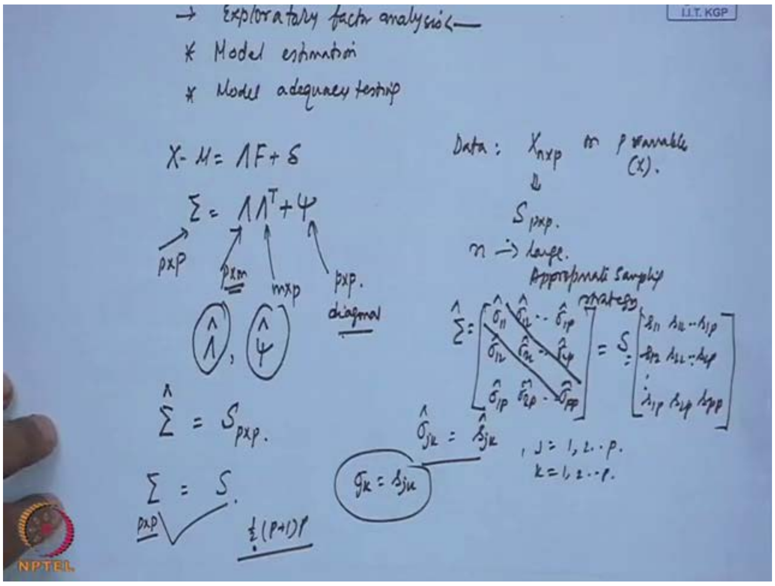
**Applied Multivariate Statistical modeling**
**Prof. J. Maiti**
**Department of Industrial Engineering and Management**
**Indian Institute of Technology, Kharagpur**

**Lecture - 34**
**Factor Analysis-Estimation and Model Adequacy Testing**

(Refer Slide Time: 00:25)



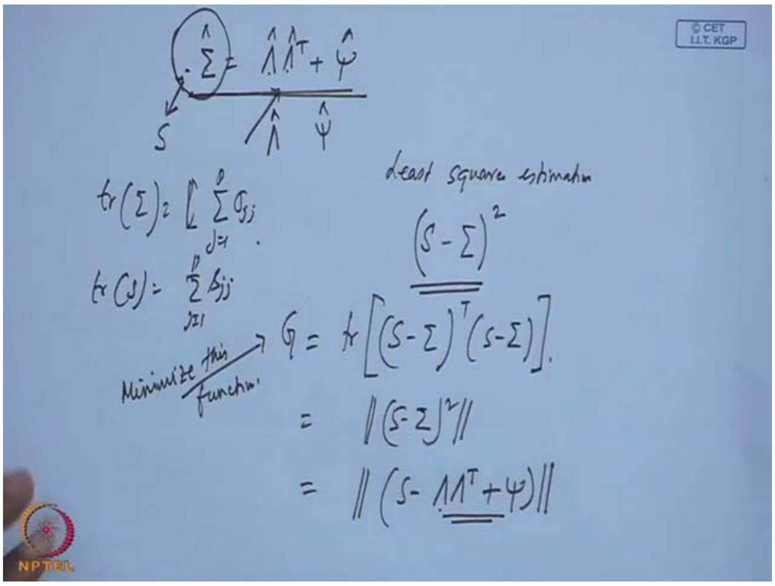
Good morning, we will continue factor analysis. Today our discussion will be on 2 issues; first one is model estimation, then followed by model adequacy testing. Both are related to exploratory factor analysis, both are related to these two. Now, what you have seen in last class that the exploratory factor model is this lambda F plus delta and the variant structure is like this, sigma equal to lambda lambda transpose plus psi. This 1 is p cross p matrix, this is p cross m, this one is m cross p, and this one is p cross p diagonal matrix. By estimation what do you mean to say, we want to estimate lambda, the estimate of lambda is lambda k.

We want to estimate psi cap, so we can say now what is known to us. We know that is capital sigma cap, this one we can assume as S, this is also p cross p matrix, then how do I go about for estimating this two, one is p cross m vector, and other one is p cross p vector, but where this specific variant term is diagonal. So, essentially, suppose you have collected a data n cross p matrix, this on p variables p x variables related to x and you compute it S also p cross p.

Now, if we think that the sample size is large and sampling strategy is appropriate strategy, then we can say that the sigma estimate, which is basically sigma 1 1, sigma 1 2 like sigma 1 p, sigma 1 2, sigma 2 2 sigma 2 p like this, sigma 1 p sigma 2 p sigma p p, these estimates or which is basically S where s each s 1 1 s 1 2 s 1 p s 1 2 s 2 2 s 2 p like this s 1 p s 2 p s p p. So, these are the estimates, which in general if I say sigma j k that estimate is s j k for j equal to 1 2 p k equal to 1 2 p. Now, we can assume that under root estimation that what will happen ultimately, if our sampling size is appropriate at recreating will have to capture properly, the behavior the covalent structure.

So, that sigma j k equal to s j k, that means what I mean to say here, I mean to say that each of the each of the element in capital sigma and this be very close very-very close. So, where p cross p matrix here and number of estimate estimating parameters will be here half into p into p plus 1 into p. These many parameters because if you see the structure here, you see these portion is mirror image of this, so ultimately it is not p cross p.

(Refer Slide Time: 06:14)




Essentially, what we estimating here or what you are assuming here, we are assuming that we want to estimate something that is sigma, which is lambda cap lambda transpose plus psi transpose, that is what you want to do. If using these estimates, we want to find out these values, estimate of these values. So, other way what you do, you also do not know this one or you know s only, so we initialize values for this lambda cap and psi cap. And then we find out this value, then we compare this value with s individual elements of the vector.

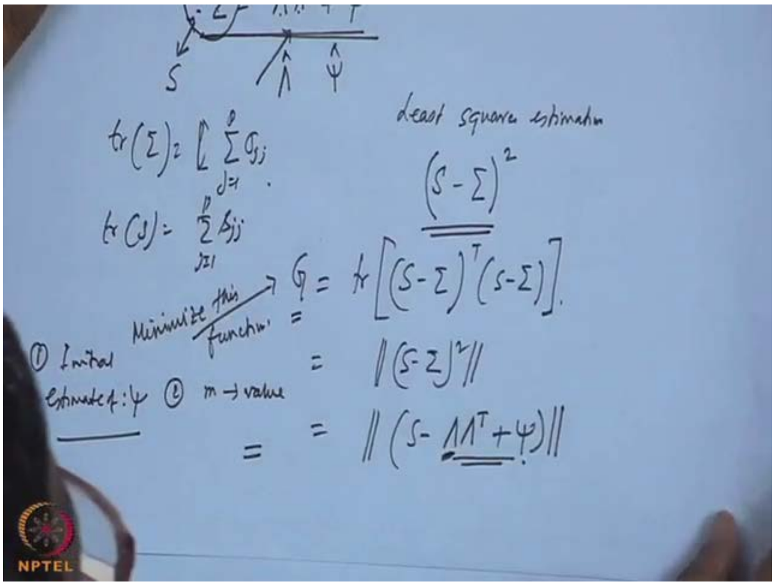
And if we find out a situation where the difference between the element of these two matrices will be as small as possible within certain threshold limit, but what is in factorial issues is, what you will find out that each variable variants, which is the issue. So, when we talk about trace of, this is nothing but the sum total of the sum total of j equal to 1 to p sum total of sigma j j that means, the sum total of variants of each other variables that means similarly, if I say trace of s that is also j equal to 1 to p s j j. So, instead of then what we require, we require function to be minimized.

