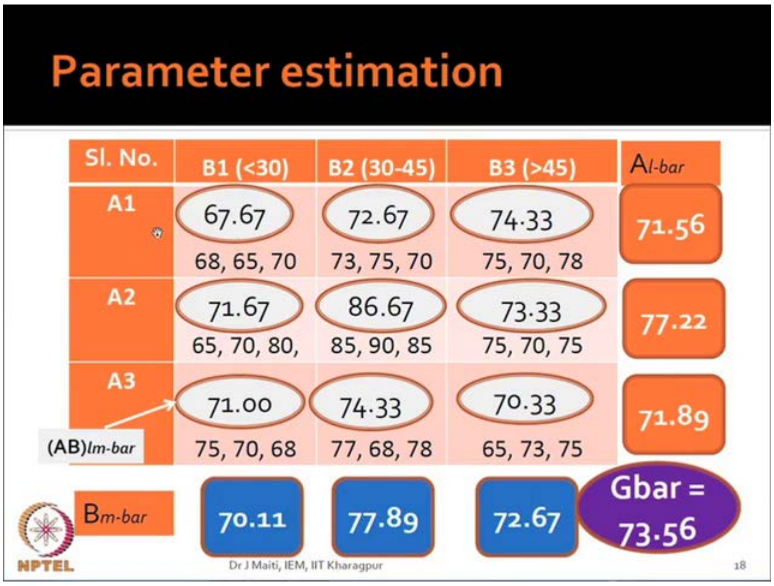
(Refer Slide Time 32:09)



How do we know the interaction effect? Basically, if I go back for example, what do you want to know, we want to know like this suppose this side is my B factor, this side is my

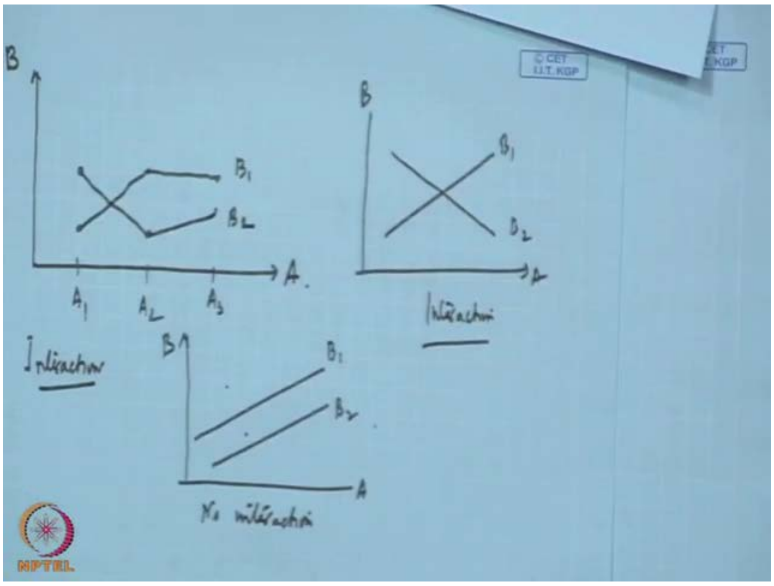
A factor. So, how many levels are there, for A 3 levels A 1, A 2, and A 3, now B also having 3 levels, so I will take the B 1 first correct, so if I take B 1 first.

(Refer Slide Time 32:51)



If I go to this figure you see if I will not consider this 2, I will consider only B 1 fix then A 1 value, A 2 value, A 3 value these are the A B L m bar. So, instead of these you will be plotting ultimately tau beta L m, so this will be this plus this minus this minus and then that value will be you will be getting, so you may get some value.

