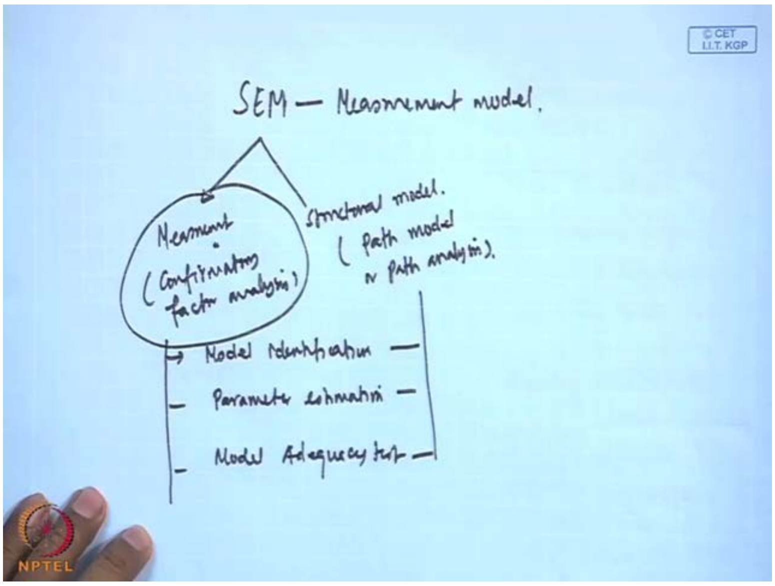
Applied Multivariate Statistical Modeling
Prof. J. Maiti
Department of Industrial Engineering and Management
Indian Institute of Technology, Kharagpur

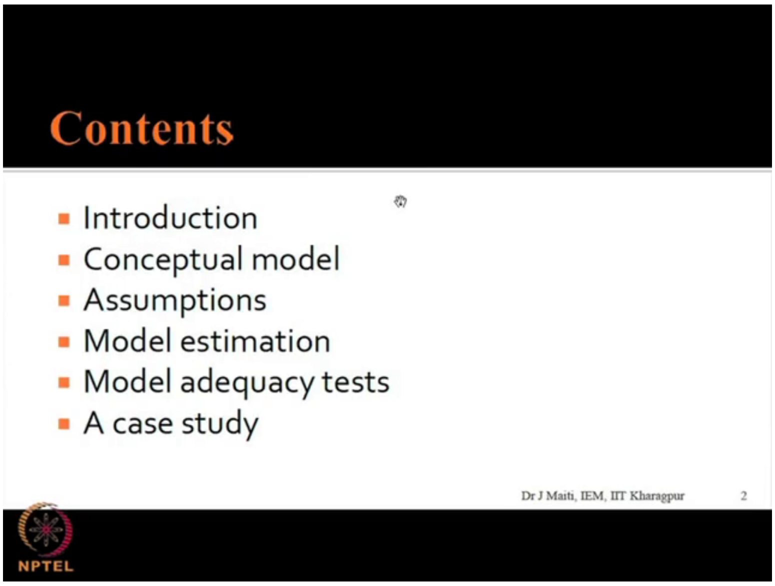
Lecture - 39
SEM-Measurement Model

(Refer Slide Time: 00:25)



Today, we will start a structural equation modelling part 1; that is measurement model. SEM, that is structural equation modelling, measurement modelling, measurement model.

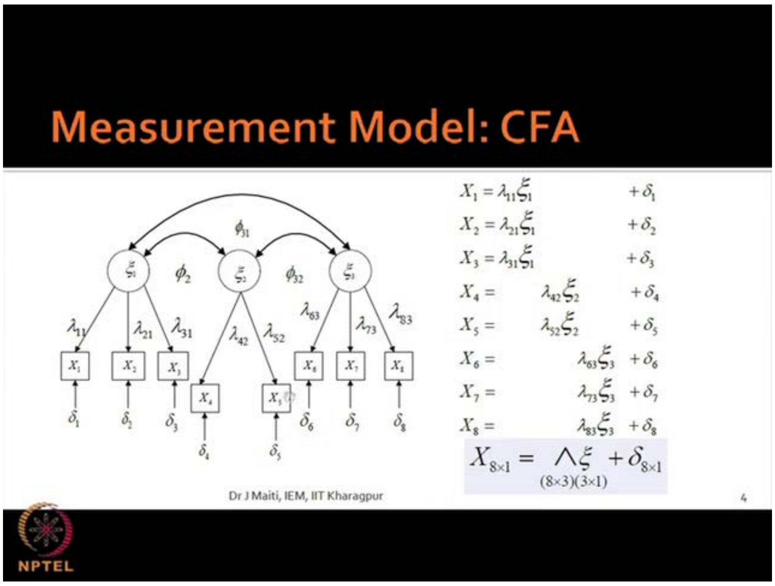
(Refer Slide Time: 00:41)



Now, let me see the content of today's presentation. That we will start with conceptual model, then the assumptions of the model, then how to estimate the model parameters and model adequacy test, followed by a case study. You see in last class, I have explained that measure structural equation modelling has two components, one is measurement component, and another one is structural component.

Measurement component is essentially a confirmative factor analysis, and structural part or we can say the structural model is equivalent to your path model or path analysis. Both the model those measurement as well as structural path, there are three important steps, one is model identification, then parameter estimation and model adequacy test. This is true for structural part, also in this lecture we will consider the measurement part which is confirmatory factor analysis. So, let us start with a conceptual model first. You see here.

(Refer Slide Time: 02:35)



In last class I have shown you similar diagram and you see there are three factors Xi 1, Xi 2 and Xi 3 and each of the factors are manifested by a different variables starting from X 1, X 2, X 3 for Xi 1, X 4 and X 5 for Xi 2 and X 6, X 7, X 8 for Xi 3. So, in confirmatory factor analysis the basis is that that there are hidden constructs which are this or latern construct. Other you can say hidden variable that Xi 1, Xi 2, Xi 3 which are the causes of some manifest variable like X 1, X 2, X 3 to X 8 by putting this arrogate. For example, from Xi 1 to X 1, Xi 1 to X 2 and Xi 1 to x 3 we are restricting the model here in such a sense that we know that Xi 1 is manifested by X 1, X 2 and X 3. This manifest variable.

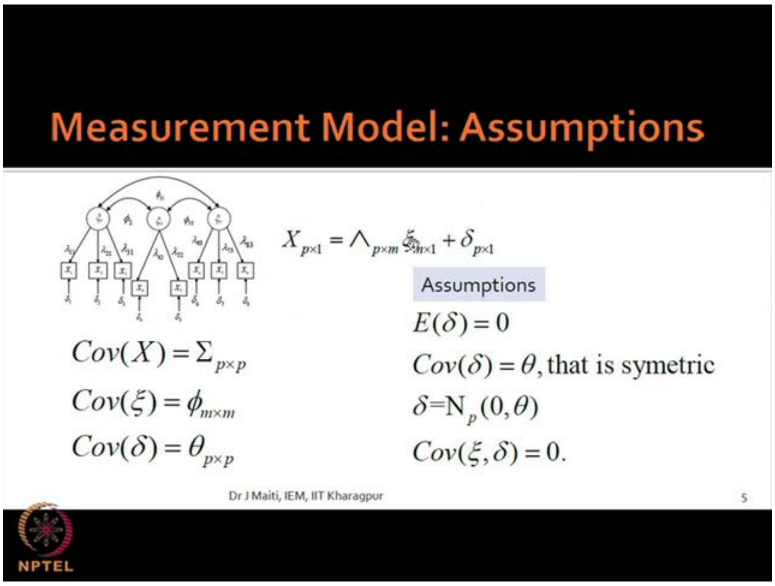
Similarly, Xi 2 is manifested by X 4 and X 5. Similarly, Xi 3 is manifested by X 6, X 7 and X 8, another issue here that this hidden construct or latren construct, latern variable they co vary in the sense that if Xi 1 change there may be change of Xi 2. There may be change of Xi 3 where the correlation components is there. So, these type of, this is a typical structure of confirmatory factor model and other issues here, apart from this correlation between the construct which is denoted by phi.

This come phi 2 1, this is phi 3 1, this curvature line, this curvature line is basically phi 3 2. Now, we are saying X 1 is caused by Xi 1 and as a result a causal linkage is given, Xi 1 to X 1 and parameter which basically depicting the relationship between Xi 1 and X 1 is lambda 1 1. Earlier, I also told you this lambda 1 1, this 1 1 this suffix comes that X 1 from X 1 this point is taken, Xi 1 1 is taken.

So, it is not possible that the variability of X 1 will be fully explained by Xi 1. So, here is possibility of some other variables or hidden causes which may affect X 1 or we can say the errors part, noise part, all those things are considered by delta 1. So, in the same manner you have to explain X 1, X 2, X 3 and up to X 8. What I said verbally, this is depicted in equation form. We are saying X 1 is represented by lambda 1 or Xi 1 plus delta 1. If you think of regression line point of view you will be getting this equation.

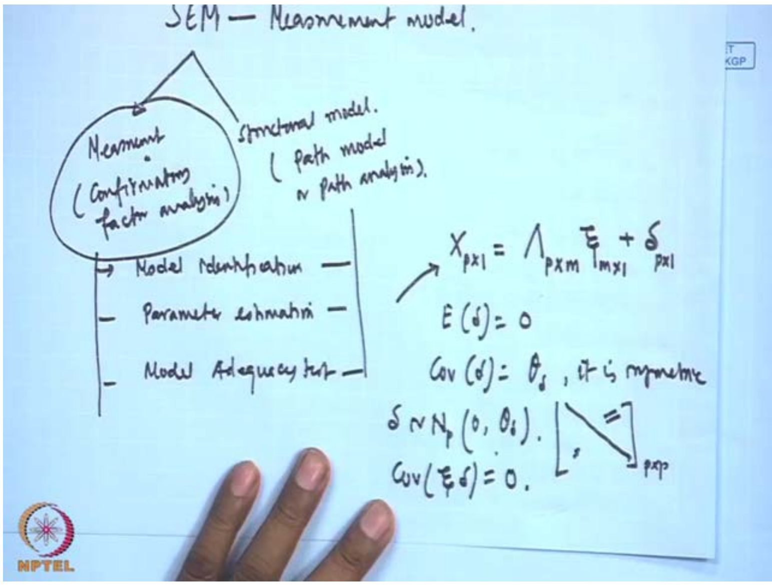
So, in the same manner there are as there are 8 X variables, so you are getting 8 linear equation and collective if you write in matrix form then that will be X 8 cross 1 equal to that capital lambda, which is a matrix of the dimension 8 cross 3. And which you can see here 8 cross 3 dimensions here, because Xi 1, Xi 2, Xi 3 that 3 and 8 X variables, this 8 plus this is Xi plus delta this one. So, this is the, this is the matrix form equation for this particular example.