Now, if you go for the least square estimates least square estimates least square estimation then what you will find out that we require minimizing the error, what is error here, the error is the difference between the two. So, our sample covariant matrix is this, population covariant matrix is this and we want the minimization function, which basically minimize the sum square errors that means sum square of the elements of this matrix.

Here we are interested in the variants components, we are considering here that sum square error like this and for this is in the matrix domain, so this will be like, and this will be this. And then we take basically we minimize a function let it be g, which is trace of this. So, essentially this leads to that second norm, this we can say instead of S if we write lambda transpose plus f psi plus psi, so this is what sigma is. We want to minimize this function, so what are the things you must know?

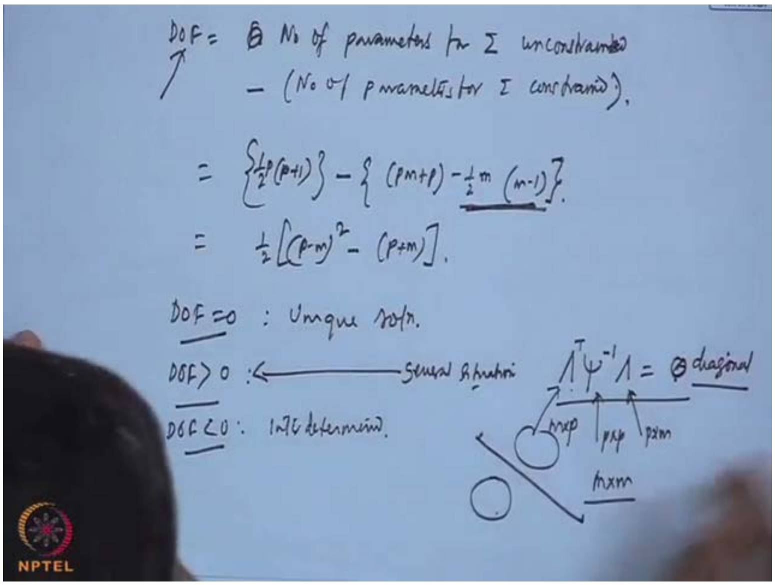
See if you want invent take the derivative that maximum like the derivative of this. And then with respect to the parameters put it into 0, then you get these values, this is the standard ways we use to do and we have the simple functions here. There are problem is that, you will find out that your capital lambda as well as psi, and these are the unknown parameters.

(Refer Slide Time: 11:32)




If I take this respect to lambda and what is required here, we require knowing particularly, we have to initialize the values. So, as a result what is required? We require an initial estimate of psi this is one, the second requirement is how many factors we are accepting in factors, what is this value of m that value also we require to know.

(Refer Slide Time: 11:34)

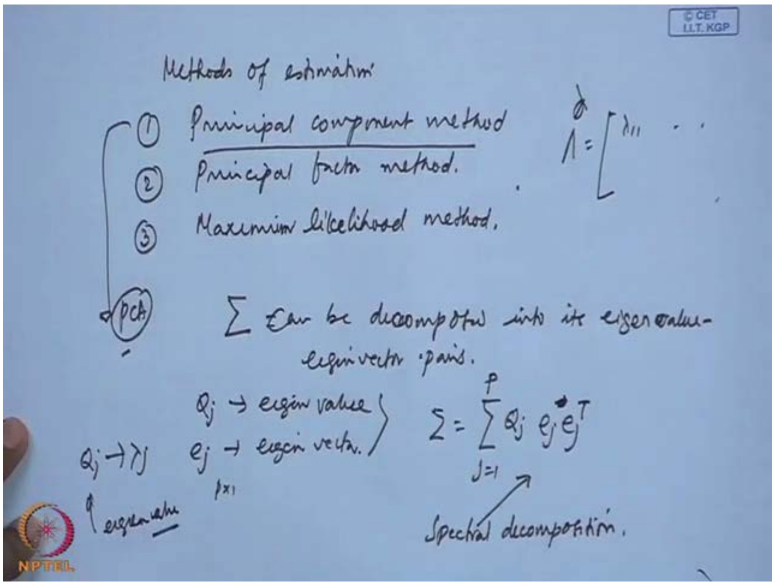



Essentially, if we look into the factor model number of parameters to be estimated as well as that, what is the degree of freedom available? Then what you will find out, we find out that degree of freedom is number of parameters for that capital sigma, this is unconstraint and minus number of parameters for sigma constraint. In the unconstraint case, the total number of parameter will be half p into p plus 1 and when you make a constraint case that is p m plus p minus half m into m minus 1, and this will lead to half of p minus m square p plus m. So, this is what the degrees of freedom available are.

Now, if your degree of freedom DOF is 0, you will get a unique solution. If DOF greater than 0, if you have several solutions and if DOF less than 0, it is not determinant. So, we can say the factor model is in determinant. The general situation is this is the general situation. So, what you want is, as we will be getting large number of solution because there is no unique solution in this case and is over estimation situation. So, you require putting certain constants that constraint will be lambda that psi inverse lambda, this transpose will become 0.

So, this will become diagonal now, these are constraints. This one is m cross p, this is p cross p, this one is p cross m, so resulting will be m cross m. So, ultimately the diagonal elements m elements will be there and of diagonal elements that will become 0. And as a result you will find out that, this is the number of parameters you are relaxing you are basically fixing to 0. And ultimately you are getting this that is what is we are saying about. So, you have to fix that m in such a manner that your degrees of freedom should not be this, as well as we do not want also this, we want this situation. Now with this background, I will now explain what are the methods available for estimations?

