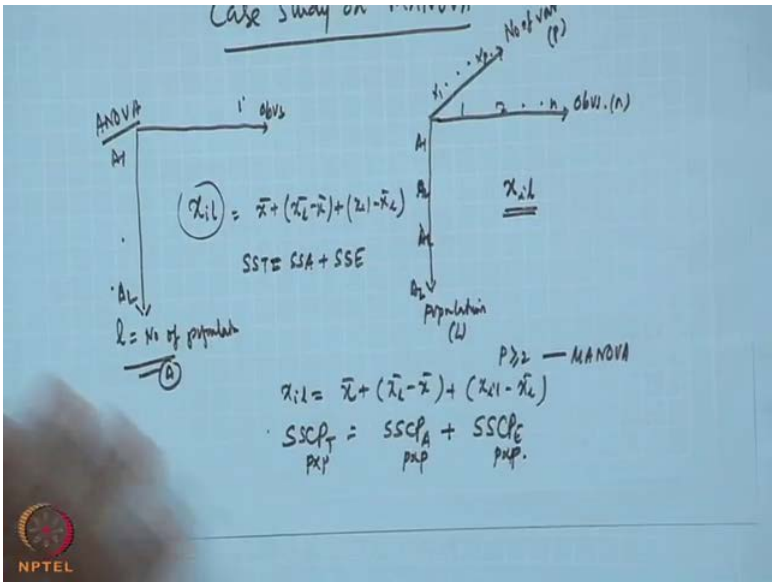
Applied Multivariate Statistical Modeling
Prof. J. Maiti
Department of Industrial Engineering and Management
Indian Institute of Technology, Kharagpur

Lecture - 20
MANOVA – Case Study

(Refer Slide Time: 00:23)



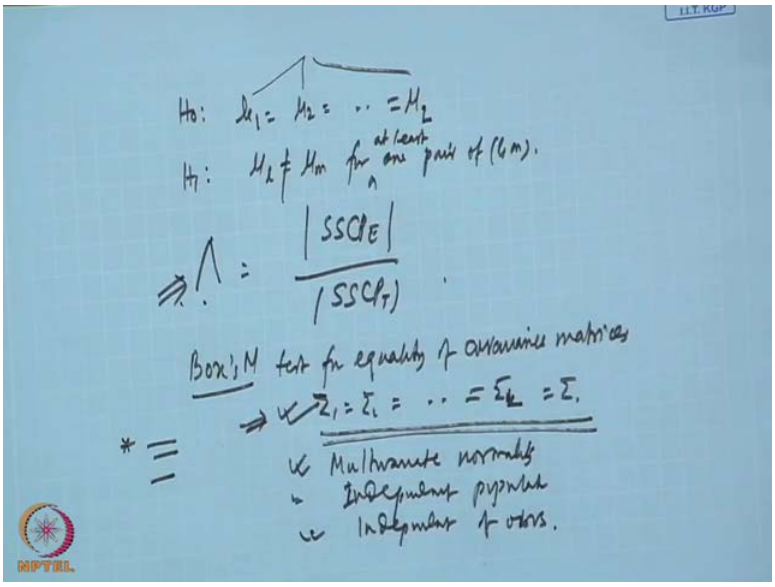
Good afternoon. Today, we will discuss case study on MANOVA. I think you have not forgotten MANOVA and I am just five minutes discussing that basic concept. Then we will straight away go to the case study, because the details of MANOVA are already given to you. Now, we have seen in ANOVA, there is l equal to number of populations, and number of and different observations i and we have taken only one observation x il. Then this observation is broken into x bar plus x l bar minus x bar plus your x il minus x l bar in ANOVA that is the case. Then we have ultimately in one way ANOVA, we have ultimately found out that SST equal to SS, I can write SSA plus SSE. If I say this population, we are denoting in terms of A, so it is A 1 to A capital L, so this one for ANOVA.

In MANOVA, we have just in one way ANOVA, we have just duplicated the same thing, but with one addition that addition is number of variables. I think you can remember this, this observations n, populations that capital L, so A 1, A 2, then A 1 to A

capital L, then 1 to n and it is x 1 to x p. So, when p is greater than equal to 1, that it is the issue is no longer ANOVA; it is a MANOVA issue multivariate analysis of variance.

So, our notation was similar like ANOVA we have kept, but this quantity is a vector quantity and we have described in the same manner like x bar plus x l bar minus x bar plus x il minus x l bar. This is again we are found out that the SSCP matrix total equal to SSCP, suppose the A factor or a population of A plus SSCP E. If there are p variables, this will be p cross p, this will be p cross p, this will be p cross p.

(Refer Slide Time: 03:26)



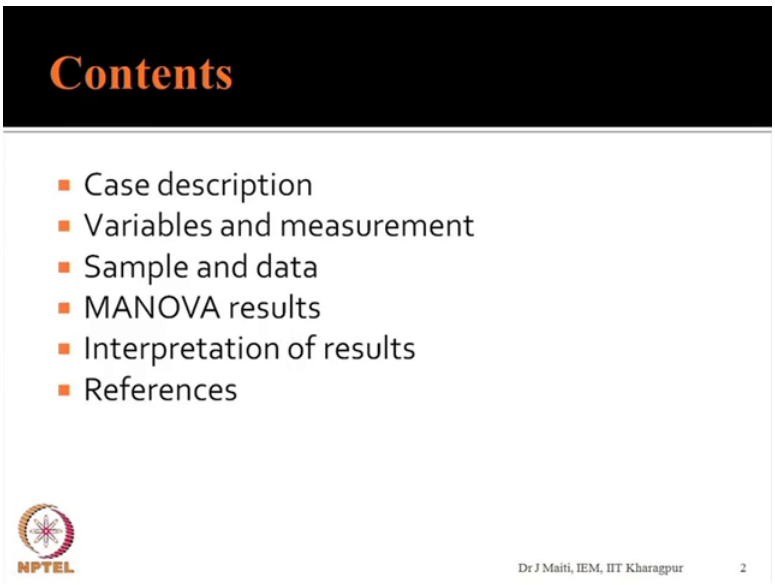
Also, what you have learnt so far in MANOVA that we will test the hypothesis. That hypothesis was that mu 1 equal to mu 2 equal to like this equal to our, we are saying mu L and alternate hypothesis was mu l not equal to mu m for one pair of, for at least one pair, for at least one pair of lm that. Then we created Wilks lambda. Wilks lambda is determinant of SSCP E by determinant of SSCP total and we created one say another statistics and that follows chi square distribution.

Then, based on this, we have seen that whether the statistics is showing that whether there is a significant difference across the mean values of the different level of the population. Before that, what we have done, we have also done box's M test for equality of covariance matrices, covariance matrices. That means sigma 1 equal to sigma 2 equal to sigma L equal to sigma. If this condition satisfies, then what happen this with other statistics values, those things will give us much accurate result compared to situation

where these are not equal. If the covariance matrices are not equal, can we still use MANOVA? We can still use MANOVA. In that case, the suggestion is try to make equal sample size and use not only Wilks lambda you go for some more statistics like pillai's trace, roy's largest characteristic root.

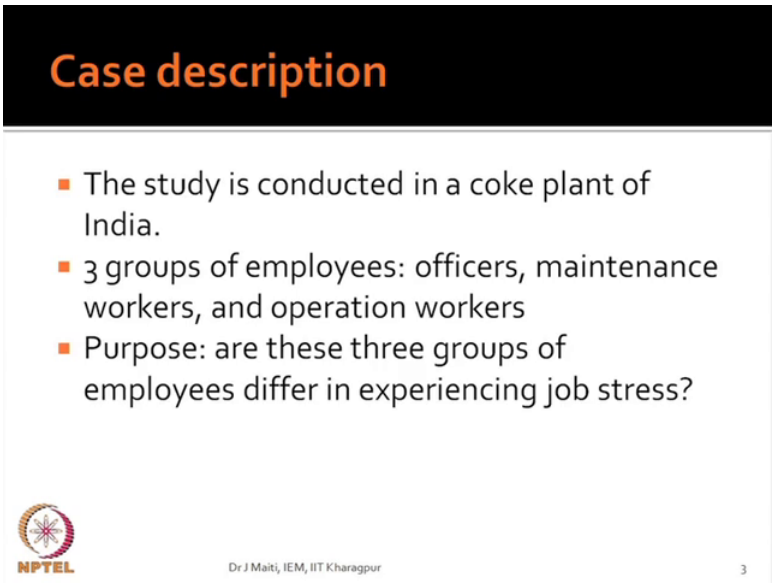
So, there are many techniques available. We will see all those techniques. If all those techniques suggest that there is difference, then only you will go for difference, but even then please keep in mind that where, when there is violations of assumptions like multivariate normality, another one and the second one is, the first one, first one is normality, second one is this violation is there and the populations are independent; independent population that is not a big issue because it is usually true independent population, then independence of observations, independent of observations. So, this thing we have already seen. So, with this background, I am presenting one case study and see the result.

(Refer Slide Time: 06:37)



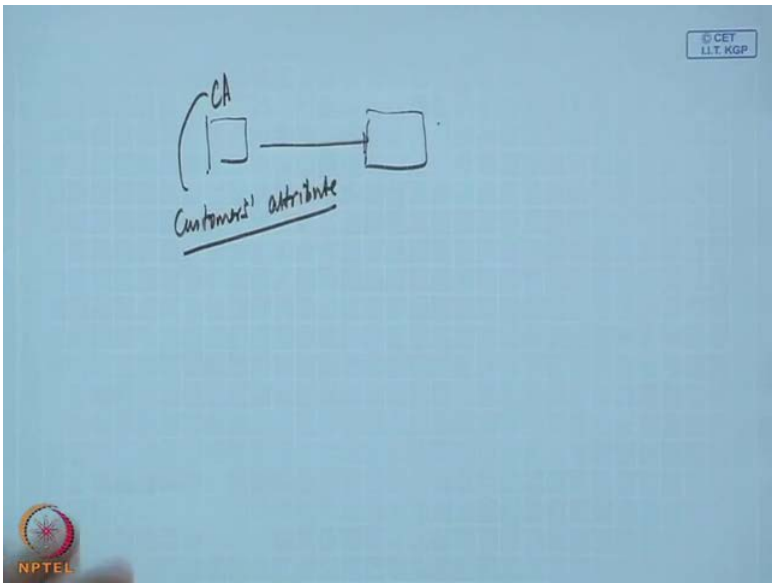
So, the content of the case study presentation is first I will describe the case, then what are the variables we have considered and how we have measured the variables, how, what data we have collected, what results we got using MANOVA and what are the interpretation of results and finally some references. This is a real case study we are doing now.

(Refer Slide Time: 07:09)



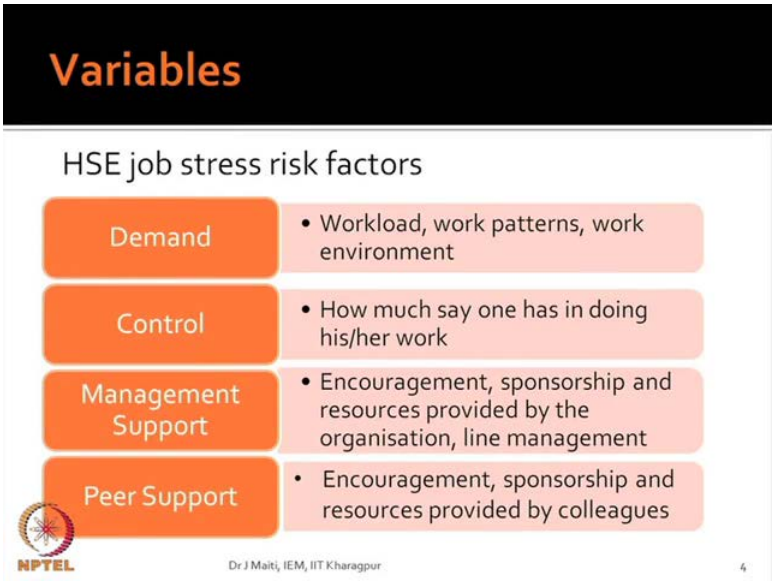
So, the study is conducted in a coke plant of India. 3 groups of employees we have considered, officers, maintenance workers and operation workers. Purpose, are these three groups of employees differ in experiencing job stress? You all know job stress is now, nowadays it is a big issue everywhere across all job titles, across all sectors starting from your hardcore heavy industry to small scale industry to office workers to IT workers; job stress is a big issue. This work we are doing for one company. They want to design for, they want us to design something for job stress that mean to manage job stress and that is also engineering design. But, as you all know that engineering design starts with customer requirements, so you must understand what is the requirement?

(Refer Slide Time: 08:12)



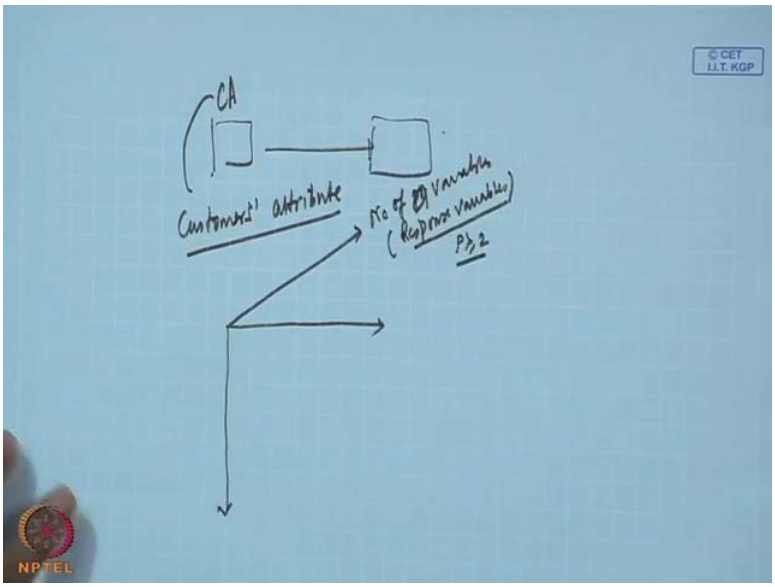
Then, accordingly you design and then the design will ultimately fulfill the requirement that is the process. So, this study is coming under customer requirements or customer attributes, customers' attributes identification or customer requirement analysis, this is our study.

