



these two. Now, what you have seen in last class that the exploratory factor model is this  $\lambda F + \delta$  and the variant structure is like this,  $\sigma = \lambda \lambda^T + \phi$ . This  $\lambda$  is  $p \times p$  matrix, this is  $p \times m$ , this one is  $m \times p$ , and this one is  $p \times p$  diagonal matrix.

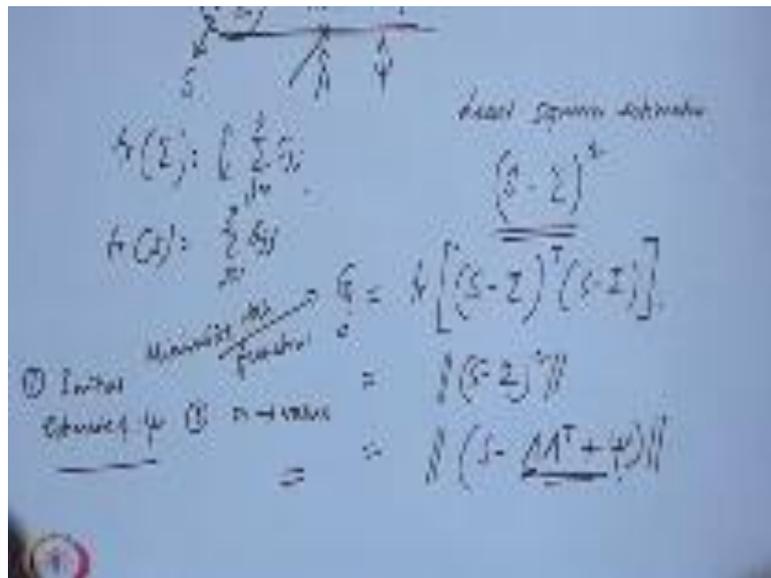
By estimation what do you mean to say, we want to estimate  $\lambda$ , the estimate of  $\lambda$  is  $\lambda_k$ . We want to estimate  $\phi$  cap, so we can say now what is known to us. We know that is capital  $\sigma$  cap, this one we can assume as  $S$ , this is also  $p \times p$  matrix, then how do I go about for estimating this two, one is  $p \times m$  vector, and other one is  $p \times p$  vector, but where this specific variant term is diagonal.

So, essentially, suppose you have collected a data  $n \times p$  matrix, this on  $p$  variables  $p \times p$  variables related to  $x$  and you compute it  $S$  also  $p \times p$ . 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  $\sigma_{11}, \sigma_{12}$  like  $\sigma_{1p}, \sigma_{22}, \sigma_{2p}$  like this,  $\sigma_{1p}, \sigma_{2p}, \sigma_{pp}$ , these estimates or which is basically  $S$  where  $s$  each  $s_{11}, s_{12}, s_{1p}, s_{22}, s_{2p}$  like this  $s_{1p}, s_{2p}, s_{pp}$ .

So, these are the estimates, which in general if I say  $\sigma_{jk}$  that estimate is  $s_{jk}$  for  $j = 1, 2, \dots, p$   $k = 1, 2, \dots, p$ . 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 behaviour the covalent structure. So, that  $\sigma_{jk} = s_{jk}$ , 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  $p \times p$  matrix here and number of estimate estimating parameters will be here half into  $p$  into  $p + 1$  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  $p \times p$ . Essentially, what we estimating here or what you are assuming here, we are assuming that we want to estimate something.

(Refer Slide Time: 06:13)



That is  $\sigma$ , which is  $\lambda \text{ cap } \lambda^T + \phi^T$ , 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  $\lambda \text{ cap}$  and  $\phi \text{ cap}$ . 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  $j = 1$  to  $p$  sum total of  $\sigma_j$  that means, the sum total of variants of each other variables that means similarly, if I say trace of  $s$  that is also  $j = 1$  to  $p$   $s_j$ .

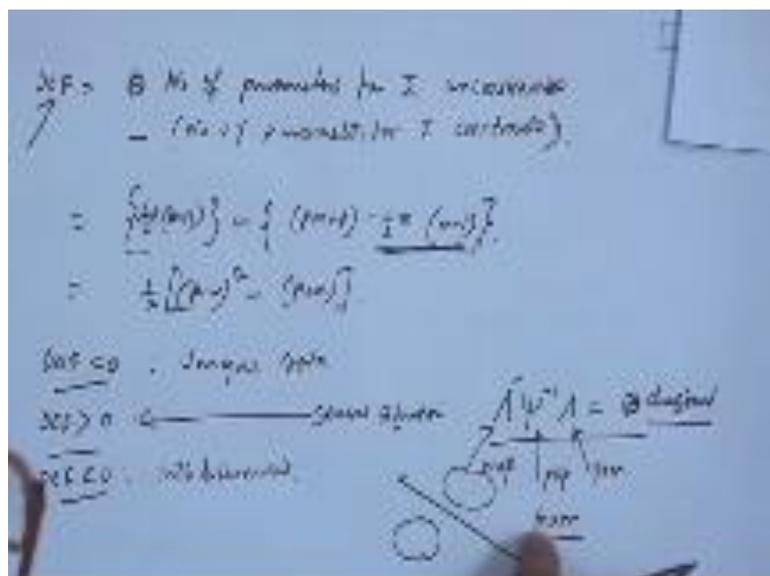
So, instead of then what we require, we require function to be minimized. Now, if you go for the least square estimates least square estimates least square estimation then what you will find out that we require minimizing the error, what is error here, the error is the difference between the two. So, 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  $\lambda^T + f \varphi + \varphi$ , 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  $\varphi$ , and these are the unknown parameters. 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  $\varphi$  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. 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.