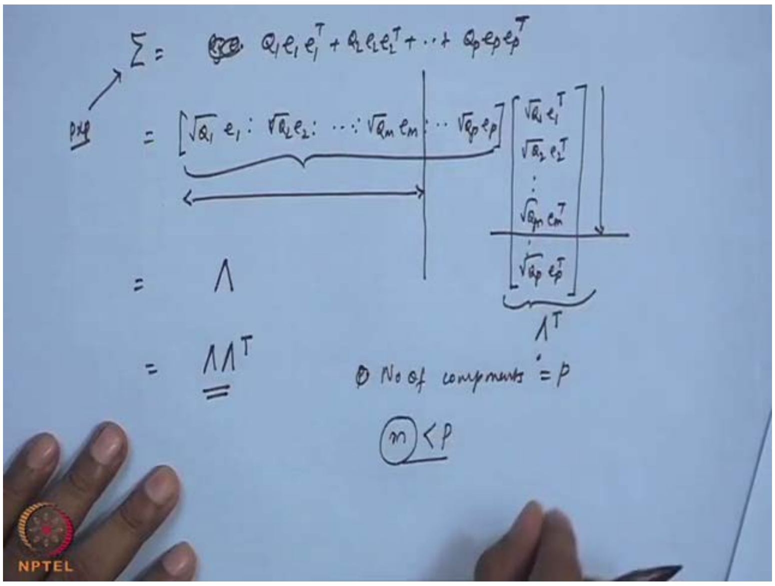
(Refer Slide Time: 15:45)



First one is principal component method, second one is principal factor method and third one is your maximum likelihood method. First we describe principal component method, it is similar that is basically what you have done in PCA. In PCA you consider the population mat covalence matrix and we say that this can be decomposed into its eigen value and eigen vectors pairs. Actually, if I say that Q j is a eigen value and e j is the eigen vector, then what you can write, you can write that sigma is Q j e j.

Suppose e j we are writing like this, this will be p cross 1 so e j e j transpose, then summation of j equal to 1 to p, this is known as spectral decomposition. In PCA when we discussed this we instead of Q j i have written this as lambda j, but in this case, we are using lambda that capital lambda, which is the component lambda 1, these are the factorial loadings, so I am using Q j for eigen values.

(Refer Slide Time: 18:43)

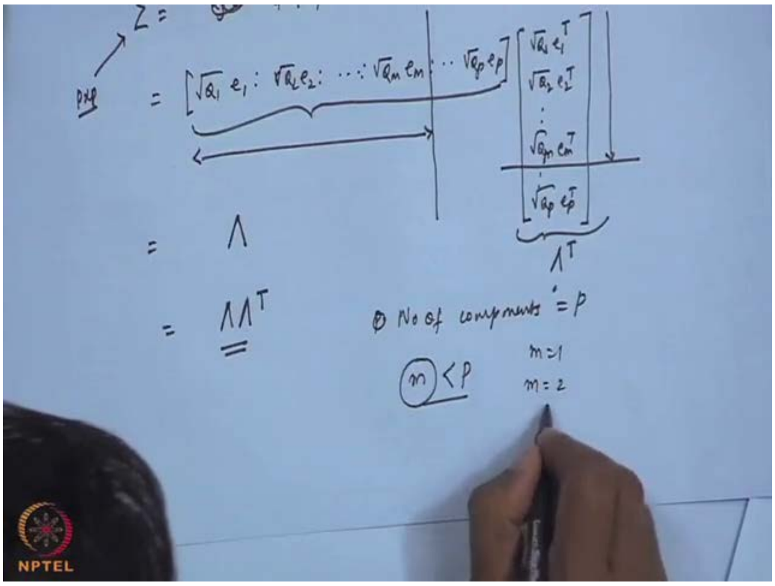


If this is the case then we cannot proceed like this that this is my Q 1 e 1 e 1 transpose plus Q 2 e 2 e 2 transpose like this Q p e p e p transpose, you can write like this one. In this manner similarly, this can be written Q 1 root e 1, Q 2 root e 2, Q m root e m then Q p root e p multiplied by root Q 1 e 1 transpose, root Q 2 e 2 transpose like root Q m e m transpose, root Q p e p transpose. So, if we denote this first portion as lambda then this will definitely lambda transpose, that means essentially we are writing sigma is lambda transpose.

And it is possible you can verify this, which is very much possible. And in this case what is happening here, so there are p so number of component extracted components extracted is p and your this matrix is also p cross p matrix so maximum possible component is accepted. And here differently, we are assuming that the rank of this matrix is differently p that is full rank.

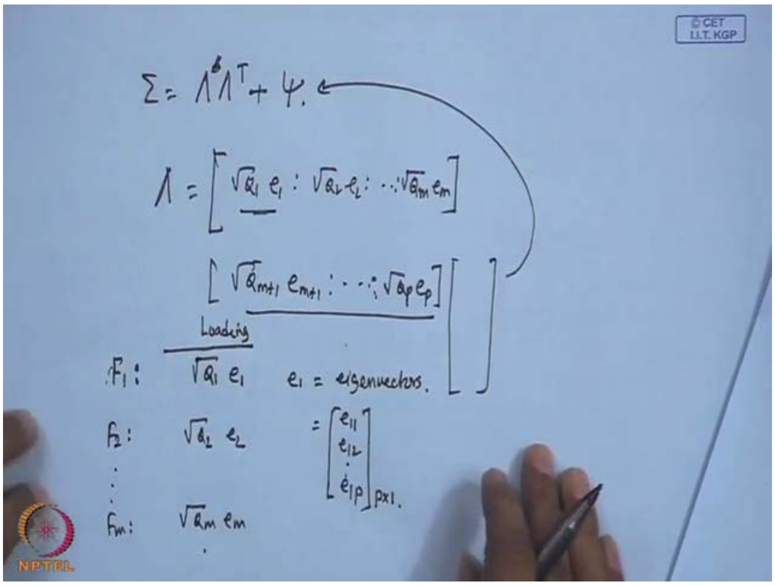
In factor analysis what is the purpose that there are many variables, which are correlated so in that case we want much less number of factors that is m, where m is less than p. So, in that case if I say that this is the domain of our m and similarly, this much is for m, then the remaining portion we can attribute to the error terms of this p variance because if we do not extract p factors, which we also do not want.

(Refer Slide Time: 22:12)




We want is subset of p that may be very small if m equal to 1, that is the base m equal to 3 or like this, but it must be the subset of this. So, we are now making little manipulation here.

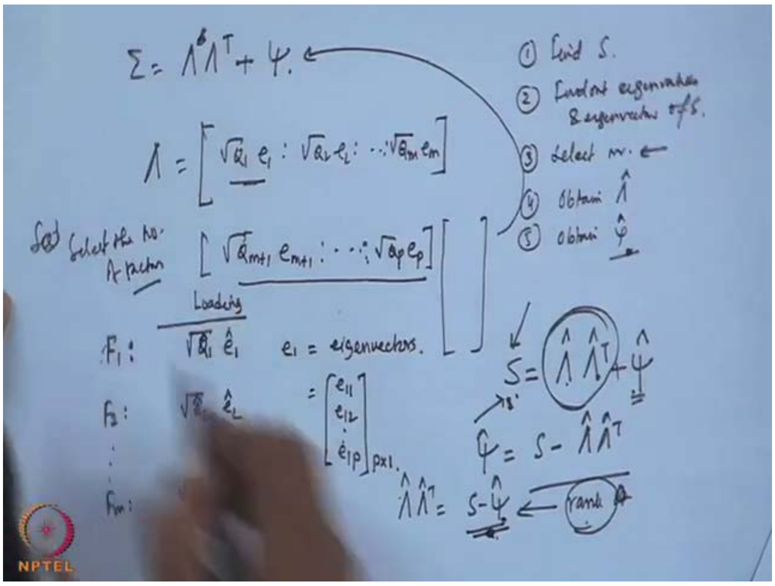
(Refer Slide Time: 22:13)




I want to write that sigma equal to this transpose plus psi where, this delta 1 is nothing but we are considering Q 1 e 1 root Q 2 e 2. So, ultimately root Q m e m the derived constraints. So, that means remaining portion, remaining portion means root over 1 Q m plus 1 e m plus 1 to your Q p and e p this portion multiplied by its transpose will lead to this psi. If this is the case then I can say that my first factor, if it is f 1 this factor is loading is root over Q 1 and e 1 where, e 1 is the eigen vectors, which is e 1 1, e 1 2. So, how many variables is there, p variables p cross 1.