(Refer Slide Time: 08:37)



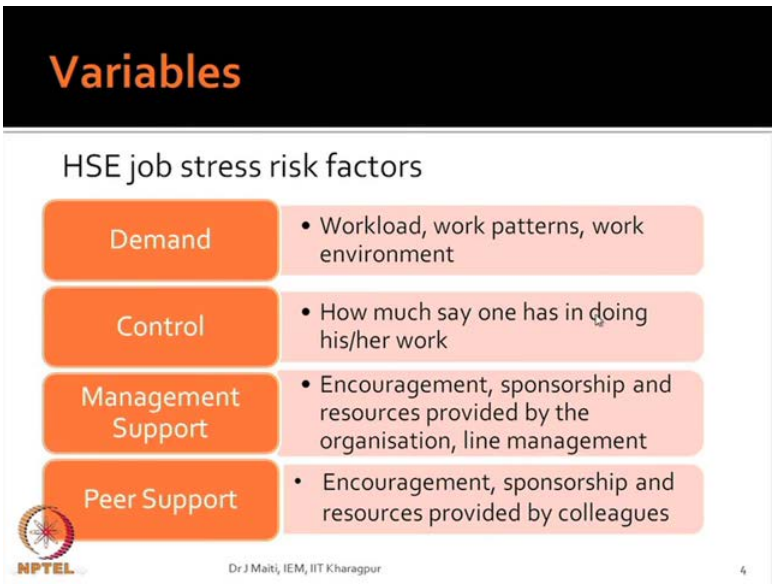
So, we have considered HSE job stress risk factors. What you require to do?

(Refer Slide Time: 8:50)



You must have response variables. In MANOVA, you have seen that number of variables; this is one of the dimensions here, number of variables. So, these variables, we are saying these are response variables plus I think yes, this is related to response variables only. Now, response variables can be 2 or more here in this case MANOVA, What are the response variable in our study?

(Refer Slide Time: 9:34)

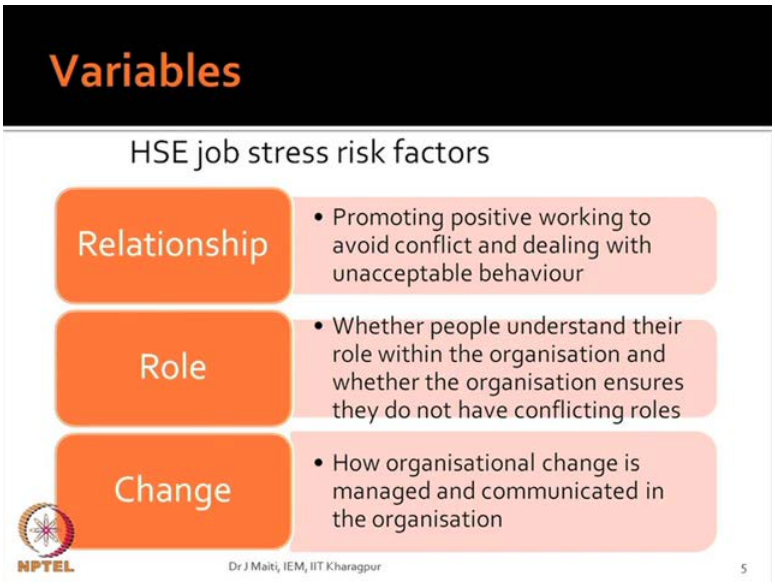


The person working is imposed or exposed to certain amount of demand because of its job that that like workload, work pattern and work environment. Then there are certain

amounts of controls because you have the responsibility, you have the job, you have to execute the job, and you must have control on your job; so how much say one has in doing his or her work that is the definition of control here.

Then management support, so encouragement, sponsorship and resources provided by the organization, line management, these are coming under management support. Everybody working in an organization requires all these things; otherwise he or she cannot work in this where support is very important. Fourth one is peer support means your colleagues must support to your work, your feelings, they must help your day to day work, and they must help you in for your well being also. If those things are absent, then job stress will become more.

(Refer Slide Time: 10:38)



So, then relationship, promoting positive working to avoid conflict and dealing with unacceptable behavior that is in the work, in the job. Then role, role whether people understand their role within the organization and whether the organization ensures they do not have conflicting roles. When you are given a job, for example, you are doing a research work, so it is a time bound definitely job, the three years, four years you will take to complete.

So, you must know what you are going to do and your guide also; at the same time, you also support that you have used the your ways of doing guide is looking for some way and you are doing in some other way, then there will be a conflict. So, similarly, in the

job the organization wants the job to be done in certain way and you are doing in some other way, so what will happen then? Conflict will be there. When there is conflict, there is stress. Getting me?

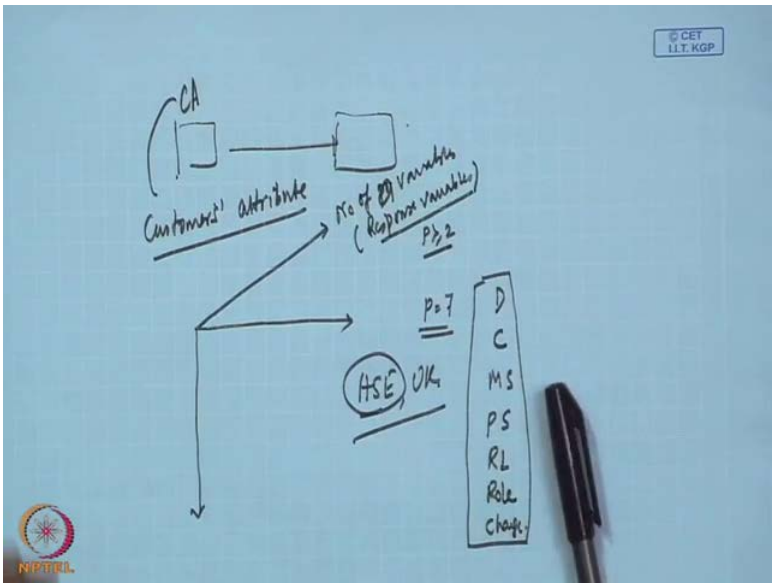
So, we are saying that roles would be crystal clear. There should not be any ambiguity in understanding the roles, what I have to do. If there is difficulty in understanding what I have to do and ultimately even my performance will be very poor and because of poor performance, the stress will accumulate and one day burn out situation will come. That is very typical issue now in all industrial sector, even in academics also nowadays, it is where the situation is like this.

Then, finally, the other variable is real change. Many a times what happen, you have seen this is a case study. So, we understand the totality in how we are going for the case study, not the MANOVA technically we will be discussed here. We will be discussing a case study with the application of MANOVA. Now, change is another. So, that is why I am explaining. So, please when you go for similar studies, either in your research or project or any consulting work or any of your professional work, you will understand the purpose of the study. The variables must be and the variables must be clearly understood.

So, you must know for the purpose, what are the key variables? Here, we are talking about response variables; they are HSE seven risk factors. So, the seventh risk factor is change. Change means what? There is a change in job profile in that sense that organization want the job to be done by supposing working to group working, or group working to some mechanization, semi mechanization, some changes is required because of improving productivity.

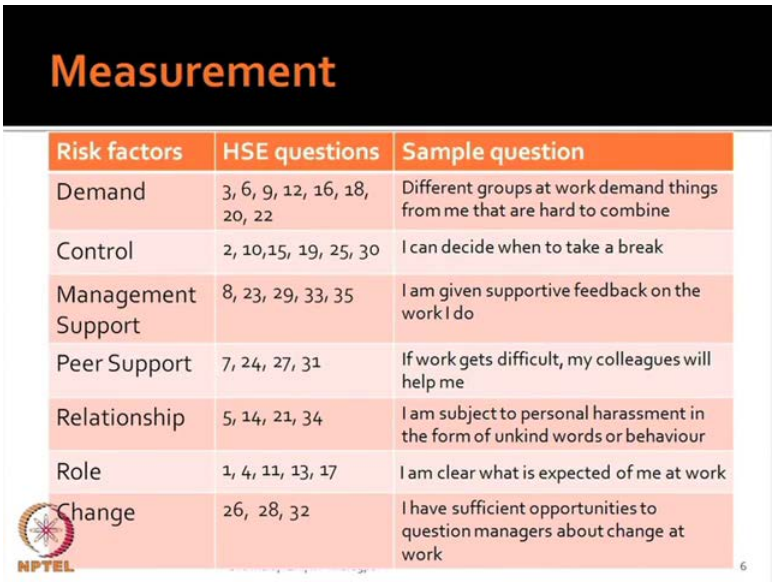
So, these changes must be properly managed by the organization, otherwise what will happen? You will find out the people will not accept, worker will not accept it and worker in the sense the employees will not accept it and ultimately, employee will be frustrated also because of this and this change if it is not well designed and well communicated, then there is a question of job stress. Getting me? So, there will be in a nut shell. In our case study, how many ps are there? p is equal to 7.

(Refer Slide Time: 14:15)



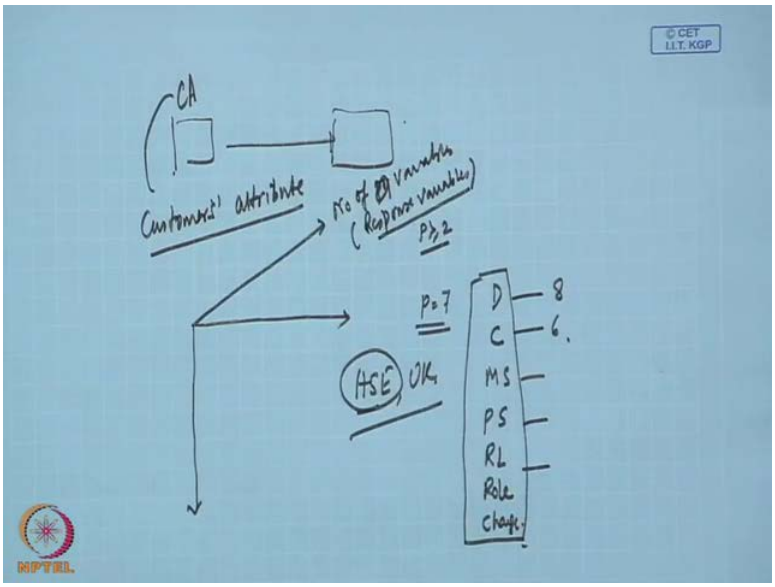
So, 7 response variables starting from demand, then control, demand control, then management support, then peer support, then your relationship, then your role, then change; these are the seven risk factors, which are or other way, I can say employees are exposed to all those risk factors. These risk factors are developed by HSE, health and safety executives UK, health and safety executives UK. So, we want to first work is we have measure those variables.

(Refer Slide Time: 15:06)



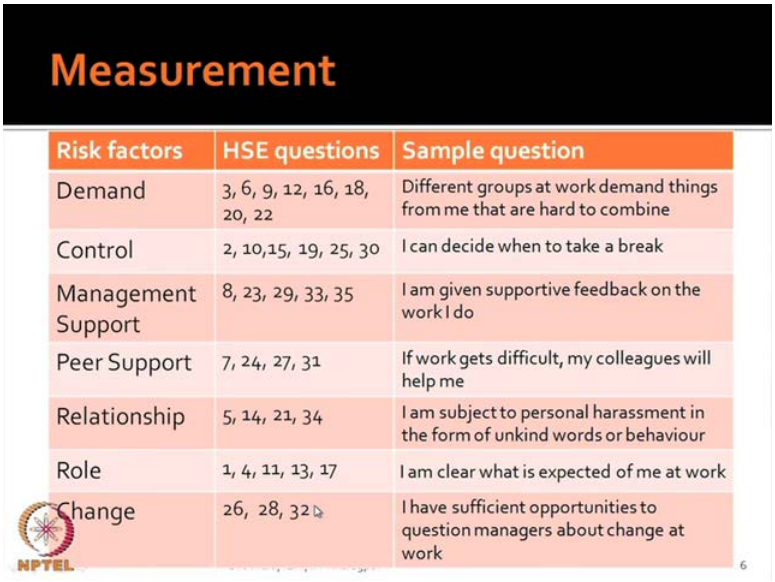
How do we measure? HSE has given questions against each of the factors.

(Refer Slide Time: 15:10)



For example, that in there are two major demands, there are 8 questions given. So, for control, there are 6 questions. So, questions are there, these questions will be administered to different group of people and their responses will be collected. Those responses will be our measured values.

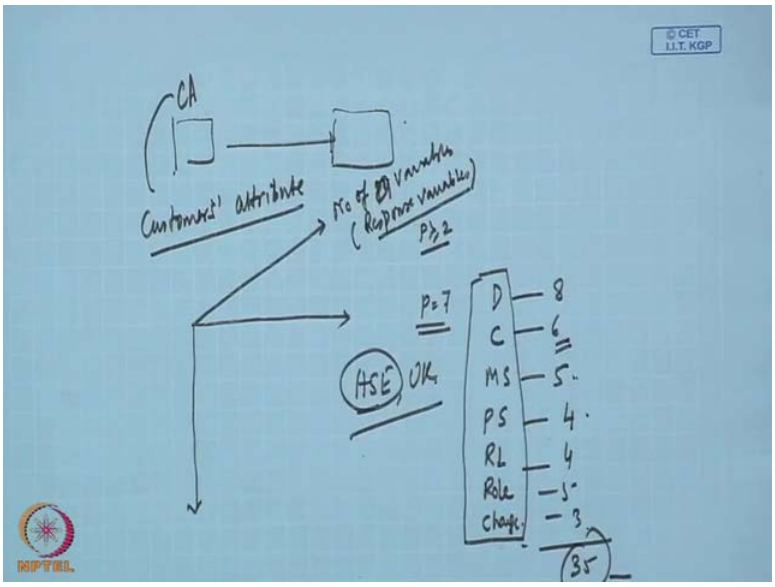
(Refer Slide Time: 15:43)



Now, you see this table here; demand, control, management support, all those things. HSE questions, first is demand 8, control 6 questions, management support 5 questions,

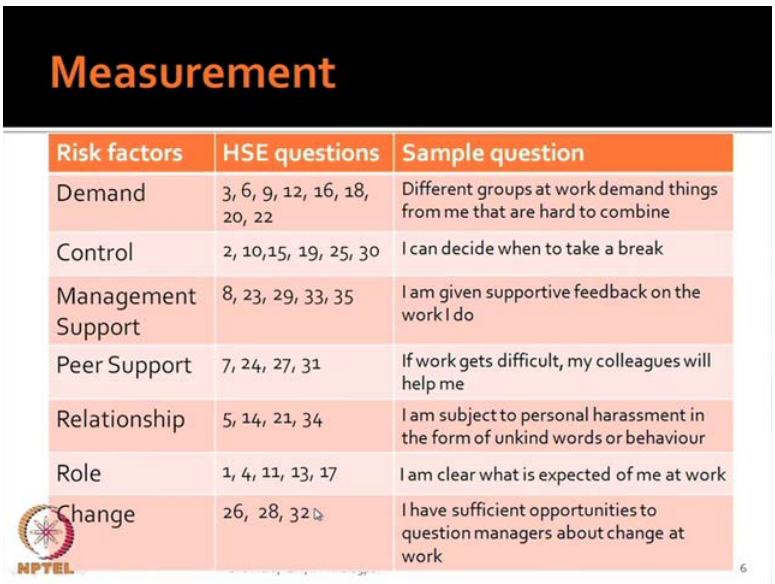
peer support 4 questions, relationship 4 questions, role 5 questions and change 3 questions, totality there are 35 questions.