(Refer Slide Time: 07:10)



Now, you can go for a general equation from there that means what I means to say here the general equation will be X p cross 1 where we are saying.

(Refer Slide Time: 07:17)

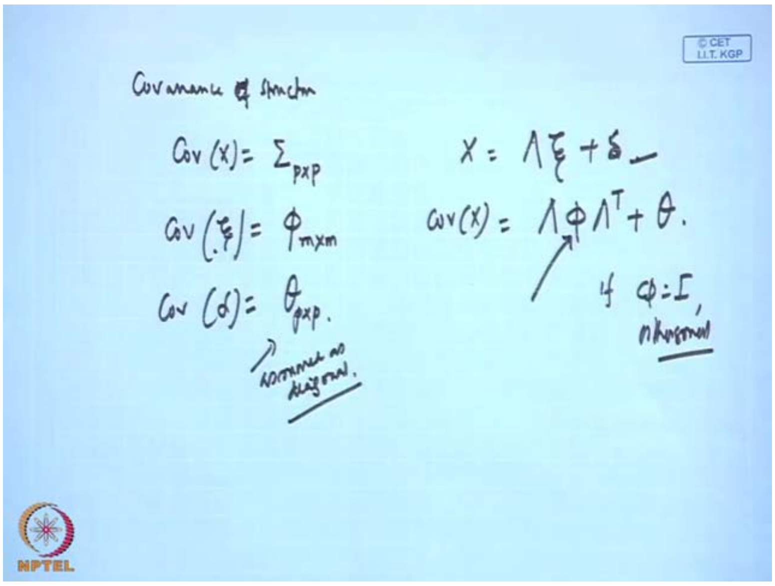


There are P number of manifest variables and which can be represented by this manner, the lambda which is a matrix of matrix relating P manifest variable to m factors. And this m factors are denoted by like this m cross 1 and definitely for every manifest variable there will be an error. So, this is the equation for confirmatory factor model, if you see recall that factor analysis you have found out there also similar relationship, but there are there is difference in the structure of the model then the model assumption in the covariance structure.

So, what are the assumption? Here we assume that the expected value of delta equal to 0, that mean the noise variable, the error terms that mean is 0 and covariance of delta, this one is your theta. Or you can write theta delta, also m times theta delta and it is symmetric. So, if I say like this will be P cross P.

So, this you variance component of delta of diagonal in the covariance that will be equal, that is why symmetric and delta is multivariate normal with mean 0 and covariance matrix theta delta. Or you can write theta also another important issue here is that assumption is that covariance between Xi, that mean manifest, sorry latern construct and the error term related to the manifest variable, they are 0. So, this is our your assumptions related to confirmatory factor model.

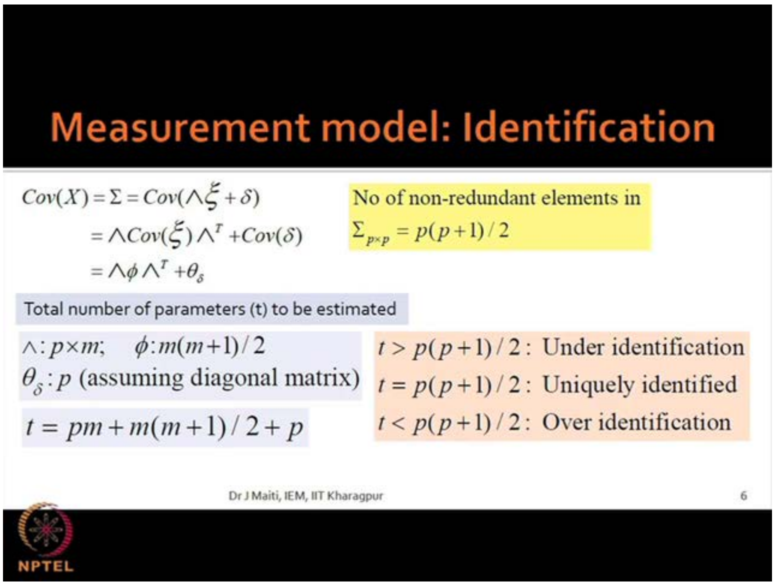
(Refer Slide Time: 09:50)



Now, there as I told you there are covariance. So, the covariance structure of your see covariance structure, first one is covariance of X, this will be capital sigma P cross P. There will be covariance structure for the Xi larern construct, this one will be phi which is again m cross m matrix diagonal, not diagonal this is symmetric matrix. Then covariance of delta which we say theta P. P it is mostly assumed as diagonal matrix, assumed as diagonal not necessary always it will be diagonal, but it is assumed like this

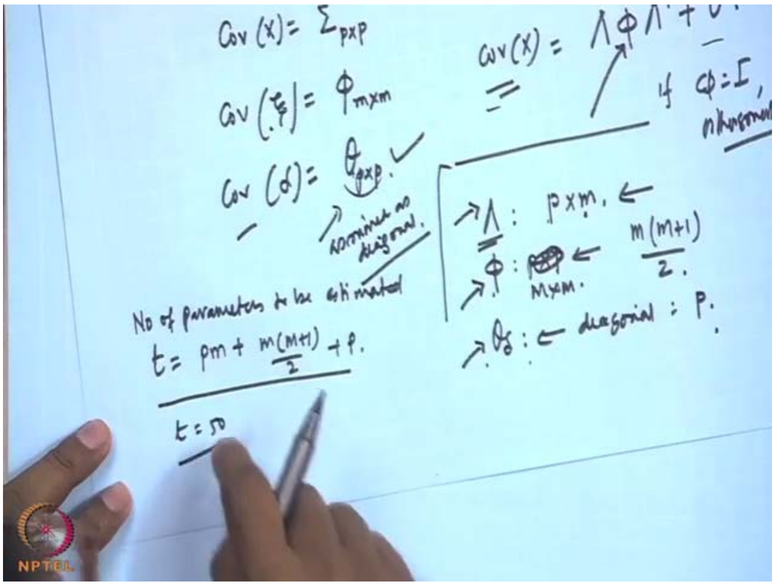
So, you have seen earlier in your exploiting factoring, also we have written X equal to that delta Xi plus, I think we had lambda Xi plus delta, means this capital lambda, replace delta. Now, if you find out the covariance of X then ultimately what you will be finding out? You will be finding out something like this plus covariance of this delta. This is theta in X and this phi was not there in excusive factor, it was I if phi equal to I then it is orthogonal factor analysis. So, this is null set that covariance structure and the relationship between this, okay?

(Refer Slide Time: 12:03)



So, now let us stick to come to the model identification part, what I mean to say here by model identification if you clearly look into the model and the parameters that to be estimated as well as the information, what is available. There must be sufficient necessity and sufficiency of the information available to estimate the parameters of the confirmatory factor model. So, in order to do so you should now let us find out that what are the parameters, we require to be estimated if you see this slide you will see that we have few parameters to be estimated from in C F A.

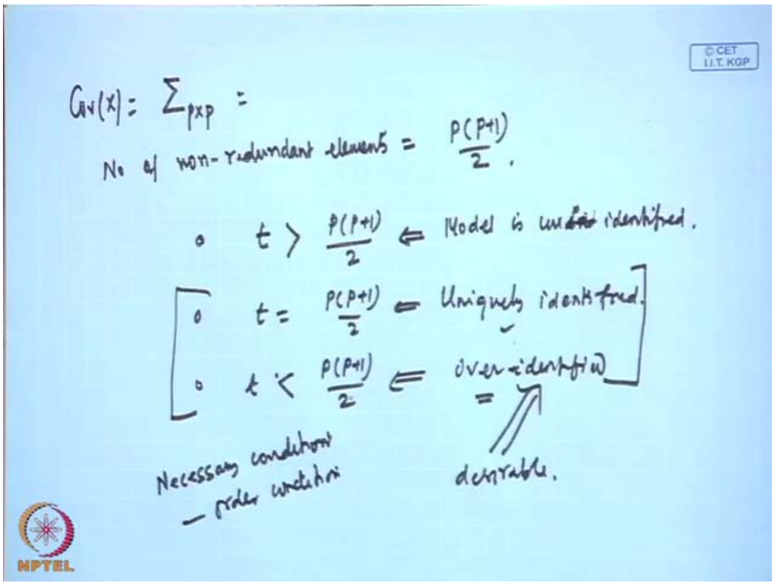
(Refer Slide Time: 12:59)



One is your lambda, capital lambda which is p cross m matrix. So, these many parameters to be estimated, there is phi which is also a p cross p matrix which is p cross p matrix, but being symmetric it has phi is m cross m matrix. I am sorry phi is m cross m matrix which is symmetric matrix. So, numbers of parameters will be m into m plus 1 by 2 to be estimated and then there is theta delta, as I told you that theta delta. What is this theta delta? This is delta, this is theta delta. So, theta delta or theta so there if we assume that it is diagonal then there will be p number of parameters to be estimated.

So. we require to estimate capital lambda, we require to estimate phi, we require to estimate theta delta. Here in case of capital lambda p cross m, this number of parameters phi m into m plus 1 by 2, this number of parameter theta can number or parameters are there. Then in total the number of parameters to be estimated, number of parameter to be estimated we can write p m plus m into m plus 1 by 2 plus p, okay? Now, what you require to know that if t, suppose t equal to 50. If we require to estimate p 50 parameters and you require at least 50 simultaneous equations, getting me? Now, what information we have in case of confirmatory factor analysis?

(Refer Slide Time: 15:31)

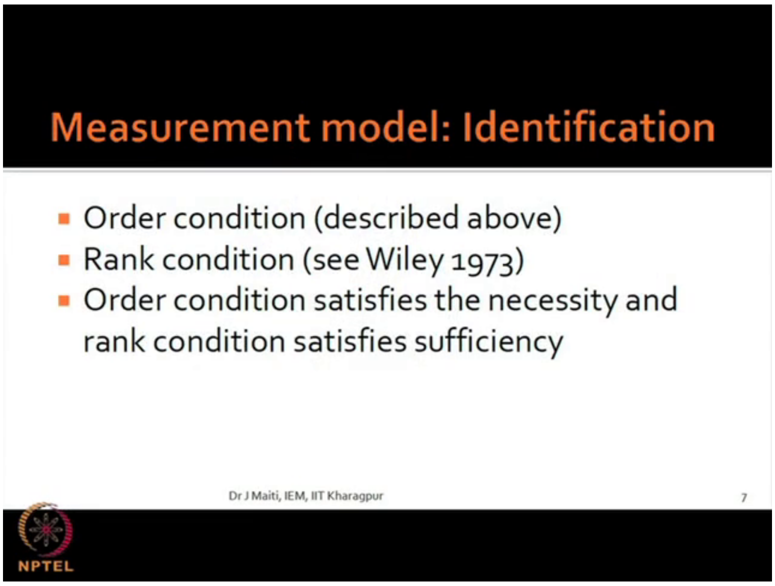


We have only one information, that is capital sigma which is the variance, co variance matrix of X, that is co variance matrix of X. Now, how many unique or non-redundant elements, this is also symmetric matrix. So, it has number of know redundant element equals to p into p plus 1 by 2. Now, see this what type of situation will occur? We require to estimate p number of parameters, it may so happen that number of parameters to be estimated is greater than number of independent non redundant elements in p. That is the available information, it may so happen that t equal to p into p plus 1 by 2. It may so happen that t less than p into p plus 1 by 2.