(Refer Slide Time: 11:33)



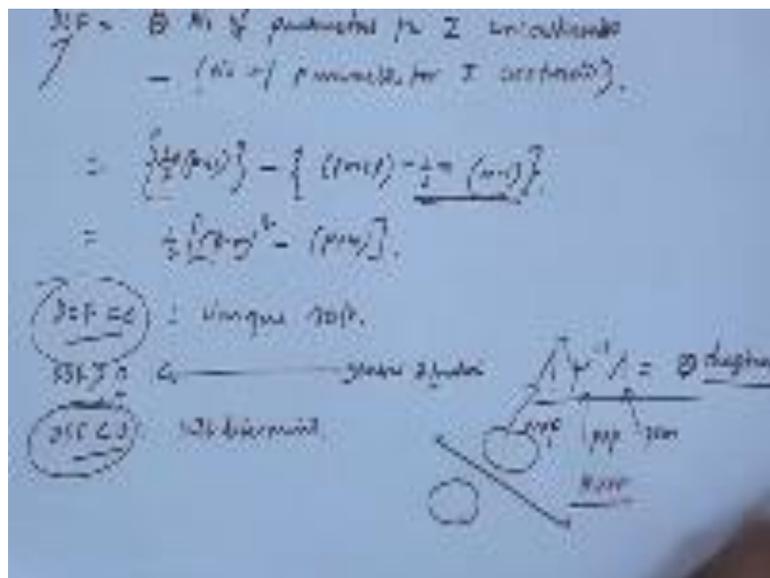
Number of parameters for that capital  $\sigma$ , this is unconstrained and - number of parameters for  $\sigma$  constrained. In the unconstrained case, the total number of parameter will be half  $p$  into  $p + 1$  and

when you make a constraint case that is  $p \times m + p - \text{half } m$  into  $m - 1$ , and this will lead to half of  $p - m$  square  $p + m$ . 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 okay. 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  $\phi$  inverse  $\lambda$ , this  $T$  will become 0.

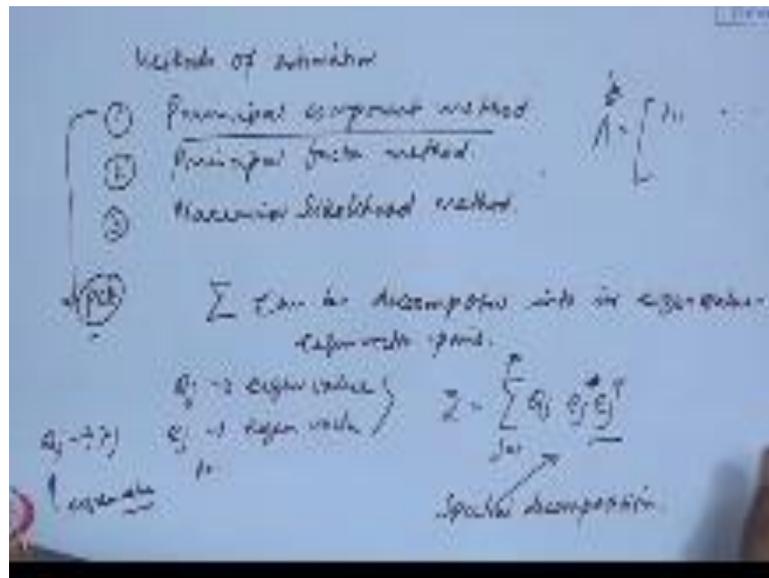
So, this will become diagonal now, these are constraints. This one is  $m \times p$ , this is  $p \times p$ , this one is  $p \times m$ , so resulting will be  $m \times m$ . So, ultimately the diagonal elements  $m$  elements will be there and of diagonal elements that will become 0. And as a result 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.

(Refer Slide Time: 15:15)



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.

(Refer Slide Time: 15:36)



Now with this background, I will now explain what are the methods available for estimations? 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 eigenvectors pairs okay.