(Refer Slide Time: 16:14)



I think 8 plus 6 plus 5 plus 4 plus 4 plus 5 plus 3, so 5 plus 8 plus 8, 16, 21, 21 plus 6, 27 plus 8, 35 questions. So, 35 questions are used to measure the job stress suffered by different people at the organization.

(Refer Slide Time: 16:41)

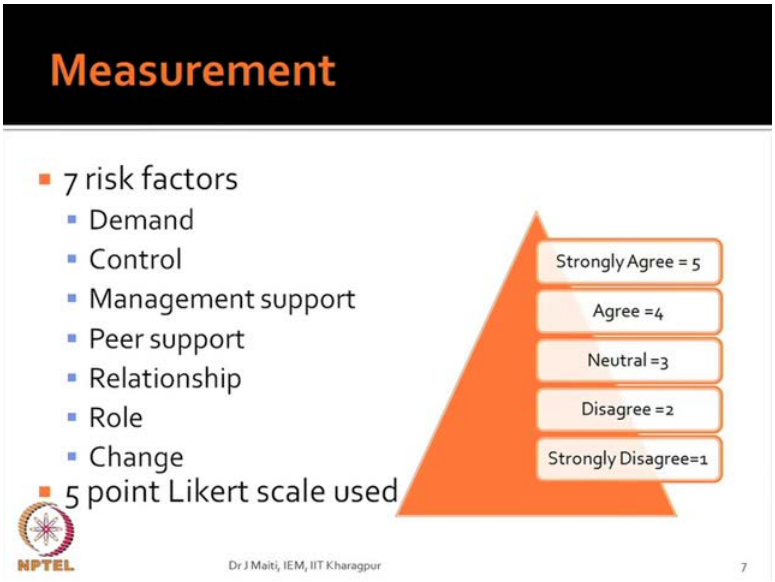


Now, what are the sample questions? For example, the question number 3, this is different groups at, you are been, the employees are being asked, this question different

groups at work demand things from me that are hard to combine. This is one of the questions related to demand imposed on the employee; control, I can decide when to take a break that means he has control; management support, one of the questions is out of this 5 question is I am given supportive feedback on the work I do; peer support.

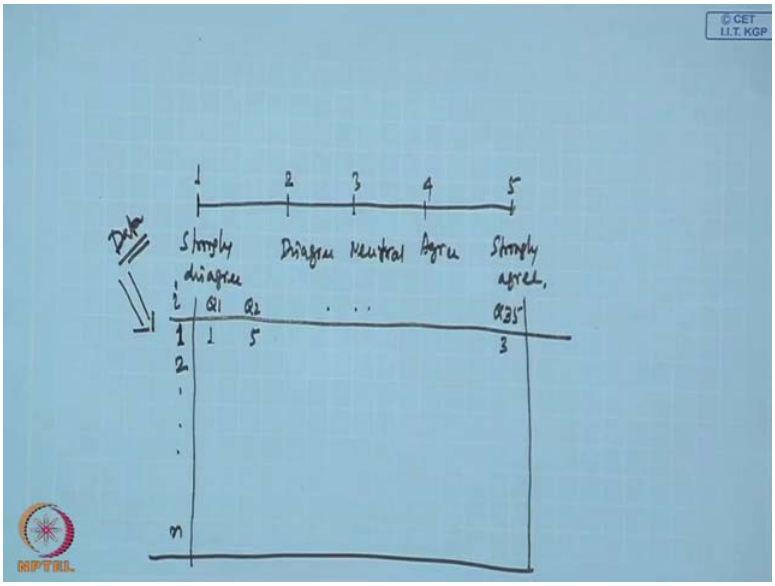
If work gets difficult, my colleagues will help me; relationship, I am subjected to personal harassment in the form of unkind words of behavior role. I am clear what is expected of me at work; change, I have sufficient opportunities to question managers about change at work. So, there are such 35 questions.

(Refer Slide Time: 17:50)



Each of the questions must be measured. What is the measurement scheme? Measurement scheme is likert scale. In the likert scale, when an employee is asked a question related to demand control or any other factors. Then for every questions, you are given a scale 1 to 5 whether the question, with respect to the question, the respondents strongly disagree the statement, disagree, neutral, agree, strongly agree. So, it is basically ordinal data in order.

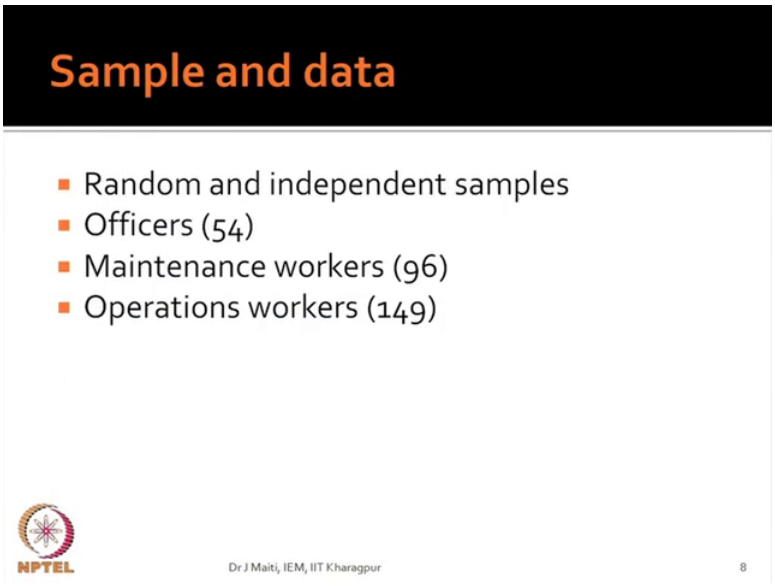
(Refer Slide Time: 18:35)



For every questions you are giving, first one that is the one it is strongly disagree. Then 2, you disagree, 3 this is neutral, not that is neither neutral neither agree nor disagree, fourth one is agree, fifth one is strongly agree. So, then how the data is collected? Data is collected in the same manner for every individual if I say I is the nomenclature for individual. So, first one, he was asked the question 1, question 2 like this question 35 and out of this 35 questions, 8 questions related to demand, 5 questions related, 6 questions related to control, 5 questions related to your management support, 4 questions related to your peer support and so on.

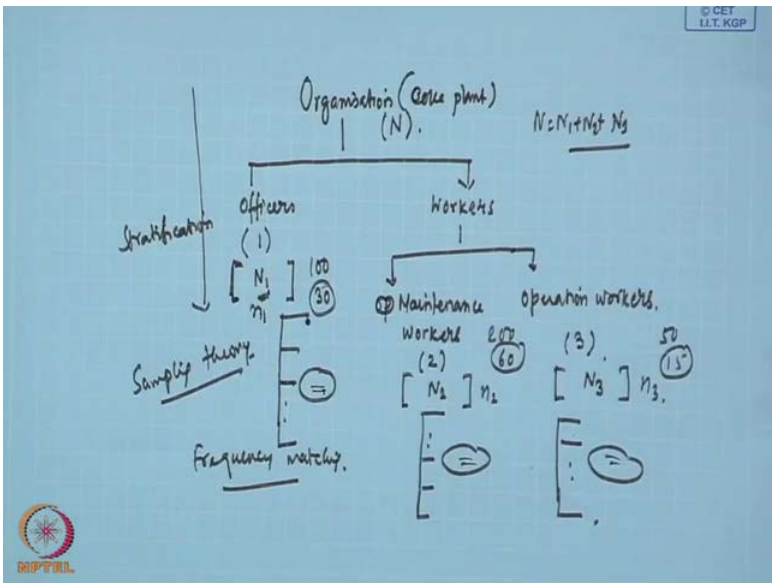
So, then everybody who are respondents there, suppose the first question, he completely disagrees, then the value will be 1. For the second question, suppose he completely agrees, then the value will be 5. Similarly, for the last question, he may be completely neither agree nor disagree neutral, so value will be like this. So, in this manner, suppose n number of persons was asked the 35 questions and they have given the values responses. So, this is the, this is what is the level data collection, first level of information that for everywhere you are collecting like this, that is away from the measurement point of view.

(Refer Slide Time: 20:40)



So, how do you collect the data? This is coming under study design now? What was our study design? Now, see application of statistics will be felt useful only when you do the real work.

(Refer Slide Time: 21:06)



So, our study design was when you have seen that this is the total organization, here organization with respect to our coke plant. So, this one is coke plant. Then what you have found out that there is officers level workers and shop floor workers. So, we are finding, writing like this officers and shop floor workers. Then in officers, workers, shop

floor workers, there are operations workers or I first one can write maintenance workers, this way we have designed, then operations workers, so this is our stratification. This is our stratification.

Now, what sampling we have used here, it is basically stratified randomized sampling. What we have done? We have taken officers as my population 1, maintenance worker as population 2 and operation worker as population 3. Then what we have found out we have seen that what is the number of people working in this category, number of people working in this category, so N 1, N 2 and N 3, these are the number of people working.

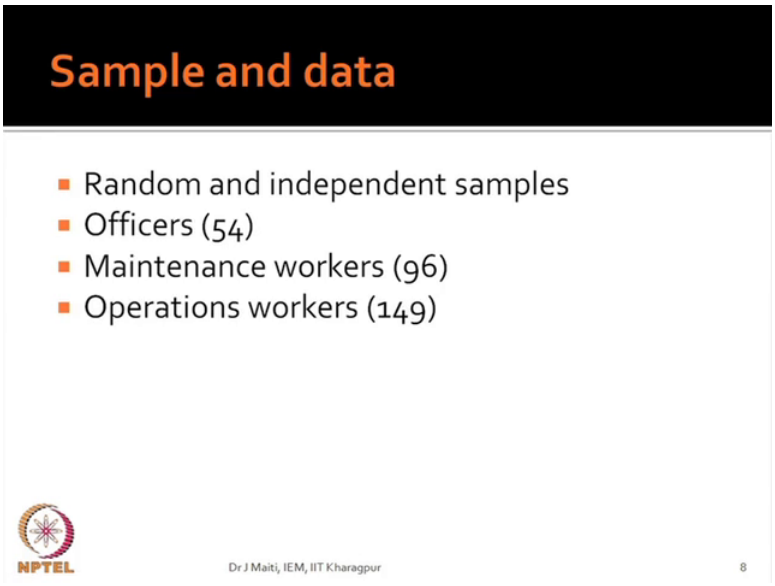
Then there is one concept called sampling theory. In sampling theory, you will find out that if N 1 people are working in this group or in totality in organization that is the N people are working, where N equal to N 1 plus N 2 plus N 3 in this case. This people are working. Now, officers, maintenance workers, operations workers, so many people are there, so again, there are seven factors.

So, you will be measuring everywhere seven factors. So, then how do we decide sample size? What do we mean, what will be the N 1 value, what will be N 2 value and what will be N 3 value, how do we decide? You have to use certain scientific methods. One of the methods I told you that frequency matching. Frequency matching mean if there are 100 workers working here and you suppose you collect 30 and here 200, so it should be 60. Suppose here it is 50, then it will be how much here, 100 hundred means 15, so 15, this is what is known as number of people working, this is mean the sample collected.

But, there is one problem here. Problem is that there are seven variables. If there is one variable fine that, this may be frequency seven variable, some of the variables having high variability compared to others. So, it is that is that is why what is required, you have to find out what is the highest variable, variability variable. Then what is the variable having highest variability?

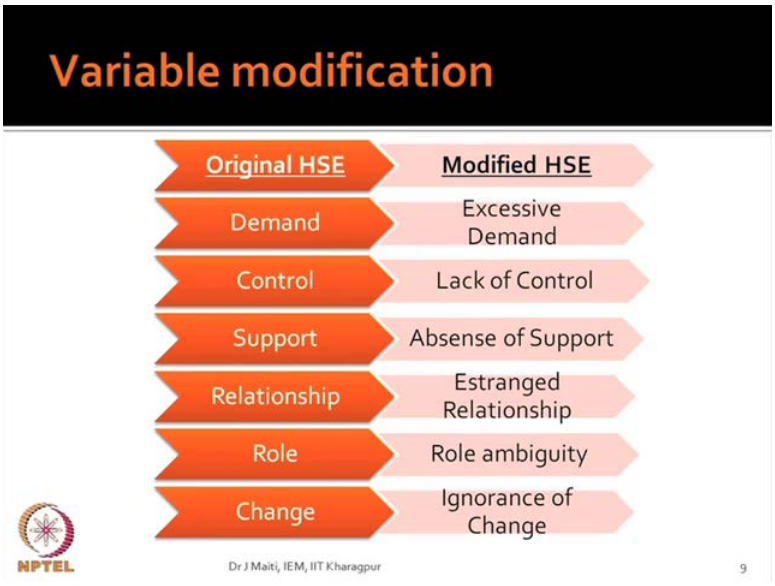
You consider all there may be simultaneously, you consider everything. You go for a pilot study irrespective of may be that stratification or with stratification, with stratification will be better one. Then with this pilot study, you find out which one is having the highest variability. Then there is formula Newman algorithm and other algorithms are there. Using this formula, you will be able to know what is the sample size required.

(Refer Slide Time: 25:14)



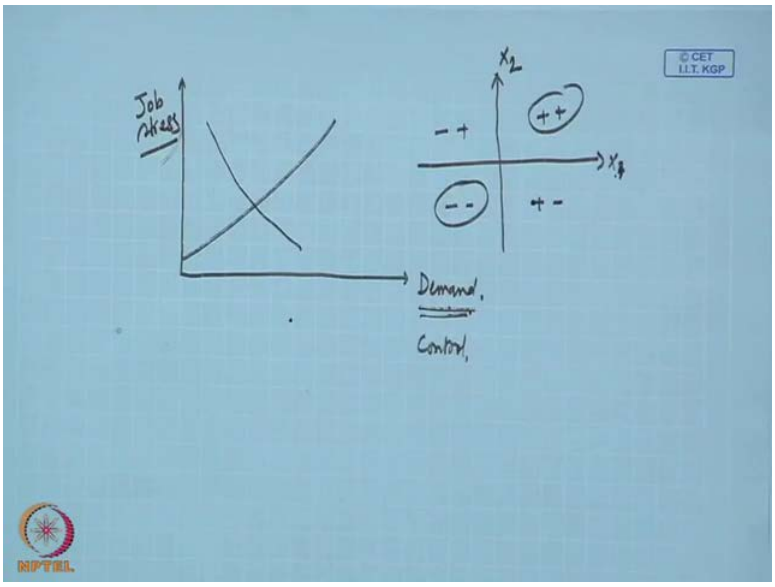
So, doing such calculations, we found out that in officers, we require 54, maintenance worker 96 and operation workers 146 because number of operation workers are more than maintenance workers and officers much less. So, this is what I can say the sample size determination and when we administer the questions to the officers, 54 officers were selected randomly. It is not that we have selected one by another. It is randomly selected. Similarly, 96 maintenance worker, 149 this workers, this is selected randomly.

