Data Analysis and Decision Making – I
Prof. Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 40
Factor analysis

A warm welcome, to my dear friends and students; a very good morning, good afternoon, good evening to all of you and as you know this is the DADM which is Data Analysis and Decision Making – I, set of lectures or course under NPTEL MOOC series. And, this course is for 12 weeks, total number of hours is 30 that is 60 lectures and we are in the 8th week and today, is the last class of the 8th week which is the 40th lecture and as you know that each week we have 5 lectures, each being for half an hour and I am Raghu Nandan Sengupta from IME Department, IIT Kanpur.

So, first I will wrap up whatever we discussed in the last two lecture. So, when we were discussing about the PC principal component method. So, I was taking the vectors. So, vectors are corresponding to the eigenvectors. So, the first value for the first vector; that means, this the 1 comma 1 would be multiplied with the x 1 random variable that is the weight. Then corresponding to 1 comma 2 for the first vector would be multiplied by X 2 continue with this in this way.

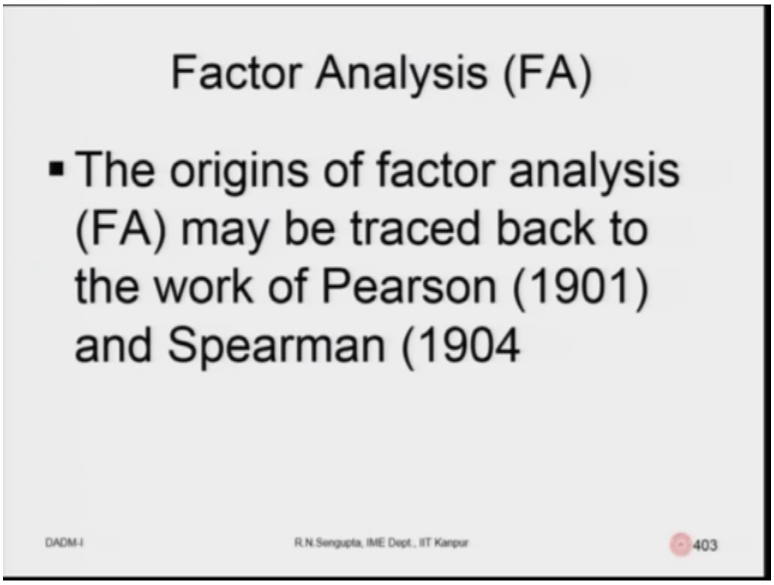
So, these are the linear combinations which you are trying to basically formulate corresponding to the Y 1 which is the first principal component. Then the corresponding cell values for the second vector first second third fourth would be multiplied by the corresponding X 1, X 2, X 3, X 4 respectively to get you Y 2 we will continue in this way. Now, we have already presupposed by the consideration of eigenvalues and eigenvectors they are orthogonal.

So, obviously, it will mean Y 1, Y 2, Y 3, Y 4 are all orthogonal point – 1. Point number – 2, you have to basically find out the variability and also check that the basic assumption based on which we are proceeding is that the sum of the squares of the deltas which you are formulating for the first stage, second stage, third set stage, stage by the word stage I mean corresponding to Y1, Y 2, Y 3 when you add up the square should be 1, we have also proved that. Then third thing which we basically found out for the variability; variability we also found out basically corresponding to Y 1, Y 2, Y 3 basically coming out with the eigen values.

So, in this way taking these three important concepts we have been able to prove the principal component method is actually based on scientific and mathematical reasons. And once we find out Y 1, Y 2, Y 3 we can safely say they are the best combinations of X 1, X 2, X 3 till X p in some proportions such that the variability and the dependent structure is given the maximum set of information from those Y 1 to Y p and they are orthogonal. So, based on that we can basically conclude the principal component method which we did is right.