So, for factor 2 it will be root Q 2 e 2 and I am talking about the loading part so similarly your factor m, this is Q m e m. Once you have this value, but how to get these values. These values will use S matrix that is a sample covalence matrix, which then will write this lambda cap lambda cap transpose plus psi cap.

(Refer Slide Time: 22:13)



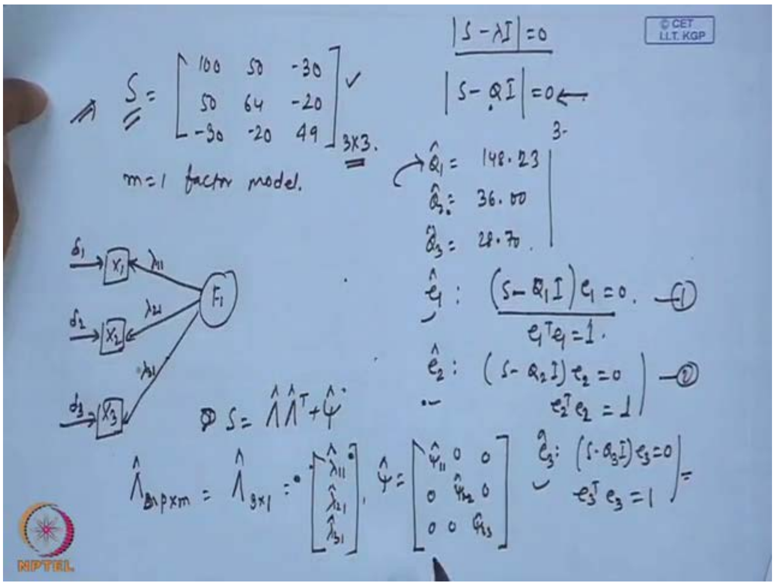

So, in this case your things will be like this, all will be cap, e 1 cap, e 2 cap, e m cap. So, usually we will estimate this lambda and that is troops vector decomposition of this S matrix, you will do the spectrum of s matrix and then how do you get the psi. You will be getting equal to S minus lambda cap lambda cap transpose, and then what is the other step. First is find S, second step find out eigen values and eigen vectors of s, then third one is you select m and then fourth one is once you selected then find out the loading, lambda cap, then fifth one is obtain psi cap.

So, how do you select m by accepting one, but there are certain suggestions so if you see this equation, you are interested to find out this value. So, this value is nothing but lambda cap, lambda cap transpose is nothing but s minus psi, what is the rank of this. What I mean to say find out the rank of this matrix, but you do not know psi, here the way we are giving you are basically finding out to this psi. So, if you go by this method initially you try first and find out this one, there are methods to first select the number of factors that will discuss.

But if you initialize as sigma zero cap then you find out this one, find out this rank or if we also you can find out what is this rank of this matrix. And then you can start with, but this principal component analysis once you get like this, then using the Eigen values, you can find out that how many factors you will keep, that you are seeing in PCA. There are different methods of identifying the number of principal component to be kept.

For example, that cumulative percentage for example, your average eigen values for example, the broken stick method for example, your screen plot for example, we have also gone for certain hypothesis statistics. So, those issues are under model adequacy test, I will explain to get more, but at present we will see that using s and simple component analysis for principal component method, how the factors can be retained.

(Refer Slide Time: 28:56)




Let us take one example here, suppose my s is this 150 minus 30 50 64 minus 20 minus 30 minus 20 49. This is 3 cross 3 matrixes, what I want? I want let m equal to 1 factor model. So what does it signify m equal to 1 factor model that means there is one factor let it be f 1, I have 3 variables x 1, x 2, x 3. I want to extract one factor out of these 3 variables and I want to know lambda 1 1, lambda 2 1, lambda 3 1 and also there is delta 1, delta 2 and delta 3.

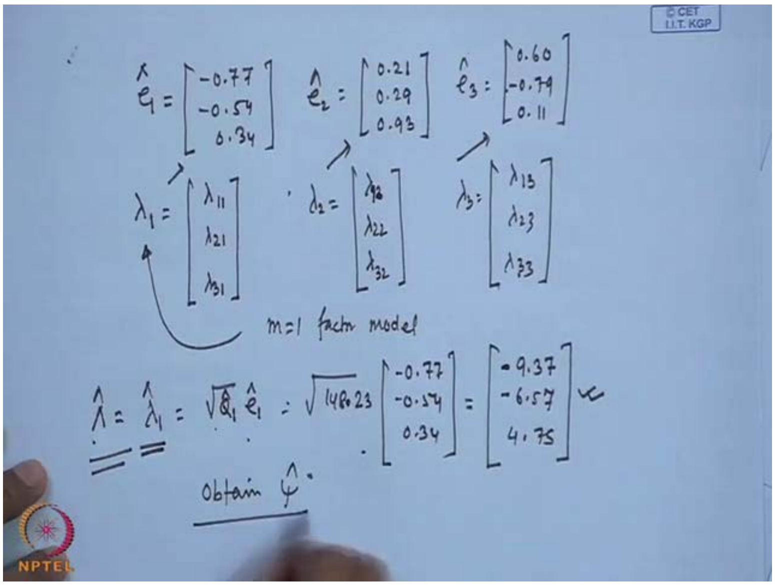
And essentially the factor model all of know this, if we use S this will be lambda, lambda transpose plus psi if we use cap, this is the situation. Now, here our lambda cap is 3 p cross m that means it is lambda curve cap 3 cross 1 and you will find out of S, this is lambda 1 1 estimate, lambda 2 1 estimate, lambda 3 1 estimate. So, we want this in addition we want psi cap, which will be your sum value as psi 1 1 cap, psi 2 2 cap, psi 3 3 cap, this will be 0 0 0 0, then 0 0.