(Refer Slide Time: 25:59)



So, this is my data. For analysis purpose, we have done little modification. What is this modification? Modification is like this.

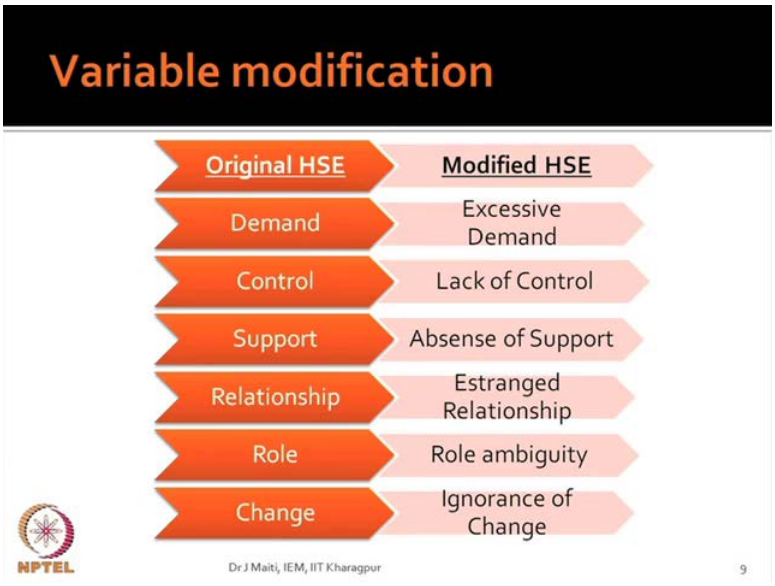
(Refer Slide Time: 26:11)



That job stress is a negative sense variable that means job stress higher means not good, but demand is also, if demand is higher, this is also not good. Then I can say they are in one quadrant, same quadrant. For example, if I say there are two quadrants for that positive positive and negative negative sense or positive negative and negative negative, negative positive that is not this is the quadrant conversion. It may help you later on. This is my positive positive quadrant means if this is my one variable x 1, this one another variable x 2, so x 1 increases, x 2 increases that both positive. This one is, write down x 2, 1 this side, then x 1 negative positive and this one is again negative negative. This is positive negative.

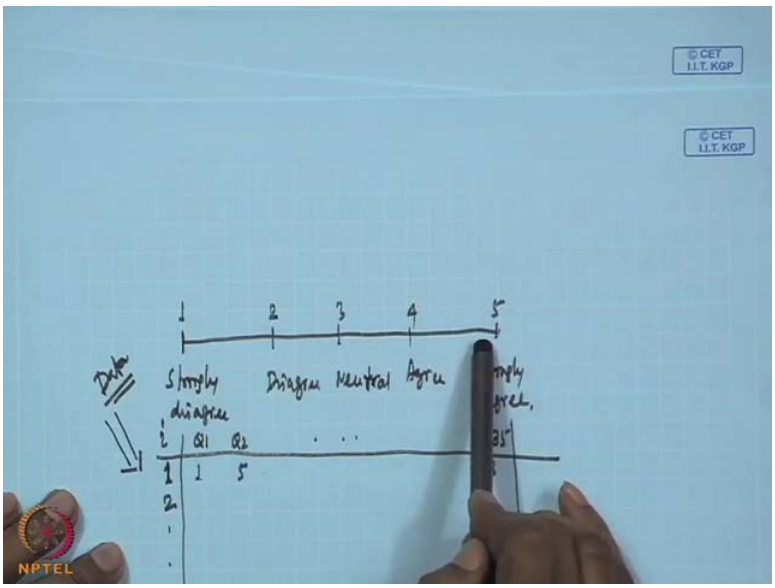
So, what you want wanted here, job stress is a negative quantity. So, we want to convert each of the predictor, the risk factors each of the risk factors in either positive positive or negative negative that sense means same sense. So, demand, if demand increases, job stress increases, so your result will be like this may be. Now, if control increases, if I write here control again, then job stress will decrease. So, that means this is one increasing, other decreasing, we do not want this. We in order for comparable analysis, we have made everything one sided, either increasing or increasing, increasing or decreasing. Here, it is actually increasing, increasing case.

(Refer Slide Time: 28:04)



So, instead of demand, we are writing excessive demand or excess of demand, lack of control. So, lack of control increases means job stress increases. Then support, absence of support, then relationship, estranged relationship, role, role ambiguity we are saying and change, ignorance of change. So, in that way, we have changed the variable and accordingly the data collected, these questions are also required to change.

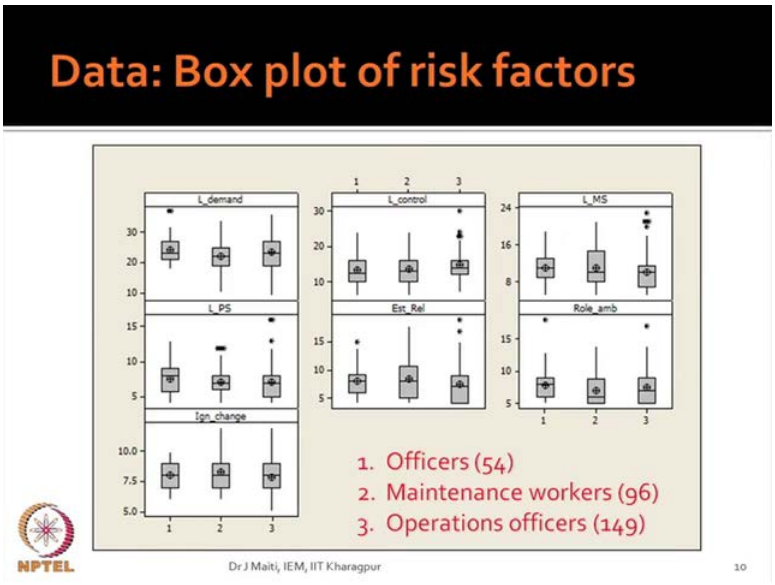
(Refer Slide Time: 28:32)



Suppose those questions, which are positive in nature; for example, controls, strongly agree that mean 5, it will be reducing the job stress. So, we have reversed the questions

also. So, reversal is required for the negative questions, in the sense we are developing. Accordingly, what happened? We have gathered the information.

(Refer Slide Time: 28:56)



Now, this is the status from box plot point of view for there are seven response variables. Variable 1, it is L demand means it is excess demand, then lack of control, then lack of management support, lack of peer support, then estranged relationship, role ambiguity, ignorance of change. We are seeing that the first one is officer's box plot. Second one is maintenance worker plot. Third one is your operations worker box plot.

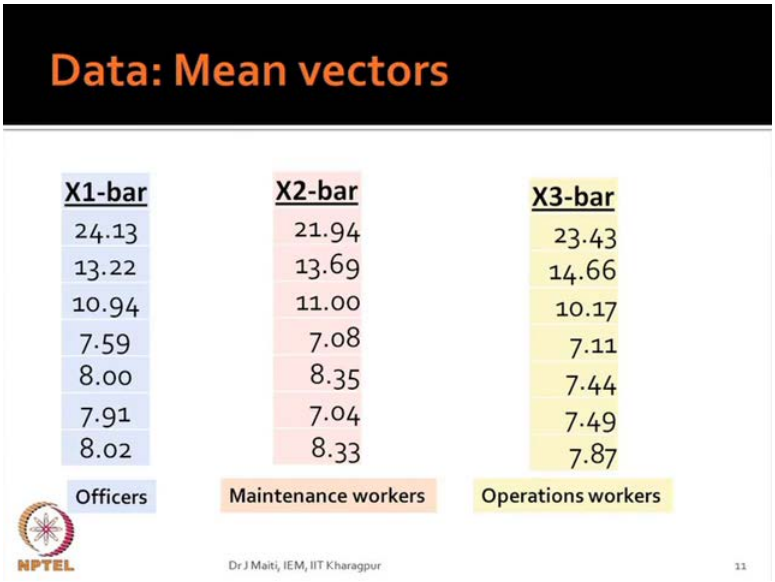
You see the mean value everywhere. There are certain what I can say outliers' values are there. It is always better you first remove the outlier and then. Now, the question is there is differences, it is almost visible that there maybe, there are some differences probably, but we are not able to explore this that how much differences is there. In addition, job stress is such an issue, which is individually oriented means although there are ten factors, all are affecting you. Getting me?

Then, what that mean one factor? At a time, if we see, we may not get the clear picture. We have to see all the factors simultaneously. If you take one factor at a time, then it will be ANOVA. So, we are taking all factors at a time. Why? This is because all factors are simultaneously occurring on the employees. One ANOVA what will happen? The ANOVA will give individual factor differences, but it may so happen that the covariance

structure of all the response variable are governing, what is the stress level imposed on or exposed to the different employees. Getting me?

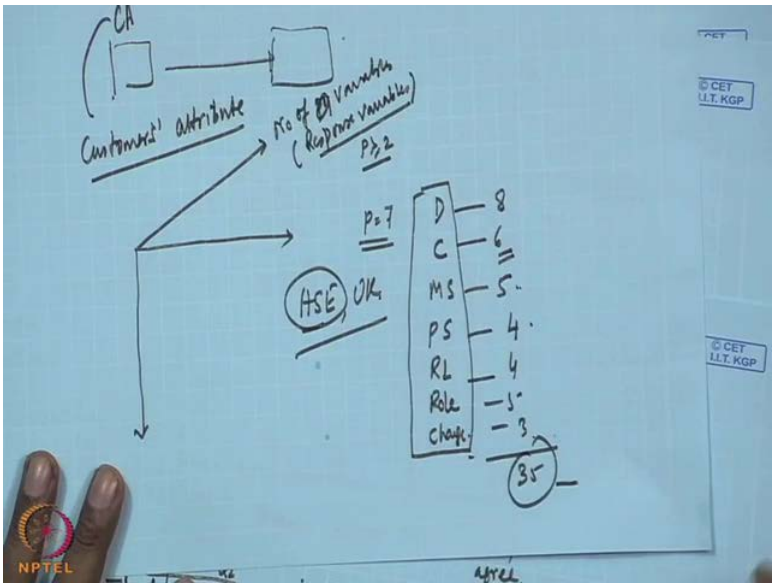
So, what is required then? It is required that we must, we must collectively assess why MANOVA that is why. This is because we will go for MANOVA because to know the collective effects. So, this is what is our first level of information and individual box plot says there is differences, but we are not sure enough that how much differences is there over, the really there is differences or not and if there is differences, which of the variables are making this differences.

(Refer Slide Time: 31:30)



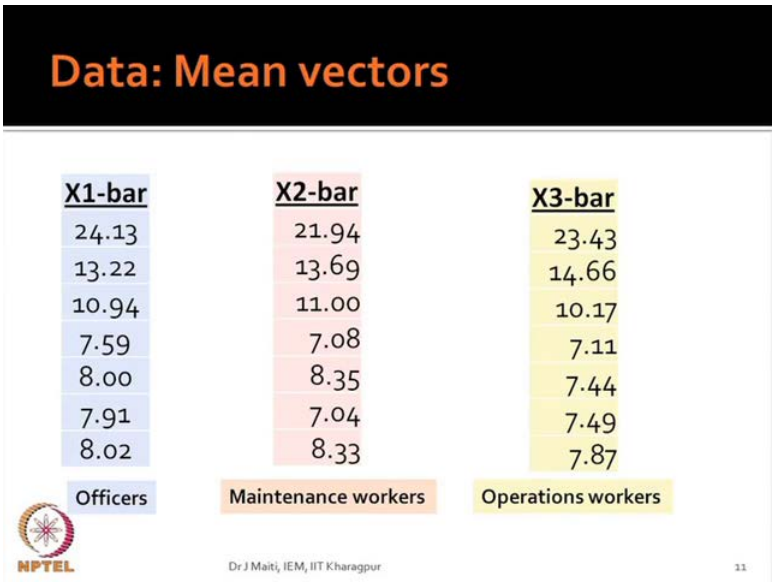
You see now the data vectors. What you require? First, you have developed questions. You have taken the responses. Then you are creating the variable values. How do create values? How many variables are there? Seven variables are there. How many questions are there against this variable?

(Refer Slide Time: 31:53)



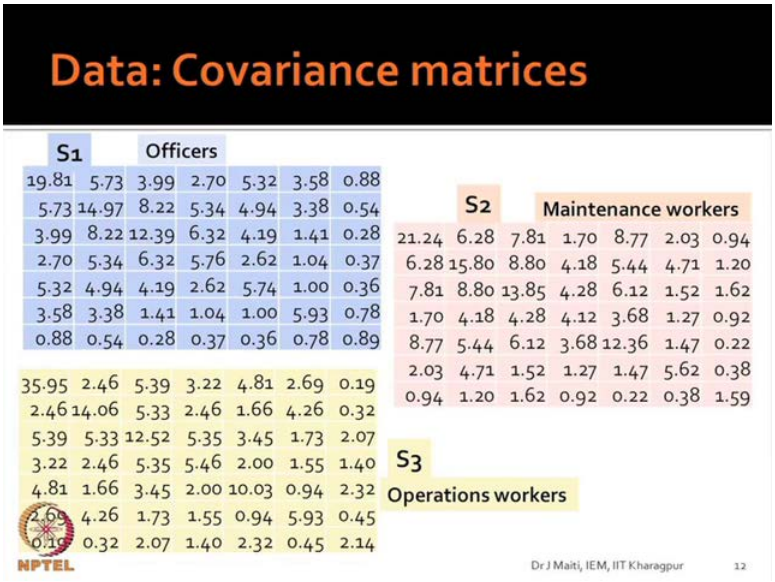
There are 8 questions again depend on all those things. What will be the demand value? So, demand value will be sum of all 8 questions values or mean of this 8 question values. So, we are considering that as 8 questions are asked against demand, we are expecting that that these 8 questions are required from the content validity point of view. So, we are not going for mean responses. We are going for submitted responses. So, that we have done.