Now, the first case, this is model cannot be identified. This is model is unidentified or under identified, un identified this case number of parameters or the number of unknows and knows are equal. This is uniquely identified case and this case this is over identified, because we have more information available over identified case. So, at least this two are necessary if you have this case type of situation which is uniquely estimated. If you have this over estimation, that is the desirable one over estimate is the desirable one.

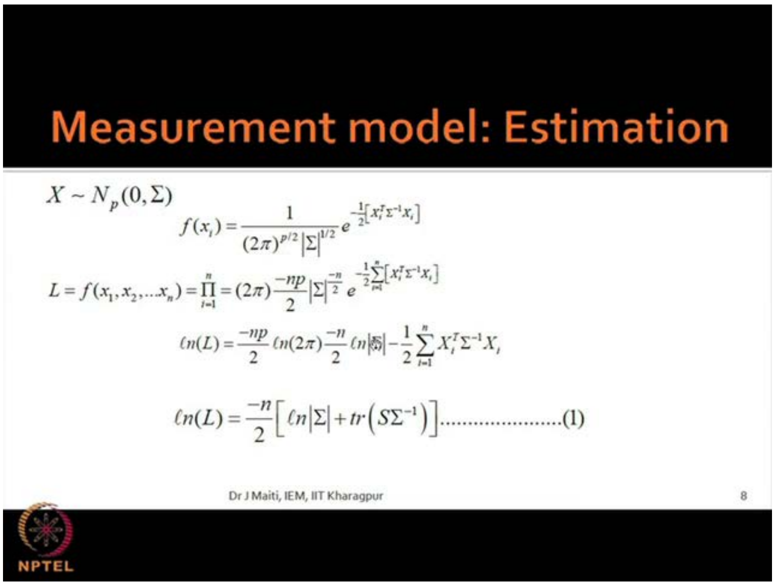
This condition, particularly this two, if this two conditions are either of the two is satisfying you are saying that necessary condition is satisfied. Necessary condition which is also known as ordered condition, okay? But ordered condition alone is not sufficient, this is necessary condition necessity is satisfied. There is another condition called rank condition, because then you will see that we have basically talking about the matrices.

(Refer Slide Time: 18:08)



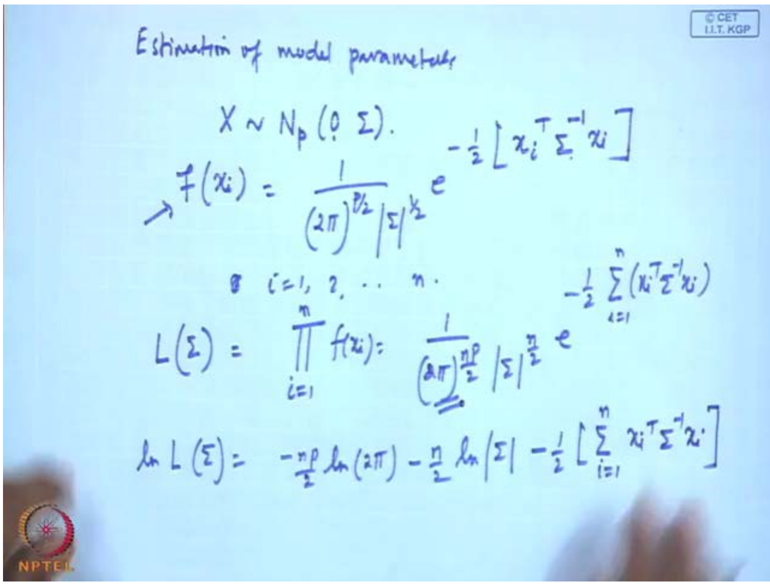
So, the rank condition, rank of my matrix is important issue here and it is little bit complicated one also. Rank conditions also verbally you have to satisfy then the rank condition is the sufficient condition and for example, is given that Willey in 1973 that is the reference. Now, order conditions satisfy the necessity and rank conditions satisfied the sufficiency. Let us assume that it is done.

(Refer Slide Time: 18:55)



In the sense model e is identified, if model is identified. The next step is how to estimate the model. So, estimation of model parameters, okay?

(Refer Slide Time: 19:06)



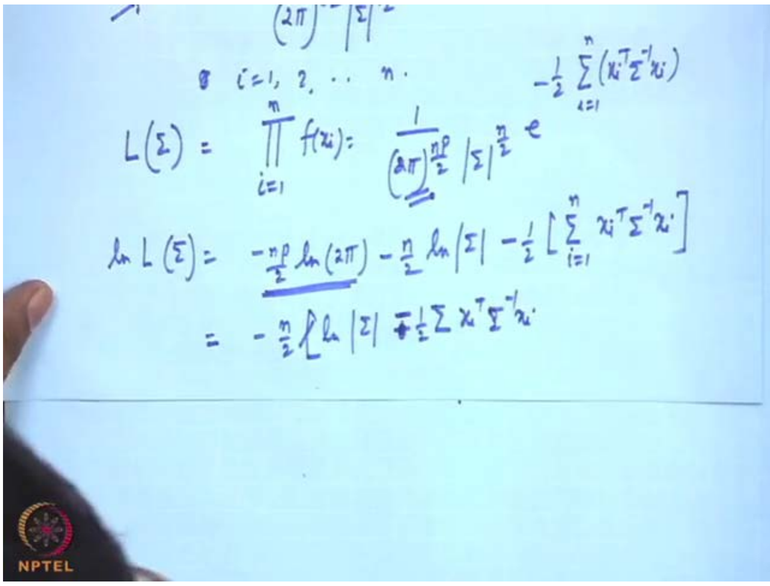
So, you have seen that we assume that ultimately that X, the manifest variable is normally distributed. It is the primitive multivariate normal p X variable p, p number of variable are there, multivariate normal with mean 0 and variance, co variance matrix sigma, capital sigma. Then for any observation multivariate observation Xi you can write that this is the PDF can be written like this, 2 pi to the power p by 2 sigma determinant to the power half e to the power minus half. Then X minus mu e, e is to write that is X i minus mu e is zero here.

So, X minus m transpose, that means Xi transpose sigma inverse Xi, this is the multivariate normal distribution per density function for a particular multivariate observation. Now, we collect in observation i equal to 1 to n, we want to know the log first the likelihood. So, likelihood if you see this equation you find, you see the there is only one parameter which is sigma and mu e is 0.

So, only one parameter is there, so we can write log of likelihood of sigma, not log likelihood of sigma which can be written as multiple equation of this, i equal to 1 to n f xi which will be multiplying i equal to X equal to 1, X 1, X 2 up to X n. Then the resultant will be like this 1 by 2 pi to the power n p by 2. Then determinant to the power n by 2, then e to the power minus half, then sum i equal to 1 to n. Then Xi transpose sigma inverse Xi that is what will be the likelihood and it is customary to take log likelihood.

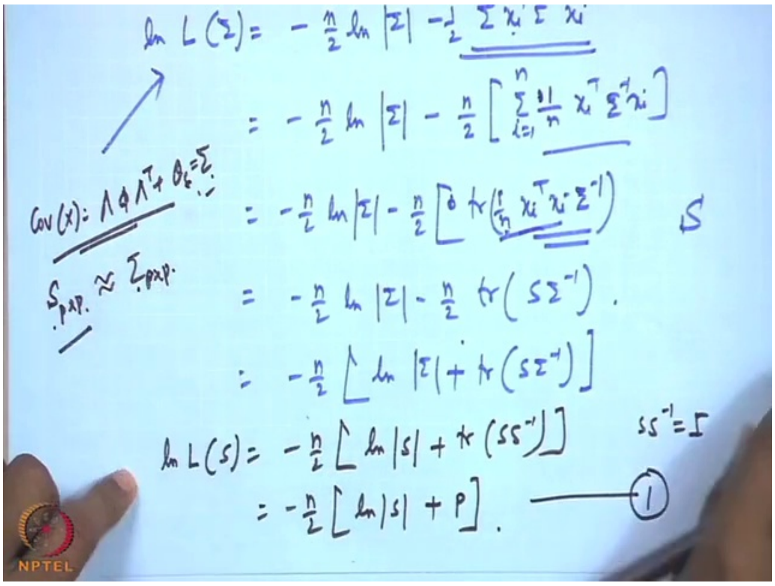
So, if you take log likelihood then what you get? You get minus m p by 2 log 2 pi for this term, minus n by 2 log this for this term minus half, i equal sum of i equal to 1 to n Xi transpose in sigma inverse Xi. So, obviously this from our we want to estimate this sigma that parameters it will be will go for some optimization root and there this constant term in the equation has whether you had keep it or do not keep it this is immaterial.

(Refer Slide Time: 22:28)



So, we remove this constant. So, you can write this as minus n by 2 log of this plus half of I can write like this, minus n by 2 into this minus. So, minus half of this Xi transpose this xi. So, let me write a phrase that log of l, this sigma equal to minus n by 2 log determinant sigma minus half Xi transpose sigma inverse Xi.

(Refer Slide Time: 23:25)

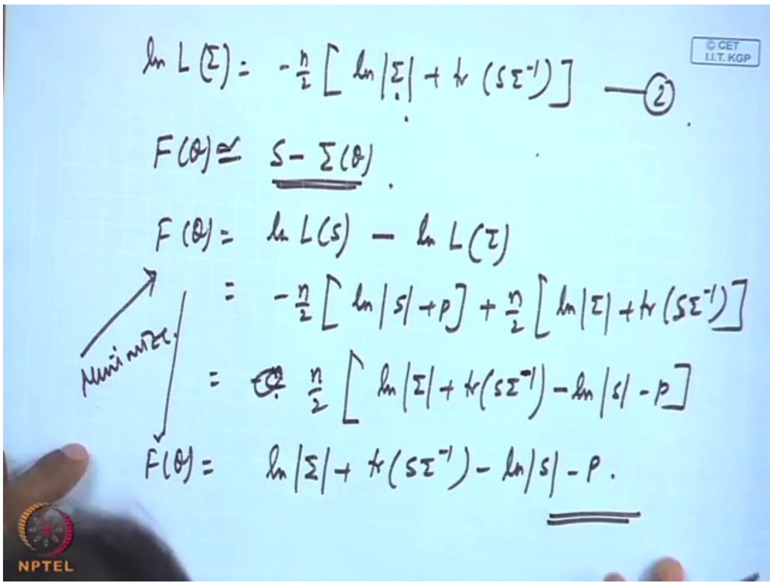


Now, this term can be written like this minus n by 2 log this minus, if I write n by 2 then summation i equal to 1 to n. This can be written like this, this can be written like this 1 by n, I am teaching because I have considered n here. So, this one can be written like this, write again, you write like this like this, then I come to R. R form n by 2, this minus n by 2 we can write this quantity as trace 1 by trace of 1 by xi transpose xi sigma inverse this, this is possible. Now, 1 by n Xi transpose Xi, this is nothing but variance, covariance matrix of the sample provided, n is large, 1 by n minus 1 and 1 by n become same.

