**INDIAN INSTITUTE
OF
TECHNOLOGY
KHARAGPUR**

**NPTEL
National Programme
on
Technology Enhanced Learning**

Applied Multivariate Statistical Modeling

Prof. J. Maiti

**Department of Industrial Engineering and Management
IIT Kharagpur**

Lecture – 39

<u>Topic</u>

SEM – Measurement Model

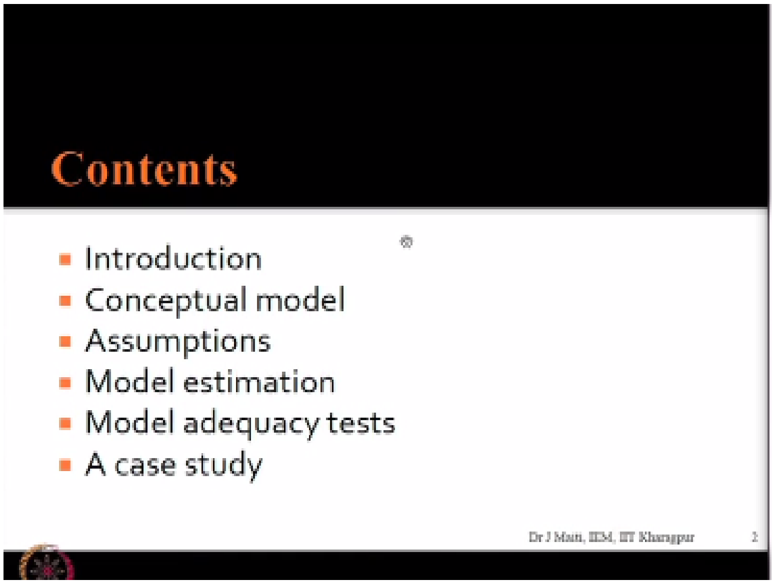
Today, we will start a structural equation modeling part 1.

(Refer Slide Time: 00:27)

That is measurement model, SEM that is structural equation modeling, measurement model.

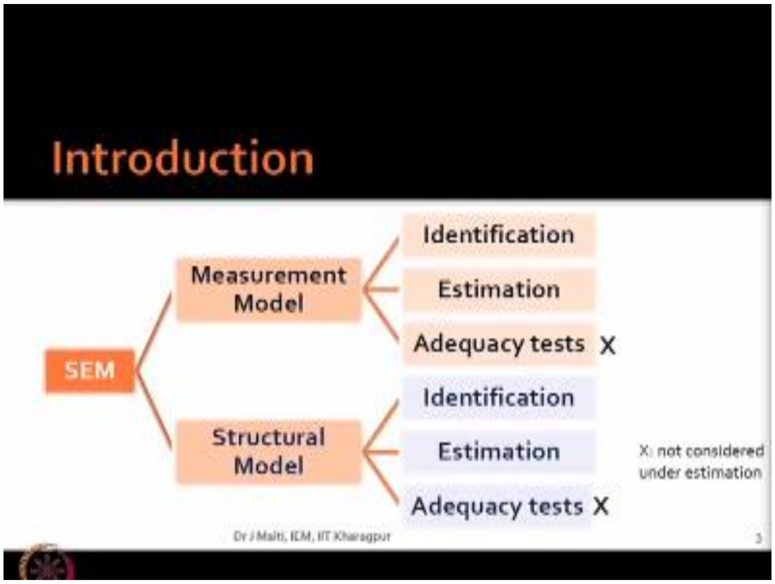
(Refer Slide Time: 00:40)

(Refer Slide Time: 00:41)



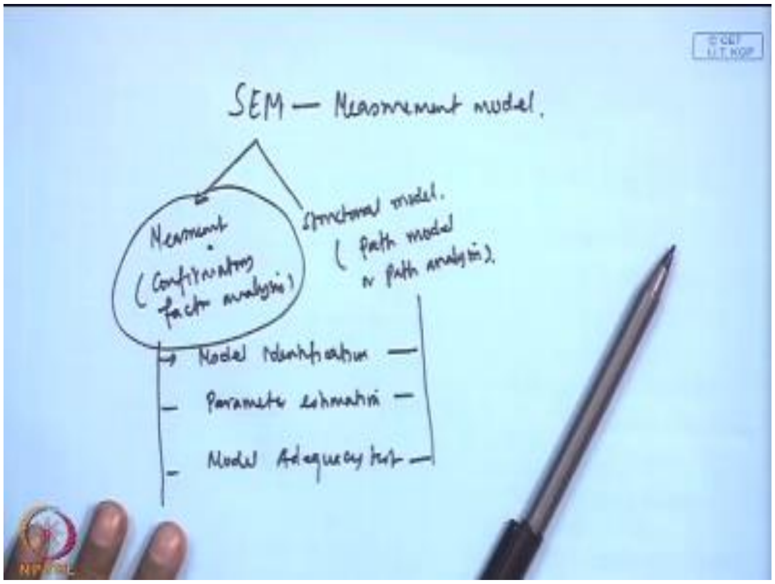
Now, let me see the content of today's presentation. That we will start with conceptual model, then the assumptions of the model, then how to estimate the model parameters and model adequacy test, followed by a case study.

(Refer Slide Time: 00:58)



You see in last class.

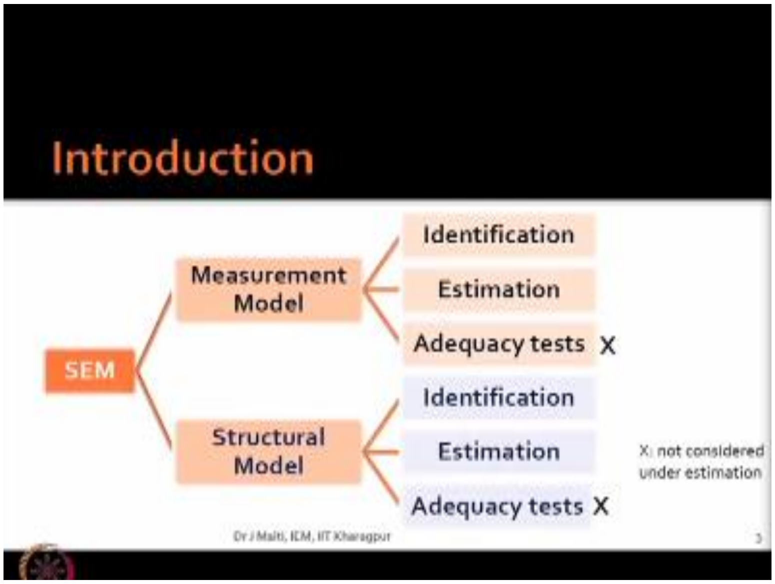
(Refer Slide Time: 01:07)



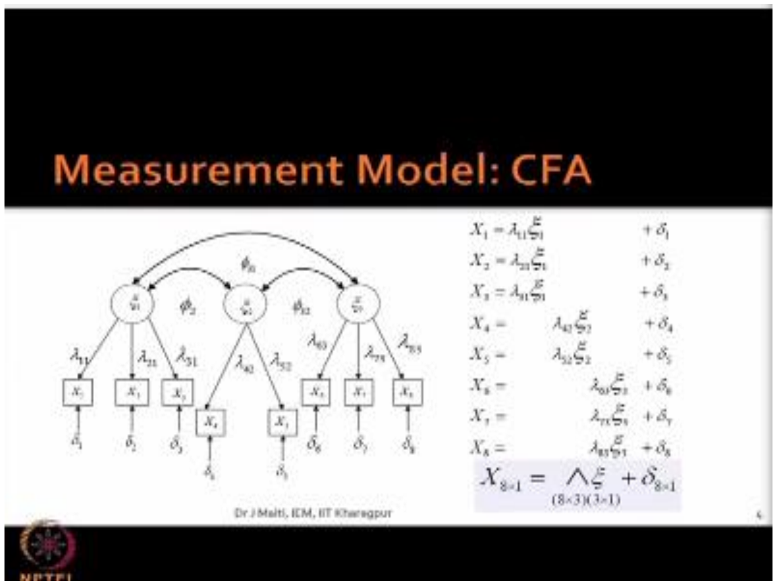
I have explained that measure structural equation modeling has two component, one is measurement, and another one is structural component. The measurement component is essentially a confirmatory factor analysis, and structural part or we can say the structural model is equivalent to your path model or 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.

(Refer Slide Time: 02:34)



(Refer Slide Time: 02:35)



So, let us start with a conceptual model first. You see here, in last class I have shown you similar diagram and you see there are three factors ζ1, ζ2 and ζ3 and each of the factors are manifested by a different variables starting from X1, X2, X3 for ζ1, X4 and X5 for ζ2 and X6, X7, X8 for ζ3. So, in confirmatory factor analysis the basis is that, that there are hidden constructs which are this or latent construct, other you can say hidden variable also that ζ1, ζ2, ζ3 which are the causes of some manifest variable like X1, X2, X3 to X8 by putting this arrogate.

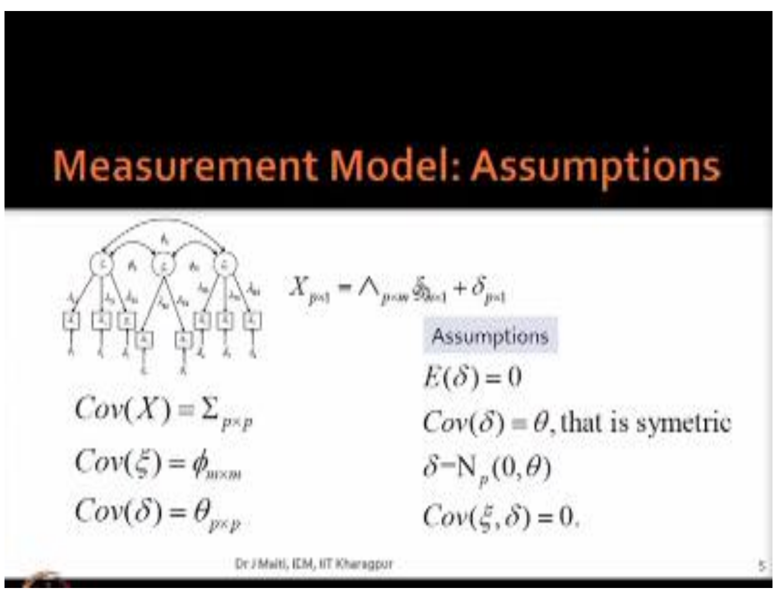
For example, from ζ1 to X1, ζ1 to X2 and ζ1 to X3 we are restricting the model here in such a sense that we know that ζ1 is manifested by X1, X2 and X3, this manifest variable. Similarly, ζ2 is manifested by X4 and X5. Similarly, ζ3 is manifested by X6, X7 and X8, another issue here that these hidden construct or latent construct or latent variable they co-vary in the sense that if ζ1 change there may be change of ζ2, there may be change of ζ3 when the correlation component 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 constructs which is denoted by ϕ this one ϕ21, this is ϕ31, this curvature line, this curvature line is basically ϕ32. Now, we are saying X1 is caused by ζ1 and as a result a causal linkage is given, ζ1 to X1 and the parameter which basically depicting the relationship between ζ1 and X1 is λ11.