(Refer Slide Time: 32:26)



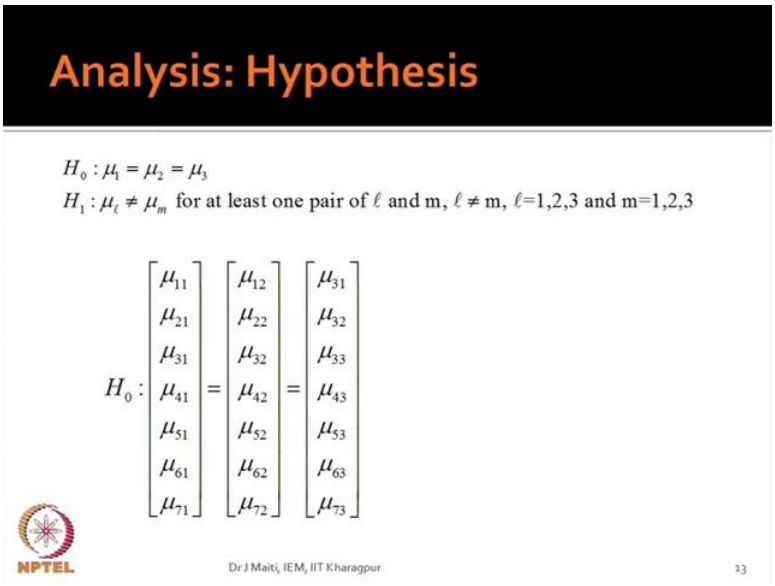
Accordingly, the result, the data whatever we got that is like this. so X 1 bar means that is basically for officer, that this is for demand mean, then control, management support, peer support, your relationship, role and change. So, the same manner officer, maintenance worker, operations workers, mean vector, you have to compute. Understood Tanmay? This is we have to do. You have collected data, all those things. Now, for MANOVA, you require the means, here means are all mean vectors because we have more than one variable once. What more you want?

(Refer Slide Time: 33:18)



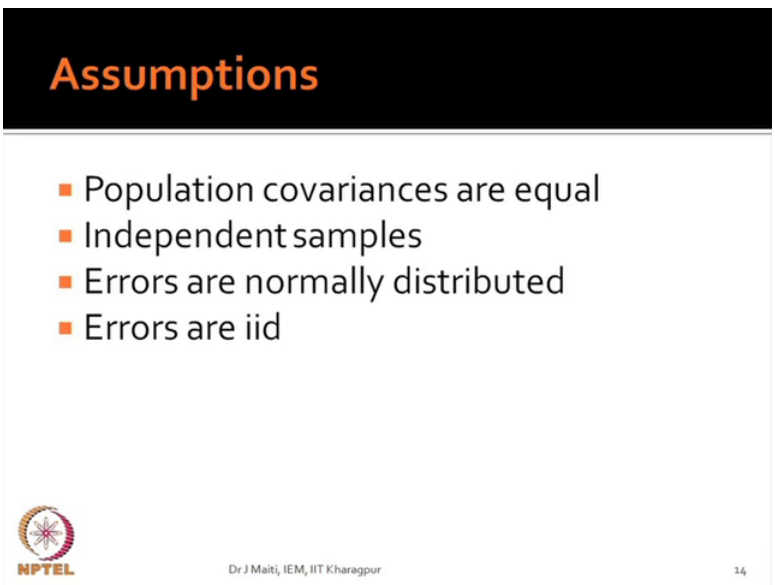
You want the covariance matrices. So, you require to use software or excel; matlab, wherever you are comfortable. So, second your work is for data point of view, you find the covariance matrices. You see that S 1 covariance matrices like this, here 54 officers observations were there. Then S 2 is the maintenance worker, 96 workers were there and S 3 is the operation worker, again 149 observations we have collected. So, your inputs to MANOVA X bar means X bar in the sense X 1 bar, X 2 bar, X 3 bar and S 1, S2, S3 mean vectors and covariance matrices.

(Refer Slide Time: 34:10)



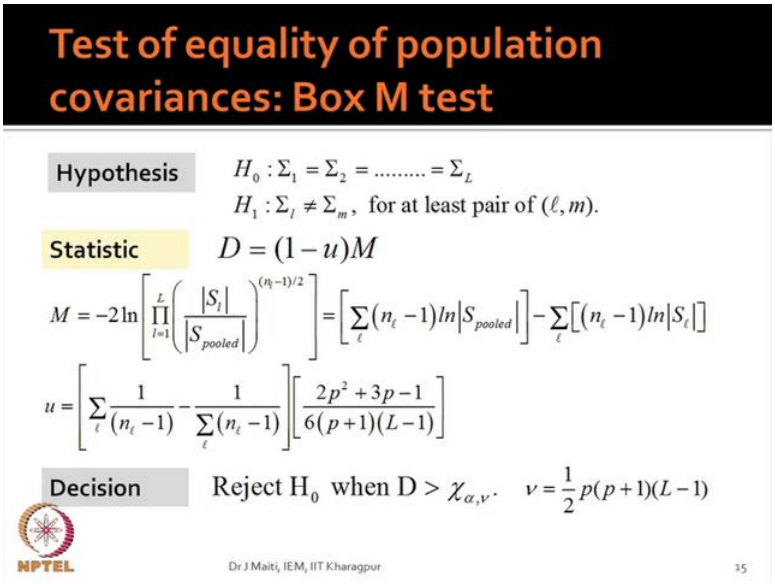
Then, your hypothesis, we are saying that all the groups are equal in the experiencing the job stress. There are seven risk factors, so seven mean values will be there. Each of the mean values across the three groups that is equal. That is why; you see what we are writing here, that first one demand, demand, demand, control, control, control like this. So, every value is equal to the corresponding values in the vectors. What is our alternate hypothesis? We are saying no, no, at least one pair may be mu 11, mu 12, they are not equal or may be mu 71 and mu 72 or mu 72, mu 73, and they are not equal. That is your hypothesis.

(Refer Slide Time: 35:17)



What you require to do? If you want to use MANOVA, you have certain assumptions. Those assumptions must be tested. What are those assumptions, population co variances are equal, independent samples, errors are normally distributed, and errors are iid. Now, independent sample means we are saying that officer and workers or officers, maintenance workers and your operations workers, they are independent. When you are capturing data from officers group, this is not influenced by the operations workers responses. They are totally independent. That is the first. Now, population co variances are equal, what statistics we will use for MANOVA? In MANOVA, box M we will use box m test. So, let us see what is this box M test?

(Refer Slide Time: 36:19)

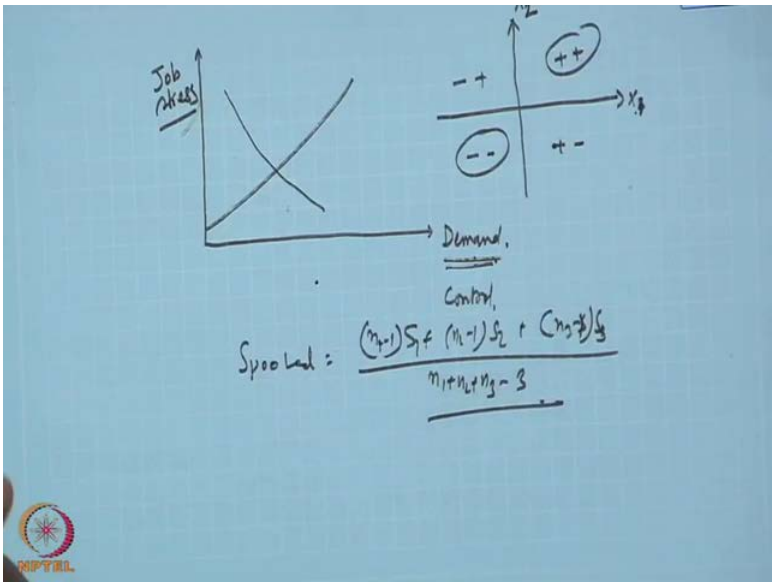


So, box M test, as we have discussed earlier that it says that there is no differences in the covariance structure across the different populations or other way. If we consider it is a factor across the different levels of the factors, where this factor is a designation group, officer, operations worker and maintenance workers. Our alternate hypothesis is yes, there is difference in at least one of the pairs.

Then, we have used D equal to 1 minus u M and M is this S pooled S l this one. Can you remember this one? You have seen this earlier. This one when you we have taken log this is, this is the formulation. So, that means what you require to do? You require calculating the determinant of each of the sample co variances, you also require

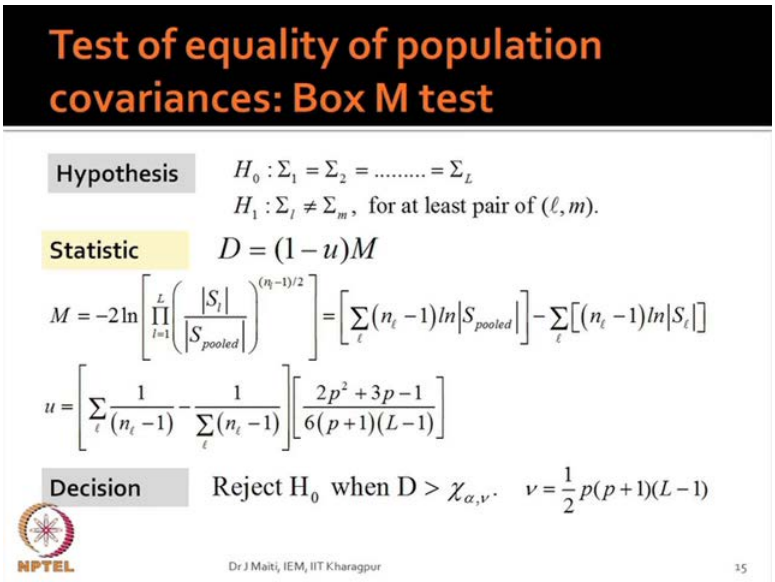
computing S pooled and take the determinant of that S pooled. There is one quantity u, which is given by this measure and you have to test what is your S pooled.

(Refer Slide Time: 37:45)



S pooled will be suppose n 1 minus 1 S 1 plus n 2 minus 1 S 2 plus n 3 minus 1 S 3 by n 1 plus n 2 plus n 3 minus 3. So, that also you have to calculate.

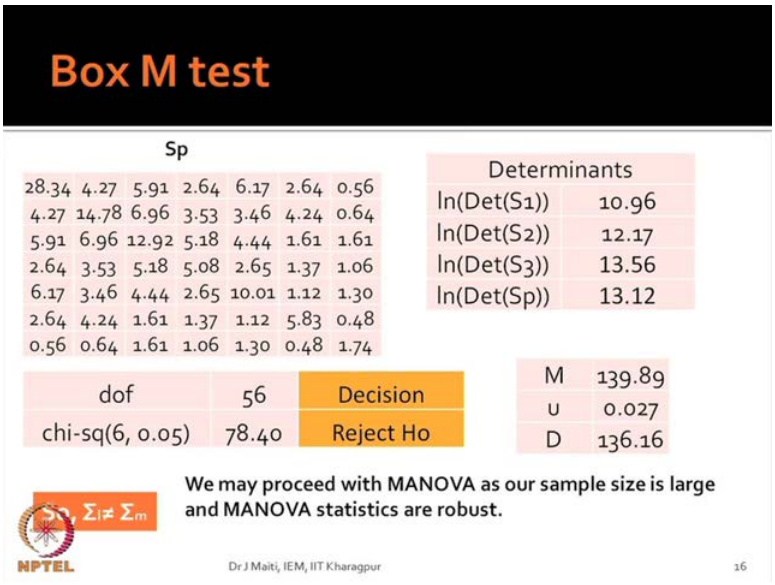
(Refer Slide Time: 38:08)



Now, see then using this D statistics and this D statistics will follow chi square distribution with nu degrees of freedom where nu is 1 by 2 p into p plus 1 into L minus 1.

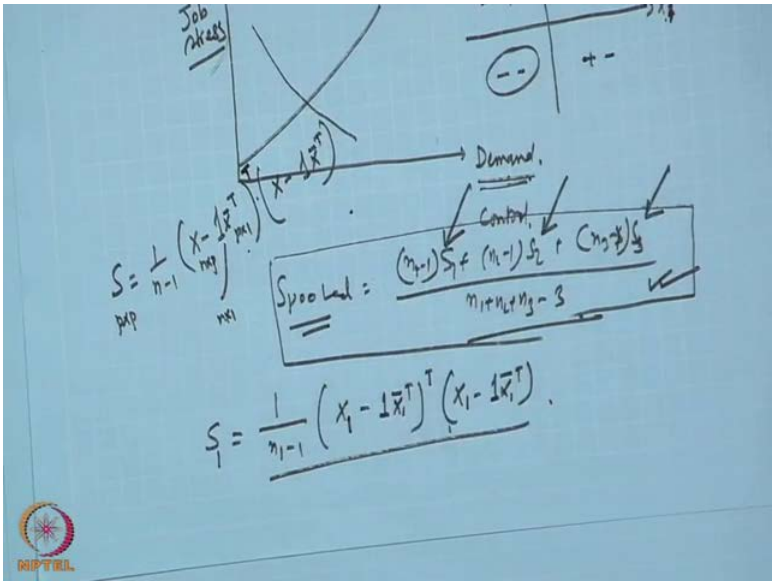
p is the number of variables and L is the number of groups or number of levels or populations.

(Refer Slide Time: 38:28)



Now, see, we will show you the box test here. So, as the data is excessive, large amount of data and we cannot manually calculate. What I have done is I have used excel. In excel, you can very well calculate all those things. So, first we have calculated S p using this formula.