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  $p \times 1$  so  $e_j e_j^T$ , then  $\sum$  of  $j = 1$  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: 18:41)

$$S = \sum_{i=1}^p \lambda_i e_i e_i^T = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \dots + \lambda_p e_p e_p^T$$

$$= \begin{bmatrix} \sqrt{\lambda_1} e_{11} & \sqrt{\lambda_2} e_{21} & \dots & \sqrt{\lambda_p} e_{p1} \\ \sqrt{\lambda_1} e_{12} & \sqrt{\lambda_2} e_{22} & \dots & \sqrt{\lambda_p} e_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\lambda_1} e_{1p} & \sqrt{\lambda_2} e_{2p} & \dots & \sqrt{\lambda_p} e_{pp} \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} e_{11} \\ \sqrt{\lambda_1} e_{12} \\ \vdots \\ \sqrt{\lambda_1} e_{1p} \\ \sqrt{\lambda_2} e_{21} \\ \sqrt{\lambda_2} e_{22} \\ \vdots \\ \sqrt{\lambda_2} e_{2p} \\ \vdots \\ \sqrt{\lambda_p} e_{p1} \\ \sqrt{\lambda_p} e_{p2} \\ \vdots \\ \sqrt{\lambda_p} e_{pp} \end{bmatrix}$$

$$= \Lambda \Lambda^T$$

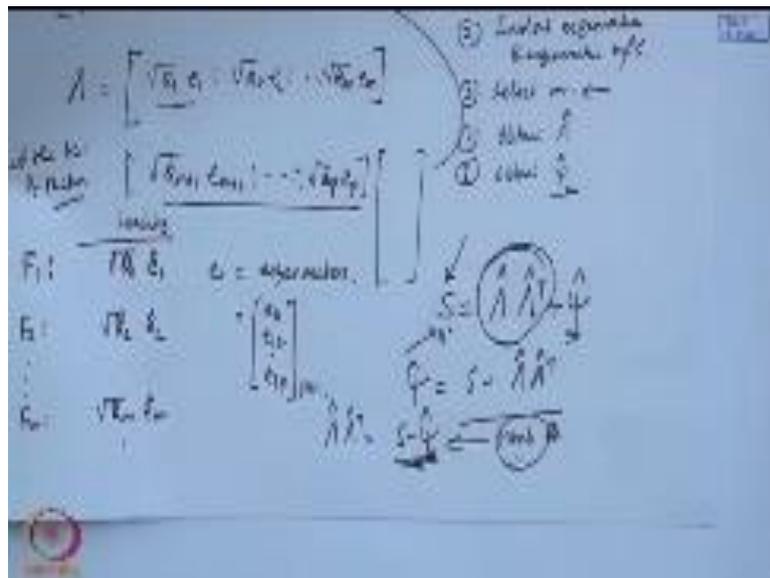
$$= \Lambda \Lambda^T$$
 No. of non-zero components = p

If this is the case then we cannot proceed like this that this is my  $Q_1 e_1 e_1^T + Q_2 e_2 e_2^T$  like this  $Q_p e_p e_p^T$ , you can write like this one. In this manner similarly, this can be written  $Q_1 \text{root } e_1, Q_2 \text{root } e_2, Q_m \text{root } e_m$  then  $Q_p \text{root } e_p$  multiplied by  $\text{root } Q_1 e_1^T, \text{root } Q_2 e_2^T$  like  $\text{root } Q_m e_m^T, \text{root } Q_p e_p^T$ .

So, if we denote this first portion as  $\lambda$  then this will definitely  $\lambda^T$ , that means essentially we are writing  $\sigma$  is  $\lambda^T$ . 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  $p \times p$  matrix so maximum possible component is accepted.

And here differently, we are assuming that the rank of this matrix is differently 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. We want is subset of  $p$  that may be very small if  $m = 1$ , that is the base  $m = 3$  or like this, but it must be the subset of this. So, we are now making little manipulation here.

(Refer Slide Time: 22:12)



I want to write that  $\sigma = \sqrt{\lambda} + \phi$  where, this  $\lambda$  is nothing but we are considering  $Q_1 \times e_1$  root  $Q_2 \times e_2$ . So, ultimately root  $Q_m \times e_m$  the derived constraints. So, that means remaining portion, remaining portion means root over  $1 \times Q_{m+1} \times e_{m+1}$  to your  $Q_p$  and  $e_p$  this portion multiplied by its  $\sqrt{\lambda}$  will lead to this  $\phi$ . 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  $Q_1 \times e_1$  where,  $e_1$  is the Eigen vectors, which is  $e_{11}, e_{12}$ . So, how many variables is there,  $p$  variables  $p \times 1$ .

So, for factor 2 it will be root  $Q_2 \times e_2$  and I am talking about the loading part so similarly your factor  $m$ , this is  $Q_m \times e_m$ . Once you have this value, but how to get these values. These values will use  $S$  matrix that is a sample covariance matrix, which then will write this  $\lambda = \sigma^2 + \phi^2$ . So, in this case your things will be like this, all will be  $\sigma^2, e_1 \sigma^2, e_2 \sigma^2, \dots, e_m \sigma^2$ . So, usually we will estimate this  $\lambda$  and that is through vector decomposition of this  $S$  matrix, you will do the spectrum of  $s$  matrix and then how do you get the  $\phi$ . You will be getting  $\sigma^2 = S - \lambda \text{cap} \lambda \text{cap}^T$ , and then what is the other step.