Earlier, I also told you this λ11, this 11 the suffix comes that X1 from X1 this 1 is taken and ζ1 1 is taken okay. So, it is not possible that the variability of X1 will be fully explained by ζ1. So there is possibility of some other variables or hidden causes which may affect X1 or we can say the errors part, noise part and all those things are considered by δ1. So, in the same manner you have to explain X1, X2, X3 and up to X8.

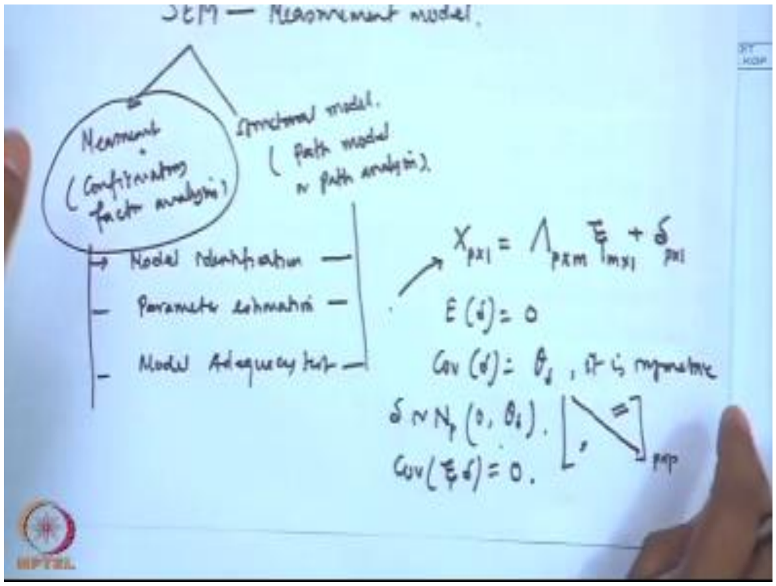
What I said verbally, this is depicted in equation form. We are saying X1 is represented by λ1 or ζ1 + δ1. If you think from the regression line point of view you will be getting this equation okay. So, in the same manner there are, as there are 8X variables, so you are getting 8 linear equations and collectively if you write in matrix form then that will be X8 x 1 equal to that Λ which is a matrix of the dimension 8 x 3 and which you can see here 8 x 3 dimensions here. Because ζ1, ζ2, ζ3 that 3 and 8X variables, this 8 plus this ζ+ δ this one. So, this is the matrix form equation for this particular example.

(Refer Slide Time: 07:10)



Now, we can go for a general equation from there.

(Refer Slide Time: 07:16)

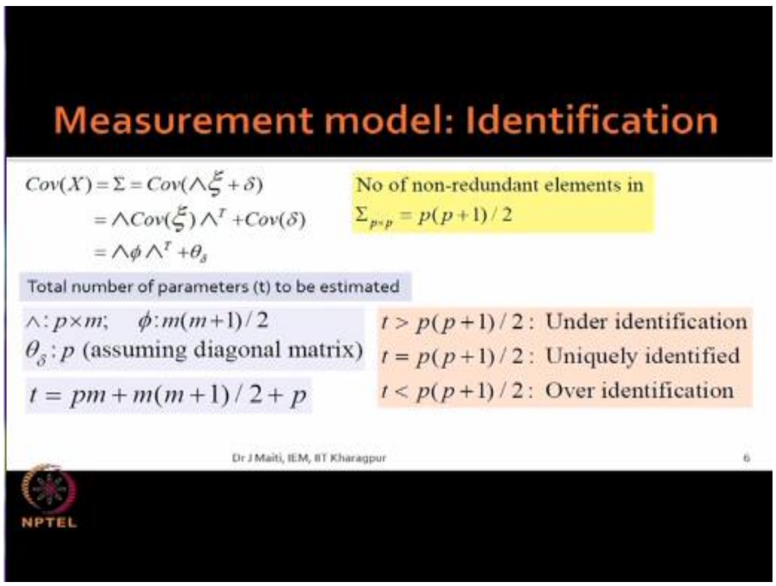


That means what I means to say here the general equation will be X p x 1 where we are saying, there are P number of manifest variables and which can be represented by this manner, the Λ which is a matrix of matrix relating P manifest variable to m factors. And this m factors are denoted by like this m x 1 and definitely for every manifest variable there will be an error. So, this is the equation for confirmatory factor model.

And in fact if you see the recall that factor analysis you have found out there also similar relationship, but there are, there is difference okay, 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 $\delta = 0$, that mean the noise variable, the arrow terms that mean is 0 and covariance of δ, this one is your θ, or you can write $\theta\delta$, also sometimes $\theta\delta$ and it is symmetric okay.

So, if I say like this will be P x P. So, this you variance component of δ of diagonal in the covariance that will be equal, that is why symmetric and δ is multivariate normal with mean 0 and covariance matrix θδ, or you can write θ also okay. Another important issue here is that assumption is that covariance between ζ the manifest, sorry the latent construct and the arrow term related to the manifest variable, they are 0 okay. So, this is our your assumptions related to confirmatory factor model. Now, there as I told you there are covariance structure.

(Refer Slide Time: 09:51)

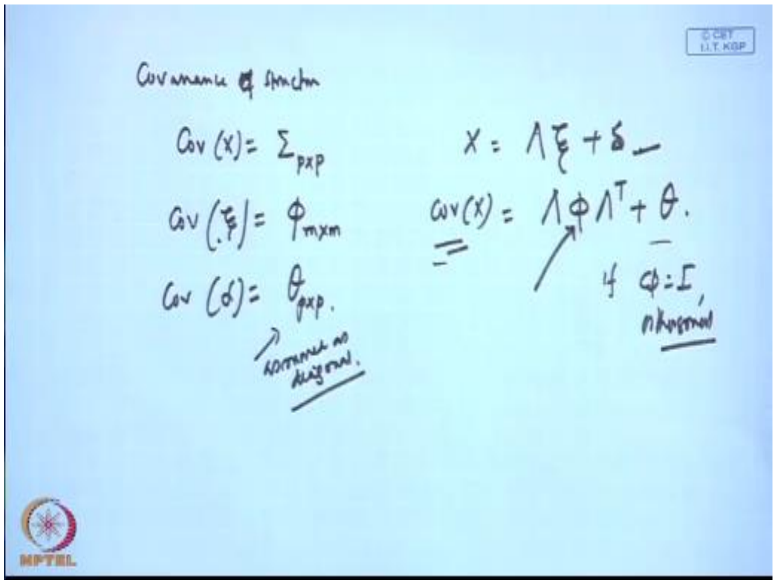


So, the covariance structure of your see covariance structure, first one is covariance of X, this will be ∑ P x P. There will be covariance structure for the ζ latent construct this one will be Ø which is again m x m matrix diagonal not diagonal this is symmetric matrix. Then covariance of δ which we say $\theta_{p \times p}$ okay 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 = δ ζ + I think we had **λ** ζ + θ means this **λ** replace δ. 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 δ. This is θ in x and this Ø was not there in excusive factor, it was I if Ø = I then it is orthogonal factor analysis. So, this is null set that covariance structure and the relationship between this, okay.

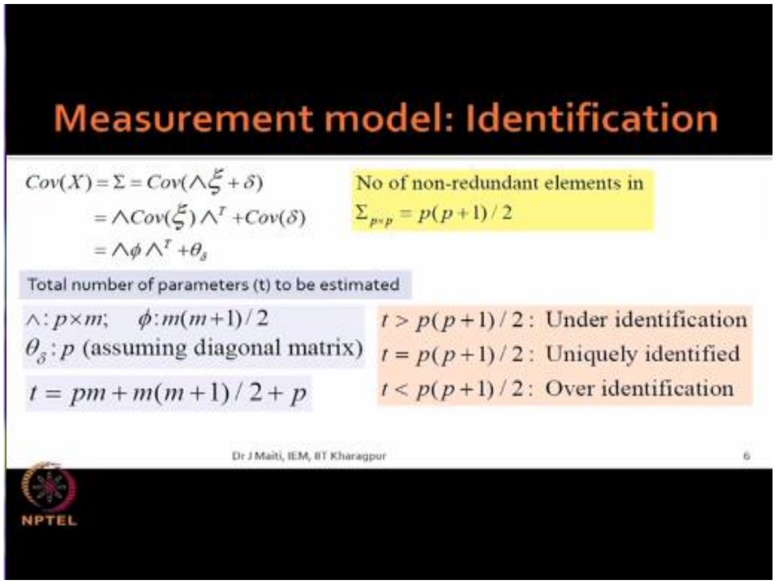
So, now let us stick to come to the model identification part, what I mean to say hereby model identification if you clearly look into the model.

(Refer Slide Time: 12:20)



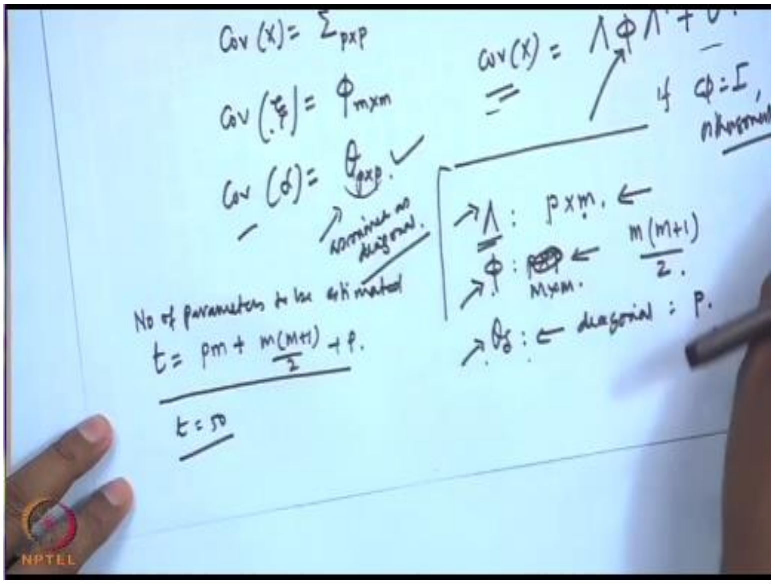
And the parameters that to be estimated as well as the information, what is available there must be sufficient necessity and sufficiency of the information available to estimate the parameters of the confirmatory factor model. Okay so, in order to do so.