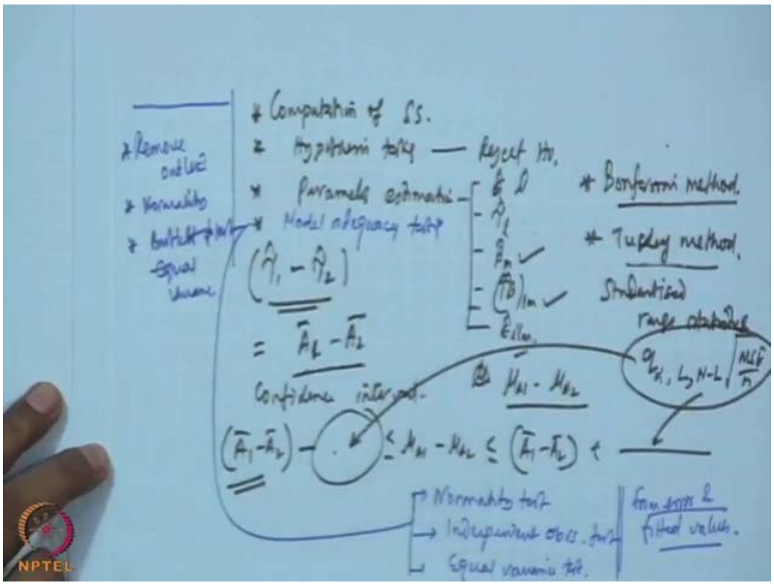
(Refer Slide Time 33:15)



Ultimately, you may get like this, so this is for B 1 you may get a situation like this, this is with respect to B 2 when you get this type of cross section intersection is there. Then you will say that interaction is significant there is interaction. So, in general if I find like this, suppose two factors are going like this suppose, this is my A and this is my B, so let it be B 1 and B 2, if two factors plots are going like this.

This is parallel no interaction and when there is interaction, you will be getting some cross when this line will not be parallel. Now, this shows interaction, now whether the interaction is significant or not you have to use the statistical formula. Using your Bon Ferroni, using your even a Tukey test also will give you or simply the first level formula we will use in Tukey test. We will test that the hypothesis, the difference is significant or not, so that means ultimately where we are standing now.

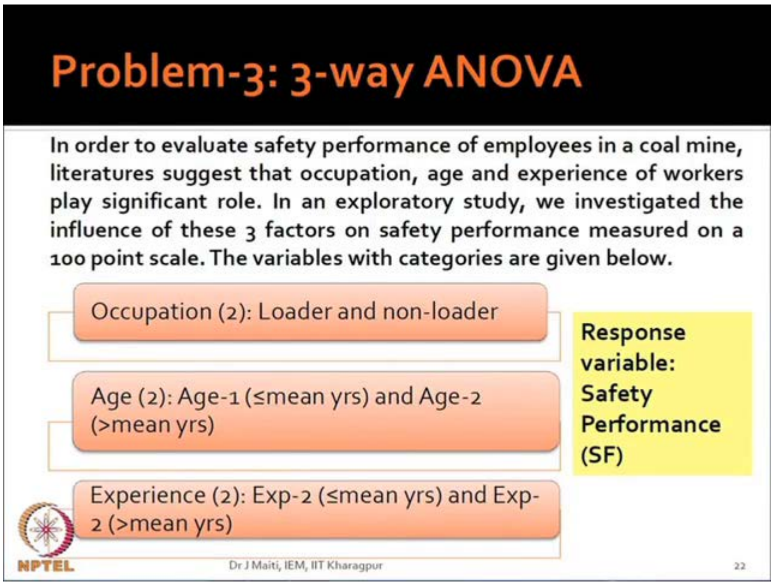
(Refer Slide Time 35:02)



You know the interactions are also known and then after that your work is the model adequacy testing. So, under model adequacy testing, you will be doing normality test then independent observation test then your equal variance test. Now, you will be doing all those tests from error, data error and filtered values wherever required. Please also remember that, it is always advisable that when you get a data set better remove outlier before fitting to models.