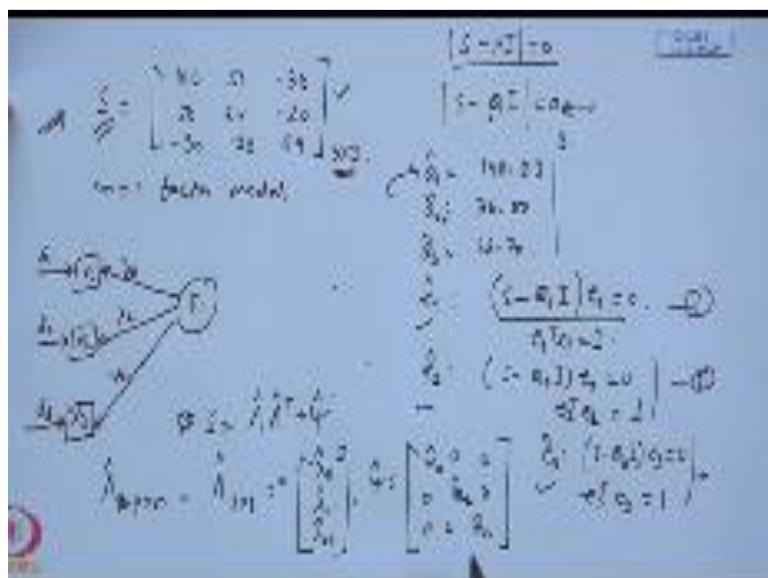
First is find  $S$ , second step find out Eigen values and Eigen vectors of  $s$ , then third one is you select  $m$  and then fourth one is once you selected then find out the loading,  $\lambda \text{cap}$ , then fifth one is obtain  $\phi \text{cap}$ . So, how do you select  $m$  by accepting one, 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  $\lambda \text{cap}$ ,  $\lambda \text{cap}^T$  is nothing but  $s - \phi$ , what is the rank of this. What I mean to say

find out the rank of this matrix, but you do not know  $\phi$ , here the way we are giving you are basically finding out to this  $\phi$ . 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  $\sigma^0$  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 sis this 150 - 30 50 64 - 20 -30 - 20 49. This is  $3 \times 3$  matrices, what I want? I want let  $m = 1$  factor model. So what does it signify  $m = 1$  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  $\lambda_{11}, \lambda_{21}, \lambda_{31}$  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  $\lambda$ ,  $\lambda^T + \phi$  if we use cap, this is the situation.

Now, here our  $\lambda$  cap is  $3 \times m$  that means it is  $\lambda$  curve cap  $3 \times 1$  and you will find out of  $S$ , this is  $\lambda_{11}$  estimate,  $\lambda_{21}$  estimate,  $\lambda_{31}$  estimate. So, we want this in addition we want  $\phi$  cap, which will be your sum value as  $\phi_{11}$  cap,  $\phi_{22}$  cap,  $\phi_{33}$  cap, this will be  $0 \ 0 \ 0$ , 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  $3 \times 3$  matrix computation it will take time. So, all of you know that determinant of  $s - Q$  that will be  $0$  in PCA analysis, I have written like this  $s - \lambda_i = 0$ . 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  $3 \times 3$  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  $= 0$ .

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  $e_1^T e_1 = 1$  similarly, for  $e_2$  second Eigen vector you find out like this  $Q_2^T e_2 = 0$  subject to  $e_2^T e_2 = 1$ . Similarly, for  $e_3$  cap will be finding out  $s - Q_3^T e_3 = 0$   $e_3^T e_3 = 1$ . 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.

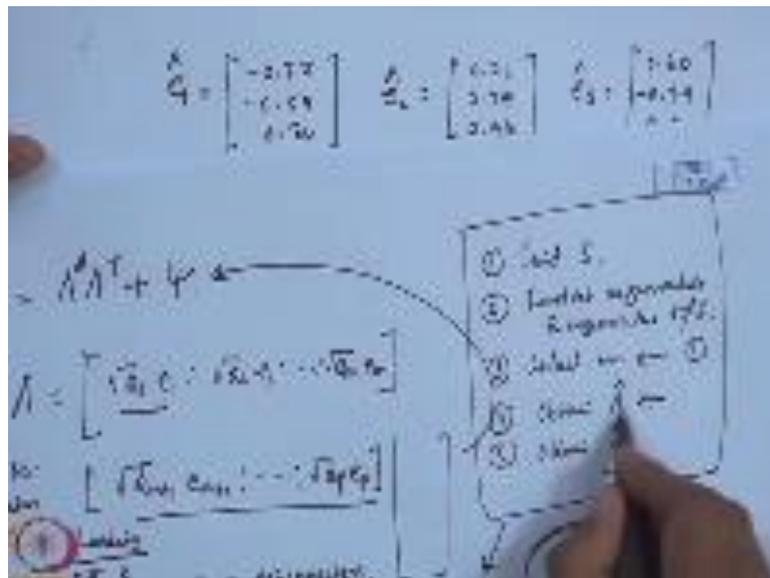
(Refer Slide Time: 34:24)

Handwritten mathematical derivation on a blue background showing the calculation of eigenvalues and eigenvectors for a 3x3 matrix. The matrix  $A$  is given as  $\begin{bmatrix} -0.77 & 0.21 & 0.29 \\ 0.21 & 0.29 & 0.93 \\ 0.29 & 0.93 & 0.11 \end{bmatrix}$ . The characteristic equation is shown as  $|A - \lambda I| = 0$ , leading to a cubic equation. The eigenvalues are found to be  $\lambda_1 = 1.4823$ ,  $\lambda_2 = -0.60$ , and  $\lambda_3 = 0.79$ . The corresponding eigenvectors are also shown.