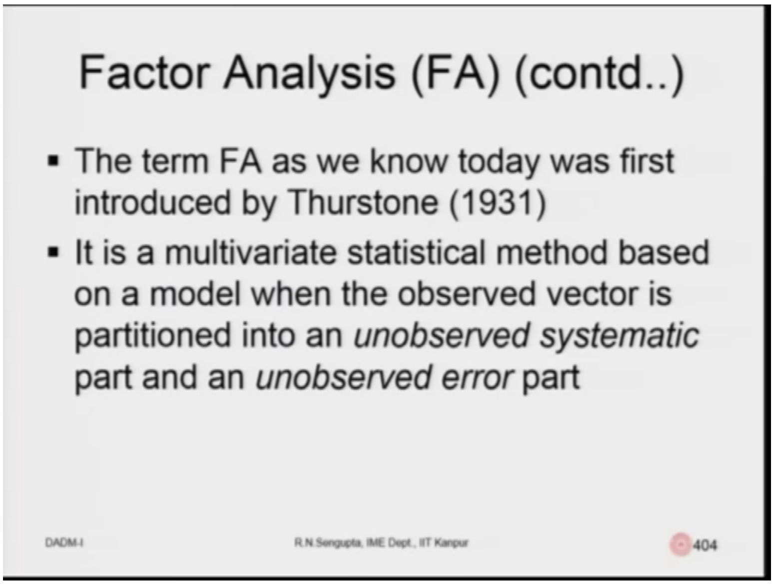
Obviously, it has been little bit long run process we spend about one and a half or about 2 lectures to find out the corresponding principal component method and solve it. So, we will basically go into the second method and proceed correspondingly take some time in order to solve that and obviously, if you remember I did mention about the multi linear regression concept also we will come to that.

(Refer Slide Time: 03:58)



Now, we will consider the concept of factor analysis; so, the origins of factor analysis which is dates back to the work of Pearson in 1901, and Spearman. So, basically based on that, we will try to understand the concept of factor analysis. Well, basically first go through the concept then give you the results.

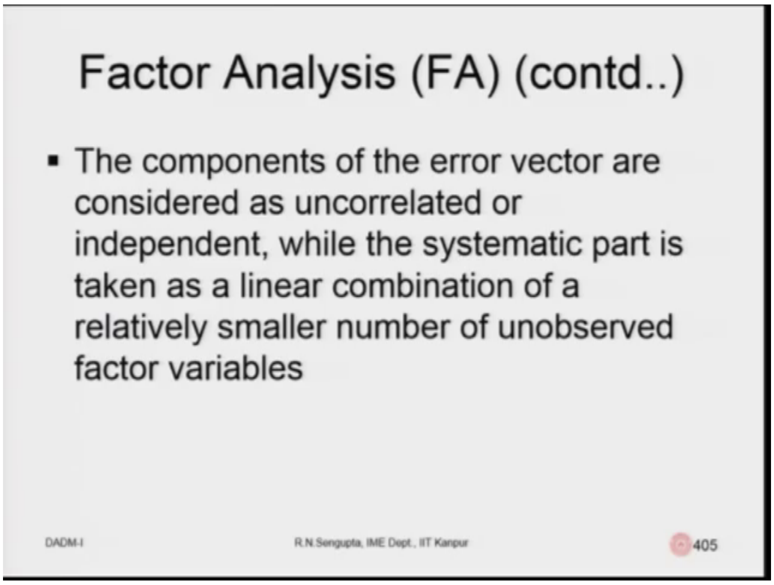
(Refer Slide Time: 04:13)



So, the term factor analysis as we know today was first introduced by Thurstone in 1931. It is basically a multivariate statistical method based on a model when the observed vector is partitioned into unobserved systematic part and an unobserved error part. So, you have X 1 to X p. So, their effects are divided into two sets; one in unobserved systematic dot type of error or type of I would not say that error systematic type of effects and another is basically white noise which we cannot control.

So, what do you want to find out is that find out those sets or combinations in such a way that we are able to give the maximum level of prediction. Prediction I am word using the word very generally level of prediction such that we are able to solve the problem and give us the maximum set of information.