You remove outliers then you also test the normality of the data, normality test also you can test then you go for Bartletts test, for Bartletts test for equal variance. So, these are the first things you will do then you fit the model then you go for hypothesis, testing parameter, estimation, interaction, plots errors all those things computation. Then finally, again you recheck with the error data that yes the assumptions are satisfied, so this is what is two way ANOVA. Now, we will go for another example explaining the three way ANOVA case.

(Refer Slide Time 37:06)



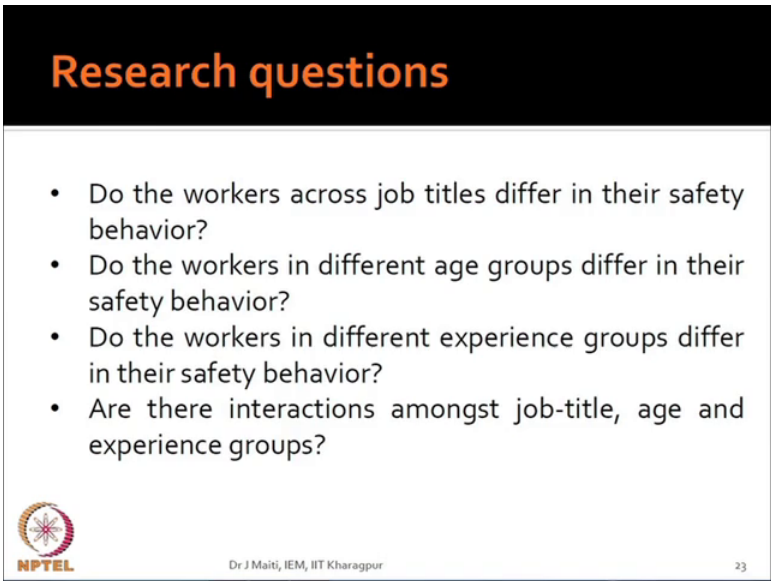
Here, the problem in order to evaluate safety performance of employees in a coal mine. This study is a real case we have done it, but here I am showing you the partial results because I have not taken all the variables only few variables I have taken here. Literature suggests that occupation age and experience of workers plays significant role. In an exploratory study, we investigated the influence of these three factors on safety performance measured on a 100 point scale.

The variables with categories are given below what we have occupation there are in mine the loader is a very dangerous groups prone to accident because of the many reasons particularly the hazardous working condition point of view. Non loader mean other than loader, all are non loader then age we have categorized into two groups we are saying that less than equal to mean age is group one, greater than equal to mean age is group two.

This is abrupt breaking of the, it can be better broken for example, young middle old like this. But, here to show, to demonstrate the applicability of three way ANOVA we have reduced this one. Similarly, experience also, we have categorized into two groups, less than mean years of experience and greater than mean year of experience and this is a standard process also.

Many times, we categorize continuous variable or dichotomize. Basically, continuous variable under less than mean greater than mean conditions, less than equal to mean greater than equal to mean, then the response variable. Here is safety performance which was given by a safety supervisor is who observes the people that he over the years he has rated that this is his performance under 100 point scale.

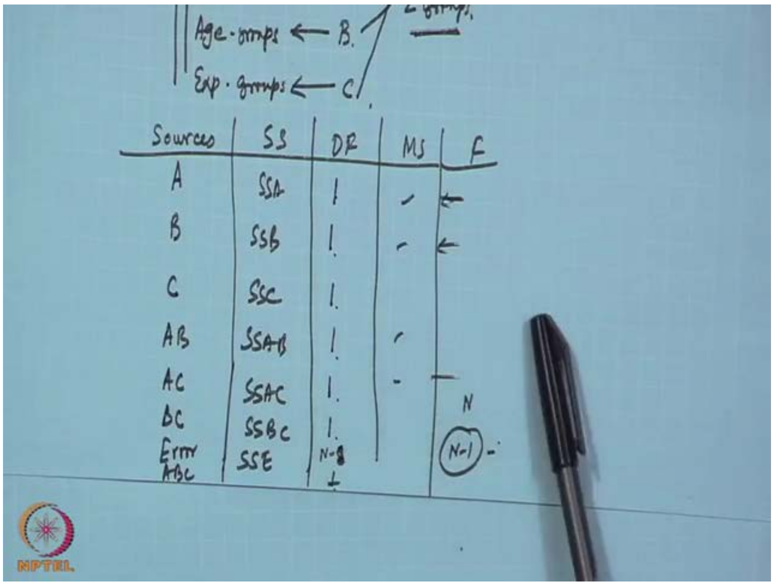
(Refer Slide Time 39:19)