So, what are the values for your  $e_1$ , you have solved  $e_1$  values are - 0.77 -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- 0.79 0.11. So, what I want, I want  $\lambda_1$ , which will be basically a vector  $\lambda_{11}$ ,  $\lambda_{21}$ ,  $\lambda_{31}$  and you have seen earlier also. Similarly, you can find out here  $\lambda_2$ , which will be  $\lambda_{21}$ , this  $\lambda_{12}$ ,  $\lambda_{22}$  then  $\lambda_3$ . Using this  $\lambda_3$ ,  $\lambda_{13}$ ,  $\lambda_{23}$ ,  $\lambda_{33}$ , our question is we want  $m = 1$  factor model, so we are concentrating only on  $\lambda_1$  not the other 2. These effects are basically giving to the  $\phi$  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 - 0.77 - 0.54<sup>0</sup> 0.34. So, if you multiply this two, what you will be getting, you will be getting - 9.37 - 6.574.75. So, this is your  $\lambda$  value, now I think if you go by the steps.

(Refer Slide Time: 37:53)



I said select  $m$ ,  $m = 1$  obtain  $\sigma$  that is  $\lambda$  cap then obtain  $\lambda$  cap means you are basically your  $\lambda$  cap is this one.

(Refer Slide Time: 37:58)

$C_1 = \begin{bmatrix} -0.77 & 0 \\ 0.17 & 1.75 \end{bmatrix}$

$\lambda_1 = 1.21$

$\lambda_2 = 0.98$

$v_1 = \begin{bmatrix} 0.60 \\ -0.78 \\ 0.8 \end{bmatrix}$

$v_2 = \begin{bmatrix} 0.12 \\ 1.15 \\ 1.75 \end{bmatrix}$

$\lambda_1 = \lambda_2 = \sqrt{0.23} \hat{v}_1 = \sqrt{0.23} \begin{bmatrix} -0.77 \\ 0.17 \\ 0.06 \end{bmatrix} = \begin{bmatrix} -0.37 \\ -0.27 \\ 0.95 \end{bmatrix}$

Obtain  $\hat{v}_1$

So, you have obtained  $\lambda$  cap now what you require to know.

(Refer Slide Time: 38:14)

2<sup>nd</sup> method

$$\hat{A}\hat{A}^T = \begin{bmatrix} -9.57 \\ -6.57 \\ 4.75 \end{bmatrix} \begin{bmatrix} -9.57 & -6.57 & 4.75 \end{bmatrix}$$

$$= \begin{bmatrix} 87.80 & 61.56 & -44.51 \\ 61.56 & 43.16 & -31.21 \\ -44.51 & -31.21 & 22.56 \end{bmatrix}$$

$$S = \hat{A}\hat{A}^T = \begin{bmatrix} 87.80 & 61.56 & -44.51 \\ 61.56 & 43.16 & -31.21 \\ -44.51 & -31.21 & 22.56 \end{bmatrix}$$

$$= \begin{bmatrix} 87.80 & 61.56 & -44.51 \\ -1.27 & 20.96 & 1.21 \\ 2.27 & -1.21 & 87.80 \end{bmatrix}$$

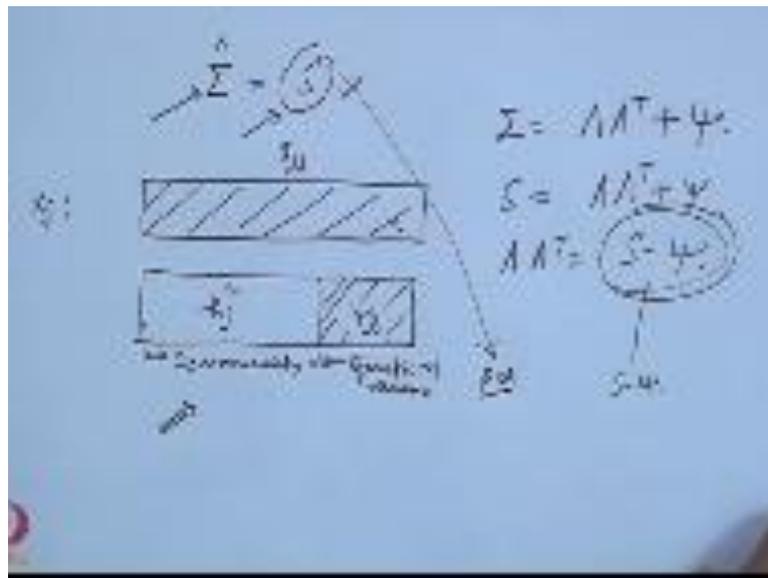
You have to find out or obtain  $\phi$  cap. In order to obtain  $\phi$  cap, you require knowing what  $\lambda$  cap  $\lambda$  cap<sup>T</sup> is... This value will be  $-9.57 - 6.57 \ 4.75$  into its<sup>T</sup>  $-9.57 - 6.57 \ 4.75$ . Now, values are when you estimate this into this, this into this, this into this like this, so ultimately we will be getting  $887.80 \ 61.56 - 44.511$ . How do you get this,  $9.57$  into  $9.57$  is this  $-9.57 -$  this one, this  $-$  this  $+$  this, this  $-$  this coming that is the way.