(Refer Slide Time: 12:42)



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.

(Refer Slide Time: 12:59)

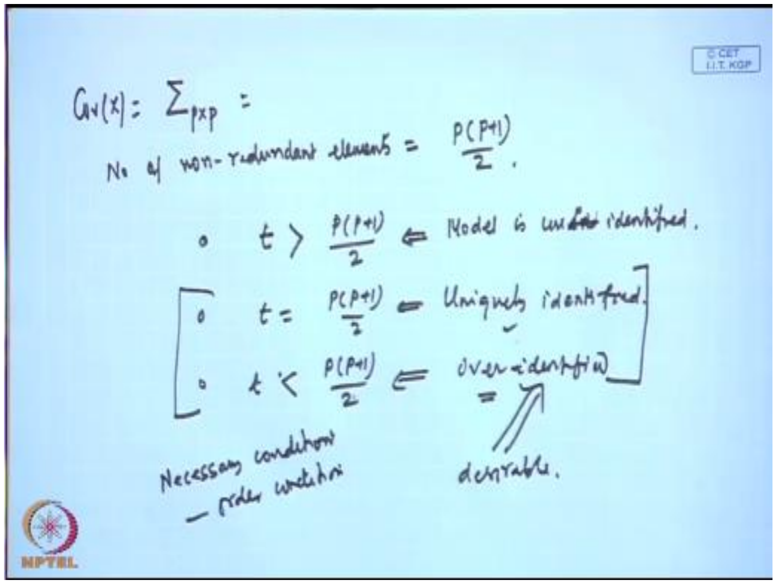


From in CFA 1 is your λ which is p x m matrix so, these many parameter sto be estimated, there is Ø which is also a p x p matrix which is p x p matrix, but being symmetric it has Ø is m x m matrix. I am sorry Ø is m x m matrix which is symmetric matrix. So, numbers of parameters will be m (m+1)/ 2 to be estimated and then there is θ δ as I told you that θ δ what is this θ δ.

This is δ, this is θ δ so, θ δ or θ 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 λ, we require to estimate Ø we require to estimate θ δ here in case of λ p x m, this number of parameters Ø m(m + 1)/ 2, this number of parameter θ δ p number or parameters are there. So then in total the number of parameters to be estimated, number of parameter to be estimated we can write p (m + 1)/ 2 +p okay.

Now, what you require to know that if t suppose t = 50. If we require to estimate p 50 parameters and you require at least 50 simultaneous equations, getting me. So now, what information we have in case of confirmatory factor analysis.

(Refer Slide Time: 15:30)

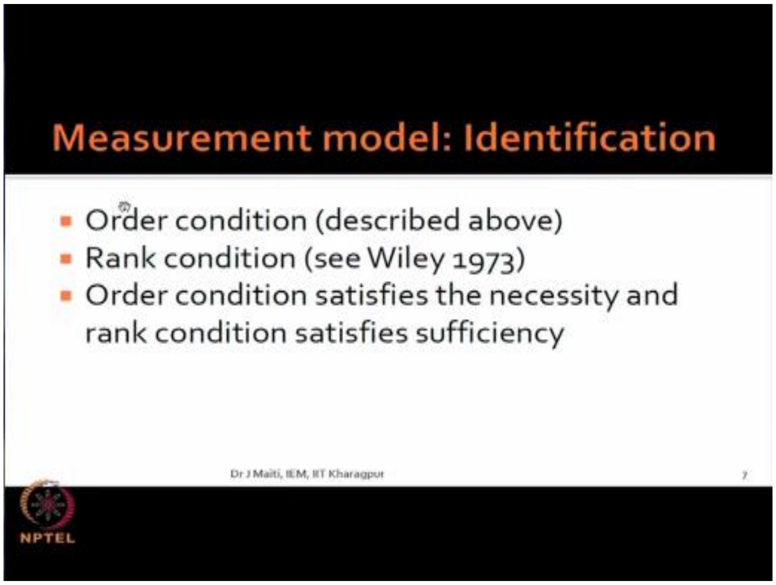


We have only one information that is Σ which is the variance, co variance matrix of X that is co variance matrix of X. Now, how many unique or non-redundant elements, this is also symmetric matrix. So, it has number of non redundant element equals to p(p +1)/ 2. Now, see this what type of situation will 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 = p (p +1)/ 2. It may so happen that t < p (p +1)/ 2.

Now, the first case, this is model cannot be identified. This is model is unidentified or under identified, unidentified this case number of parameters or the number of unknowns 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 overestimate 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.

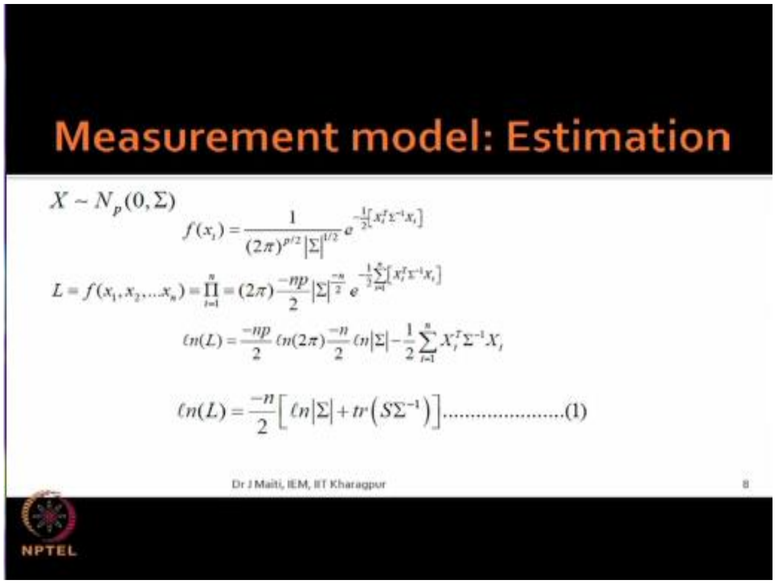
Necessary condition which is also known as ordered condition, okay but ordered condition alone is not sufficient.

(Refer Slide Time: 18:08)



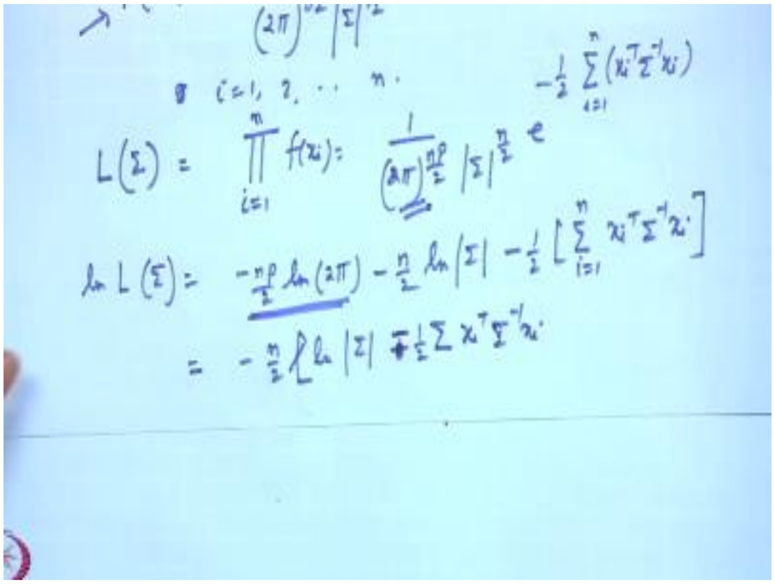
This is necessary condition necessity is satisfied there is another condition called rank condition, because then you will see that we have basically talking about the matrices. So, the rank condition, rank of my matrix is important issue here and it is little bit complicated one also. Rank conditions also per value you have to satisfy then the rank condition is the sufficient condition and for example, is given that Wiley in 1973 that is the reference. Now, order conditions satisfy the necessity and rank conditions satisfied the sufficiency.

(Refer Slide Time: 18:56)



Let us assume that it is done in the sense model e is identified, if model is identified. The next step is how to estimate the model.

(Refer Slide Time: 19:07)



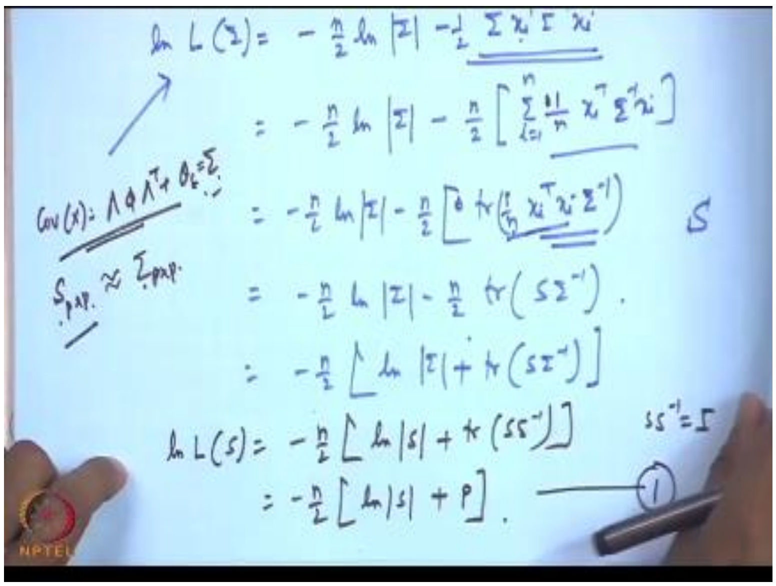
So, estimation of model parameters, okay. So you have seen that we assume that ultimately that x, the manifest variable is normally distributed. It is the primitive multivariate normal p_x variable p, p number of variable are there, multivariate normal with mean 0 and variance, co variance matrix Σ. Then for any observation multivariate observation X_i you can write that this is the PDF can be written like this, 2 $\pi^{p/2}$ Σ determinant to the ½ $e^{½}$

Then X- e, e is to write that is Xi – μ is zero here. So Xi – μ^{-1} that means Xi transpose σ^{-1} Xi this is the multivariate normal distribution per density function for a particular multivariate observation. Now we collect in observation I = 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 σ and μ is 0.