Then question the research questions, here do the workers across job titles differ in their safety behavior, do the workers in different age groups differ in their safety behavior. Do the workers in different experience group differ in their safety behavior, are there interactions amongst job title age and experience groups may be title very significant questions. These are because if we find out that there is difference across job titles, so the prevention action will be accordingly taken if there is no difference across the job titles.

It may so happen that the working conditions are not affecting, it may be something else is affecting or all the working conditions are equally hazardous because the safety performance in the mine studied are not good in the sense. You can answer all those questions, so in order to answer these questions if I ask you what the sources of variation are, you have to use ANOVA.

(Refer Slide Time 40:35)



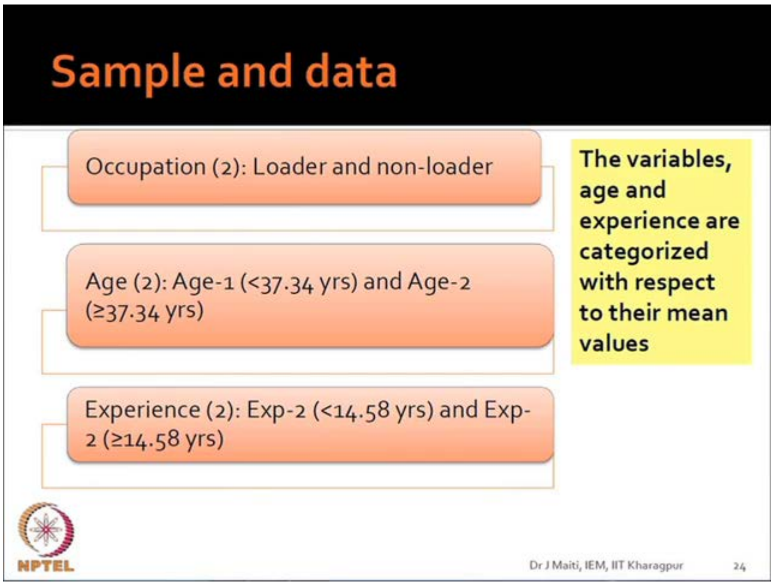
Definitely, we will say first is job title, what will happen now. Job title factor A, then second one is our age group, so age groups that is your factor B, third one is experience group this is your factor C. So, we have three factors each with two groups we are saying it three way ANOVA because three factors are considered. So, if I say that what are the sources of variation then you can write, here the sources is first is factor A, then factor B, then factor C, interaction A B interaction, A C interaction, B C and error. Is it problematic, very simple, so this is the beauty of ANOVA, ANOVA is not conceptually that difficult, it is easy to understand also?

So, S S you will be calculating SSA, SSB, SSC, SSAB, SSAC, SSBC and SSE then definitely the last one is that is the total one. So, your degrees of freedom you write down then you write down the M S, you write down F, getting me, we have taken two groups and each of the variables having two levels. So, 2 minus 1 1, again 2 minus 1 1, 2 minus 1 1, 1 into 1, 1 1 into 11, this is 1 then there is one more sources of variation that is what A B C three way interaction.