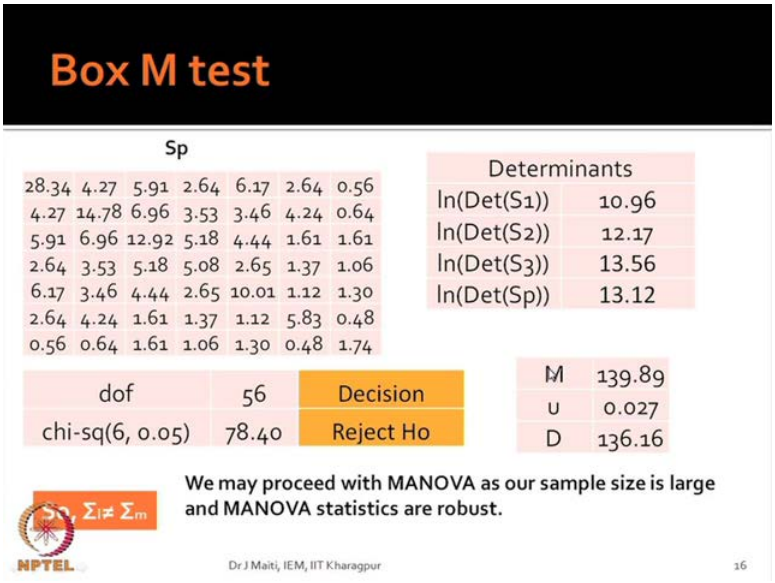
(Refer Slide Time: 38:48)



Using this formula, we have calculated S p and you know, S 1, S 2 and S 3 are also computed. So, what will be the S calculation 1 by n minus 1, then X minus, what is this S will be p cross p, X is n cross p. So, it should be subtracted by mean. Mean is what mean is p cross one

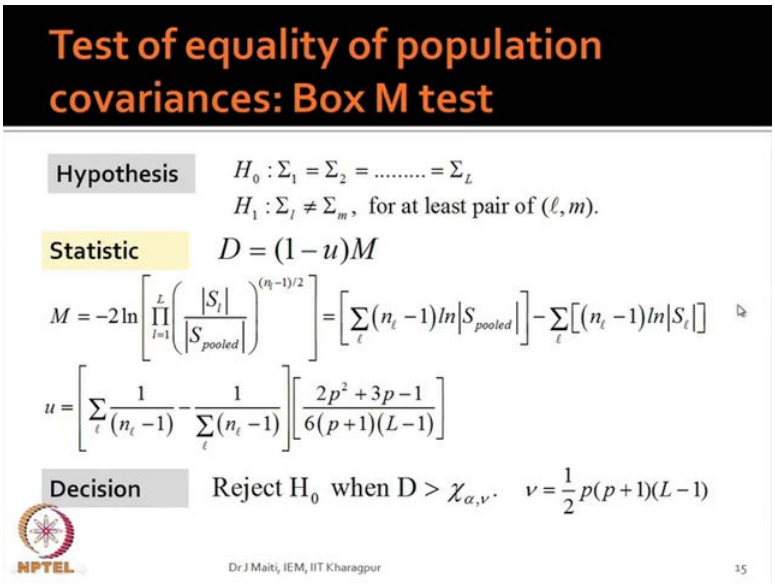
So, you want n cross. So, here one mean transpose, this one is n cross 1, this one into x minus 1 X bar transpose and here this will be transpose. So, I am writing again for everywhere when you calculate S, suppose you are calculating S 1, then you are writing like this n 1 minus 1 X 1 minus 1 X 1 transpose, transpose X 1 minus 1 X 1 transpose, this is the case. So, using this, you will be getting S 1, you will be getting S 2, you will be getting S 3. Once you know S 1, S 2, S 3, you will be getting S pooled and this S pooled value is given here.

(Refer Slide Time: 40:12)



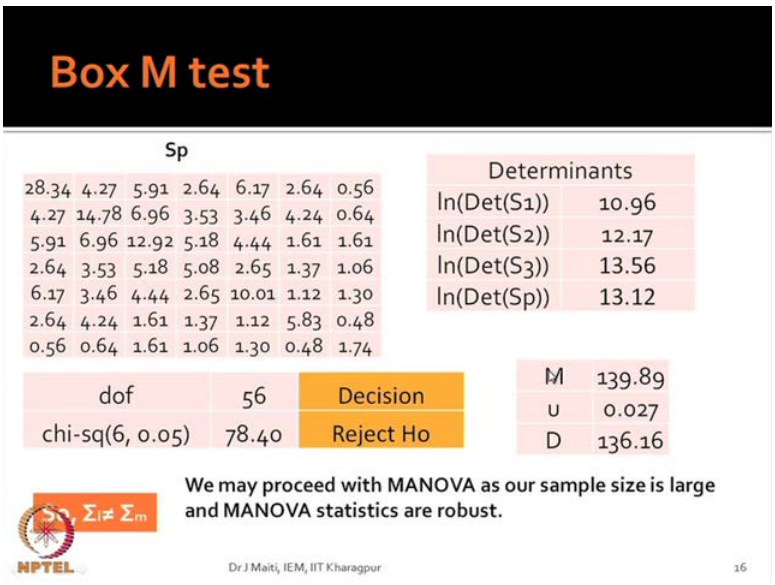
So, it is 7 cross 7 matrix. Now, then you find out log determinant of S 1. What is already there? S 1 is there. So, in this case, this is 10.96, log determinant of S 2 is 12.17, log determinant of S 3 is 13.56, log determinant S p is 13.12. Then we calculated M. How we have calculated M?

(Refer Slide Time: 40:38)



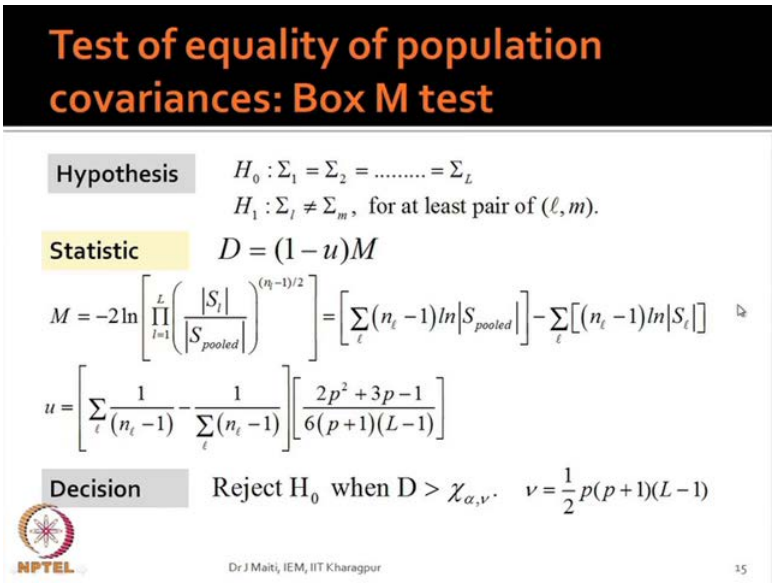
We have first calculated this that l equal to, summation l equal to 1 to capital L n l minus l, this one, this quantity is first computed followed by this quantity, then we have subtracted this value.

(Refer Slide Time: 40:54)



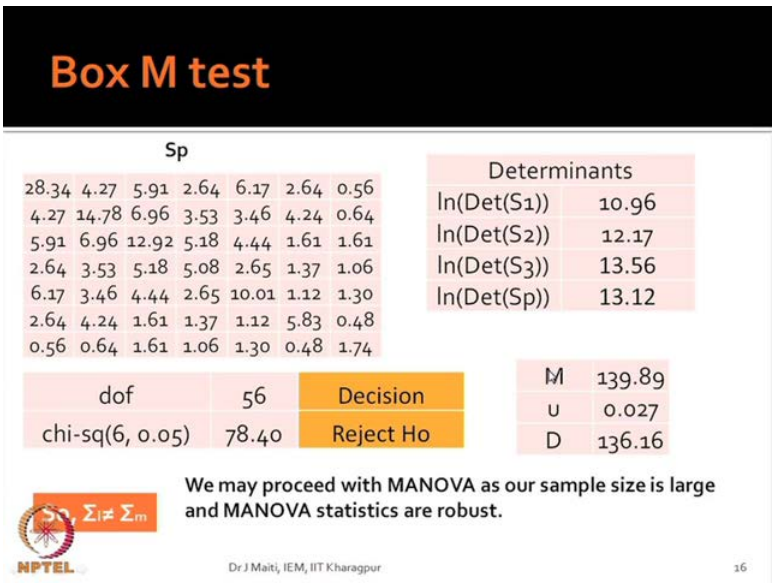
This M value is coming here 139.89.

(Refer Slide Time: 40:58)



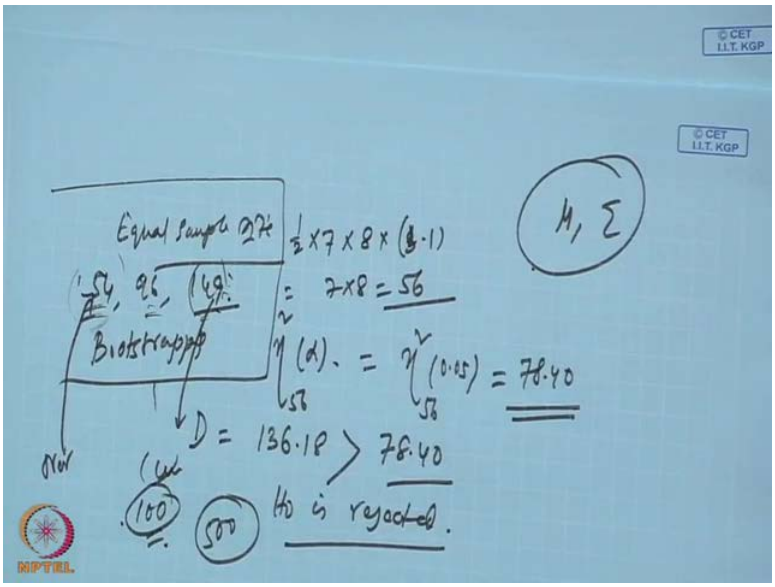
Similarly, we have computed the u value, this minus this and this.

(Refer Slide Time: 41:04)



This also comes to the value like this 0.027. D is equal to 1 minus u into M that is 136.16. Now, what will be its degrees of freedom? Degrees of freedom is half p into p plus 1 into L minus 1.

(Refer Slide Time: 41:22)



So, degrees of freedom is half p means 7 into 7 plus 1 into L minus 1 that is 3 minus 1, so 7 into 8 that means 56. So, our case is chi square 56. Now, what will be your alpha value? You choose if you choose alpha value equal to 0.05, then chi square six alpha 56, 0.05 is 78.40, you will get from table. Now, this 78.40 and your computed statistics D is 136.18, which is greater than tabulated 140, that means null hypothesis is rejected.

So, H 0 is rejected. What does it mean? It means co variances are not equal. So, what should we do? One of the assumptions is co variances, co variances must be equal, but we are unhappy because covariance matrices are not equal and we are not able to go for ANOVA that is MANOVA traditionally, but what we have searched? I have searched the literature. I found out that still MANOVA can be done, but for the test statistics of MANOVA, you have to rely on, not only the wilks lambda, we have to go for some other statistics. You compare four five statistics are available, if all the statistics or majority or the statistics says that there is difference in the mean vectors, you can consider the result. Getting me?

Also, the suggestion given that it is better always go for equal sample size. How do we go for equal sample size? Here, I have 54 observations from officers, 96 from maintenance workers and 149 from this. What I mean to say that as we found out this is the case, so the next step should be why cannot we go for bootstrapping. It is a beautiful

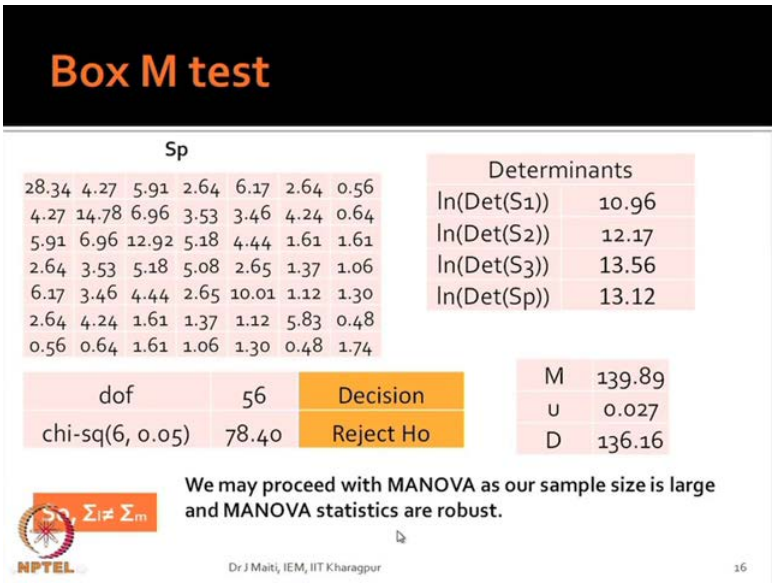
concept. This is because you all will be using data collecting from different sources, and you will be facing the problem of unequal sample size also.

So, what I mean to say can you not go for bootstrapping, bootstrapping what will happen? You will sample from this 54 with an assumption that 54 is the representative of total population. In that case, you here, you do over sampling and here, you do under sampling. Then what will you do, you fix to may be such as let 100 data points. By doing over sampling, make it 100, by doing under sampling, make a 100 or if you want to say no, I want to make everything 500 that also is possible. Then in every case, you must go for over sampling and make it 500.

So, once equal data points are available, then it is suggested in literature that even when the covariance structure is not equal, you can still go for MANOVA. Then what if covariance structure is not equal also mean values are not equal, so mu and covariance structure are not equal. It also gives us one advantage. What is the advantage? We may not go about go for a MANOVA, but what will happen; because of this, we can say that these populations are distinctly different. So, separate model for each of the population we can do, some other model, not MANOVA model.

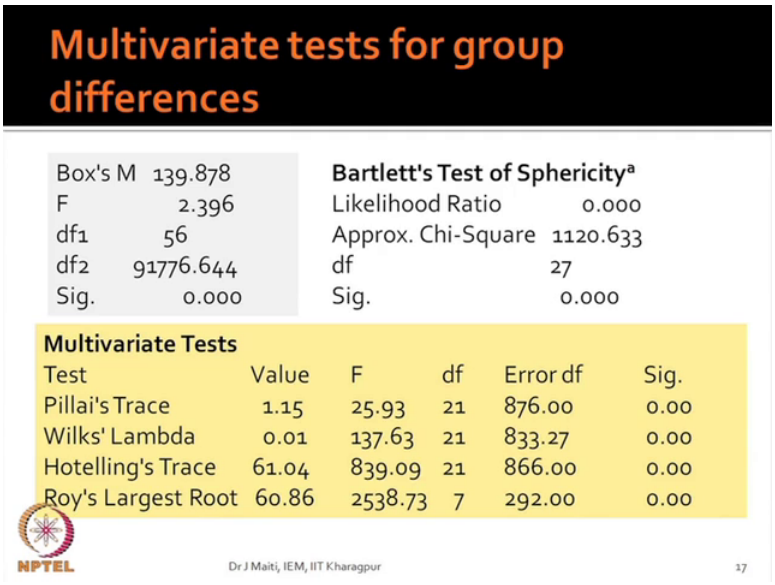
Suppose I want to do regression. Now, this box M test shows that covariance structures is different and mean values also, we are that from the box plot, individual box plot, we found that mean values are also different. We have not tested collectively. Now, then I can go for suppose I will, I want to see that that how the demography, can work place variables are affecting job stress, then I will develop the different model for different groups. That is another advantage you will get. This will come.

(Refer Slide Time: 45:55)



Anyhow, what I have written here, you may proceed with MANOVA as sample size and MANOVA statistics are robust.