So, with modelling equation we can write like this minus n by 2, then this is trace of s sigma inverse. This one is minus n by 2 log of this plus trace of s sigma inverse. So, this is your log likelihood. Now, see what is the condition here in our estimation here, actually this procedure is like this, you will from the model from the model covariance of X, your lambda phi, lambda transpose plus theta delta. From this you will get covariance X sigma, that is sigma in terms of model parameter you collect sample, then you get the covariance matrix.

Also, from sample there will be S. S is the again p cross q that sample covariance matrix, there what we want to do? We want to match this two, this two suppose a condition is such that S is equal to sigma, then if I put here what I can write? Log of l of s, this can be written like this, this log of S determinant of S plus trace of S S inverse. Now, S S inverse is I, so you can write like this. This log of S plus sum of the diagonal elements of the matrix I, that is p this is your equation number 1. Now, another one is we have already seen.

(Refer Slide Time: 27:02)

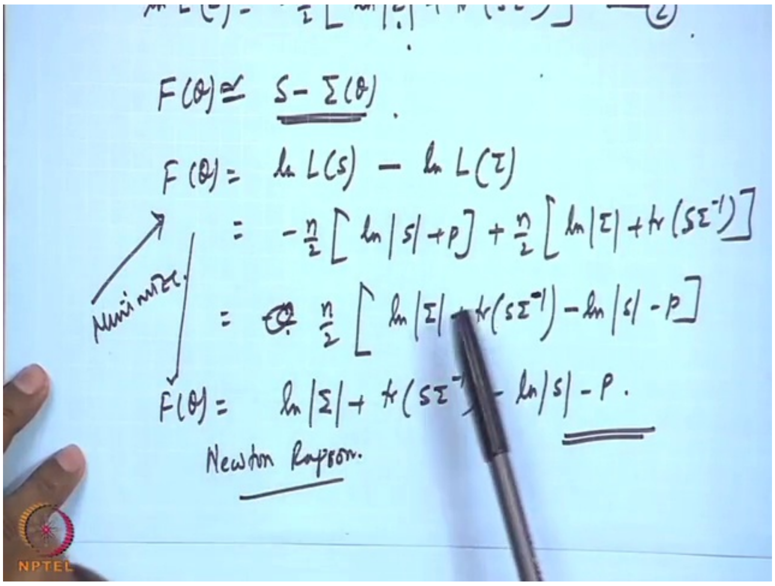


The likelihood one, this equal to minus n by 2 log of this plus trace of this, this is our equation 2. So, what you want? We want to find out parameters. Now, this sigma this will be in terms of model parameter S here, and here when we are talking about S it is basically the numerical values and here it is in terms of model parameters like lambda phi and all those things.

So, we will create a function now that we want to minimize that we are saying F theta which is nothing but S minus sigma theta of this nature which I am saying, not exactly which will be of this nature. So, then we can write like this F theta equal to log of l S minus log of l sigma, which if you write this is minus n by 2 log of determinant of S plus p plus n by 2 log of determent of sigma plus trace of S, this trace of S sigma inverse.

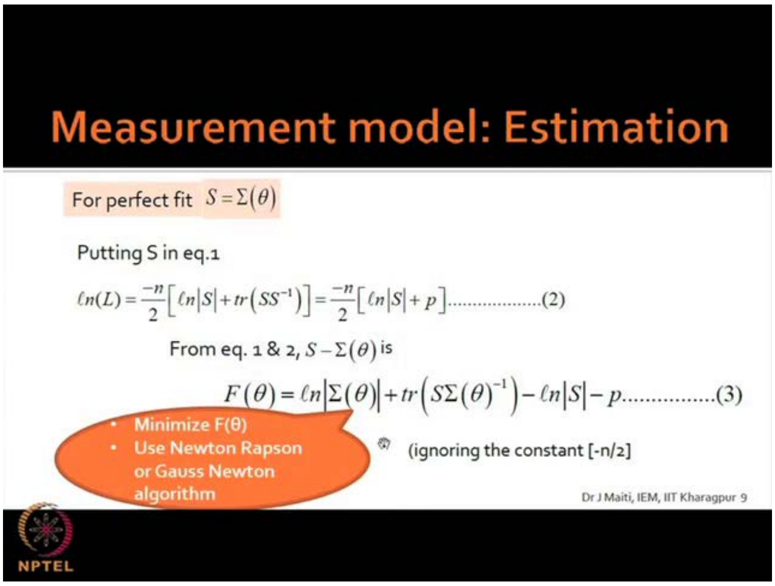
So, this one you can write minus that n by 2. Then log of determinant of sigma plus trace of S sigma inverse minus log of S determent S minus p, we want to minimize this function. So, keeping this constant n by 2 again is of no use. So, final equation will be for our estimation, is this log of determinant of sigma plus trace of S sigma inverse minus log of S minus p. This is the equation, which want to minimize, okay?

(Refer Slide Time: 29:38)



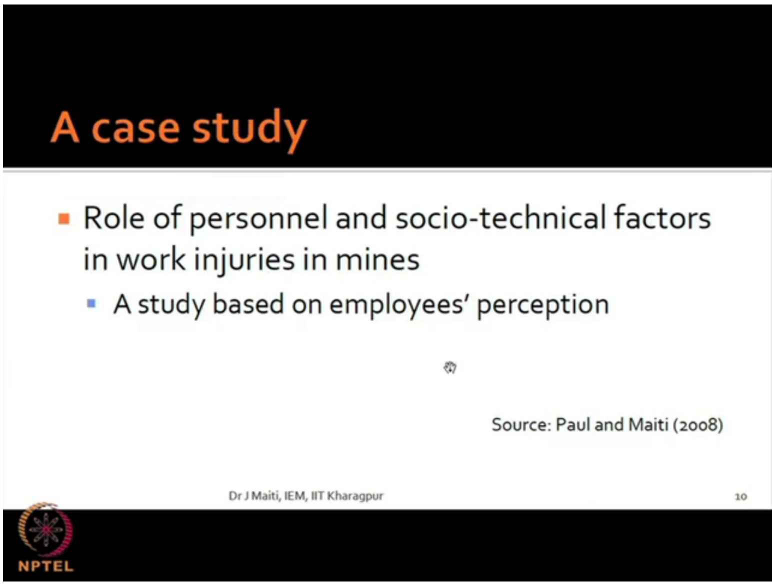
So, its null issue you have to use Newton Rapson or similar method, Newton Rapson similar method of numerical that optimization part. So, this is what is in the nut shell, the parameter estimation in place of confirmatory factor analysis, which is basically our measurement model, you see here.

(Refer Slide Time: 30:05)



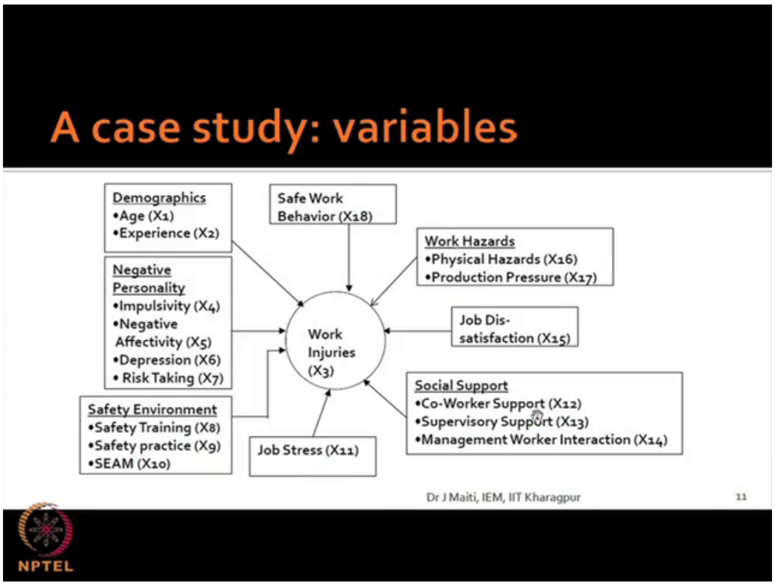
This is the theta log sigma theta trace of this, this is the case ignoring the constant. So, ultimately minimize this one using Newton Rapson or Gauss algorithms, okay?

(Refer Slide Time: 30:23)



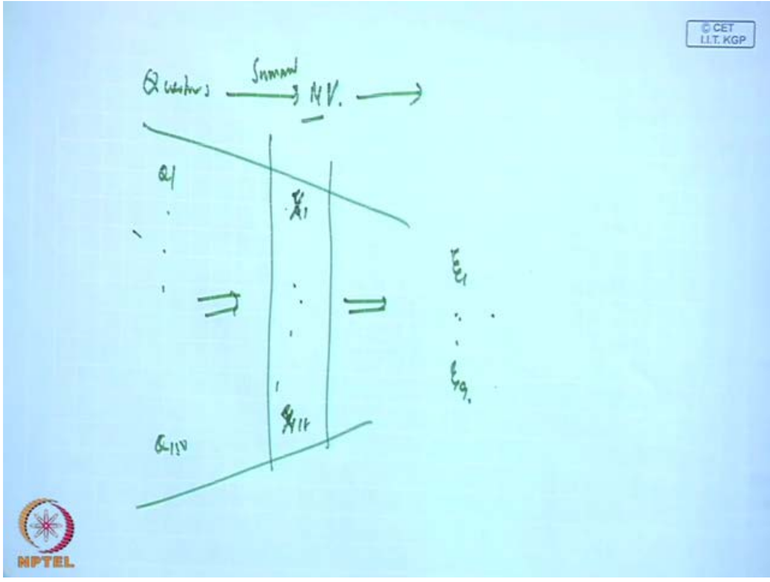
Now, let us see that whatever mathematics we have described now can it be put into a case study as a real life example. I will show you the example here which I have shown you earlier also in the in last, in the first class of structural equation modelling. I have shown you this one, but there what I have done actually, I have given you a glimpse of this things, just like scrawling down the slide. Because just the glimpse of what is this not will describe in detail that what is this measurement model and with the same case study, okay? So, the case study as you know it is a role of personal and socio technical factors in work injuries in mines and a study based on employee's perceptions and you can see that the source is Paul and Maiti, 2008. It is published in the ergonomics, okay?