So, I repeat again - this into - this is this - this - this, this - this. So, in the same manner this  $\times$  this, this  $\times$  this, this  $\times$  this, you will be getting  $61.56 \ 43.16 - 31.21$  then  $-44.51 -31.21 \ 22.56$ . So, your  $\phi$  cap is  $S - \lambda$  cap  $\lambda$  cap<sup>T</sup>, which is  $150 \ -30 \ 50 \ 64 - 20 - 30 - 20 \ 49$  - this matrix, - 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: 42:03)



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  $\sigma_j^2$  and we have seen in last class that this can be partitioned into 2 half's, one is  $\phi_j^2$ , which is unexplained or PCP variants and other one is this  $h_j^2$ .

Now, this one is communality, this is your specific variance. Now, also you have seen that your  $\sigma$  is  $\lambda$ ,  $\lambda^T + \phi$  or  $S$ , we can write  $s$  be  $+$   $\phi$ , so that this, this, this into this  $\lambda$  into  $\lambda^T$  is  $- \phi$ . So, instead  $S$  matrix, if we now factorize  $S - \phi$  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  $S - \phi$ . Now, if you do little more manipulation here,

(Refer Slide Time: 45:09)

Handwritten mathematical derivation on a blue background:

$$S = T \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_p \end{bmatrix} T^T$$

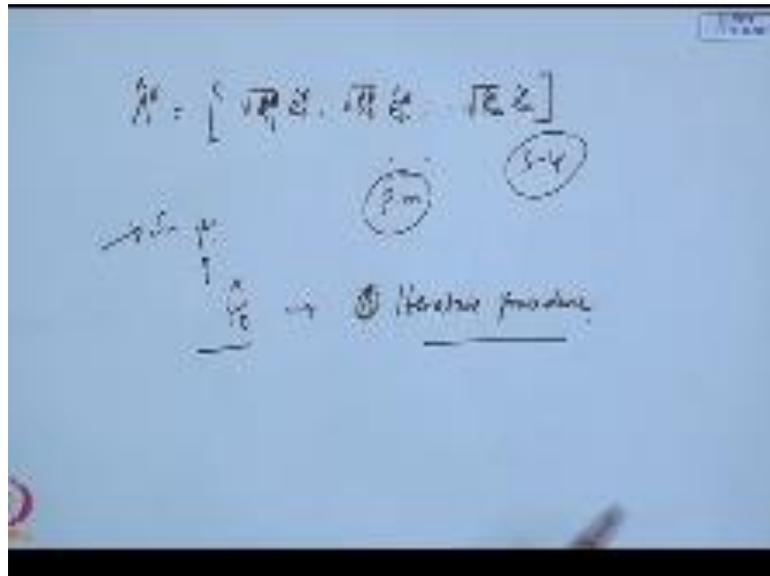
$$S - \varphi I = T \begin{bmatrix} \lambda_1 - \varphi & & \\ & \lambda_2 - \varphi & \\ & & \ddots \\ & & & \lambda_p - \varphi \end{bmatrix} T^T$$

where  $\sigma_j = \lambda_j - \varphi$  and  $\sum_{j=1}^p \sigma_j^2 = \sum_{j=1}^p (\lambda_j - \varphi)^2$ .

So our  $S$  is this  $s_{11}, s_{12}, s_{1p}, s_{2p}$  and if I say my  $\varphi$  is, so I create  $S - \varphi$  means, the diagonal elements will be changed. Now, using this is happening here, we are writing that  $\sigma_j = \lambda_j - \varphi$ . If you take the estimate  $s_{jj} = \lambda_j - \varphi$  then  $s_{jj} - \varphi = \lambda_j - 2\varphi$ . So, you are writing here  $\lambda_1^2 - 2\lambda_1\varphi + \varphi^2 + \lambda_2^2 - 2\lambda_2\varphi + \varphi^2 + \dots + \lambda_p^2 - 2\lambda_p\varphi + \varphi^2$  that is  $\sum_{j=1}^p (\lambda_j - \varphi)^2$ .