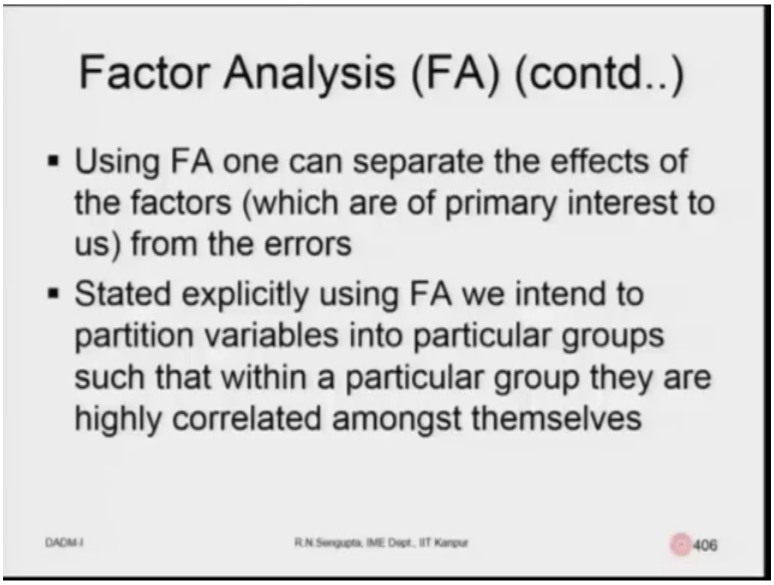
(Refer Slide Time: 05:10)



Now, the components of the error vectors are considered as uncorrelated. So, obviously, if there are X 1 to X p the errors which we will have we will consider that for the best case the number of errors are also p., But, obviously, it can be more or less, but we will consider these errors in such a way that they would uncorrelated and independent which may not be true. We will try to basically assume very simplistic assumption.

So, for the concept of factor analysis you will basically have X 1 to X p. So, one set would be the variables in minimum number of quantities of X 1 to X p such that as I mentioned they would be linear combinations of the effects coming out from relatively small number of unobserved factor variables. Another would be the errors which we will consider they are uncorrelated and independent. So, we will have basically two sets of combinations coming out. But, if you remember for the principal component analysis we did not consider any error sort of thing.

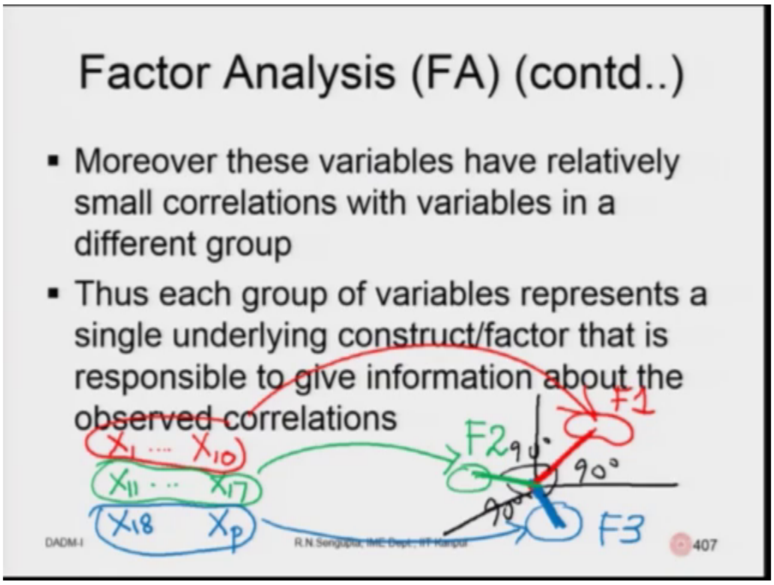
(Refer Slide Time: 06:18)



So, using the concept of factor analysis one can separate the effects of these factors which are a primary interest to us from the errors. So, what we are interested in, take out the errors what is left is the pure effect and you want to find out from the factors in the minimum number of first factors which will give us that total number of maximum effect.

Stated explicitly using factor analysis we intend to partition variables into particular groups such that within a particular group they are highly correlated among themselves, such that the effect of the dependent structure we are able to study it with the maximum accuracy. I am using the word accuracy again in the general sense.

(Refer Slide Time: 07:01)



Moreover these variables have relatively small correlations. So, what technically we are doing is that the X 1 to X p will basically try to divide into say for example, groups F x which is there between each group interrelationship would be maximum point – 1. Point number – 2 the effects within the groups or they or the correlations or so called the dependency structure between the groups would be minimum.