We want this from this data using principal component analysis. So, I will not going for deriving that how that principal, components analysis is done, to find out this like eigen values and eigen vectors that will not describe here, straight way I am going to give you the results. It is 3 by 3 matrix computation it will take time. So, all of you know that determinant of s minus Q i that will be 0 in PCA analysis, I have written like this s minus lambda i equal to zero. Here I am using Q I, I told you earlier this equals to differentiate lambda as loading not the eigen values, so if you do like this, ultimately it is 3 by 3 matrix.

You will be getting it as equation here, which will be 3 or third order polynomial. So, your roots will be Q 1, Q 2 and Q 3 and these values are 148.23 36.00 28.70. Now, if you want to know e 1 cap what you require to do, you require to put Q 1 into this equation then make this equal to 0. So, first it is find S is given, second set is find eigen values, these are the eigen values then find out the eigen vectors.

So, you put this and subject to e 1 transpose e 1 this equal to 1 similarly, for e 2 second eigen vector you find out like this Q 2 i e 2 this equal to 0 subject to e 2 transpose e 2 equal to 1. Similarly, for e 3 cap will be finding out s minus Q 3 i e 3 equal to zero e 3 transpose e 3 this will be 1. If you solve this equation you will be getting this, solve this equation you will be getting this and you solve this equation and you will be getting this.

(Refer Slide Time: 34:28)

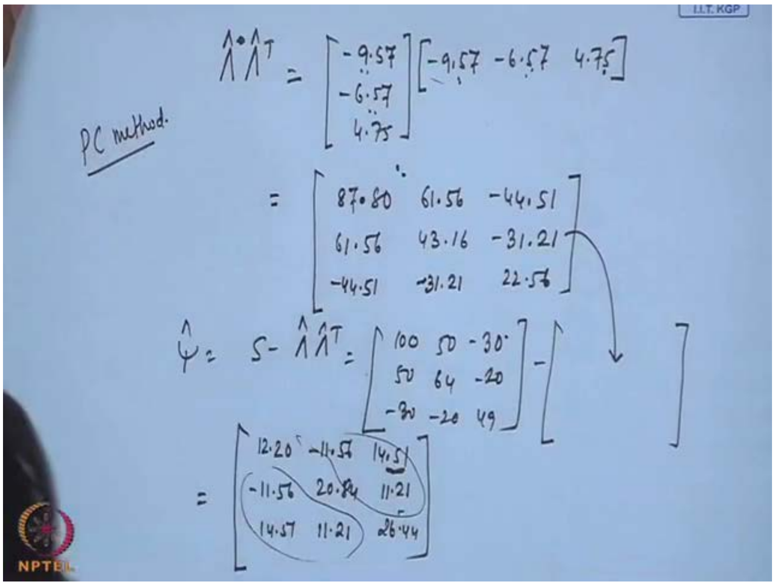



So, what are the values for your e 1, you have solved e 1 values are minus 0.77 minus 0.54 0.34. So, if you write like this your e 2 cap 0.21 0.29 0.93 and e 3 cap is 0.60 minus 0.79 0.11. So, what I want, I want lambda 1, which will be basically a vector lambda 1 1, lambda 2 1, lambda 3 1 and you have seen earlier also. Similarly, you can find out here lambda 2, which will be lambda 2 1, this lambda 1 2, lambda 2 2 then lambda 3 2. Using this lambda 3, lambda 1 3, lambda 2 3, lambda 3 3, our question is we want m equal to 1 factor model, so we are concentrating only on lambda 1 not the other 2.

These effects are basically giving to the psi and we are saying that the other one factor is subset, so then lambda 1 estimate, what you have said this is Q 1 e 1 estimate. So, what is our Q 1, Q 1 I think we found out 148.23 and what is our e 1 estimate, e 1 estimate is minus 0.77 minus 0.54 zero 0.34. So, if you multiply this two, what you will be getting, you will be getting minus 9.37 minus 6.57 4.75.

So, this is your lambda value, now I think if you go by the steps I said select m, m equal to 1 obtain sigma that is lambda cap then obtain lambda cap means you are basically your lambda cap is this one. So, you have obtained lambda cap now what you require to know, you have to find out or obtain psi cap. In order to obtain psi cap, you require knowing what lambda cap lambda cap transpose is...

(Refer Slide Time: 38:22)

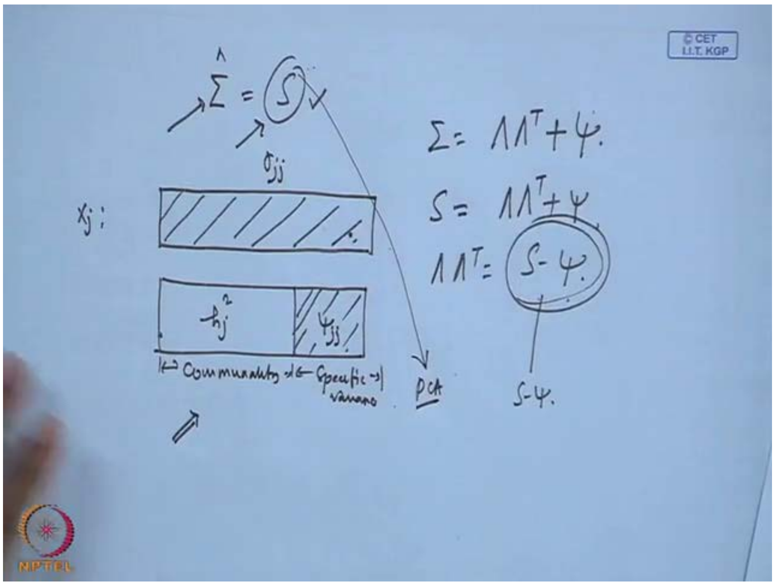



This value will be minus 9.57 minus 6.57 4.75 into its transpose minus 9.57 minus 6.57 4.75. Now, values are when you estimate this into this, this into this, this into this like this, so ultimately we will be getting 88 87.80 61.56 minus 44.511. How do you get this, 9.57 into 9.57 is this minus 9.57 minus this one, this minus this plus this, this minus this coming that is the way. So, I repeat again minus this into minus this is this minus this minus this, this minus this.

So, in the same manner this cross this, this cross this, this cross this, you will be getting 61.56 43.16 minus 31.21 then minus 44.51 minus 31.21 22.56. So, your psi cap is S minus lambda cap lambda cap transpose, which is 150 minus 30 50 64 minus 20 minus 30 minus 20 49 minus this matrix, minus this matrix, so ultimately you will be getting a matrix like this. So, obtained element we want 0, there is quite large values.

Now, how much large that to be tested, but by seeing this simple values we will not be able to judge it that is significantly large or not, it is difficult. So, this is what our principal component method is, PC method, second method is principal factor method. I think we have discussed somewhere a second method is principal factor method.