So, that will also become 1 and what is the number of it all depends on the number of observations you will be taking that N minus. What are the degrees of freedom 1 2 3 4 5 6 7, so 7 that will be N minus 7, N minus 1 will be the total. So, N minus total minus 7 N minus 8 N minus 8 will be the error degrees of freedom, so you will be finding out M S.

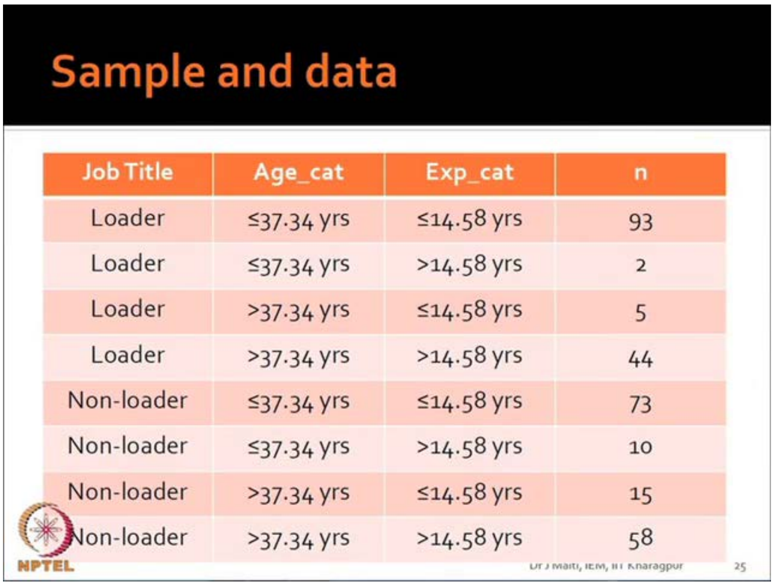
Accordingly, then for every factor you will be finding out the F value and you have to see that whether there is difference or not ok, let us see that the study.

(Refer Slide Time 43:46)



Now, we have taken these are basically that group and there are 300 observations, so total observation is 300.

(Refer Slide Time 43:58)



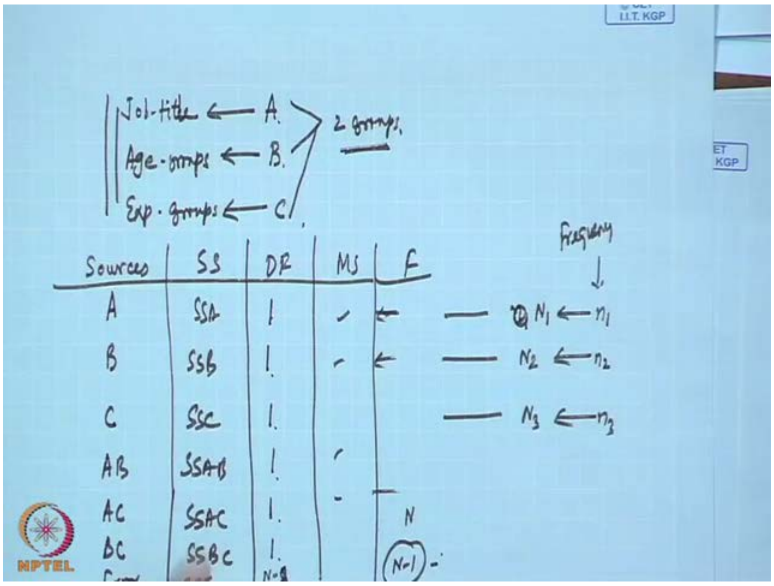
Job Title	Age_cat	Exp_cat	n
Loader	≤37.34 yrs	≤14.58 yrs	93
Loader	≤37.34 yrs	>14.58 yrs	2
Loader	>37.34 yrs	≤14.58 yrs	5
Loader	>37.34 yrs	>14.58 yrs	44
Non-loader	≤37.34 yrs	≤14.58 yrs	73
Non-loader	≤37.34 yrs	>14.58 yrs	10
Non-loader	>37.34 yrs	≤14.58 yrs	15
Non-loader	>37.34 yrs	>14.58 yrs	58

Now, see the observations that job title wise this is basically the category wise we have given. Suppose, what I mean to say loader less than 37.3 years of age and experience less than these 14.5 years, these are the mean age and mean experience value 93. So, the distribution is very skued, you see only two observation available 2 in this combination loader less than 37.34 years of age and greater than 14.5 years of experience, this is quite obvious. But, this low value they definitely distort the total authenticity of the estimation and this is one concern.