So only one parameter is there so we can write log of likelihood of σ not log likelihood of σ which can be written as multiple equation of this i = 1 to n f(xi) which will be multiplying I = X = 1, X 1, X 2……X n, then the resultant will be like this 1/(2Π $)^{np/2}$ .Then determinant to the power n by 2 then $e^{-1/2}$ then sum I = 1 to n. Then Xi^T σ^{-1} 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 –np/2 log 2Π for this term, -n/2 log this for this term -1/2 I = 1 to nXi^T σ^{-1} Xi. So obviously this from our we want to estimate this σ 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. So we remove this constant term, so you can write this as –n/2 log of this plus half of I can write like this –n/2 into this minus. So minus half of this Xi^T this xi, okay.

(Refer Slide Time: 23:25)



So let me write a phrase that log of l, this σ = -n/2 log determinant σ – ½ Xi^T σ^{-1} Xi. Now this term can be written like this –n/2 log this minus if I write n/2 then summation i = 1 to n. This can be written like this, this can be written like this 1/ 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/2 – n/2.

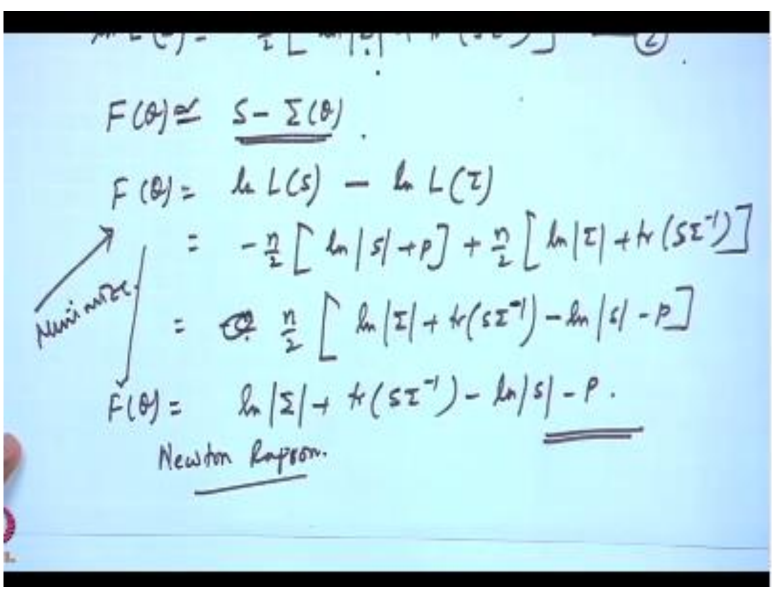
We can write this quantity as trace1 by trace of $1/xi^T$ xi σ^{-1} this, this is possible. Now $1/nXi^T$ this is nothing but variance, covariance matrix of the sample provided, n is large, 1/n – 1/n become same. So with modeling equation we can write like this –n/2 then this is trace of s σ^{-1}. This one is

–n/2 log of this plus trace of s σ^{-1}, okay. 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 λ Φ, λ^{-1} + θ δ. From this you will get covariance Xσ that is σ in terms of model parameter you collect sample then you get the covariance matrix also, so from sample there will be S. S is the again p x 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 = σ then if I put here what I can write, Log of l(s) this can be written like this, this log of S determinant of S plus trace of SS^{-1}. Now S S^{-1} I, so you can write like this, this log(s) plus sum of the diagonal elements of the matrix I that is p, okay. So this is your equation number 1.

(Refer Slide Time: 27:02)



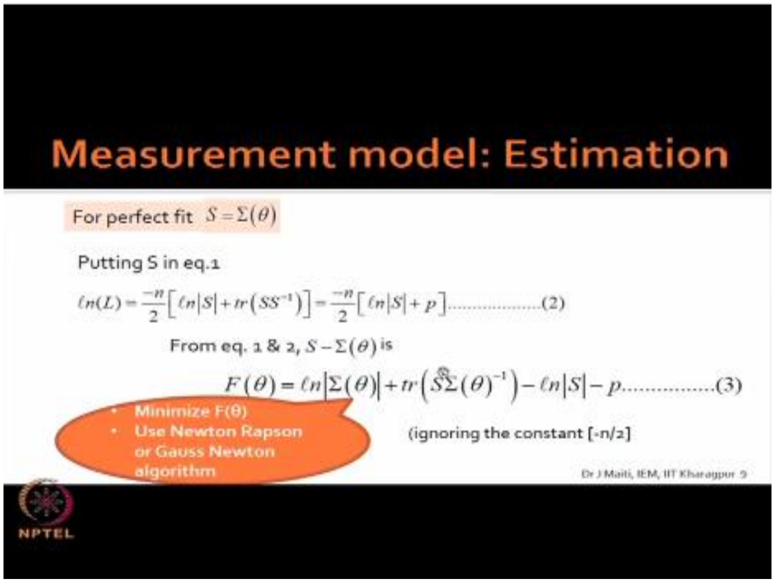
Now another one is we have already seen, the likelihood one this equal to –n/2 log of this plus trace of this, this is our equation 2. So what you want, we want to find out parameters. Now this σ this will be in terms of model parameter S here and here when we are talking about S it is basically the numerical values and here it is in terms of model parameters like λ Φ and all those things.

So we will create a function now that we want to minimize that we are saying F θ which is nothing but S - σ θ of this nature which I am saying not exactly which will be of this nature. So then we can write like this Fθ = log L(s) - log L(σ) which if you write this is –n/2 log of determinant of S + p + n/ 2 log of determent of σ plus trace of S this trace of S σ^{-1}. So this one you can write minus that n/2.

Then log of determinant of σ + trace of S σ^{-1} – log(S) determent S - p we want to minimize this function. So keeping this constant n/2 again is of no use. So final equation will be for our estimation is this log of determinant of σ + trace of S σ^{-1} – log (s – p). This is the equation which want to minimize, okay. So its null issue you have to use Newton Rapson or similar method, Newton Rapson similar method of numerical that optimization part, okay.

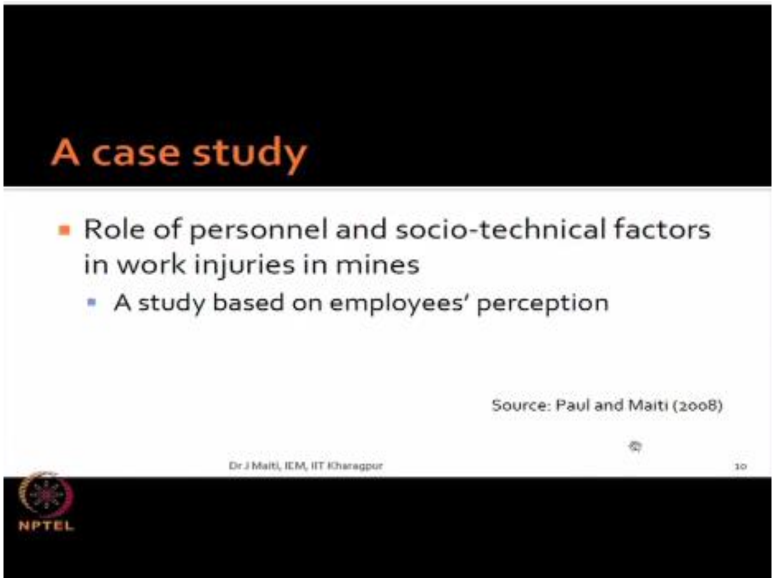
So this is what is in the nut shell the parameter estimation in place of confirmatory factor analysis which is basically our measurement model.

(Refer Slide Time: 30:05)



You see here this is the theta log σ theta trace of this is the case ignoring the constant so ultimately minimize this one using Newton Rapson or Gauss algorithms okay.

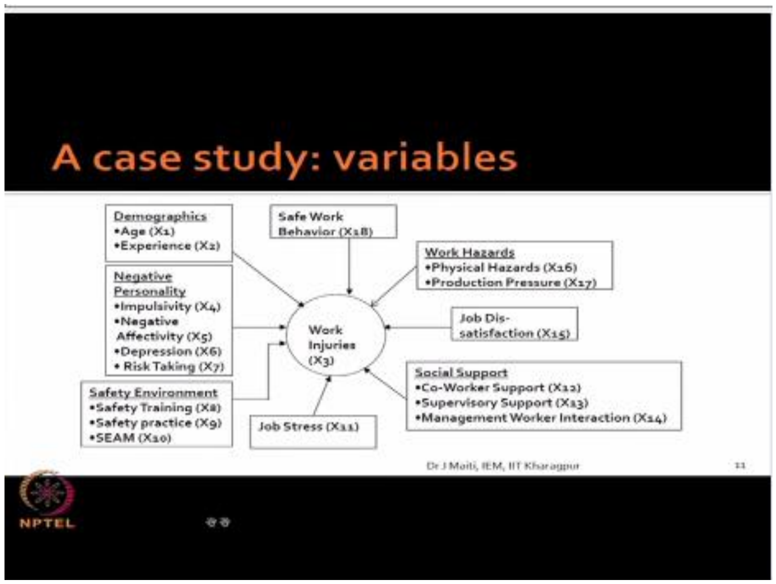
(Refer Slide Time: 30:23)



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 modeling 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.

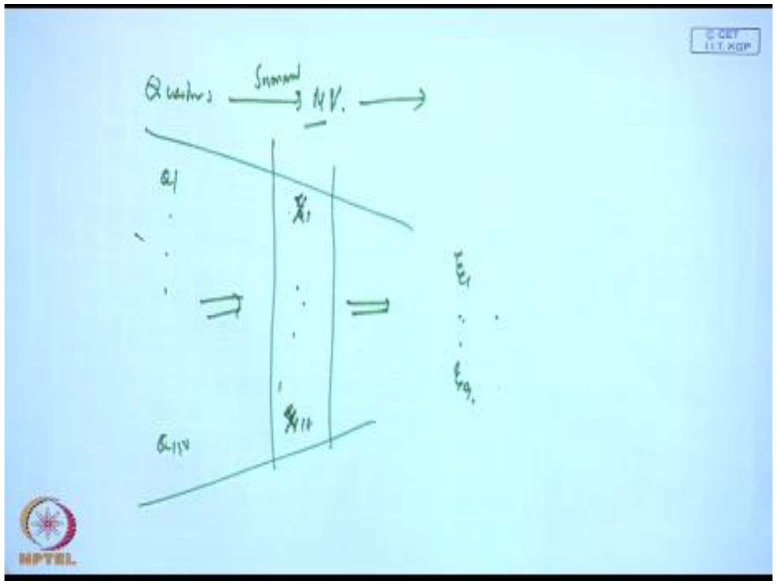
(Refer Slide Time: 31:27)



Okay 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.