(Refer Slide Time: 42:11)

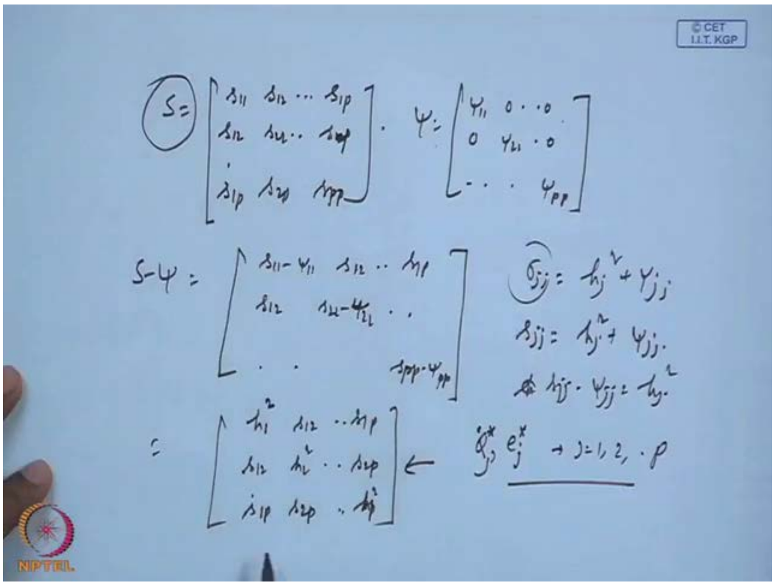



In the first method, the principal component method what you have done, we have basically taken this matrix or you can say that estimate of this matrix basically, when you are talking in sample domain, we are talking about this matrix. And our aim was to extract as minimum number of significant factors and what is that minimum value factors that we have not discussed leave I will show you. In principal component analysis we have seen that what are the important factors that component can be kept.

So, in principal factor that factor in case, what will we do? You take any variable x j, this variable variability is sigma j j and we have seen in last class that this can be partitioned into 2 half's, one is psi j j, which is unexplained or PCP variants and other one is this h j square. Now, this one is communality, this is your specific variance. Now, also you have seen that your sigma is lambda, lambda transpose plus psi or S, we can write s be plus psi, so that this, this, this into this lambda into lambda transpose is minus psi.

So, instead S matrix, if we now factorize S minus psi matrix that means, we are not in taking this one, we are basically concentrating the communality that is the common variants explained by the factors. So, whatever the method you adopted so far that mean we have seen that principal component. So, in principal component you first decomposed S in PCA based. Here what we are saying you decomposed S minus psi.

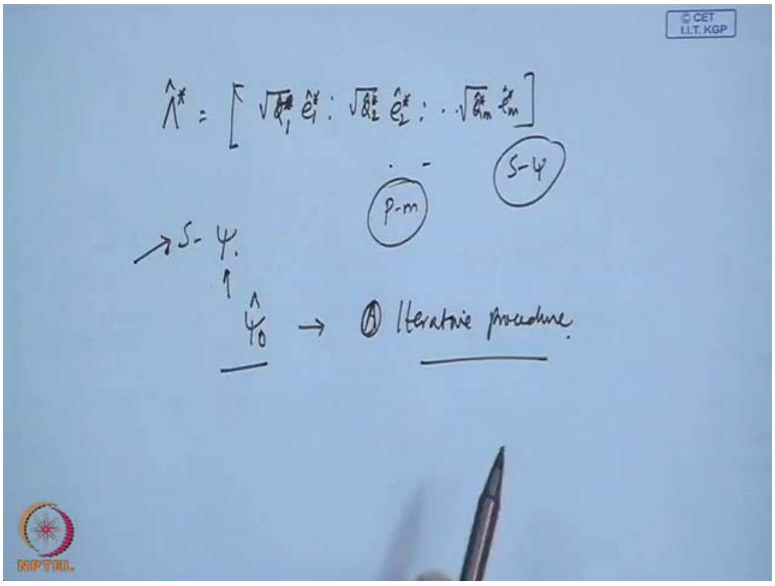
(Refer Slide Time: 45:11)




Now, if you do little more manipulation here, so our S is this s 1 1, s 1 2, s 1 p, s 2 p and if I say my psi is, so I create S minus psi means, the diagonal elements will be changed. Now, using this is happening here, we are writing that sigma j j equal to h j square plus psi j j. If you take the estimate s j j equal to h j square plus psi j j then s j j minus psi j j equal to h square. So, you are writing here h 1 square s 1 2 1 p s 1 2 s 2 square s 2 p 1 p 2 p s that is s p square.

Now, you can try so again if we can say suppose the eigen value and eigen vector, they are like this. Suppose Q star if you say Q star and e star or Q j star or e j star j equal to 1 to p, you will be getting exactly or the same thing what we have done in PCA using this, just now what you have developed that spectral decomposition of this, in the same manner you have to find out.

(Refer Slide Time: 47:38)

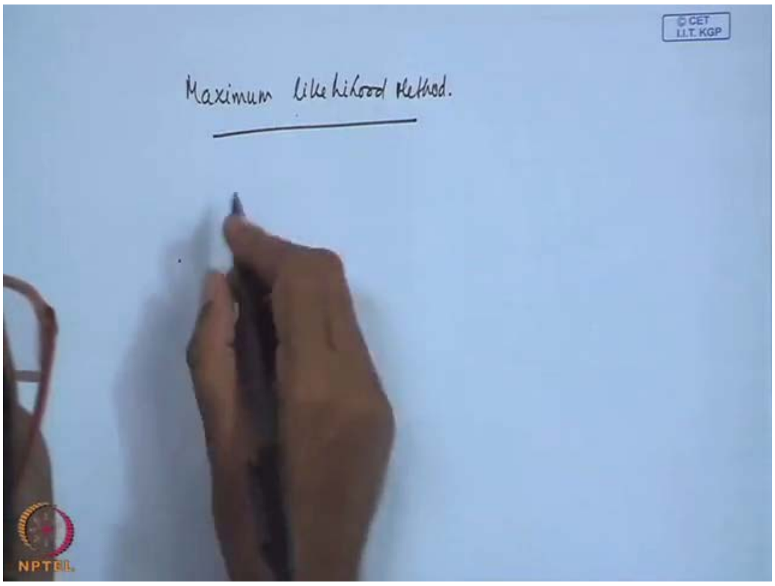


So, in this case if I say that lambda star estimate gradually using this star and this one, so this will be nothing but your root estimate Q star that one e 1 star estimate, Q 2 star estimate, e 2 star estimate, Q m star estimate. So, you can go up to p, but essentially what will happen that this is called a structure, this matrix will find out much less than p may be that angle be less than p that is 1 that you have to test it. Otherwise also that p minus m this factors, which we are not considering they will have much-much less that contribution in explaining the variability of the x variables.

So, this is the second method, but please keep in mind that in the second method case that you are factorizing S minus psi, you know the value of s, but you do not know the value of psi. So, you have to initialize this psi, so you have to find out this psi zero and initialization. One of the initialization I told you that delta means it will be iterative procedure, please keep in mind.