(Refer Slide Time: 31:27)



So, let us start like this we have several manifest variables here. There are 18 manifest variables for example, age, experience, impulsivity, negative, affectivity, depression, risk taking, safety training, safety practice, safety equipment, availability, maintenance, job stress. Like this we are wondering that how I am saying these are manifest variable although, most of the things cannot be observed. So, actually what happen for every of the variable, we ask several questions and then those questions are summed into a particular quantity. And then that summed up values we have taken as the value for each of the variable, for each of the observations or individuals who participated in this study.

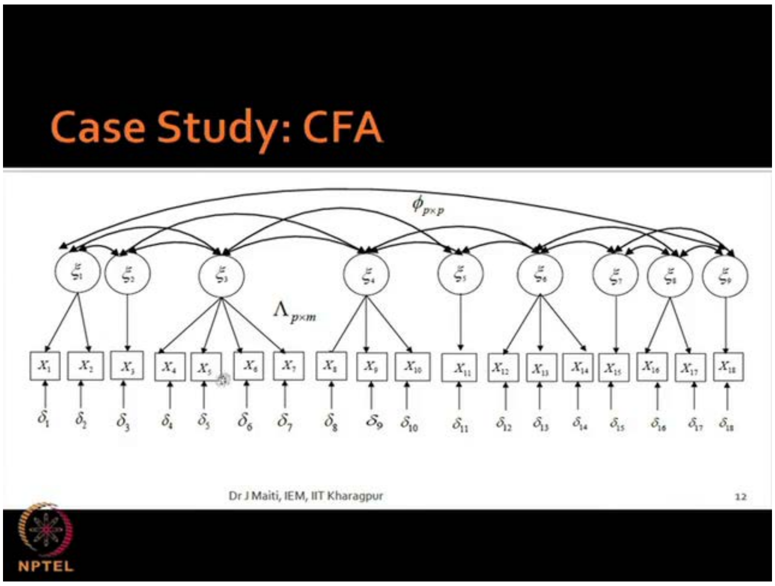
(Refer Slide Time: 32:36)



So, in that sense it is manifested, means observed in that sense. Otherwise, its two layer questions actual it was like this, only one questions, then there sum to this. This manifest variables what we are saying there sum then further level of aggression actually. Suppose, there are question 1 to let it be 150 then there are manifest variable like Xi 1, like there are 18. Then this again it is aggregated into what I can say these are X 1, sorry these are X 1 to X 18 related to Xi 1 to some Xi, let it be Xi 9.

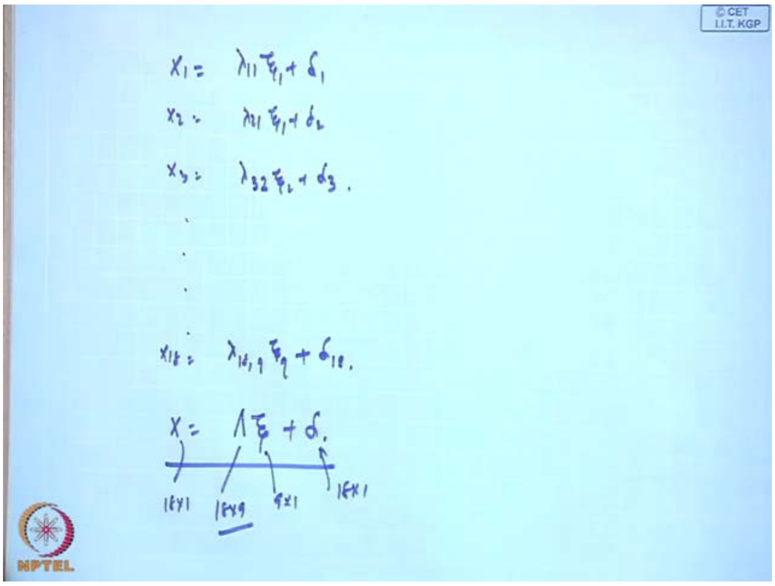
So, this level of aggression is done here. So, we are taking in this level of aggregation, we are considering here this is manifest variable, but you may start from here to here. That will be combos some we have this. So, the same thing if you write in the confirmatory factor analysis form, it will be something like this.

(Refer Slide Time: 33:38)



You see all the this covariance structure between this 9 gita Xi variables, it is not pictorially shown because of space concept. Otherwise, this will be this and then what will happen? We will immediately, you can go for the equations also for this, getting me? So, like I am giving one equation only here if I want to know what is X 1? Then this is nothing.

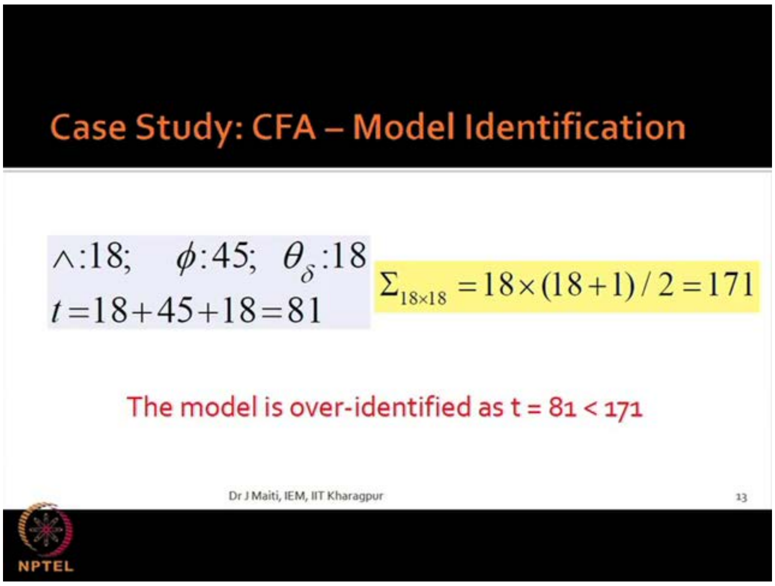
(Refer Slide Time: 34:18)



I can write lambda 1 1 Xi 1 plus delta 1 if you consider this. So, here lambda 1 1 Xi 1 plus delta 1. Similarly, this one lambda 2 1 Xi 1 plus, so x 2 will be lambda 2 1 Xi 1 plus delta 2, if you consider X 3. So, X 3 is ultimately it is the single indicator manifest variable for the constant Xi 2. So, X 3 can be written like this that 3 2 xi 2 plus delta 3.

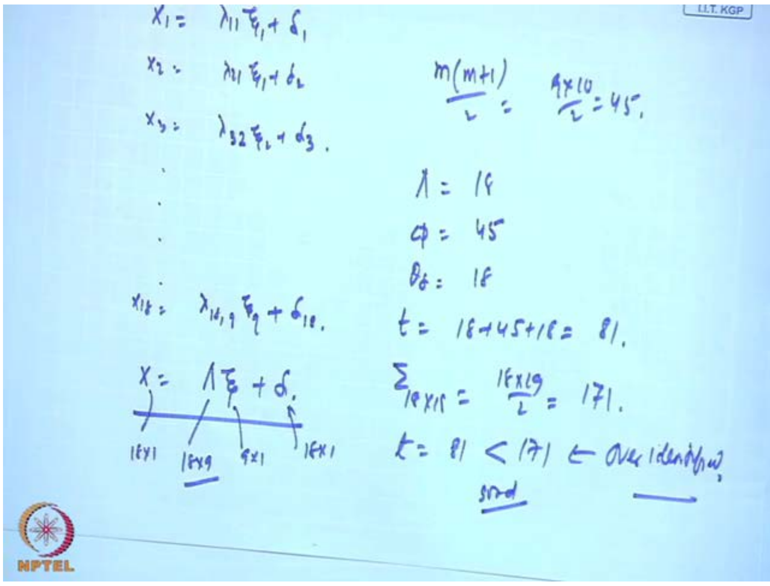
So, in the same manner as there are X 18 so you will be able to write X 18, come to this one, X 18 is again a single indicator for Xi 9. So, this is your lambda 18 9 xi 9 plus delta 18. So, you can write in matrix form, when you write in matrix form you will be getting a equation in matrix form, equation we said X equal to lambda Xi plus delta. This equation you can find out here, here we have 18 cross 1. This will also be 18 cross 1, this is our 9 cross 1. So, this will be 18 cross 9, so this type of equation you can find out, okay?

(Refer Slide Time: 36:09)



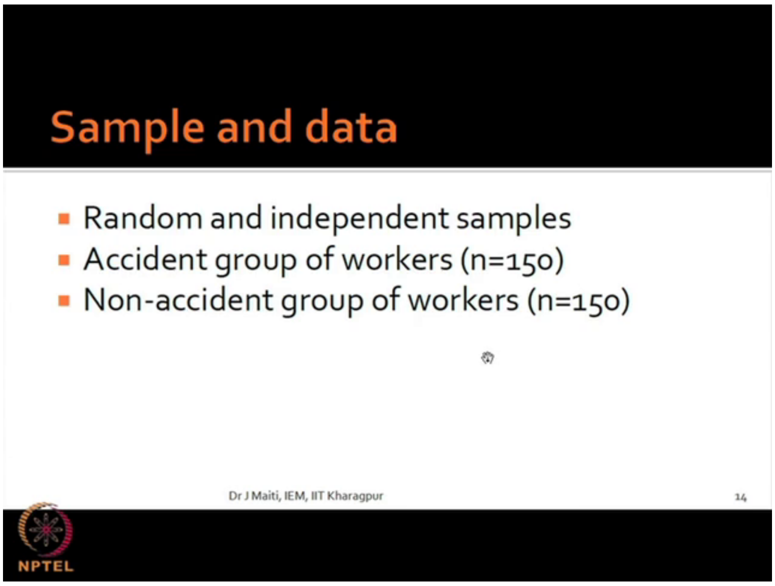
Now, let us see the model identification for this case. You just see that lambda part that how many lambdas are there. You count 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18. So, 18 lambdas, so we have written lambda 18 because others are 0. For example, lambda 3 1 if you give one linkage here lambda 3 1 that is 0, because it is a confirmatory. We know what are the manifest variable coming out of the hidden constructs phi. There are how many X 18. So, how many Xi 9 xi. So, m into m cross m into m plus 1 by 2 m into m plus 1 by 2. So, 9 into 10 by 2, that will be 45.