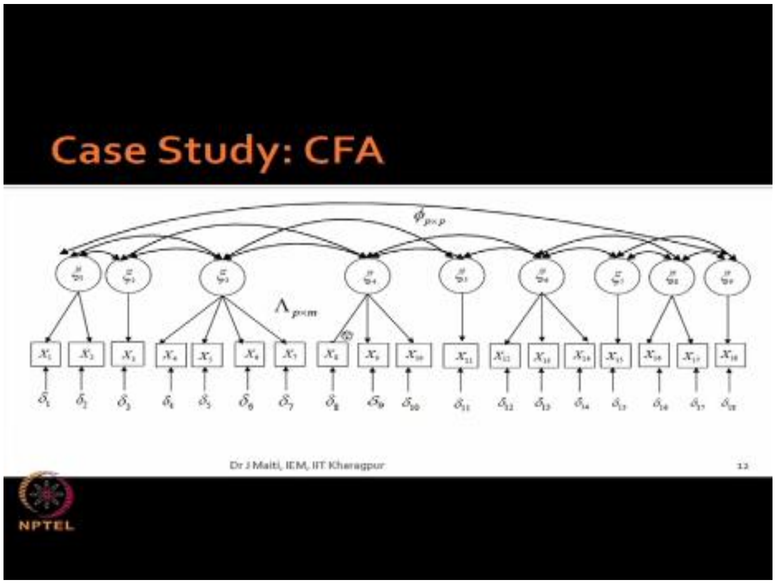
For each of the observations or individuals who participated in this study so in that sense it is manifested means observed in that sense otherwise its two layer questions actual it was like this only one questions.

(Refer Slide Time: 32:37)



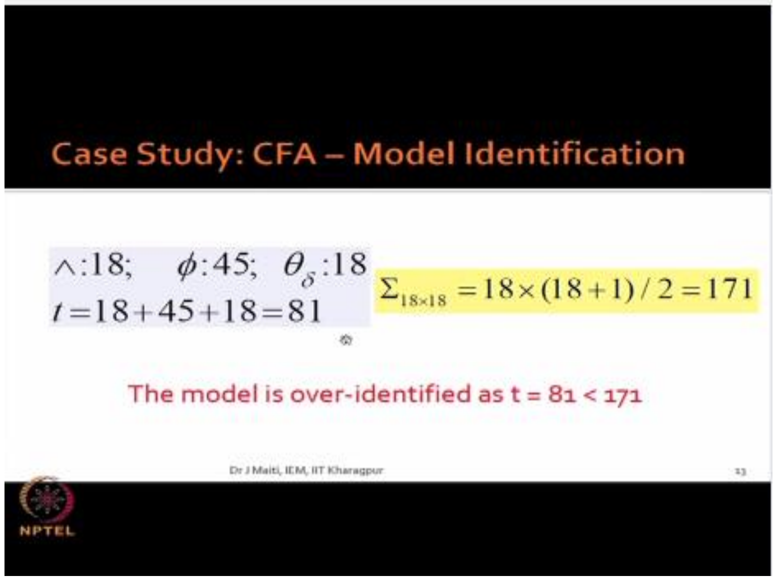
Then there sum to this manifest variables what we are saying there sum then further level of aggression actually suppose there are question 1 to let it be 150 then there are manifest variable like ζ1 like there are 18 then this again it is aggregated into what I can say these are X_1 sorry these are X_1 to X_{18} related to X_{i1} to some X_i 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.

(Refer Slide Time: 33:39)



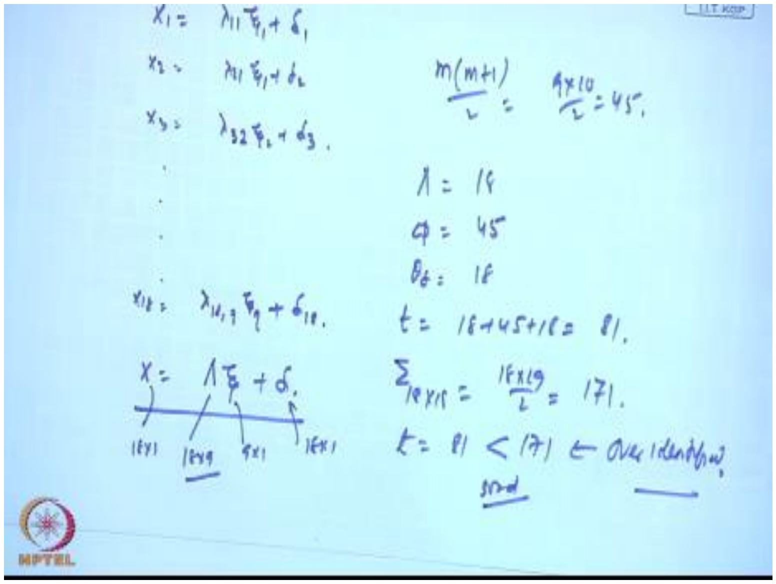
If you write in the confirmatory factor analysis form it will be something like this you see all the this covariance structure between this 9 ζ, ζ variables it is not pictorially shown because of space concept otherwise this will be this and then what will happen we will immediately you can go for.

(Refer Slide Time: 34:05)



The equations also for this getting me so like I am giving one equation only here if I want to know what isX_1.

(Refer Slide Time: 34:19)



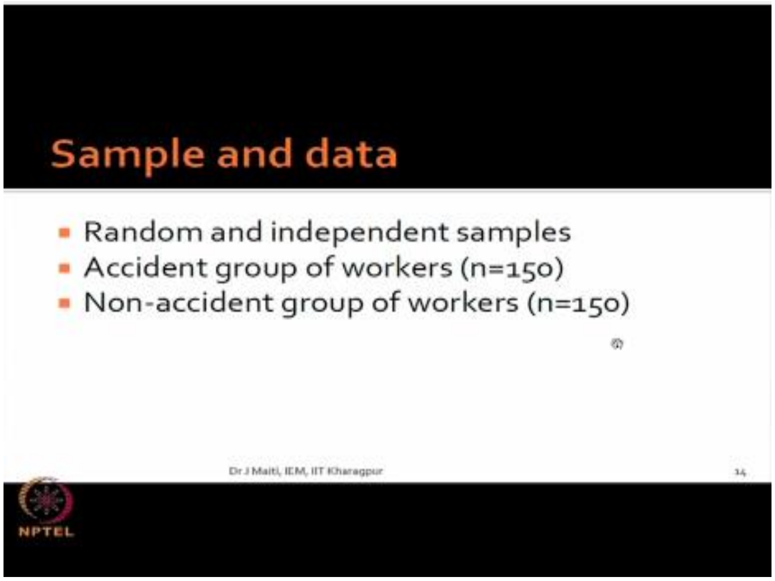
Then this is nothing I can write λ 1 1 ζ1 + δ_1 if you consider this here λ 11 ζ1 + δ_1 similarly this one λ_2 1 ζ1 + so x_2 will be λ_2 1 ζ_1 + δ_2 if you consider X_3 so X_3 is ultimately it is the single indicator manifest variable for the constant ζ 2 so X_3 can be written like this that 3 2 ζ 2+ δ 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 ζ_9 so this is your λ 18 9 ζ_9 + δ 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 λ ζ + δ this equation you can find out here we have18 cross 1 this will also be 18 x 1this is our 9 x 1 so this will be 18 x9 so this type of equation you can find out okay now let us see the model identification for this case you just see that λ part that how many λ s are there you count 1, 2,3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,16, 17, 18 so 18 λ so we have written λ18 because others are 0 for example λ3 1 if you give one linkage here λ3 1 that is 0 because it is a confirmatory we know what are the manifest variable coming out of the hidden constructs π.

There are how many X_{18} x so how many Φ 9 Φ So m into m x m into m +1 by 2 m into m + 1 by 2 so 9 into 10 by 2 that will be 45 so you have λ that is your 18 then your π related variables will be 45 π related parameters will be 45 theta δ again 18 δ 1 to δ 18 so our t is18 + 45 + 18 that is

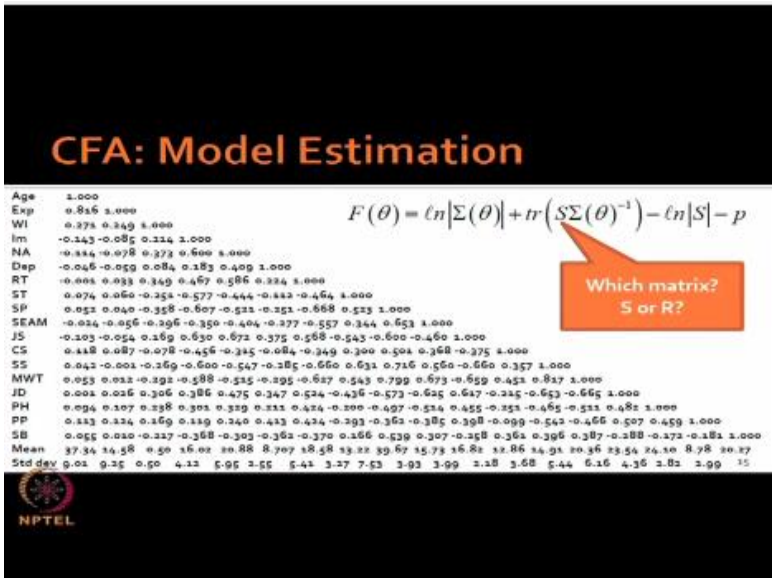
81 now what is the unique elements there we have 18 manifest variable cross 18 so 18 x19 by 2 this will be 171 now t equal to 81 it is much less than 171 the model is over identified it is a good case so and necessity.