So, I will try to draw it in the sense it is like this which may not be the actual way I have been able to do for the principle component, but this is just a try. So, let me use the accesses as so, consider you have the 3-dimensional 3 effects and you want to have the effects. So, I will basically group them consider the actual grouping is 3 because it is easy for me to make you understand in the 3-dimension and consider that the X 1 to X p number of variables.

So, we will consider the first set as this, the second set as say for example, this and the third set as this. So, what are these red, green and blue colors. So, the red, green, blue colors would be basically X 1; I am basically using the nomenclature as X 1 to X p and then divided into three groups. So, this is X 1 to say for example, X 10, then sorry I should use X 11 X say for example, 17 this 10 17 does not make any sense. I mean this is I am just giving numbers and the last one is X p.

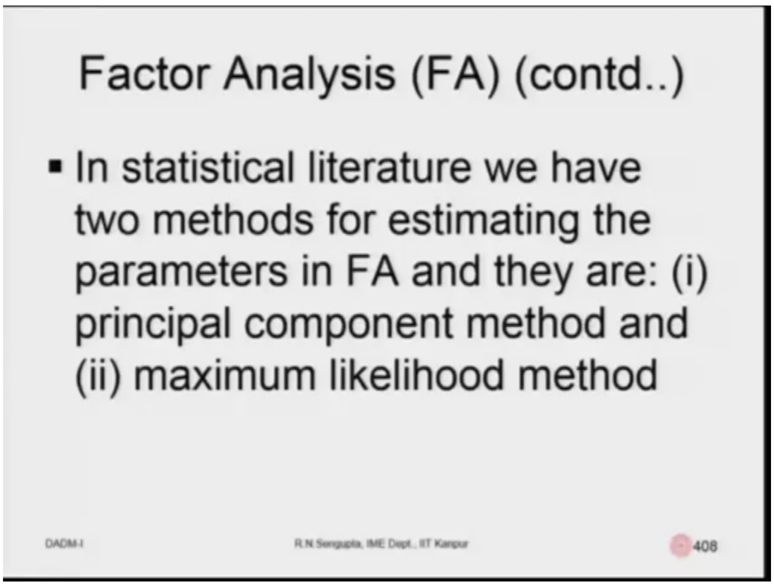
So, what I have been able to do is the first set have been moved into the first factor, I will denote it as F 1 not formula 1, F 1, the second one is mapped on to the F 2 factor 2 third one has been done into factor 3.

Now, what we meant actually the sets X 18 to X p their correlation amongst themselves with the highest what I also made it corresponding to F 2 the F a correlations between X 11 to X 17 would be the highest and finally, when I look into F the factors one set X 1 to X 10 the correlations is the highest dependence structure is the highest. But, also F 1 let me draw the line so, it will be easy in the sense.

So, consider this is the. So, called effects of as a group F 3 on the on the read readings. This is for F 1 this is for F 2, but we have been able to divide F 1, F 2, F 3 into three independent groups such that they would be orthogonal that will work; that means, that orthogonal is the theoretical sense, but what we will try to orthogonal means there is no dependence between this group.

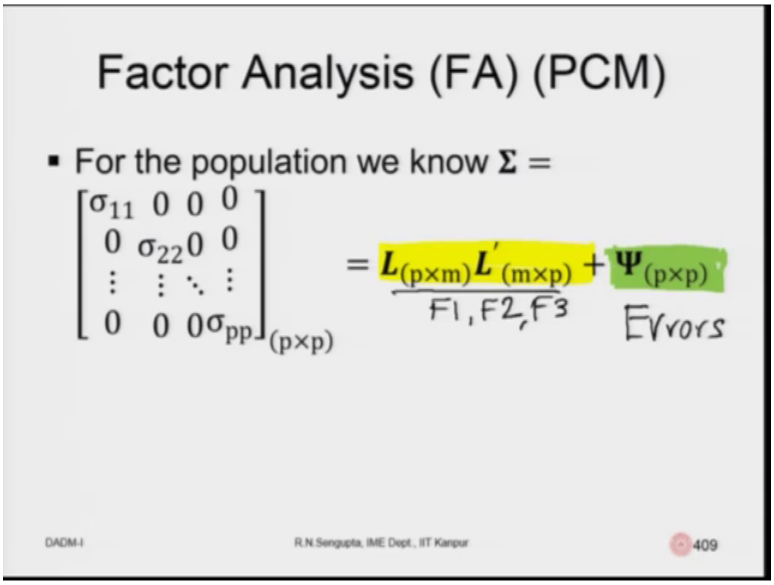
So, say for example, this is X 1 and this is X 2 inside X 1 the dependence structure is the maximum, inside X 2 the dependence structure is maximum, but when we combine them the relationship as a group the group interrelationship is as minimal as possible that is why I am trying to utilize the concept of orthogonality. Even though that may not be true 90 degrees, but they would be as high as possible such that the dependence strategy is minimum.