Another concern is this is 5 only 5 but 5 and more we consider acceptable, so only this portion is difficult than other cases. There are reasonably good amount of observations available, in reality this is a big problem while you go for real study and you will find this type of problem. So, what will be the alternative for to avoid this problem, your studyagan must be good enough?

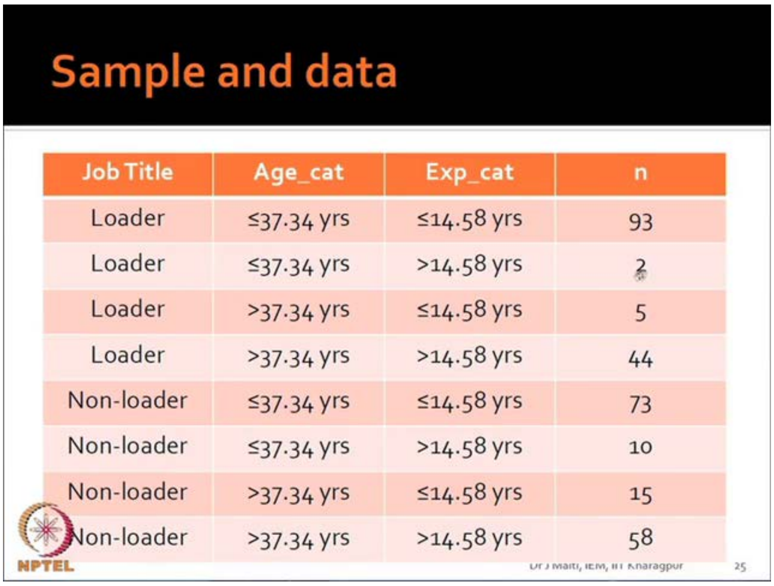
What do you mean by studyagan, before collecting data if you know that this is my structure? Now, you have to know in your system, for in this particular combination how many people are working. See in all combination what is the present level of people or number of people working getting me, so then you randomize based on this what I mean to say.

(Refer Slide Time 45:50)



Suppose my first combination there are N one people working or I can say that capital N one people working, second condition N two working, third condition like this N three working. So, then you will sample N, one from first N, two from second N, three from third in such a manner that there will be frequency matching. But, in no case so one is which frequency matching.

(Refer Slide Time 46:24)