Now, you can try so again if we can say suppose the Eigen value and Eigen vector, they are like this. Suppose  $Q^*$  if you say  $Q^*$  and  $e^*$  or  $Q_j^*$  or  $e_j^*$   $j = 1$  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: 47:37)

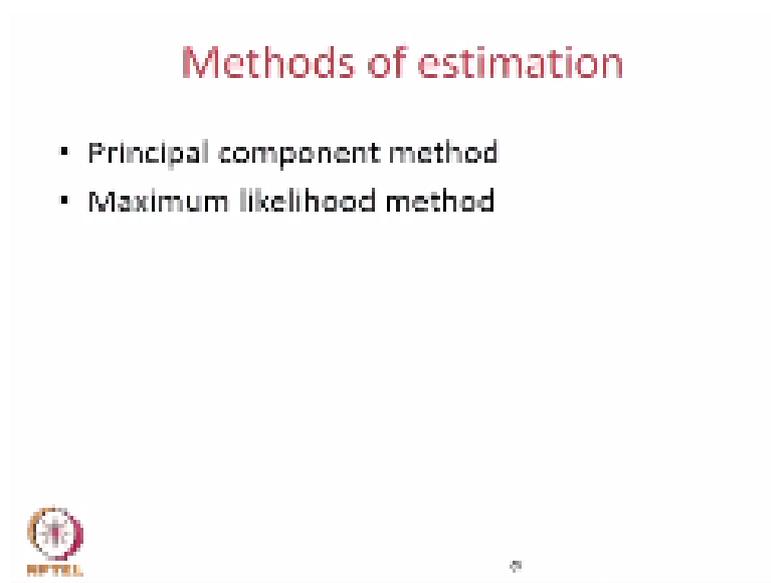


So, in this case if I say that  $\lambda$  star estimate gradually using this star and this one, so this will be nothing but your root estimate Q star that one e 1 star estimate, Q 2 star estimate, e 2 star estimate, Q m star estimate. So, you can go up to p, but essentially what will happen that this 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 - 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  $S - \phi$ , you know the value of s, but you do not know the value of  $\phi$ . So, you have to initialize this  $\phi$ , so you have to find out this  $\phi^0$  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. Okay So, I think the third method what you had said, third one will be the

(Refer Slide Time: 50:26)



## Methods of estimation

- Principal component method
- Maximum likelihood method

---

Maximum likelihood 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. 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. Here I want to tell you one example that is taken from Johnson and Wichen book.

(Refer Slide Time: 51:27)

### 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

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: 51:55)

### Results

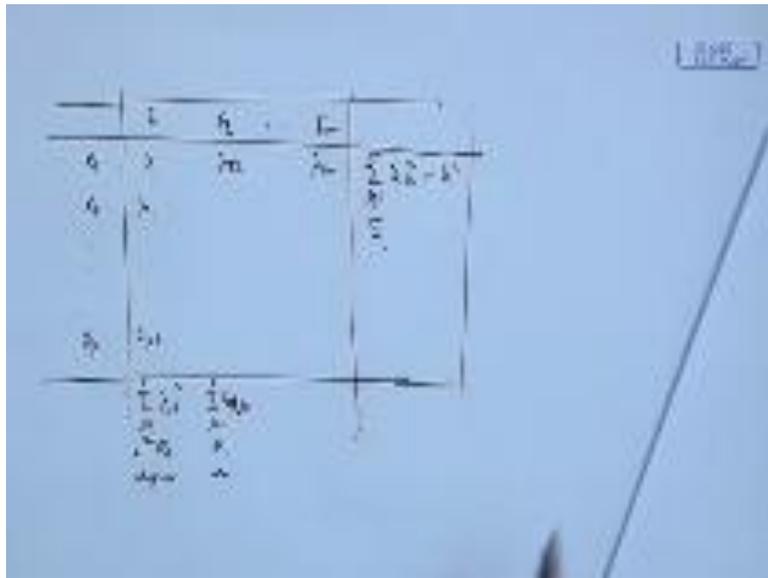
Variables	Factor-1	Factor-2	Communi- lities
Gaelic	0.553	0.425	0.48
English	0.568	0.288	0.41
History	0.302	0.450	0.38
Arith-matic	0.740	-0.273	0.53
Algebra	0.724	0.211	0.59
Geometry	0.595	-0.133	0.37
% variance explained	0.37	0.10	

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 these 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  $x_1 \times x_2$  then  $x_p$ , then your  $f_1, f_2$  like this  $f_m$  then you will be getting  $\lambda$  values here. So  $\lambda_1$   $\lambda_2$  like this  $\lambda_p$ , so then when you take that  $\lambda_j = 1$  to  $p$  so  $1 \ 1 \ 1 \ j \ 1^2$ , this is nothing but your  $q_1$  what you have developed. Similarly, the second one  $j = 1$  to  $p$   $\lambda_{j2}$  that will be your  $q_2$ . So, Eigen value here also Eigen value and similarly if you take sum here that  $\lambda_1$ , this is  $\lambda_1^2$ , this is  $\lambda_1 m$ . If I say this is also  $j =$  let it be  $k \ k = 1$  to  $m$  then your  $\lambda_{1k}$  square, this is what your  $h_1$  square cumulative.

And similarly, like this, this is what estimation is? I think I will give you little inputs to the maximum likelihood estimation, but in maximum likelihood estimation, the assumption.