(Refer Slide Time: 46:02)



Now, see that what happened? We have done MANOVA. You have a look on the slide and you see that box M test, it is basically very significant that 0 values. It says that null hypothesis is rejected. Now, we have also done the Bartlett's test of sphericity that means what you are trying to say whether the seven factors are independent or not, but

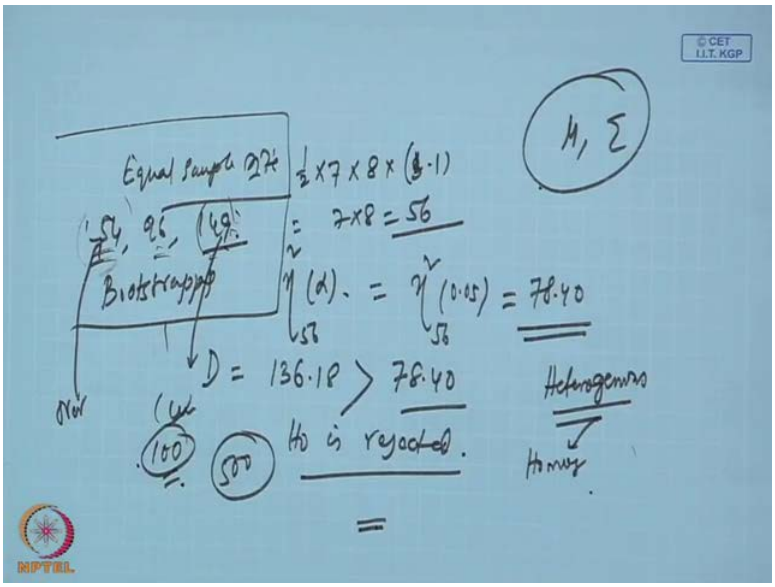
what you have found out that significant value 0 0 this that means there are they are correlated. See how the issues are coming here.

So, as they are correlated, their covariance structure is different. So, if correlation is there, can we not go for some other model? Can we not do principal component analytics and extract may be lower number of dimensions and then we will go for MANOVA? Getting me? In principal component analysis, I will show you this one. I think you please remind me, many a times I forget. Suppose we will go, first do the principal component all the seven, then we will see that what will happen ultimately whether again the box M test will give us what type of I think, you have to check because you from the application point of view, we are considering all the things.

Then, what is there using this MANOVA how I got this values? This I have used SPSS, you can use SPSS, you can use SAS, you can use mini tab, you can use statistica, so there are many softwares or you can simply use your matlab, just develop program. Then here the pillai's trace, wilk's lambda, hotelling's trace, roys largest root, so four statistics we have considered. If you see the significant level, everything, all are almost 0, almost 0.

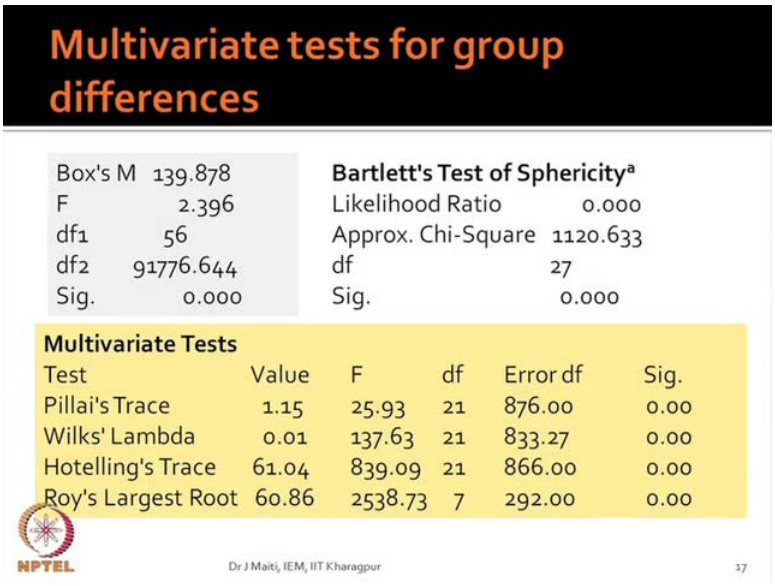
What does it signify? It signifies that there are mean differences across the three groups. Although I said that box M statistics says that there is violation, but if I consider let us go for MANOVA and do this, then the all the tests are showing that there are differences, whether you will accept it or reject I, it all depends on your requirement. But, when there is violation in the assumptions, violations is there, if you think that the results of MANOVA is still useful that means you must be conservative in making bold decisions. The other way what is required, you go for some other data collection and some other test another issue. Then that is the heterogeneous case now.

(Refer Slide Time: 49:03)



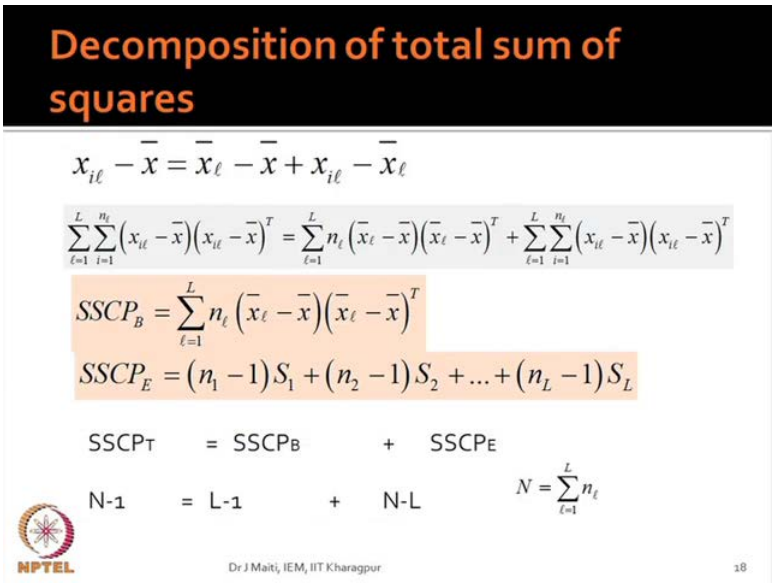
So, you all know that how that heterogeneous case, how we will make it to homogeneous one? So, there are transformations. So, before doing MANOVA, you can go for transformation, power transformation you do and I am sure I have not tested the multivariate normality here, but as it is heterogeneous one, the multivariate normality collectively, if I see that I think it will not be a multivariate normal case. It is likely. So, you, we have to remove outliers. We have seen already there are some outliers.

(Refer Slide Time: 49:45)



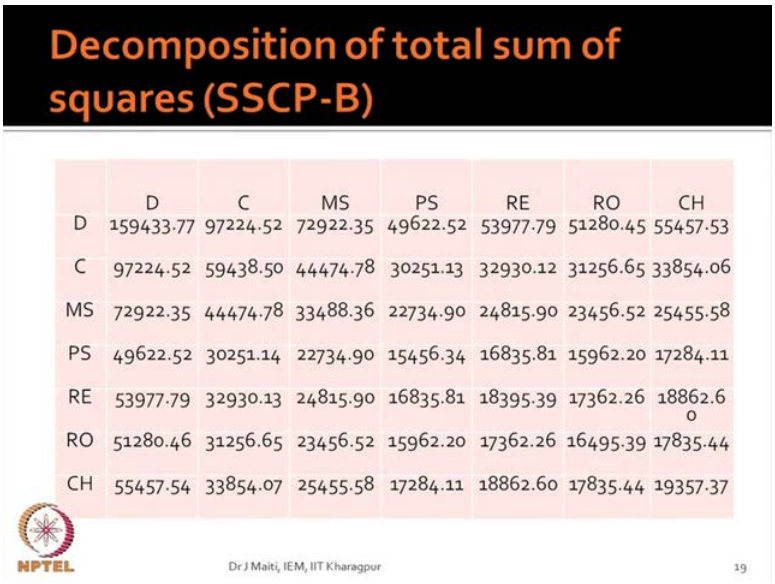
So, coming back to MANOVA, I can say yes there are differences.

(Refer Slide Time: 49:50)



If there are differences, then you go for I think this is the, you go for finding out where is the difference. So, these are basically what I am showing that earlier the computation part and you all know how to compute, although I have shown you the result that all multivariate test, this hotelling, you are not hotelling that wilks lambda. The wilks lambda will be SSCP E determinant by some the SSCP T determinant, so those things, but these are the computational formulas you will be using and these are the data.

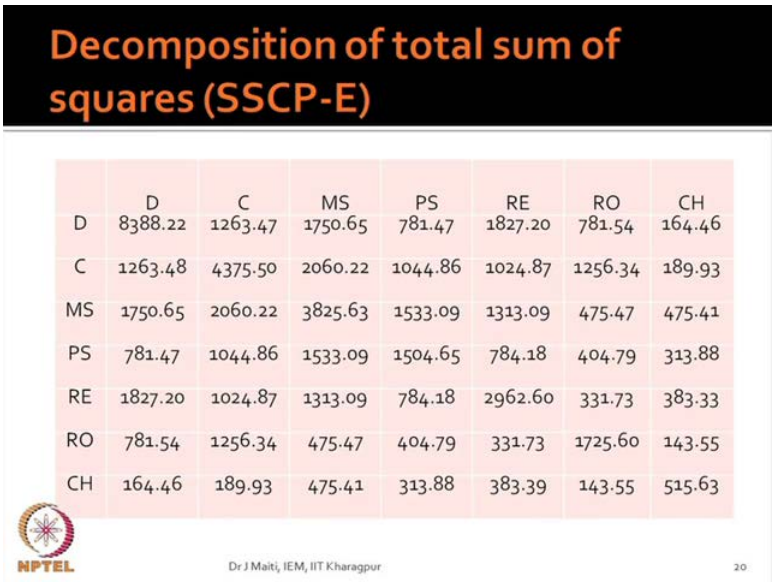
(Refer Slide Time: 50:36)



	D	C	MS	PS	RE	RO	CH
D	159433.77	97224.52	72922.35	49622.52	53977.79	51280.45	55457.53
C	97224.52	59438.50	44474.78	30251.13	32930.12	31256.65	33854.06
MS	72922.35	44474.78	33488.36	22734.90	24815.90	23456.52	25455.58
PS	49622.52	30251.14	22734.90	15456.34	16835.81	15962.20	17284.11
RE	53977.79	32930.13	24815.90	16835.81	18395.39	17362.26	18862.60
RO	51280.46	31256.65	23456.52	15962.20	17362.26	16495.39	17835.44
CH	55457.54	33854.07	25455.58	17284.11	18862.60	17835.44	19357.37

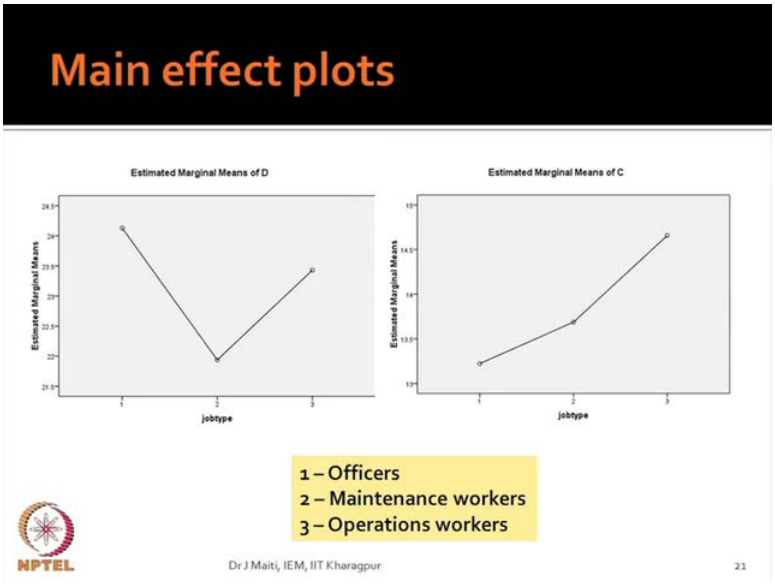
These are the data decomposition that means this is SSCP B that is between the population.

(Refer Slide Time: 50:44)



Then, this is the error covariance matrix.

(Refer Slide Time: 50:49)



Then, those two will be combined to this one. Now, interestingly when I go for marginal means plot for officers, maintenance worker and operations worker with respect to all the variables, the peculiar results are we found out some results are interesting. Interesting in the sense, one is officer, two is maintenance worker. When we talk about that demand

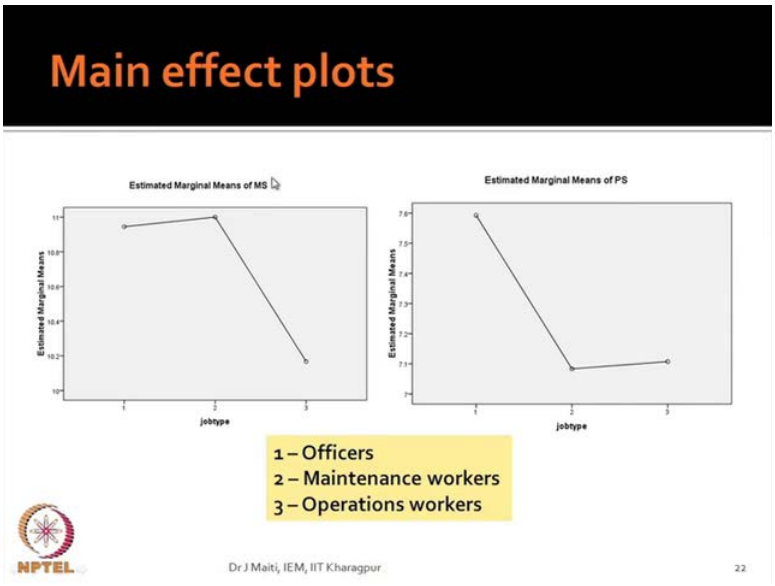
path, it is seen that maintenance worker, they are saying they are stress related to job, demand to their job is less compared to officers and worker operations worker demand is more.

So, if you go for now for the difference that which one is different, whether 1 versus 2, 1 versus 3 and 2 versus 3, I have shown you already that bonferroni intervals we can find out. For each of the difference, you can find out the confidence region. Then from confidence region to simultaneous confidence intervals, you can use bonferroni approach. The marginal plot is showing like this, but when you are talking about case study, that mean there, our, we all those things will be at the back end basically the bonferroni and all those things.

Your answer, your interpretation of this required for the interventions, what is intervention here that job stress intervention, how better can you design the job, so that the job stress will be reduced. Here, from demand point of view officers are having the highest level of job stress. In this case, from control point of view, operations workers are having the maximum. Getting me that how interesting is this one see.

It is that officers are always having better control because they are more empowered meant that operations worker, they are doing day to day routine work and most of the things are routine already structured, sop standard operating procedure is known. So, there is highly any chance to have the discretionary power. So, control point of view, this operations group, that is why, they are saying we are the job stress is more.