(Refer Slide Time: 38:15)



Condition you satisfied and sufficiency we have not tested here that says the software they test all those things 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.

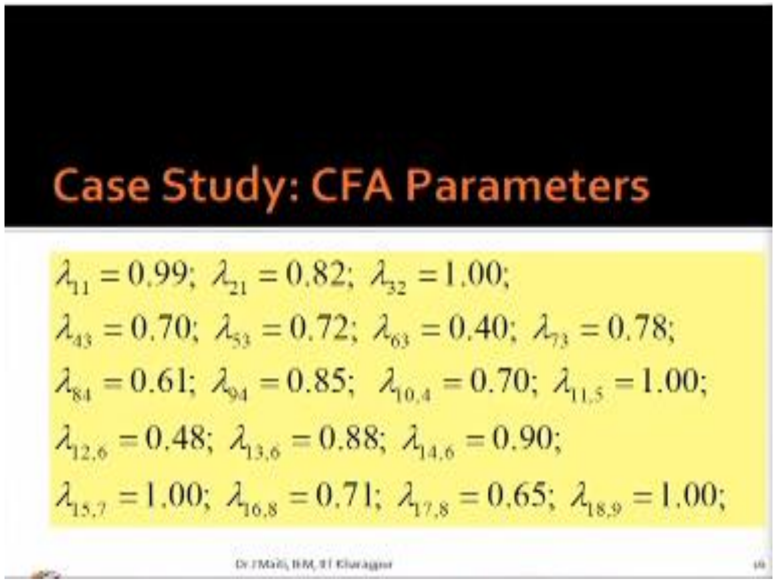
(Refer Slide Time: 38:43)



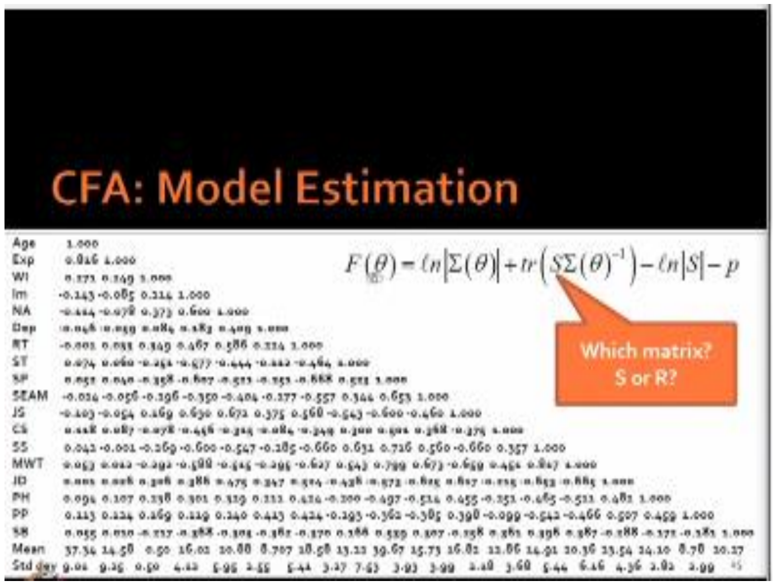
So immediately as I told you that how many independent non redundant element in your σ matrix this is the case this is from sample these are all co relation matrix now question is what we want we want basically to minimize this function and in this matrix is basically for S that is sample co variance 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)



(Refer Slide Time: 40:02)

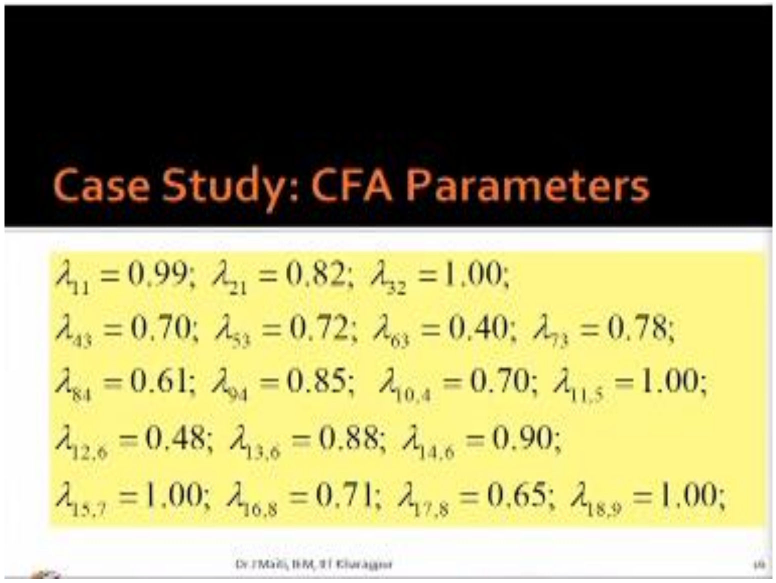


Now, through definitely this is F θ, this is the function which is to be minimized and we have used this software, in this case LISREL linear structural relations.

(Refer Slide Time: 40:20)

So this software we use and ultimately this is what is the all the parameters.

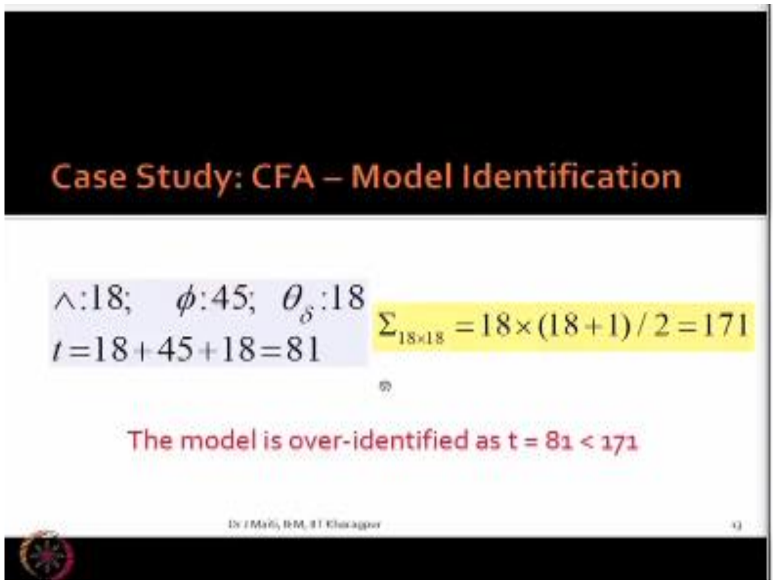
(Refer Slide Time: 40:36)



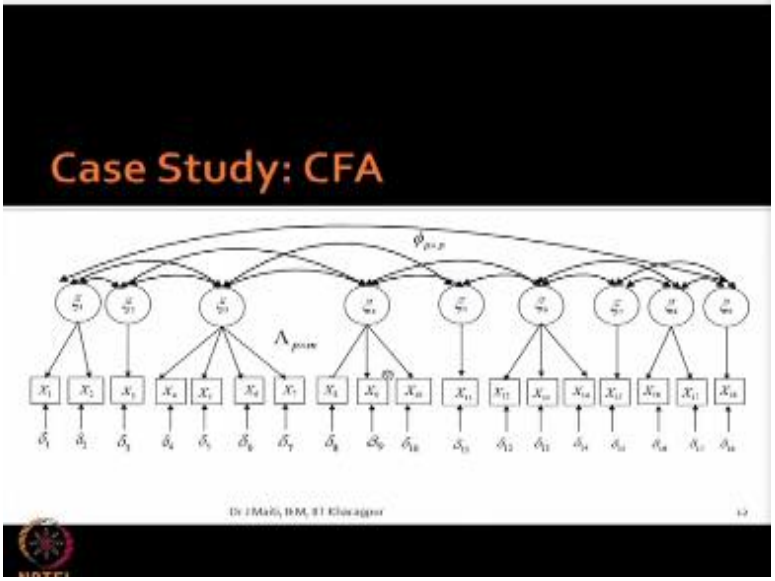
Which is estimated you can see.

(Refer Slide Time: 40:39)

(Refer Slide Time: 40:40)

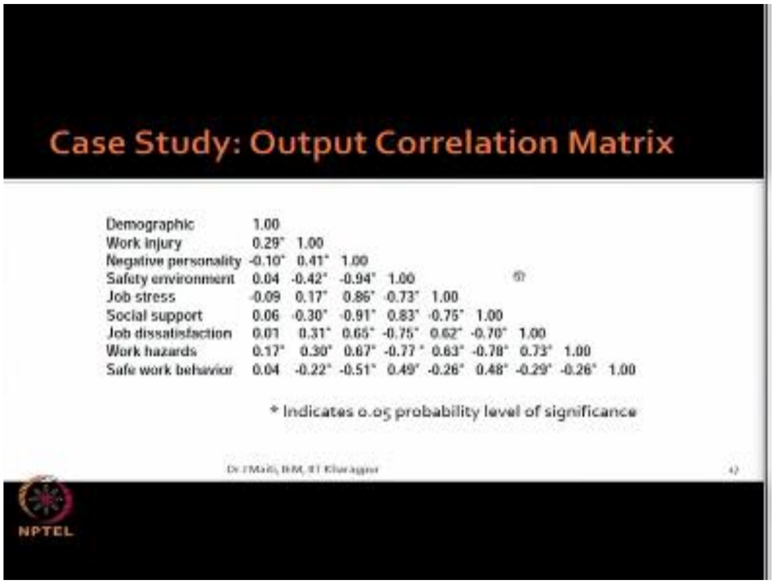


(Refer Slide Time: 40:41)



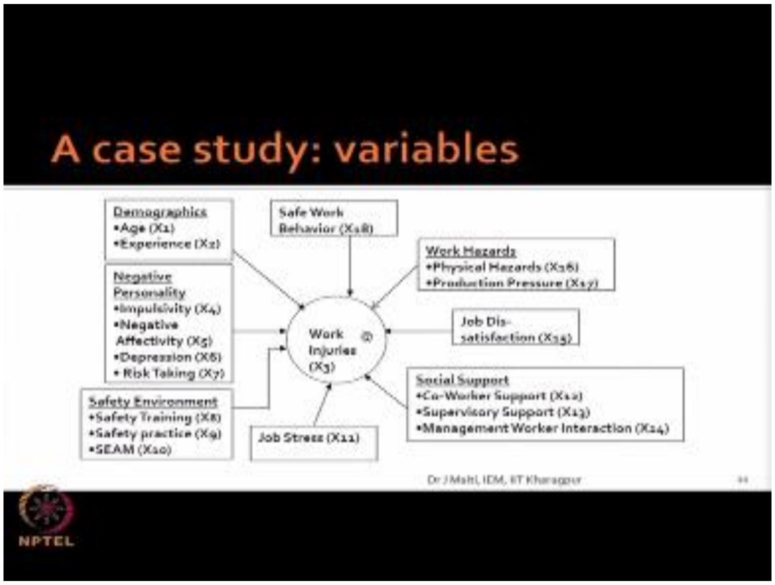
If you go back λ1, λ 2 1, λ 3 2, λ 4 3 like this and you see that λ 3 2. What is the value of λ 3 2? Here, λ 3 2is 1 because this one indicator with one constant, we assume that this is the manifest variable itself is the construct, understood?