(Refer Slide Time: 37:12)



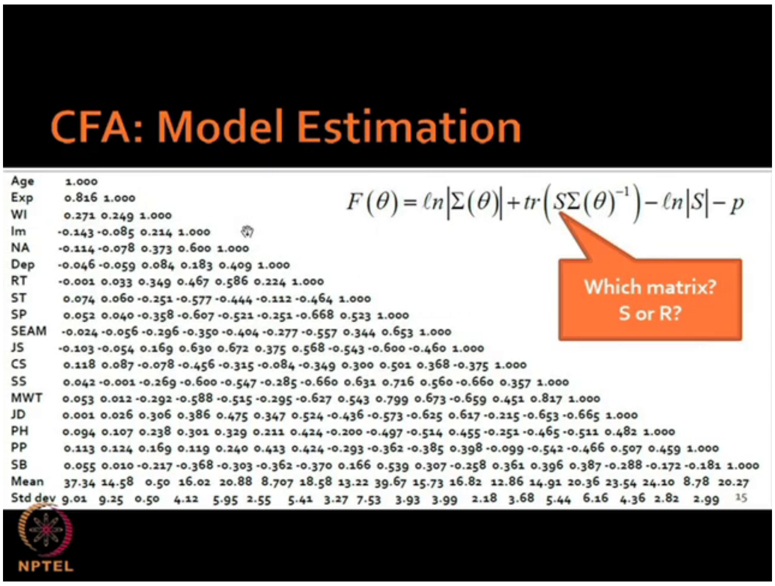
So, you have lambda that is your 18, then your phi related variables will be 45 phi, related parameters will be 45 theta delta, again 18 delta 1 to delta 18. So, our t is 18 plus 45 plus 18 that is 81. Now, what is the unique elements? There we have 18 manifest variable cross 18. So, 18 into 19 by 2 this will be 171. Now, t equal to 81 it is much less than 171, the model is over identified, it is a good case so and necessity condition you satisfied and sufficiency we have not tested here that says the software they test all those things.

(Refer Slide Time: 38:14)



Now, let us see that data part. The data part is actually random independent sampling. First we have taken accident group of workers followed by with frequency matching, non accident group of workers where all together 300 observations.

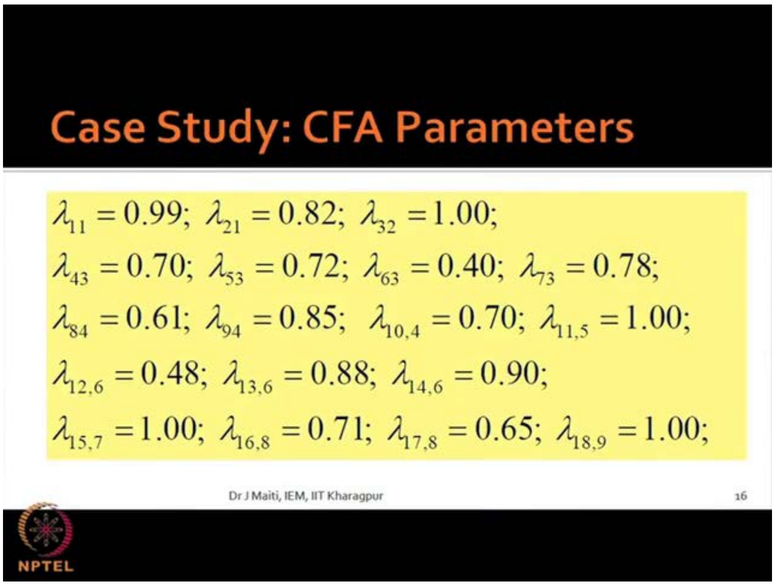
(Refer Slide Time: 38:42)



So, immediately as I told you that how many independent non redundant element in your sigma matrix. This is the case, this is from sample, these are all co relation matrix. Now, question is what we want? We want basically to minimize this function and in this matrix is basically for S, that is sample co variance matrix, but actually we have taken correlation.

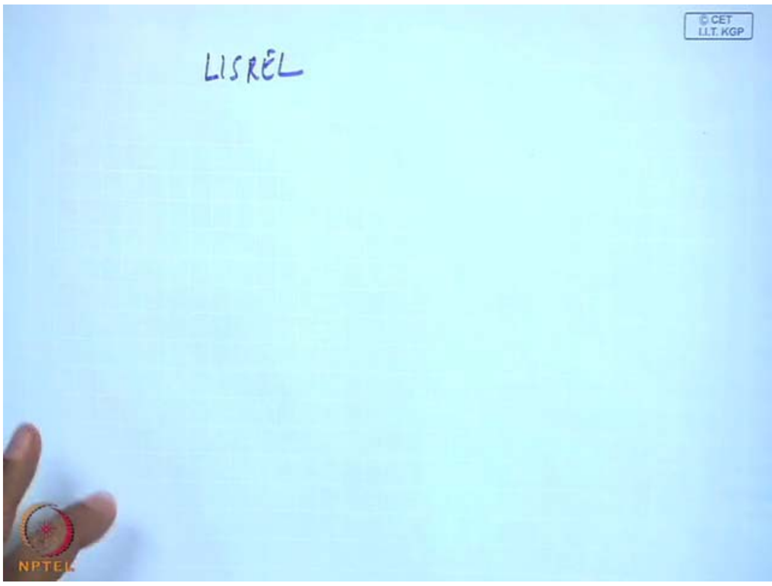
Now, question comes whether co variance or correlation it all depends on the purpose of the study. In our purpose of the study we are more interested in the pattern of the relationship, then the original strength of relationship between latent variable and your manifest variable. We are more interested in the pattern of the relationships and not the original value. So, when you are interested in the pattern of the relationship R is a variable, R matrix should be used, that is correlation matrix should be used.

(Refer Slide Time: 40:00)



Now, through definitely this is F theta, this is the function which is to be minimized and we have used this software, in this case LISREL linear structural relations.

(Refer Slide Time: 40:20)

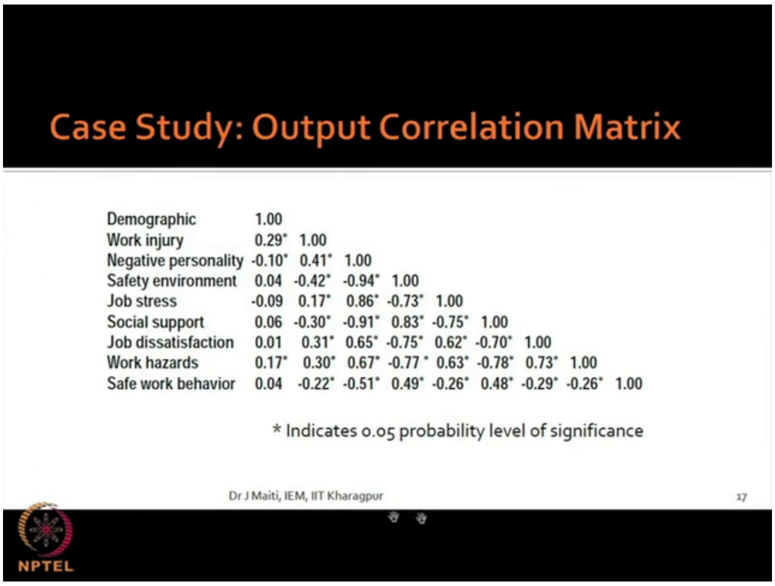


So, this software we use and ultimately this is what is the all the parameters which is estimated. You can see if you go back lambda 1 1, lambda 2 1, lambda 3 2, lambda 4 3 like this and you see that lambda 3 2. What is the value of lambda 3 2? Here, lambda 3 2 is 1 because this one indicator with one constant, we assume that this is the manifest variable itself is the construct, understood?

So, that will be the what is the output of this measurement model. See ultimately we are talking about long back I think this one, these things when we have clubbed into this factors, these are all the factors or constructs latent constructs. This Xi 1 to Xi 9 these are not arbitrary Xi 1 to Xi 9, they have some meaning. Actually X 1, X 2 if you see that age and experience then demographics, this is the this is Xi 1 impulsivity negativity all four are clubbed to the var and Xi 2 value. Actually the negative personality is given here.

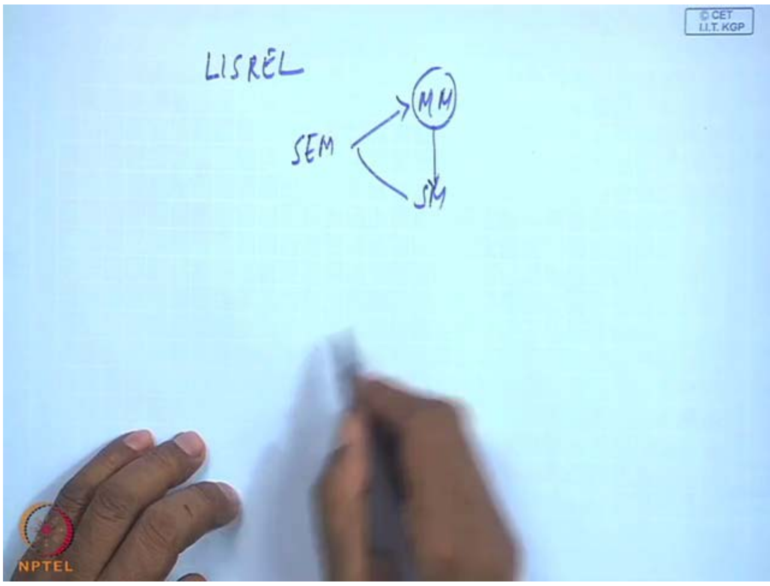
I think this Xi 3, X 1, X 2, X 3. X 3 is work injuries. It is kept as it is what X 3, Xi 3 is the negative personality, then Xi 4 is your safety environment. Again Xi 5 is job stress, Xi 6 is social support, Xi 7 is job dissatisfaction, Xi 8 is work hazards, Xi 9 is safe work behaviour. So, these are all latent variables in this sense now we also want to know the co relation matrix between latent variables which is the output of this measurement model.

(Refer Slide Time: 42:47)



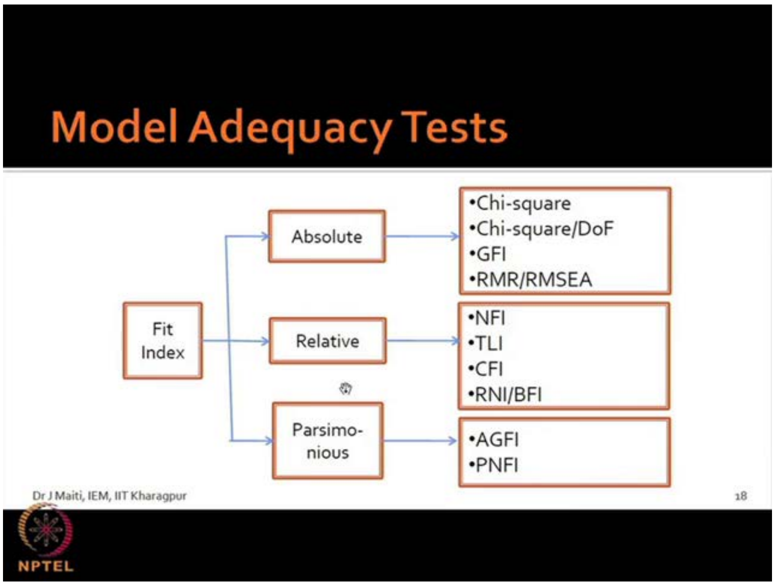
You see that demographic work injuries negative personality, these are the latent constants and this is your correlation matrix in then there are little star is there. This star indicates point 0.05 probability level of significance. I think all are significant here except this value 0.04, 0.01 some other value, but essentially we are interested from measurement model to know that what is the correlation matrix of the latent constructs or factors what we are going to evaluate or estimate, getting me? That is what we have done and we have done this with the help of this, this lambda values and the Xi and the error term you are getting this values. And this will be this is a value, you want this is very, very important one because this will be used in structural model as input to structural model in structural equation modelling, I told you.