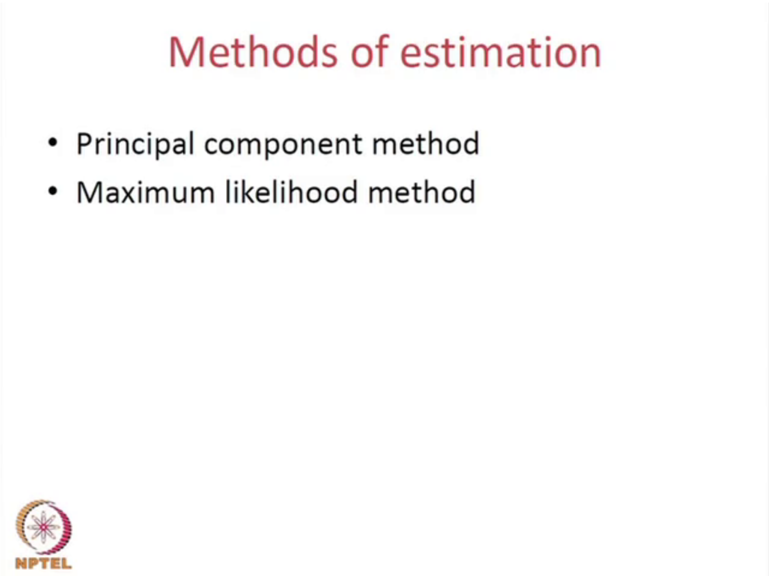
This one iterative procedure we initialize, then find estimate, then again you see that what is that the size coming, and then size that that matrix can be acceptable or not, what is the variance explained, what is not explained, then again you go on repeating these things and finally, some contrivance will take place. And then you will be able to accept that if at all that the model is to be accepted. So, I think the third method what you had said, third one will be the maximum likelihood maximum.

(Refer Slide Time: 50:15)




This is this little complicated one, what it is required that you have to find out the likelihood function, getting the likelihood function is not easy thing, but you have to find out this.

(Refer Slide Time: 50:54)




So, see what we have discussed here, we discussed here that maximum likelihood principal component method. We have already discussed that is related to principal component analysis and principal factor analysis, then there is maximum likelihood estimation.

(Refer Slide Time: 51:17)

An example (contd.)

- The correlation matrix

| | X1 | X2 | X3 | X4 | X5 | X6 |
|---|---|---|---|---|---|---|
| X1 | 1.00 | 0.44 | 0.41 | 0.29 | 0.33 | 0.25 |
| X2 | | 1.00 | 0.35 | 0.35 | 0.32 | 0.33 |
| X3 | | | 1.00 | 0.16 | 0.19 | 0.18 |
| X4 | | | | 1.00 | 0.60 | 0.47 |
| X5 | | | | | 1.00 | 0.46 |
| X6 | | | | | | 1.00 |

Here I want to tell you one example that is taken from Johnson and Wichen book. Last class we have discussed that 0.44 .35 .41, these are the basically the correlation coefficient values, which are more than 0.3. So, when more than 0.3 correlation coefficient is there, you can go for factor analysis that is the suggested one, but that number should be large in the sense, many of the correlation coefficient will be more than this.

(Refer Slide Time: 51:58)

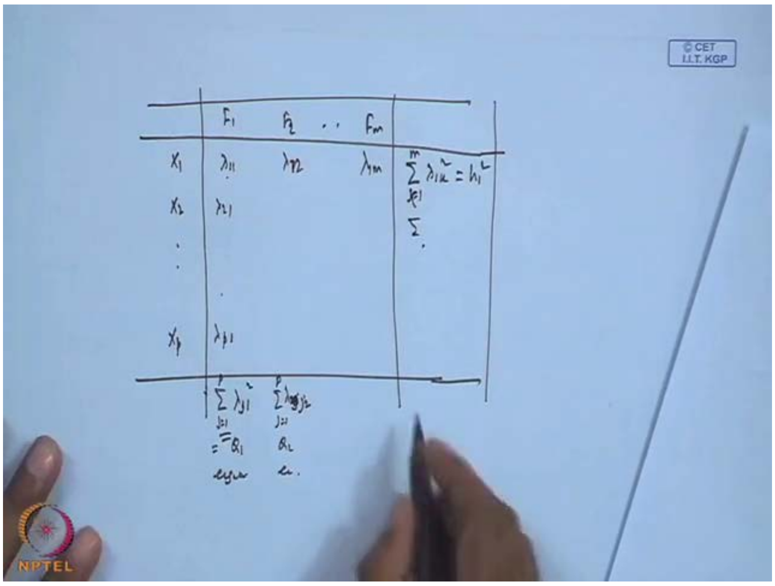
## Results

| Variables | Factor-1 | Factor-2 | Communa-lities |
|---|---|---|---|
| Gaelic | 0.553 | 0.429 | 0.49 |
| English | 0.568 | 0.288 | 0.41 |
| History | 0.392 | 0.450 | 0.36 |
| Arith-matic | 0.740 | -0.273 | 0.63 |
| Algebra | 0.724 | -0.211 | 0.59 |
| Geometry | 0.595 | -0.132 | 0.37 |
| % variance explained | 0.37 | 0.10 | |

What are these values in the columns?

Then using a principal component factor analysis for 2 factors, you will be getting something like this factor 1 and factor 2. And if you sum up the squares of these values across the rows for a particular column, you will be getting, you will be getting percentage that the lambda value or otherwise I can say that the Eigen values not lambda values. We are talking about Q values similarly, for this if you square, you will be getting the second Q value that Eigen value. And if you square this values and take some across basically you are for each row then you will be getting the communalities.

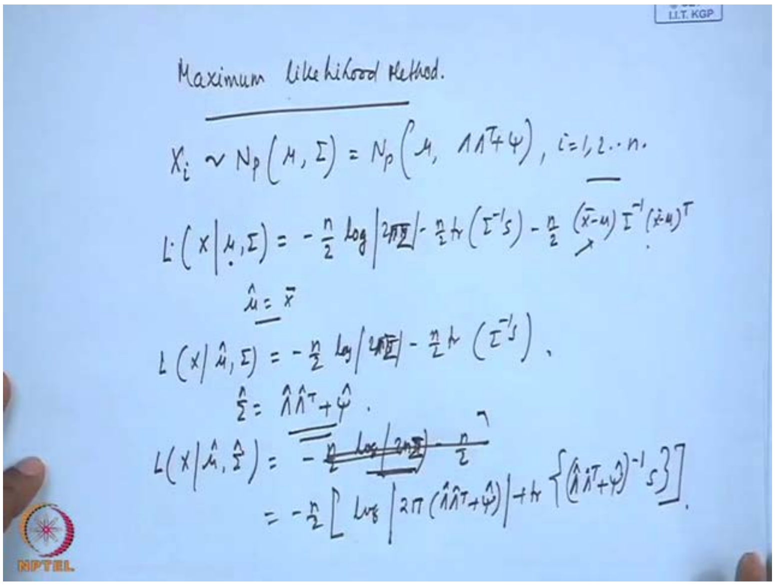
(Refer Slide Time: 53:00)



So, what I mean to say that when you are doing factor analysis and finding out suppose this is x 1 x 2 then x p, then your f 1, f 2 like this f m then you will be getting lambda values here. So lambda 1 1 lambda 2 1 like this lambda p 1, so then when you take that lambda j equal to 1 to p so 1 1 1 j 1 square, this is nothing but your q 1 what you have developed. Similarly, the second one j equal to 1 to p lambda j 2 that will be your q 2.