(Refer Slide Time: 41:14)



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,

(Refer Slide Time: 41:23)

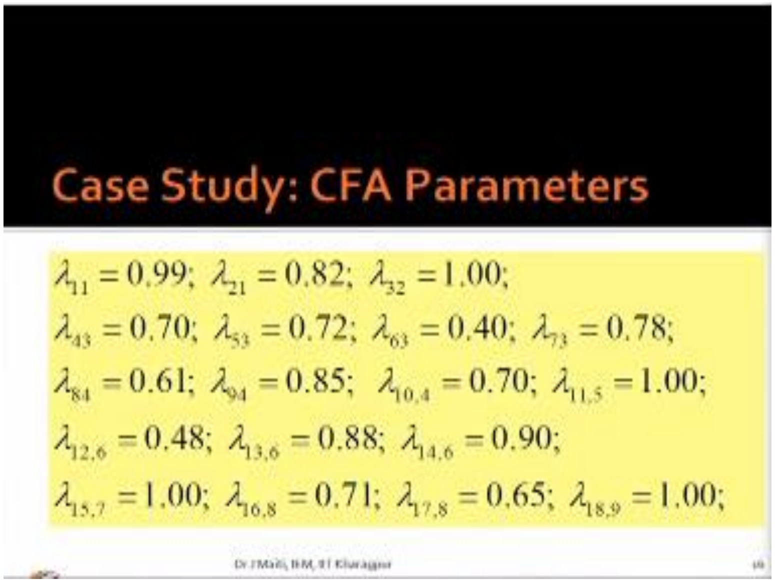


These things when we have clubbed into this factors these are all the factors or constructs latent constructs this Xi 1 to Xi 9 these are not arbitrary Xi 1 to Xi 9, they have some meaning. Actually X_1, X_2 if you see that age and experience then demographics, this is the this is Xi 1 impulsivity negativity all four are clubbed to the weather 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, and then Xi4 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 behavior. 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.

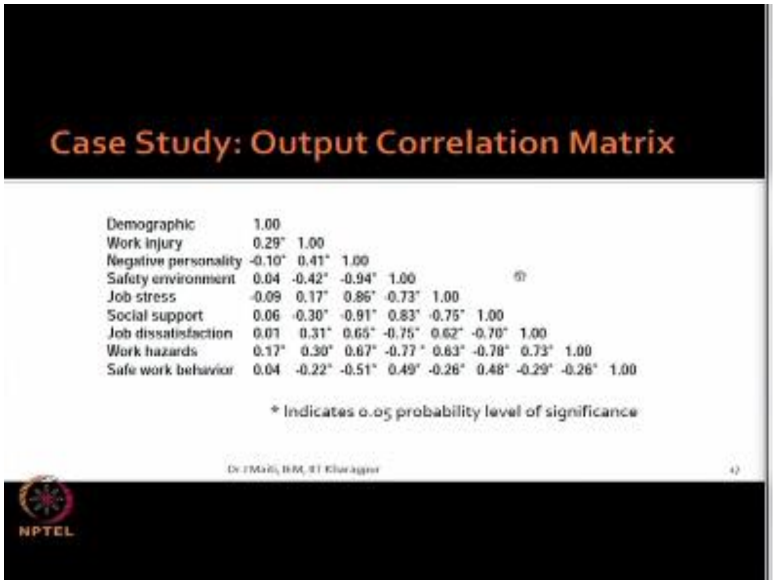
You see that demographic work injuries negative personality, these are the latent constants and this is your correlation matrix in then there are little star is there. This star indicates point 0.05 probability level of significance. I think all are significant here except this value 0.04, 0.01 some 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?

(Refer Slide Time: 43:35)



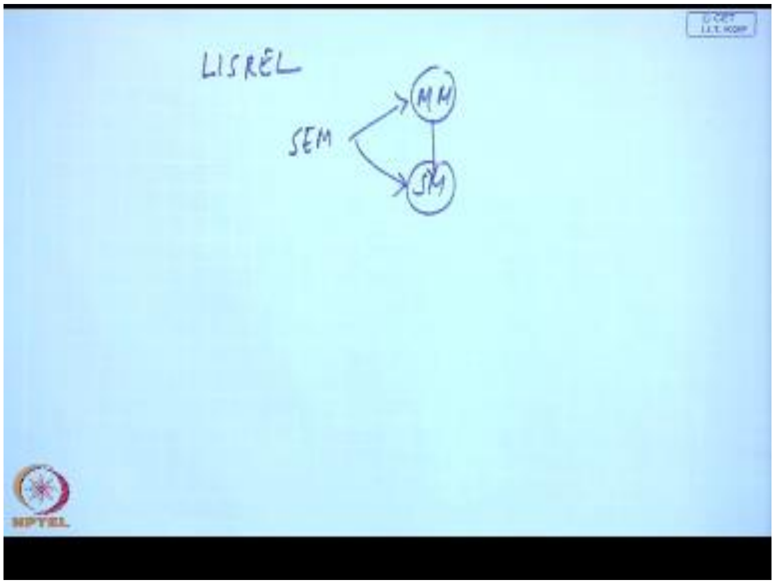
That is what we have done and we have done this with the help of this, this λ values and the Xi and the error term.

(Refer Slide Time: 43:41)



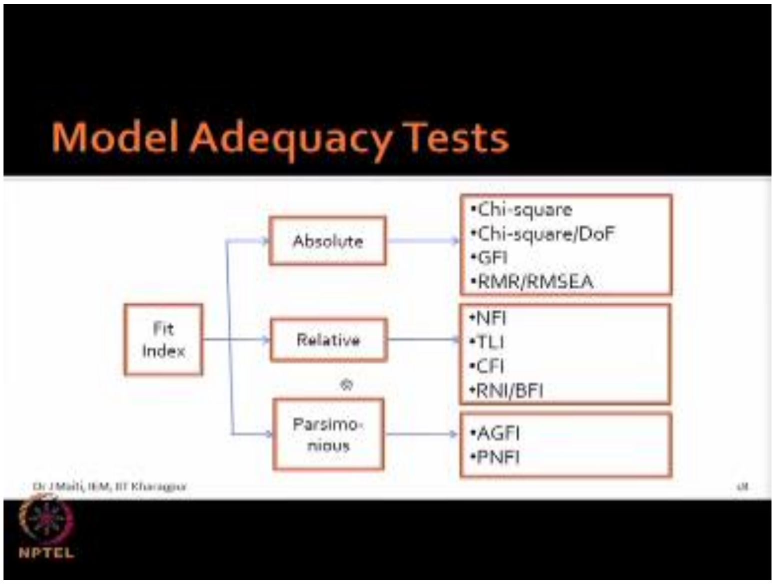
You are getting these values. And this will be this is a value, you want this is very, very important one because this will be used in structural model as input to structural model in structural equation modeling, I told you. That structural equation modeling two parts SEM has two parts.

(Refer Slide Time: 44:04)



Measurement model and structural model the output of this will be input to this, fine?

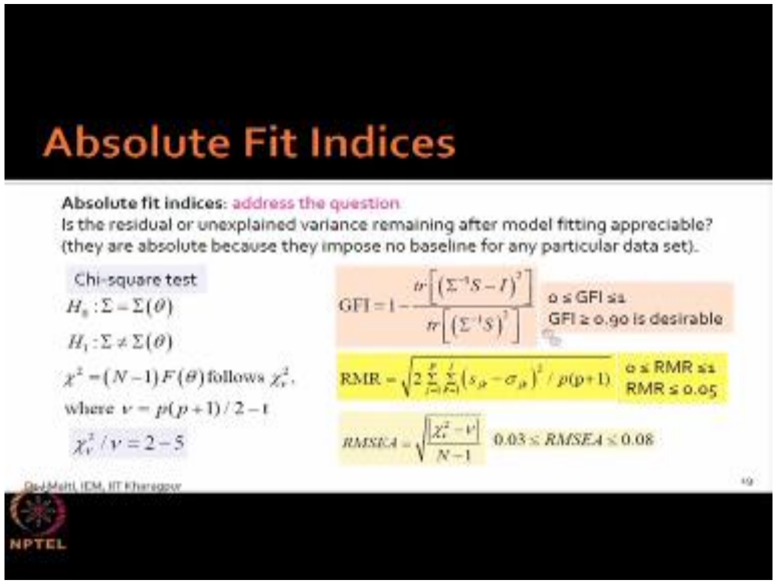
(Refer Slide Time: 44:20)



But I would consider. 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 hen 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 χ^2, 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 χ^2, χ^2/by degree of freedom, goodness of fit index, root mean square, residual root means 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 modeling 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: 45:50)



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 $\sigma = \sigma\theta$ 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 χ^2, which is n - 1 x $F\theta$, that $F\theta$ you have seen that the minimization function, so that value you have after estimation. So, χ^2 that n - 1 $F\theta$ this follows this χ^2distribution with new degrees of freedom, where new can be estimated like this, p into p + 1 / 2 - 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 χ^2value you get that should be as small as possible, because if for perfect fit $F\theta$ will be0, the n -1 x $F\theta$ that it will be 0. So, 0 is the ideal value, but it will all depends on sample size, also n - 1.Now, see that you will never get this your θ will be 0, because you are doing the numerical way of optimization, numerical optimization where some convergence value will be there.