(Refer Slide Time: 12:10)



In statistical literature we have two methods for estimating the parameters in factor analysis and they are basically the principal component method and the maximum likelihood method. So, we will consider only one of them and basically try to solve the problems accordingly.

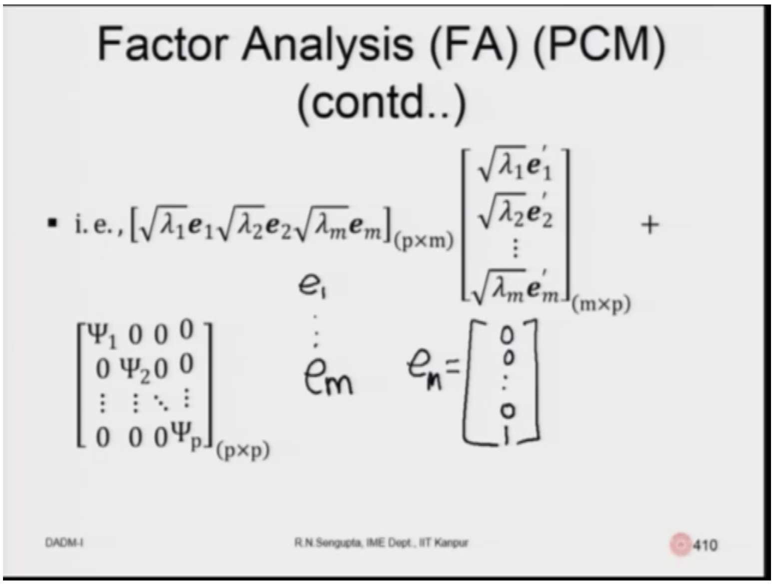
(Refer Slide Time: 12:27)



Now, the for the factor analysis principal component method which we will use so, for the population technically we will try to analyze this problem this. So, we will basically have p number of actual X 1 to X p of the variables and we will basically put partition in them into two groups. So, the two groups partitioning would be done like this do not pay attention on the variance covariance matrix on the left hand side what I am want to basically highlight is this.

So, what you want to do is, the first set would basically be the number of factors which will be which will give us the maximum set of informations such that each of the groups they would have maximum correlation amongst inside them, but when we compare their interrelationship it will be the least.

(Refer Slide Time: 13:35)



And, other effect would be wait, sorry would basically be the error terms which are coming out. So, they are basically as I mentioned this is for the case of F 1, F 2 this factors remember F 3 and these are basically related to the errors.