(Refer Slide Time: 44:02)



That structural equation modelling two parts SEM has two parts, measurement model and structural model. The output of this will be input to this, fine? That I would consider.

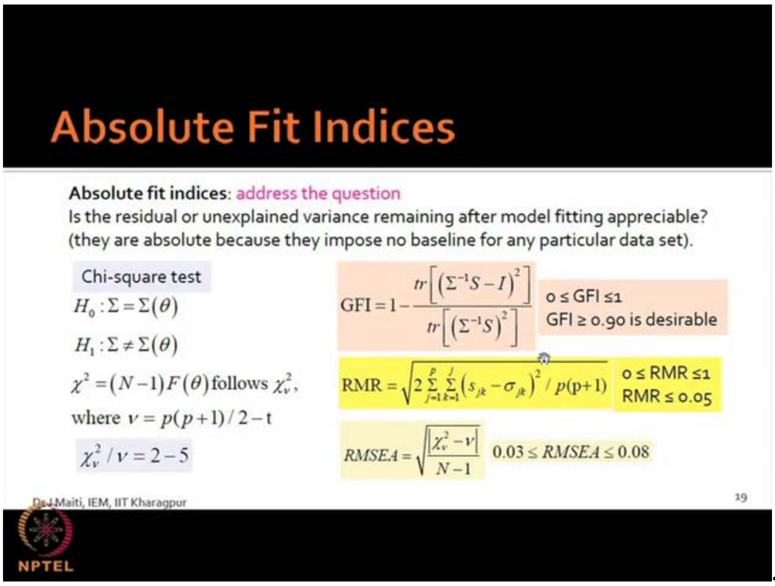
(Refer Slide Time: 44:19)



The structural measurement model would I consider measurement model as I said or not. If the model is not fit, if it is not adequate enough then the correlation matrix between the constructs generated they are not good. Also, we have doubt about those correlation values, we cannot abruptly accept this one. Now, in model adequacy test, in last class also I told you that fit index there are three types of fit index, absolute fit index, relative fit index and parsimonious fit index. Under absolute fit index chi square, chi square degree of freedom.

So, absolute fit index, relative and parsimonious we are discussing about and absolute fit index there are many indices like chi square, chi square by degree of freedom, goodness of fit index, root mean square, residual root mean square. That is RMR RMSEA standard error approximation, then relative fit index. These are the standard indices available in any literature related to structural equation modelling and most of the, why most, I think almost all the indices are based on chi square value. So, we will discuss little of this.

(Refer Slide Time: 45:48)

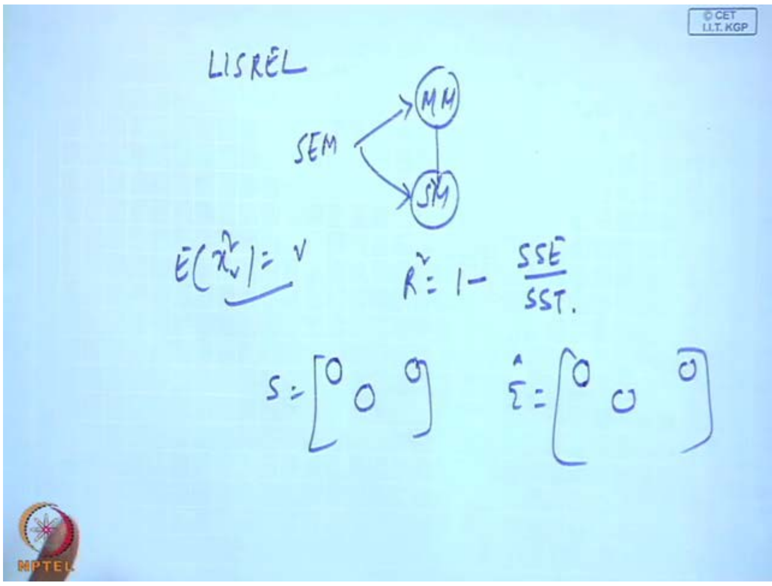


For example, absolute fit index what it does? It answers this question is the residual or unexplained variance remaining after model fitting appreciable. So, we do something like this. There will be two that hypothesis null and alternate hypothesis. Null hypothesis is we are saying that sigma equal to sigma theta, that actually that what you have estimated, that is correct and alternatively you are saying no, they are not correct. So, then this we will define one quantity called chi square, which is n minus 1 into F theta, that F theta you have seen that the minimization function, so that value you have after estimation.

So, chi square that n minus 1 F theta, this follows this chi square distribution with nue degrees of freedom, where nue can be estimated like this, p into p plus 1 by 2 minus t, that is number of non-redundant elements minus number of parameters to be estimated. That is what more degrees of freedom available here and you will find out that what chi square value you get that should be as small as possible, because if for perfect fit F theta will be 0, the n minus 1 into F theta that it will be 0. So, 0 is the ideal value, but it will all depends on sample size, also n minus 1.

Now, see that you will never get this your theta will be 0, because you are doing the numerical way of optimization, numerical optimization where some convergence value will be there. Now, if n is sufficiently large what will happen? This value will become large. So, what you want to how do then justify that whether the model is fit or not. One way is that this value should be as small as possible and other one is you go by chi square by degree of freedom.

(Refer Slide Time: 50:26)



So, it is recommended in the later lecture that essentially the chi square distribution is such that the expected value of chi square nu is nu, because that is the degree of freedom. Because it is a this parameter in chi square we use the degree of freedom only. So, actually the chi square by the degree of freedom should be 1, but is not recommended what is said that 225 is the recommended value above 35 constraints. Now, if you use G

F I that is goodness of fit index which is similar to R square in multiple regression, you can remember or recollect that R square equal to 1 minus SSE by SST.

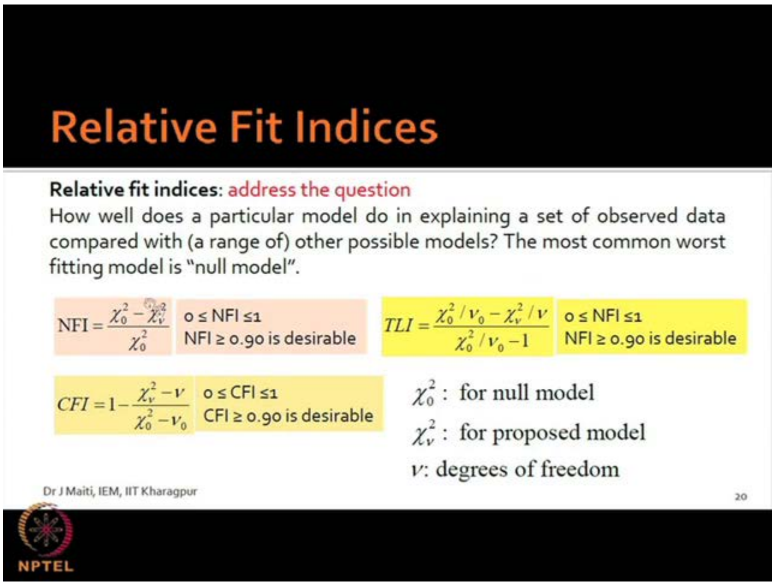
Now, you see this formulation here, that way we have written here that 1 minus trace of this by this. So, this is total variability and this one is the error term. It is similar to R square and this G F I value varies from 0 to 1 and it is desirable that G F I is greater than 0.90. So, in your model when develop a measurement model the software will give you the G F I value, if you find out that the G F I value is 0.9 or more, that it is good. It is desirable, but if it is less than 0.9 what you will do? You will not consider the error model, it all depends on the system for which you are developing the model. I am telling you even 0.8 also you can consider, absolutely no problem. If you think that the dynamics and is huge the volatility is more, many other issues you have to take into consideration.

Now, another index is RMR, I think this is something where each of the value of the S matrix and each of the corresponding value of the estimated matrix, that sigma basically you take S then you estimate sigma and by that process in between the parameters are also estimated. Now, the values sigma and S values, S this values and sigma estimated, this values are here, getting me?

Now, if you take this and this what is the difference? Take this, this what is the difference? So, we will take this, this what is the difference? These differences are squared here, you see what we have done S J K minus sigma J K this is the estimated one square by p into p plus 1, that is the non-redundant part. 2 is given because that twice of this, this by 2 p into p plus 1 by 2.

So, this quantity should be also as low as possible. It varies from 0 to 1 and RMR should be less than 0.05 and for RMSEA root mean square error approximation, this is the modulus of chi square minus its degrees of freedom by n minus 1. It is seen that 0.03 to 0.08 in this range this lies and this range also says substantial increment then relative fit index.

(Refer Slide Time: 51:33)

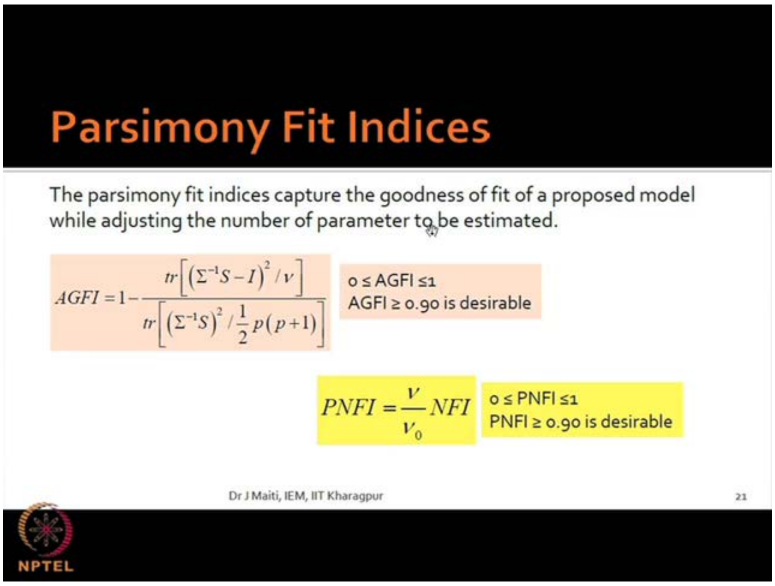


Now, how well does a particular model do in explaining a set of observed data compared with a range of other possible models. Here what you do? You creates nested model, several models and then you compare one model with other and then you say which model is better. And based on this you create a index and that index talks about your model adequacy or otherwise we can say improvement in terms of model adequacy. Here, most of mostly we will consider null model, that is the worst fit model which is known as null model, where we think that the covariance matrix is diagonal. Only diagonal means variance part is there of diagonal elements as 0.