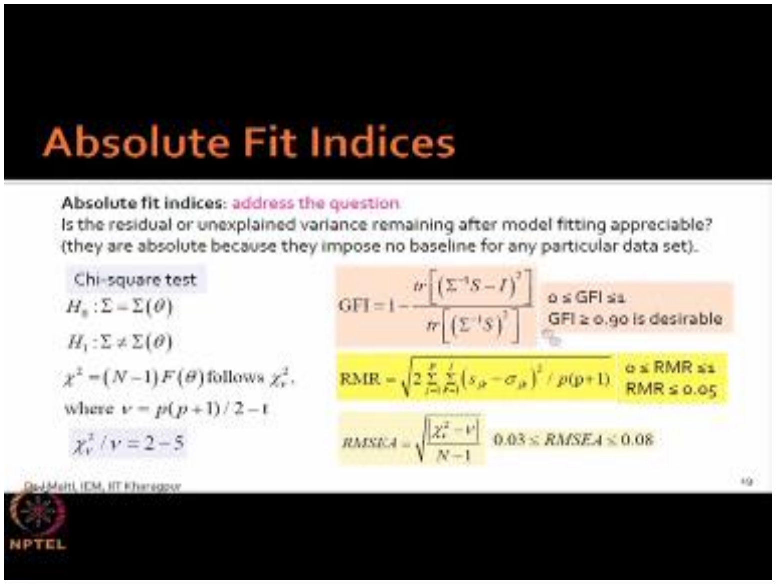
Now, if n is sufficiently large what will happen? This value will become large. So, what you want to how do then justify that whether the model is fit or not. One way is that this value should be as small as possible and other one is you go by χ^2 / degree of freedom. So, it is recommended in the later lecture that essentially the χ^2 distribution is such that the expected value of χ^2 nu is nu because that is the degree of freedom.

Because it is a parameter in χ^2 we use the degree of freedom only. So, actually the χ^2 by the degree of freedom should be 1, but is not recommended what is said that 225 is the recommended value above 35constraints. Now, if you use G F I that is goodness of fit index which is similar to R^2 square in multiple regression, you can remember or recollect that $R^{2\,=}$ 1 - SSE / SST.

Now, you see this formulation here, that way we have written here that 1 - trace of this by this. So, this is total variability and this one is the error term. It is similar to R^2 and this G F I value varies from0 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 than0.9 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.

(Refer Slide Time: 49:54)



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?

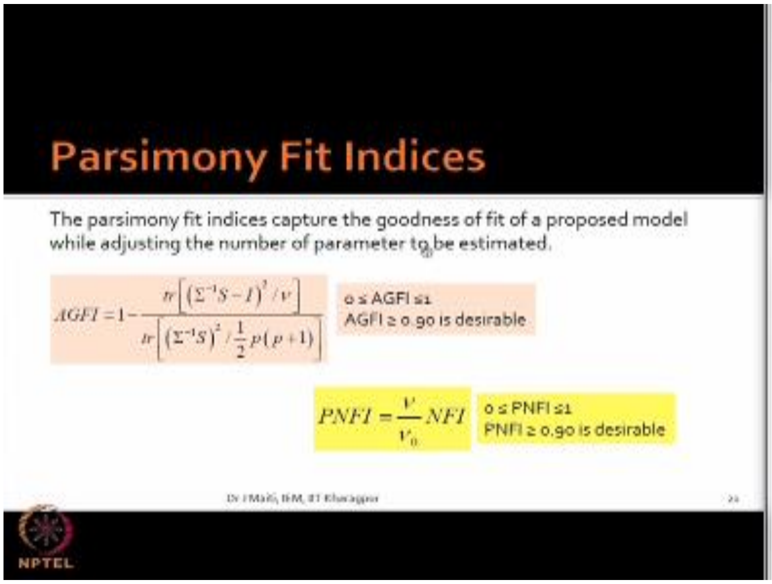
Now, if you take this and this what is the difference? Take this, this is the difference? So, we will take this, what is the difference? These differences are squared here you see what we have done S J K minus sigma J K this is the estimated one square by p into p + 1, that is the non-redundant part.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

It is seen that 0.03 to 0.08in this range this lays and this range also says substantial increment then relative fit index. 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 χ^2 for the null model is χ^2 0 and χ^2 for the proposed model is χ^2 nu, where nu is the degrees of freedom. Then you are in a position to develop or I can say quantify, this indices like NFI is χ^2 0 - χ^2 nu / χ^2 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 - c χ^2 nu - nu / χ^2 0 - nu 0.

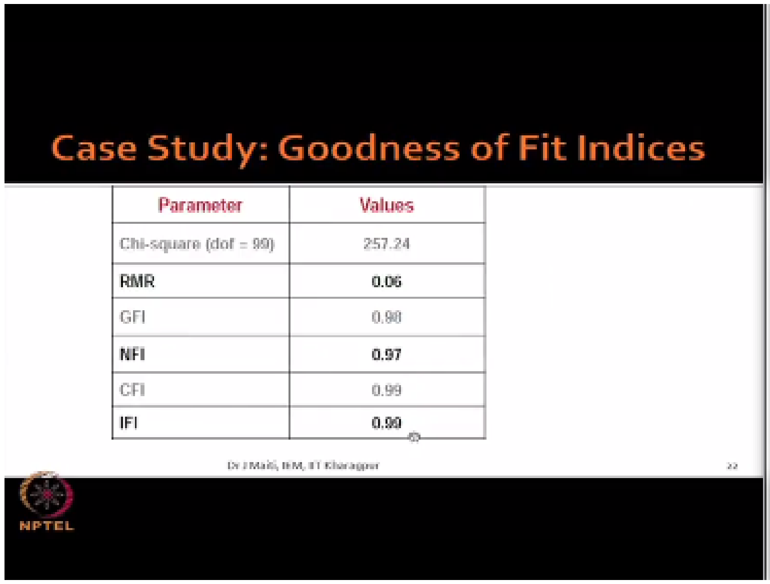
And TLI you have seen this, this also in the similar manner you see. Ultimately they take into consideration the chi square value of the proposed model and a worse pit model which is known as null model and the comparative indices are developed and higher the index value, it is better 0.9 or more is 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 for 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 than0.90 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: 54:29)



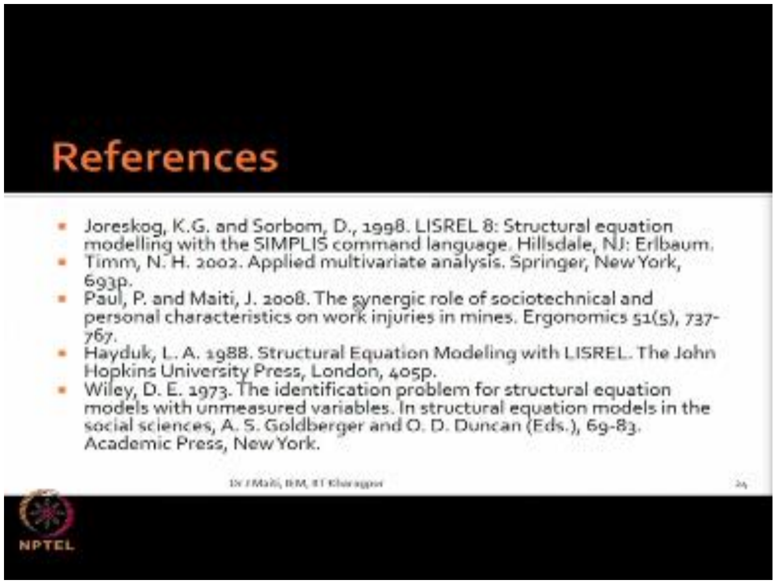
Now, let us see the goodness of fit indices for the case studies here, some of the fit indices I have given there are others. So, χ^2 degree of freedom of 99, χ^2 value is 257.24, if you divide by 99, this is almost100. So, it will be around χ^2 by degree of freedom is around 2.6. So, it is good because aim is to do 5 root mean square residual 0.06which is little more than 0.05, CFI is 0.98very good more than 0.90, NFI 0.97 more then0.90, CFI is 0.99 and IFI is 0.99.

So, essentially then χ^2 is 257.24 χ^2 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: 55:33)

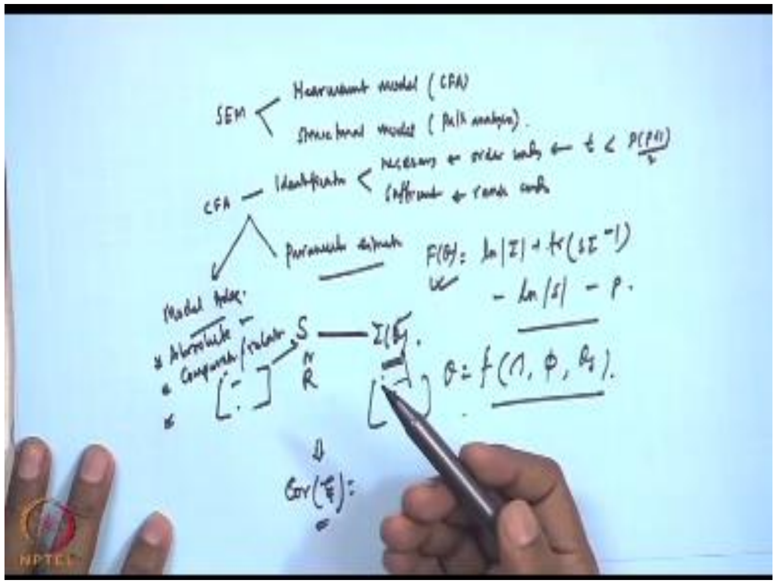
Now, see that who has basically worked in this, who are the pioneers that Karl Joreskog and Dag Sorbom, I think in around 1978 probably they have developed this software. First that listener came, I think in 1989 and its remarkable development in this field and we all are tremendously benefited.

(Refer Slide Time: 56:06)



What I can tell you further that for you have to understand the structural equation modeling. I might say that Joreskog, Sorbom this LISREL 8, structural equation medaling 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 modeling. So finally let me just summarize to my today is lecture.

(Refer Slide Time: 56:46)



We said that structural equation modeling has two parts, one is measurement model measurement model, and 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 σ plus trace of S σ inverse - log of determinant of S -p. This function is minimized through Newton Rawson 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 σθ. We try to match this two and using this function the better, the best match is considered and then you corresponding the θ value, these are used.

Now, θ is function of many things like λ, like your ψ like your θ δ. So these are θ, means θ means so many things are there. Any combinations that are 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, and 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 is 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 what are the model adequacy tests. 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^2 RA^2, 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.

NPTEL Video Recording Team

NPTEL Web Editing Team

Technical Superintendents

Computer Technicians

A IIT Kharagpur Production

www.nptel.iitm.ac.in