(Refer Slide Time: 54:47)

Maximum Likelihood Method

$$x_i \sim N_p(\mu, \Sigma) = N_p(\mu, \sigma^2 I + \phi), \quad i=1, 2, \dots, n$$

$$L(x|\mu, \Sigma) = \frac{1}{(2\pi)^{np/2}} |\Sigma|^{-n/2} \exp\left(-\frac{n}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)\right)$$

$$\hat{\mu} = \bar{x}$$

$$L(x|\hat{\mu}, \Sigma) = \frac{1}{(2\pi)^{np/2}} |\Sigma|^{-n/2} \exp\left(-\frac{n}{2} (x - \bar{x})^T \Sigma^{-1} (x - \bar{x})\right)$$

$$\hat{\Sigma} = \hat{\sigma}^2 I + \hat{\phi}$$

$$L(x|\hat{\mu}, \hat{\Sigma}) = \frac{1}{(2\pi)^{np/2}} |\hat{\Sigma}|^{-n/2} \exp\left(-\frac{n}{2} (x - \bar{x})^T \hat{\Sigma}^{-1} (x - \bar{x})\right)$$

$$= \frac{1}{(2\pi)^{np/2}} \left[ \frac{1}{\sigma^2} \left( \frac{1}{2\pi} \right)^{p \times p} \right]^{-n/2} \exp\left(-\frac{n}{2} (x - \bar{x})^T \hat{\Sigma}^{-1} (x - \bar{x})\right)$$

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 +  $\phi$ , that is the estimation and definitely it is for  $i = 1, 2, \dots, n$ , 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  $2n\pi - n/2$  trace of this  $-n/2 (x - \bar{x})^T \Sigma^{-1} (x - \bar{x})$ .

Now what will happen when you estimate  $\mu$  cap is as  $\bar{x}$ , if you put here this will become  $0$ , so then your this likelihood  $\mu$  kept  $\sigma$ , this will be  $-\log 2n\pi - n/2$  trace this  $\sigma$  inverse this. Now, if we say that  $\sigma$  cap is  $\lambda$  cap,  $\lambda$  cap  $= \sigma^2 I + \phi$  cap and then this function will become this cap  $= -n/2 \log 2n\pi - n/2$ . So,  $2n\pi$  we have written here  $\log 2\pi \sigma$  I have written wrongly here that is  $\log 2\pi \sigma$ , then  $\log 2\pi \sigma$ , then  $\log 2\pi \sigma$  this was changed. So I will change this one to like this  $-n/2 \log$  of that determinant of  $2\pi \Delta$  cap  $\phi$  cap + 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.

(Refer Slide Time: 58:14)

Lawley (1940)

Normal eqn:  $S \hat{\lambda}^{-1} \hat{\lambda} = \hat{\Lambda} (I + \hat{\Lambda}^{-1} \hat{\Lambda})$

$G = \text{diag}(\hat{\lambda} - \hat{\Lambda})$

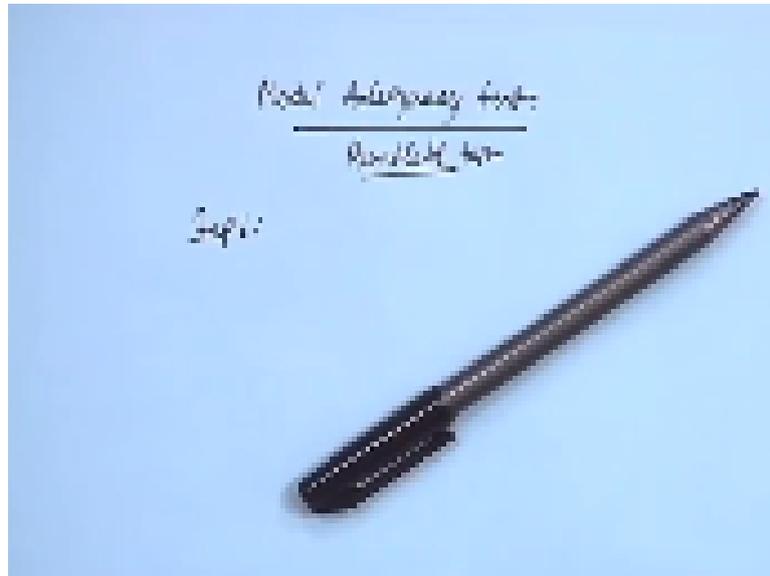
The simplified normal equation by Lawley, simplified Normal equation given by Lawley 1940, he says giving some normal equation for solving  $S \hat{\lambda}^{-1} \hat{\lambda} = \hat{\Lambda} (I + \hat{\Lambda}^{-1} \hat{\Lambda})$ . And definitely  $\hat{\Lambda}$  is diagonal matrix of this is the 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 okay 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  $\hat{\Lambda}$  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  $\hat{\Lambda}$  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,

(Refer Slide Time: 01:01:10)



That is what I can say model adequacy test. I will explain to you that Bartlett test here. 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 grab. Thank you very much.