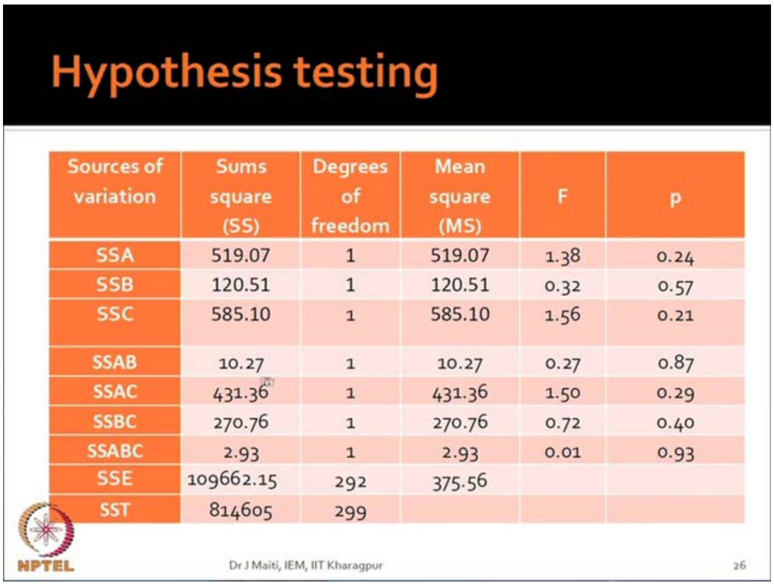
Sample and data

Job Title	Age_cat	Exp_cat	n
Loader	≤37.34 yrs	≤14.58 yrs	93
Loader	≤37.34 yrs	>14.58 yrs	2
Loader	>37.34 yrs	≤14.58 yrs	5
Loader	>37.34 yrs	>14.58 yrs	44
Non-loader	≤37.34 yrs	≤14.58 yrs	73
Non-loader	≤37.34 yrs	>14.58 yrs	10
Non-loader	>37.34 yrs	≤14.58 yrs	15
Non-loader	>37.34 yrs	>14.58 yrs	58

NPTEL Dr J Maiti, IEM, IIT Kharagpur 25

So, if I know that in reality there is not only few, very few people are working in this group, so this is not a good group. But, anyhow what you have happened we have first collected the data and later on we have seen and that is why this problem occur. This problem will occur to you also if you do not do the studied event correctly, what I have discussed in first class study design anyhow with this data we have x.

(Refer Slide Time 46:51)



Hypothesis testing

Sources of variation	Sums square (SS)	Degrees of freedom	Mean square (MS)	F	p
SSA	519.07	1	519.07	1.38	0.24
SSB	120.51	1	120.51	0.32	0.57
SSC	585.10	1	585.10	1.56	0.21
SSAB	10.27	1	10.27	0.27	0.87
SSAC	431.36	1	431.36	1.50	0.29
SSBC	270.76	1	270.76	0.72	0.40
SSABC	2.93	1	2.93	0.01	0.93
SSE	109662.15	292	375.56		
SST	814605	299			

NPTEL Dr J Maiti, IEM, IIT Kharagpur 26

We have done the ANOVA analysis this is three way ANOVA, you see SSA SSB SSC SSAB SSAC SSBC SSABC SSE and SST. So, these many sources of variations and these are the sum square 519.07 coming from factor A that is the occupation or job title. In this case then 120.51 is contributed by age 585.10 is contributed by experience the job title versus age interaction contributing very less 10.27. But, job title these are your experience that is large amount contribution is there then your age versus experience interactions.

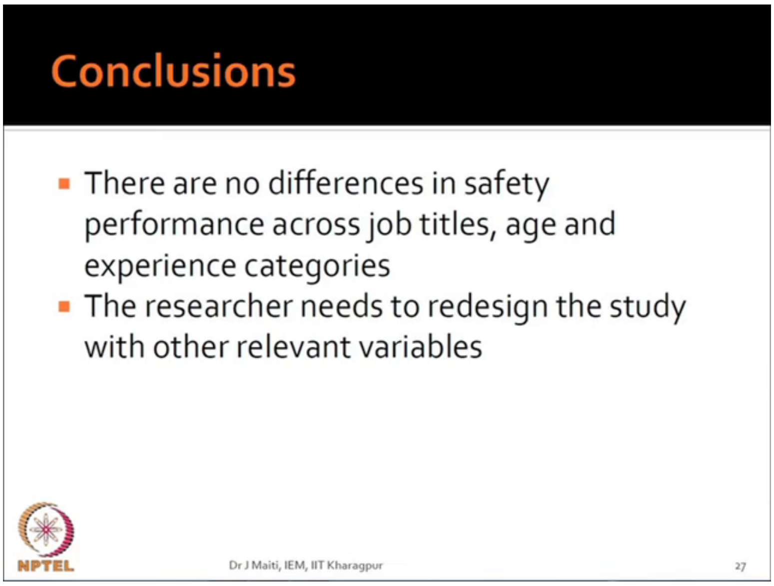
The value is 274.76, but the three factor interaction case the sum square is very very less 2.93 and the say large amount of what is this error one very large. SST will be ultimately the sum, total of all those will be the SST and as I told you there are 300 workers. So, 300 minus 1 this is the degrees of freedom for this and everywhere we are getting 1 degree of freedom for all the other sources. So, error term is 292 the degrees of freedom and in the usual manner you have computed MS and you have computed F.