Now, if you say that X chi square for the null model is chi square 0 and chi square for the proposed model is chi square mu, where mu is the degrees of freedom. Then you are in a position to develop or I can say quantify, this indices like NFI is chi square 0 minus chi square nu by chi square 0 and all these indices this values lie between 0 to 1. And it is desirable that they will be greater than 0.09, sorry 0.90 then CFI comparative fit index, that is 1 minus chi square nu minus nu by chi square 0 minus nu 0.

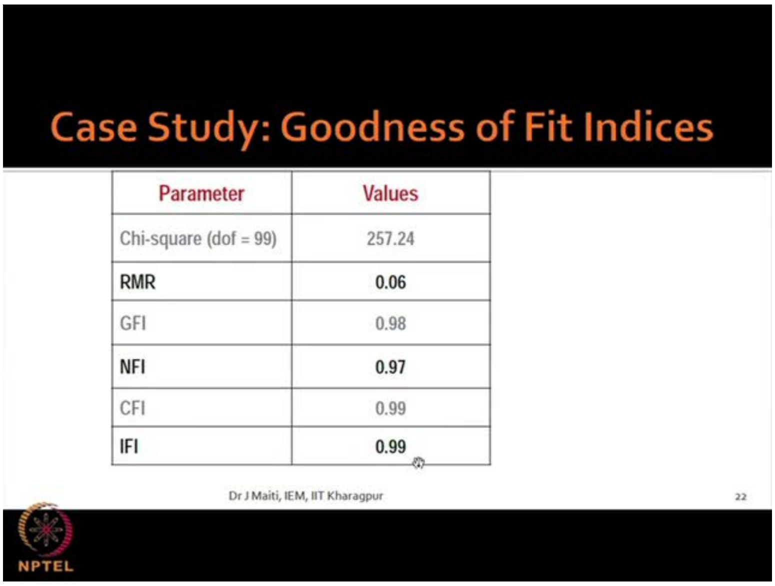
And TLI you have seen this, this also in the similar manner you see. Ultimately they take into consideration the chi square value of the proposed model and a worspit model which is known as null model and the comparative indices are developed and higher the index value, it is better 0.9 or more is required.

(Refer Slide Time: 53:39)



Then your parsimony fit index, it is similar to adjust R square S A square in you regression and it basically talks about the par parameter fit. Par parameter is estimates and AGFI here you just see that the top upper portion or the denominator here is divided by the degrees of freedom and numerator is also divided by the degrees of freedom. What we have done in calculating R A square, this value should lay between 0 and 1 and AGFI greater than 0.90 is desirable. Then parsimonious non fit index which is nu by nu zero NFI, and where this is nu is the degree of freedom proposed model and like this, okay?

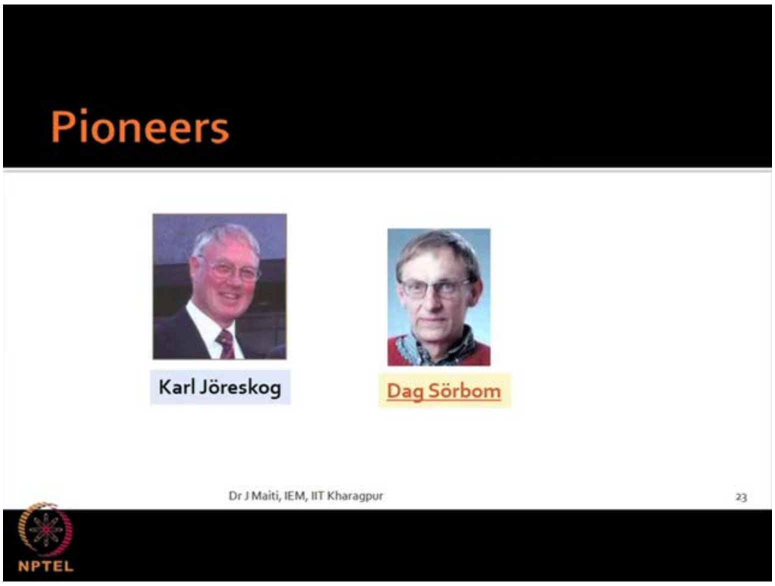
(Refer Slide Time: 54:27)



Parameter	Values
Chi-square (dof = 99)	257.24
RMR	**0.06**
GFI	0.98
NFI	**0.97**
CFI	0.99
IFI	**0.99**

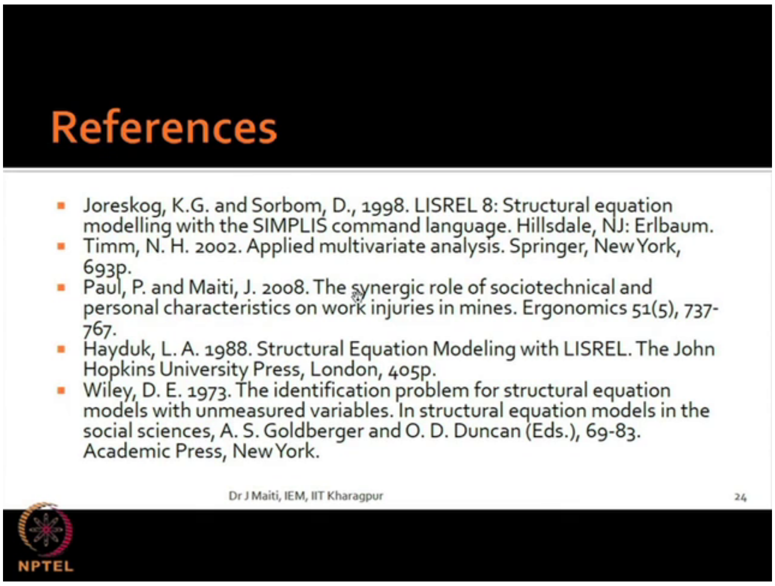
Now, let us see the goodness of fit indices for the case study here, some of the fit indices I have given there are others. So, chi square degree of freedom of 99, chi square value is 257.24, if you divide by 99, this is almost 100. So, it will be around chi square by degree of freedom is around 2.6. So, it is good because aim is to do 5 root mean square residual 0.06 which is little more than 0.05, CFI is 0.98 very good more than 0.90, NFI 0.97 more then 0.90, CFI is 0.99 and IFI is 0.99. So, essentially then chi square is 257.24 chi square by degree of freedom is around 2.6, RMR is 0.06, GFI is 0.98 like this. This model is very good, fit model.

(Refer Slide Time: 55:32)



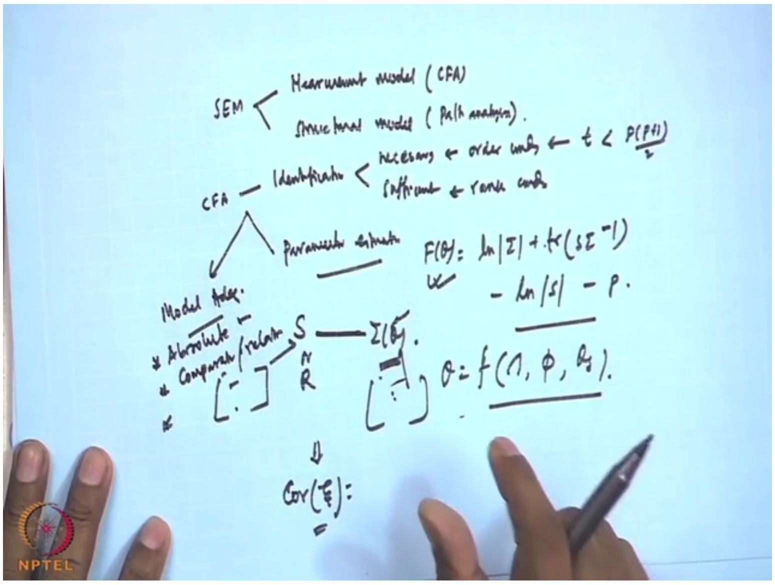
Now, see that who has basically worked in this, who are the pioneers that Karl Joreskog and Dag Sorbom, I think in around 1978 probably they have developed this software. First that listenery came I think in 1989 and its remarkable development in this field and we all are tremendously benefited.

(Refer Slide Time: 56:06)



What I can tell you further that for you have to understand the structural equation modelling. I might say that Joreskog, Sorbom this LISREL 8, structural equation medalling with SIMPLIS command language. This man, this manual is very good and you can go through and is a lot of publications by Joreskog, Soorbom. Others is the Hayduk is one person who has written a book, this and this is a very good book. Also in addition there are many other books available in structural equation modelling.

(Refer Slide Time: 56:48)



So finally, let me just summarize to my today's lecture. We said that structural equation modelling has two parts, one is measurement model measurement model, another one is structural model. So, measurement model is nothing but confirmatory factor analysis structural model is actually path analysis. Now, we have discussed details of CFA in terms of its identification, what I say that under identification there will be necessary condition.

There will be sufficient condition, this two must be satisfied in necessary condition is known as order condition and this one is known as rank condition. Other one sufficient condition is known as rank condition and in this order condition we say the number of parameters to be estimated must be less than number of non-redundant elements in the covariance matrix.

Then this is over identification case and it is a desirable case, then we have shown you the estimation parameter. Estimation, now I said that parameter estimation it is basically a function you minimize, which is basically log of determinant of sigma plus trace of S sigma inverse minus log of determinant of S minus p. This function is minimized through Newton Rapson or similar method and then the parameters are estimated and the actually we have sample data in terms of S or R. And we have the population value in terms of sigma theta. We try to match this two and using this function the better, the best match is considered and then you corresponding the theta value, these are used.

Now, theta is function of many things like lambda, like your phi, like your theta delta. So, these are theta, means theta means so many things are there. Any combinations that is what you are trying to estimate because from co relation matrix to here, co relation matrix or covariance to co variance matrix, one to one, this correspondence you are doing.

This is parameter estimation, once parameters are estimated then you can test the parameters values, this lambda using simple t test whether it is significant or not, but apart from this the another important output from this CFA is after parameter estimation is your co relation matrix of the latent correlation. Or covariance latent construct which is very, very important, because this will be the input to the structural model. Then what I have given you? I have given you the what are the model adequacy test.

So, under model adequacy we have seen that absolute test and then your comparative test or relative test, another one is parsimonious, absolute is similar to R square R A square, where the actual variance explain is considered in comparative case. We compare with different model and parsimonious, it is basically fit part parameter estimated and finally I have shown you a case study for all of you. The case study is there, if you are interested please go through Paul P and J Maiti, the synergic role of socio technical and personal characteristics in mines, published in ergonomics in 2008.

Thank you very much.