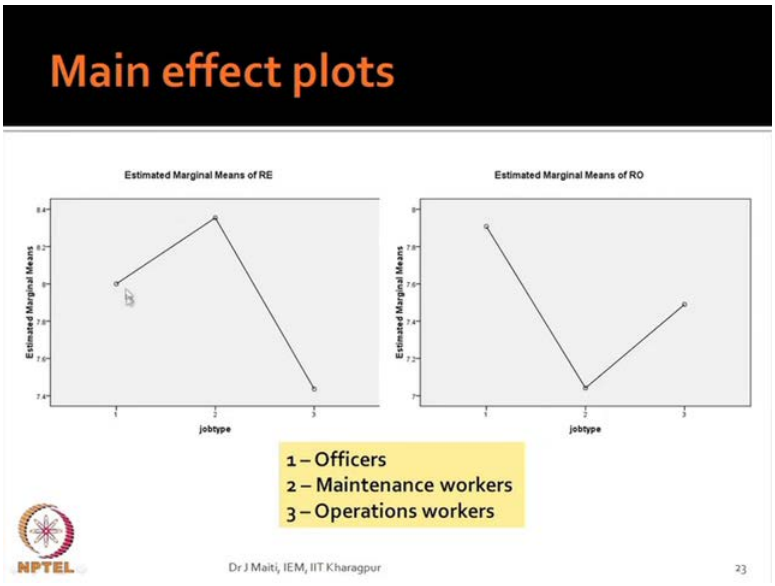
(Refer Slide Time: 53:26)



Then, you go the third one again. If you see that from the management support point of view, operations workers are saying they are better placed. That mean they are for this particular component, they are getting better support, but whereas maintenance workers and officers are still at the higher level compared to definitely operations worker.

This higher level it is not absolute higher level. It is a comparative higher level. Similarly, if you go for that marginal means for peer support, in peer support point of view also, officers are not happy because the intellectual people never united. There is always your conflict in the sense, but lower level for the survival, there you have that group. So, peer support point of view operations and maintenance workers at the same level.

(Refer Slide Time: 54:28)

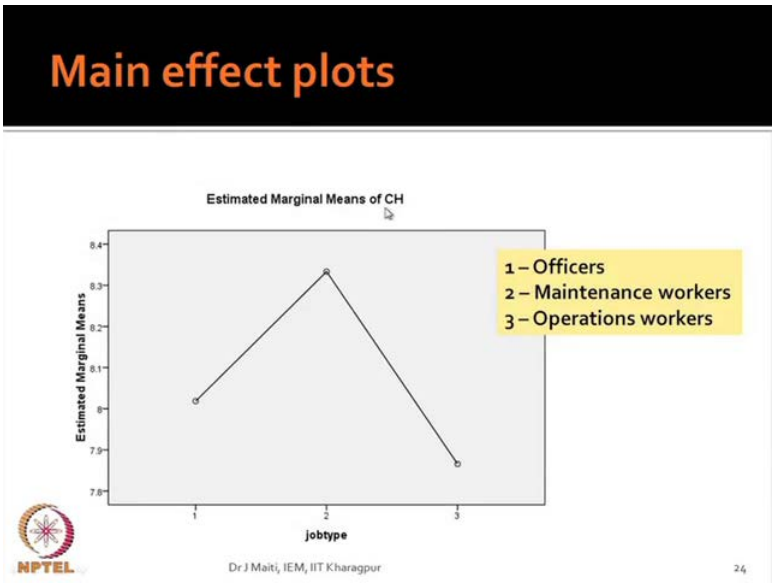


Then, officers, maintenance, then you go by the relationships. From relationship point of view, maintenance workers are saying their relationship is little poor. A reason is there that maintenance is unstructured job. For example, suppose there is a machine, which is under breakdown. So, that machine requires electrical people, may be mechanical people together. Now, see the electrical people department is somewhere, maintenance, mechanical department somewhere and they are mixing when; when it is required. So, it is not that every day they are looking after their jobs simultaneously.

So, it is quite obvious that their relationship has not built up to that level compared to your operations and officers. Operations are having low job stress with respect to relations and this is, this all are matching, all those. You see box M test does not satisfy that there is the equality of co variances, but still, we have gone for MANOVA. The results are interestingly giving us conceptually correct result and which we can test through, we go to the field and you have to test with the talking to the people who are actually working there.

Now, you see that the role, it is the role clarity point of view also; officers are saying that there is difficulty. But, maintenance workers they are saying we know what to do and operations people are having little more, but that is that may not be statistically significant.

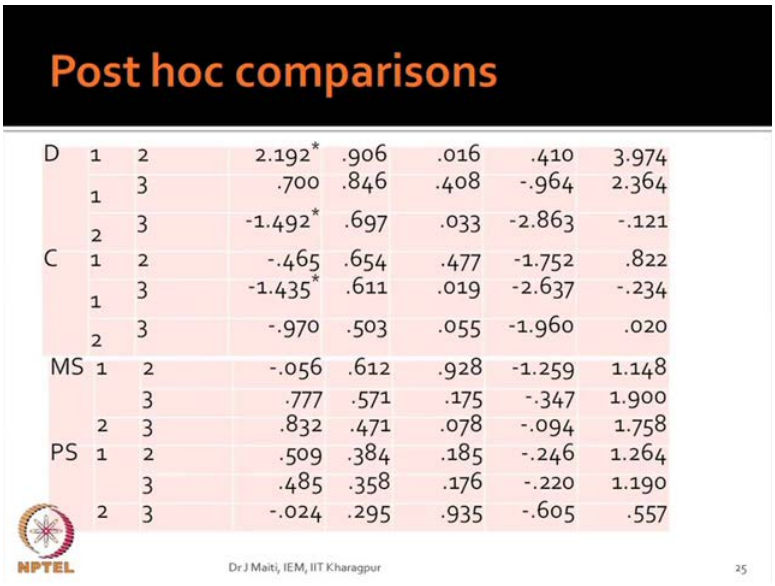
(Refer Slide Time: 56:07)



Now, come to your change. Change again, it is supporting the concept. See as maintenance people we go somewhere to maintain, that is not their department. For example, in CET, there is a maintenance problem. The people will come from maintenance department of IIT Kharagpur. That is not their department. So, whatever changes are taking place here, it may not be, may not be noticed by the maintenance people there. So, ignorance of change is definitely more for maintenance workers and interestingly our study is saying that.

Now, if so many good points coming out of this study where even this thing what I am saying that box test is not giving you this result, required results. That is why what you will do? Should we go for by this conceptually correct results or we will go for one more study? Definitely to be in the safer side, you repeat the study or you increase the sample size using box box plot, you remove the outliers. These are all things have to be done initially and transform the variables, so many things you do and then you see that all the assumptions are satisfied. If it is marginally not satisfied, then there is no problem.

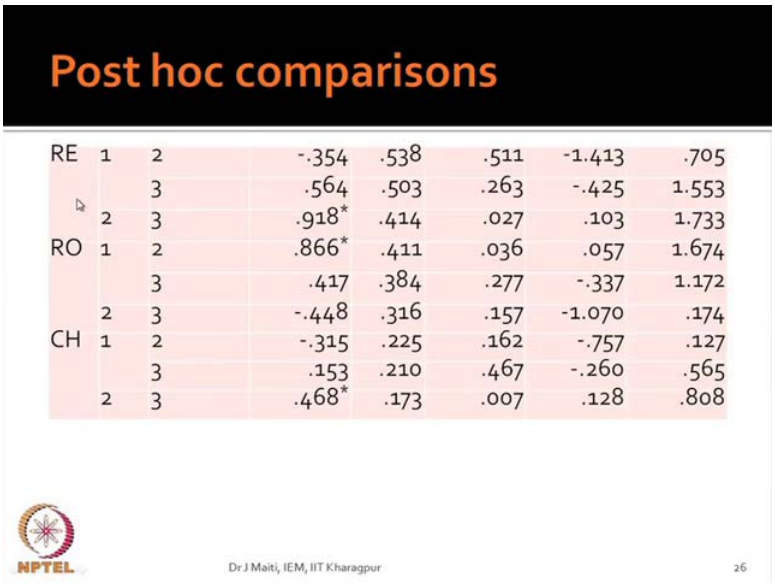
(Refer Slide Time: 57:49)



So, I think you got an interesting, interesting study I think, this interesting study that is because it is showing, it is basically conforming the concept conforming the situation where the people are working. So, all statistics most of the time statistics confirms. What you know, you want to confirm through data and that is what is happening. Now, here what we required that some other because post hoc comparisons are required, what are those things basically, you want to test there is difference, but which group are different here, wherever you are getting the star marks, these are different.

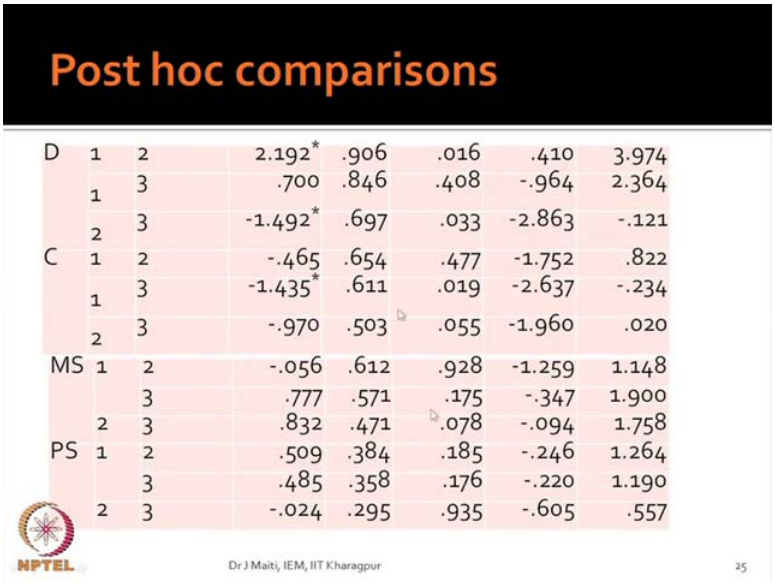
So, that means from demand point of view, officer and maintenance, there is difference and maintenance workers and operations worker there is difference, but officers and operations worker, the difference is not significant. So, similarly, from control point of view that officers and operations worker, the difference is there, management support and peer support point of view. Although we found out that there is absolute some differences in values, but statistically none of them are significantly different from alpha that 0.05 probability level of significance.

(Refer Slide Time: 59:15)



Then, in relationship point of view, you see that the maintenance workers and operations workers, they are different and role point of view that officers. And maintenance worker difference is significant, change point of view maintenance workers and operations workers are different. What are those other values?

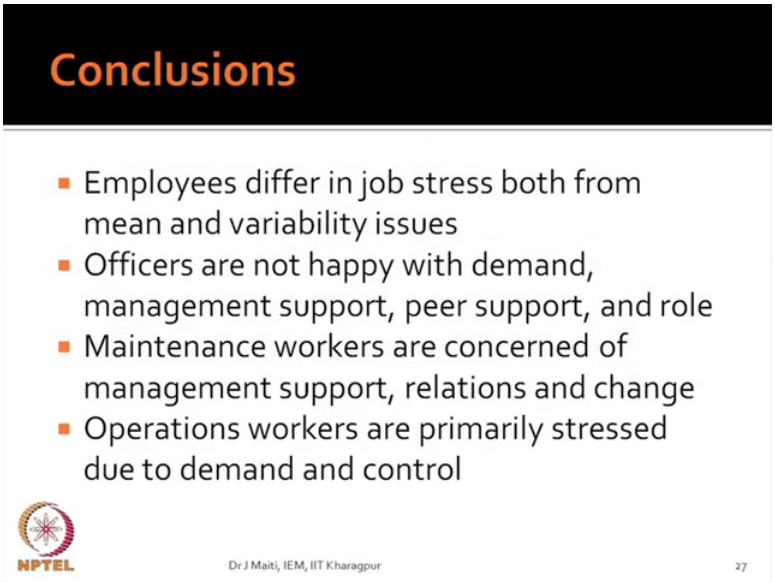
(Refer Slide Time: 59:40)



This values, this value, this value, this value, all those values, this is what is basically the significance level? This is the standard deviation of the estimate. This is the mean value in the mean difference. This is the standard error. This is the significance level. This is

the lower confidence range value and upper confidence value. You see wherever star is there, there is not 0 in between 0.41, 3.9, no 0, star is not there minus to plus. Then star is not there, minus to minus to this. Getting me?

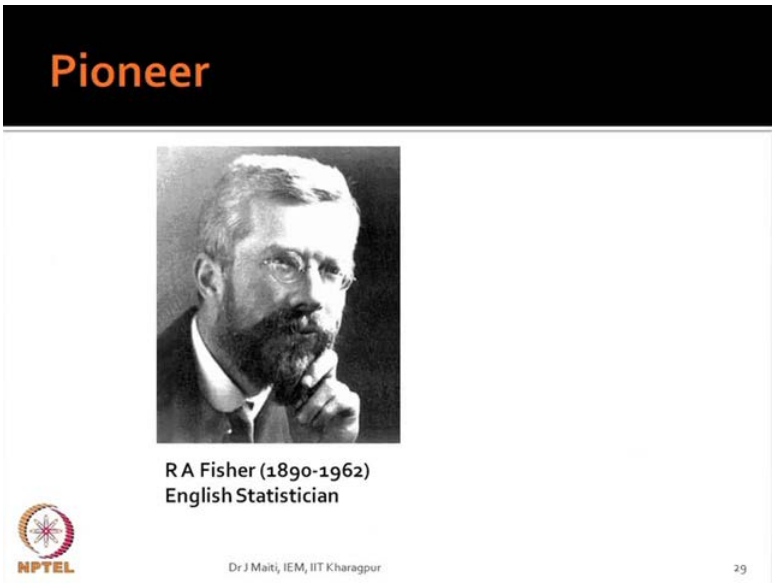
(Refer Slide Time: 01:00:28)



So, then my conclusion is employees differ in job stress, both from mean and variability issues. Officers are not happy with demand, management support, peer support and role. Maintenance workers are concerned of management support relation and change. See you are, there is job stress. Absolutely, it is there and it is no study is required to know this because it is seen in the face. It is also possible to understand people are under stress.

But, what this study is giving that the job stresses are because of many reasons, not that officers job stress vis-a-vis are this operations, worker job stress are caused by the same causes. It is clear from this study that officers from demand management support and peer support and role point of view, there is problem. Whereas maintenance worker case, management support, relations and change and operations are primarily stressed due to demand and control. Getting me?

(Refer Slide Time: 01:01:27)



So, this is what is the beauty of our statistical modeling and it is giving you clear cut answer that where you will stress and even though, my our study, there is little problem. But all the four statistics we have taken for what I can say that to check whether the differences are there or not, everyone supported, every statistics supported.

So, with this, I conclude that MANOVA is very useful. This is only one way MANOVA we have seen. There will be two way because here what we have taken? We have taken only the designation part. You can take something else, for example, job stress is not the designation problem only. People are coming from with certain socioeconomic status, heavenly burden and other things. So, many other things are there you can check. So, if you find out more variables, may be your study will give you better, much better results. Thank you very much. Next class, we will start multiple regressions.