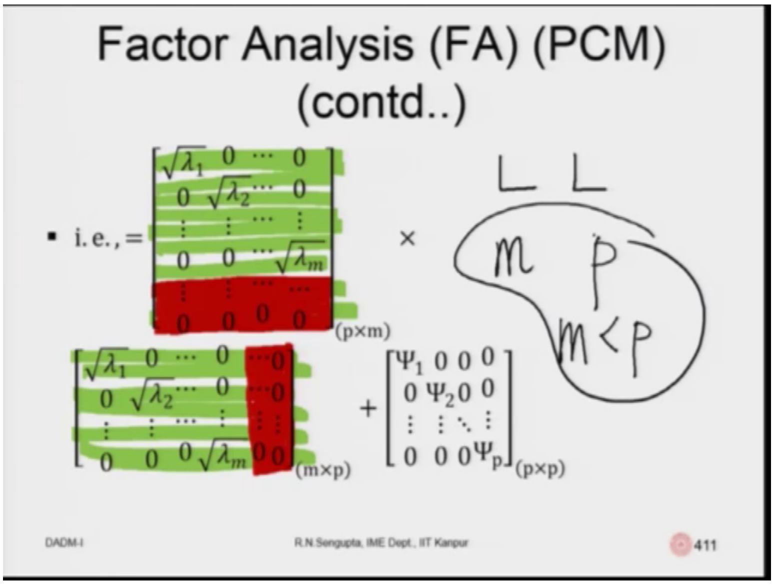
So, obviously, again if you remember it is it is intuitive that the concept which will try to utilize in order to solve this problem would be again the eigenvalues and eigenvectors principle. So, what we will try to do is that basically break them into the eigenvalues, eigenvectors concept, but with the notion that the orthogonality between the eigenvalues, eigenvectors are being considered using the unit vectors concept.

That means, this e 1 to e m which we have so, obviously, m and p would be related I will come to that so, these are the orthogonal vectors unit vectors will be there in the directions of first second third fourth till the m-th one. So, e 1 technically would be the vector such that the elements would be the first is 1, the next values are all 0 when I considered so, this would be e 1. When I considered e 2 so, let me write it.

So, the first element is 0, second element is 1, third element is 0 the last one. When I consider e 3 first is 0, second is 0, third is 1, fourth is 0, till the last one. So, when I come to the e e suffix m first 0, second 0, all are 0 except the last one. So, we are trying to basically multiply the corresponding eigenvectors in the using the unit vector methods and trend and try to find out the so called factors.

Now, if you remember what we have done is that the sets or the relationship is divided into factors, that relationship is maximum within themselves and between themselves we will basically try to bring it to as low as possible which is 0 which is the orthogonality concept which I mentioned.

(Refer Slide Time: 16:59)



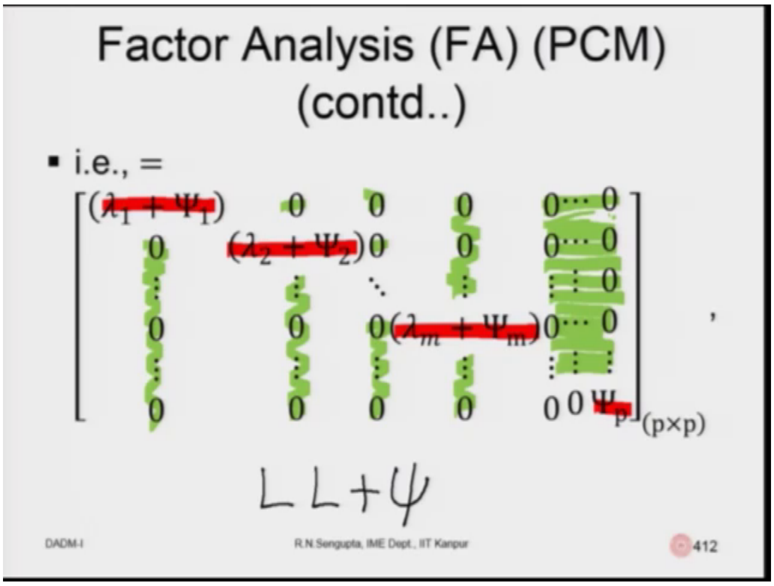
Now, once we basically find out the concept of the in the factor analysis method what actually we are doing is that dividing the whole set. So, p number of vectors the random variables were there out of that we are taking m and these m we are dividing two groups each group would have the maximum amount of interrelationship inside them between them is as low as possible.

So, using the concept basically the L and the L transpose values would be given by and if you remember one thing I will come to that. The set of values in the sense that there are p number of rows here and m number of columns, but in this case the value of p minus m number of rows which are in between here which try to these values after the m-th one m plus 2, m plus 3, m plus 4, till p 1 there are all 0's remember that, that is important.

And, obviously, if it is happening from l or l transpose the corresponding values here it is a m cross p. So, m number of rows p number of columns so, this m cross m values are all nonzero in the sense nonzero means principal diagonal is there, but after that the columns m plus 1 m plus 2 till the p-th one all values are 0 and the error terms are given by my.