Interestingly, you find out that none of the factors are significant getting me, so far what I have shown you that yes there are some factors significant, some interaction effects also significant. But, here what happened we found out that hypothesis testing is saying that none of the factors contributing then what we will do ends your work, ends here. What is the point in carrying forward, it is no not required.

So, why this has happened because there may be two three reasons, one of the reasons could be that the variables you have chosen they are not contributing. Second one, I show you that the study design might have been wrong, so concept literature says there must be difference. For example, particularly the young versus old workers, the less experience versus the more experienced workers, this literature says. But, literature is not says that the less than equal to mean and greater than equal to mean years of experience or mean years of age that may not be distinguishing this one.

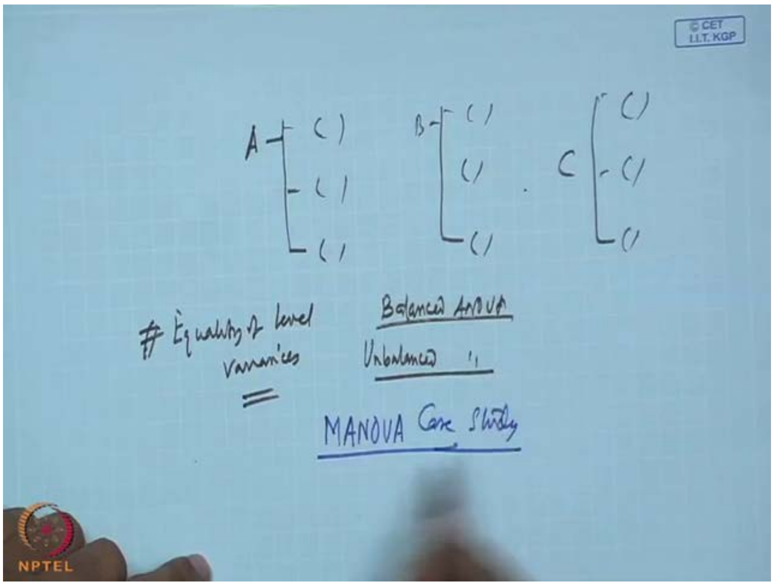
So, if we go for different categorization it will give you, it may give you better result, so ANOVA as a technique fine this is one way you are doing. But, when you come to this type of situation please remember I am now stressing it because we are now entering into the modeling part ANOVA, MANOVA. These are all models then multiple aggression will be another model.

(Refer Slide Time 50:33)



So, that is why what is your conclusion there are no differences this is called the three way ANOVA part in safety performance across job title age. The researcher needs to redesign the study with other relevant variables ok. So, I hope that, now for three way or for more way ANOVA this is unbalanced case. What is the balanced and unbalanced case balanced means the equal sample size unbalanced means across different groups there are different sample size.

(Refer Slide Time 51:13)



So, our factor A, so when you are making them different groups sample size are different, similarly factor B that may be different. Similarly, your factor C that may be different when they are equal that is known as balanced ANOVA when they are different that is known as unbalanced. Unbalanced ANOVA is problematic because you will find out one of the issue will be that equality of error, your population variance level variances. It is most of the time from real data you will find out that there is difference. Any question related to this? So, thank you very much and next class I will show you one case study on MANOVA, so for ANOVA I think we have done enough discussion. MANOVA case study is the second, the next class topic MANOVA case study.

Thank you.