So, Eigen value here also Eigen value and similarly if you take sum here that lambda 1 1, this is lambda 1 2, this is lambda 1 m. If I say this is also j equal to let it be k k equal to 1 to m then your lambda 1 k square, this is what your h 1 square cumulative is. And similarly, like this, this is what estimation is?

(Refer Slide Time: 54:58)

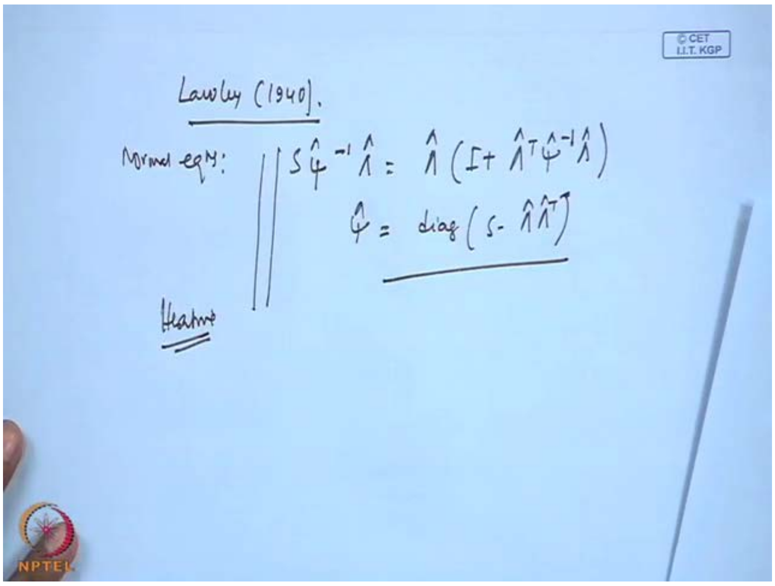



I think I will give you little inputs to the maximum likelihood estimation, but in maximum likelihood estimation, the assumption is x i is multi variant normal N p that mu and sigma. So, we can say that is N p mu sigma is this plus psi, that is the estimation and definitely it is for i equal to 1 2 n, you will be collecting n observations. So, the maximum likelihood that how we can find out the log likelihood, so you first find out the maximum likelihood then the log likelihood of x leaving mu and sigma.

This will be the simplified form like this log of log of 2 n pie minus n by 2 trace of this minus n by 2 x bar minus mu, this sigma inverse x bar minus mu transpose. Now what will happen when you estimate mu cap is as x bar, if you put here this will become zero, so then your this likelihood mu kept sigma, this will be minus log 2 n pie n by 2 trace this sigma inverse this. Now, if we say that sigma cap is lambda cap, lambda cap transpose plus psi cap and then this function will become this cap equal to minus n by 2 log 2 n pie minus n by 2.

So, 2 n pie we have written here log 2 pie sigma I have written wrongly here that is log 2 pie sigma, then log 2 pie sigma, then log 2 pie sigma this was changed. So I will change this one to like this minus n by 2 log of that determinant of 2 pie delta cap psi cap plus trace of if you are using this we are writing. Now this big commutation is there and Lawley has given some formulation for this which, I will write further the simplified normal equation by Lawley.

(Refer Slide Time: 58:16)

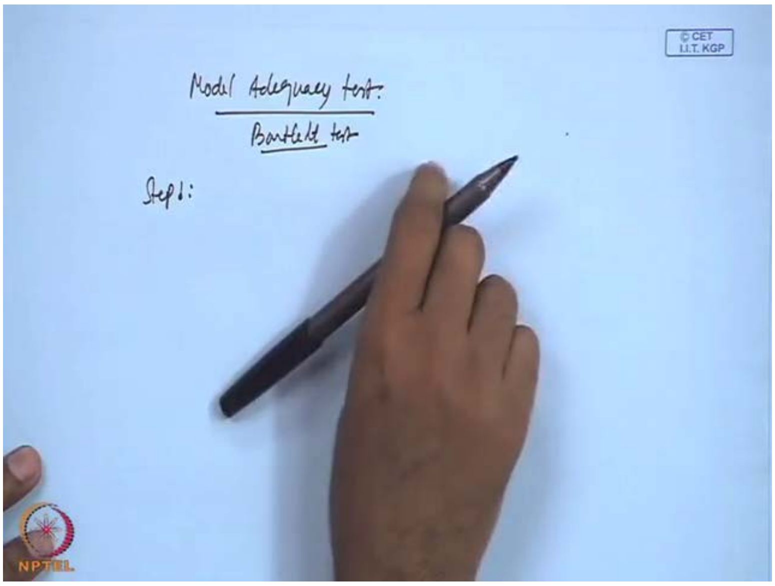



Simplified Normal equation given by Lawley 1940, he says giving some normal equation for solving S psi inverse lambda cap equal to lambda cap I plus lambda cap transpose psi inverse lambda cap. And definitely psi is diagonal matrix of this is the issue. Now, you have to find out some how to solve it because this is not a closed form solution, you require that procedure and then you will be able to solve this problem. So, these are the different estimations and as I told you that these instinct estimations principal components analysis method is easy to understand, because you are just decomposing the matrix using spectral decomposition.

In case of co principal factor method that psi value if it is not known then you cannot go for that method, but it should be initialized. The initialized point is that if you can take the diagonal element; assume that also possible take the value element of s initial value than you interpret this process. And in maximum likelihood method that is told this is preferable one because there are lot of statistical tests, which are possible, but it is the derivation part is little difficult.

As in applied engineer or a scientist, what you have to require knowing? You require knowing the basics usual formulas and how to use all those formulas, when data is given to you. So, what I have tried to give certain basics, but not the way the statistician will give you because that is not my cup of tea also. And final one I think that another one minute of time what I want to complete that you have to do certain tests, that is what I can say model adequacy test. I will explain to you that Bartlett test here.

(Refer Slide Time: 01:01:02)




Here the step 1, step 2, step 3 different steps are there, but the time is up now. What I will do next class, I will start explaining the hypothesis test that Bartlett adequate test and then will go for factor rotation, then factor scores then followed by one simple case I will show you, so that whatever the theoretical portion I have discussed here, it will be easier for you to grabs.

Thank you very much.