So, this is what is we are trying to highlight this is the m and p as such that m obviously, you should be less than p based on that we will try to do the calculations.

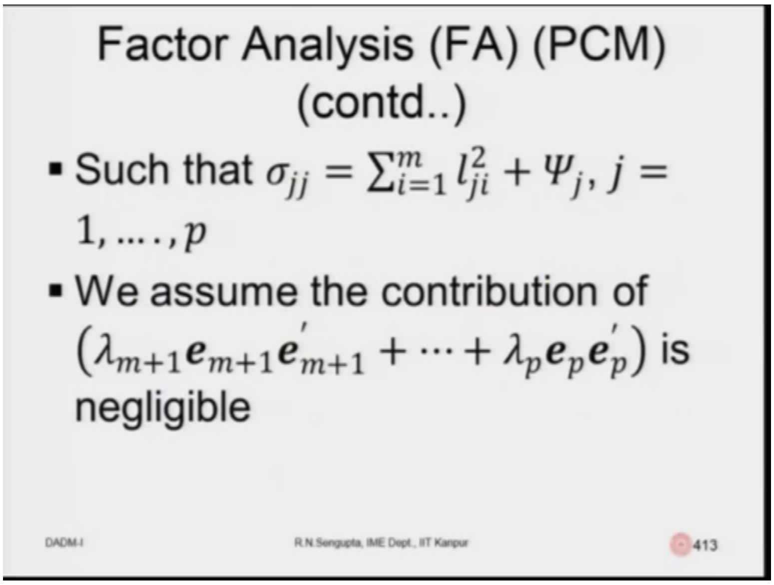
(Refer Slide Time: 19:37)



So, once we put them in the total matrix based on which we are trying to find out the relationship. So, it was what the value was given as L and L to the capital phi, these are all vectors and matrices remember that. So, when we multiplied that you will basically have the sets are like this is should not be red in color let me be use too green. So, these are corresponding to the values. So, these are all the what I am talking about these are the off-diagonal elements, this is the.

So, the diagonal elements are so, these are the orthogonality coming from the eigenvalues, eigenvalues multi plus the phi component for the first error then eigenvalues plus the phi component for the second error till the p cross p. So, if you consider the principal diagonal first element second element m-th element p-th element. So, they would basically give me the actual value corresponding to the error terms of the variance covariance matrix which you are trying to utilize.

(Refer Slide Time: 21:12)

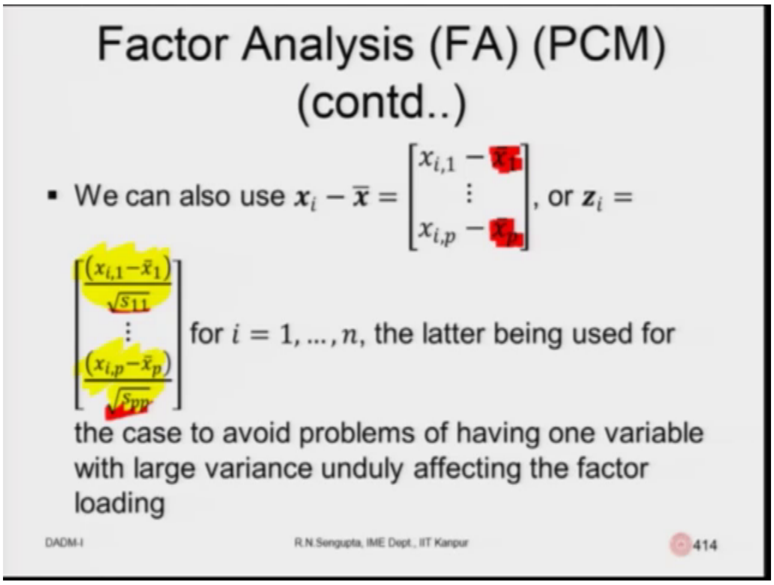


Now, our main concern; obviously, would be you want to find out the principal diagonal which is there in such a way that the values of multiplying the eigenvalues eigenvectors with errors would give us the actual component of the variance of the or the first which is the covariance of the first with itself then with the second with itself third with itself all will become out. So, when I put it equal to 1 and then find out and n consider j is equal to 1, so, the values of 1 so, this would be L 1 square and L 2 square all will be found out such that it will give me the maximum value of this errors plus the correlated terms which should be equal to the variance covariance values.

We assume the effects which is going to come out from the remaining terms. So, that means, there would be one effect which is coming out from the interrelationship between them and one is the effect which is coming when we compare between themselves between the groups between the groups errors would be as low as possible dependent structure is as low as possible and inside the groups the effects would be maximum which is coming from the variances.

Now, when we consider that obviously, the assumptions would be we will consider the best estimate for the population mean and corresponding to that we will proceed also try to utilize the best estimate for the population variance also.

(Refer Slide Time: 22:52)

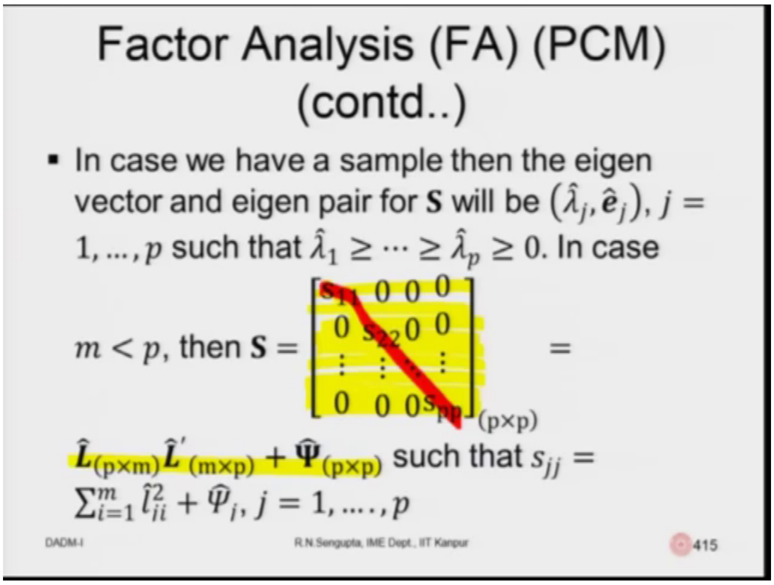


Best estimate for the population mean would obviously, with the mean value of the sample which is this for the first random variable till the p-th random variable and S 1, S 2, S 3, S 4 provided that the mean values are not known obviously, you will use one degrees of freedom we will be try to utilize S without the stars.

So, this would basically be the so called standard error value for the first random variable similarly the standard error value for the p-th variable. So, we standardize them. So, the standardized value for the first value which is the random variable minus the mean value divided by standard deviations with the say for example, a very simple z value for the first similarly for the p-th one. And, we need to find out the corresponding values correspond considering that the reading numbers are given 1, 2, 3, 4 till n.

We will use this case to avoid problems for having one variables with large variance unduly affecting the factors of the others. So, what we will try it what we are trying to do is trying to normalize them with the standard values or the standard variance such that any high and low values would basically be eliminated. And, any relationship of dependence coming on to the variability would be standardized and reduced as low as possible.

(Refer Slide Time: 24:28)



In case we have a sample when the eigenvalues vectors and eigenvalues are given; so, we will consider the eigenvalues and eigenvectors as lambda for the eigenvalues and e values unit not the now not the unit one depending on the calculations which we do the e 1, e 2, e 3, e 4 values which you find out for till the values of p because there are p number of random variables for X 1. So, we will basically be able to find out the case corresponding the fact that the standard errors should be calculated for the factor analysis based on the fact that we need to find out the eigenvalues and eigenvectors.

So, once we do that considering because eigenvalues and eigenvectors are not orthogonal. So, once we find out the standard error matrix or the variance covariance counterpart from the sample would be given, where the principal diagonal and if we use a different color principal diagonal values diagonals are the standard error whole square for first, second, third, fourth correspond in the case that we are trying to put them and find out the values 1 comma 1, 2 comma 2, 3 comma 3 for the cell. And obviously, the of the diagonal elements would be 0.

So, once we have that again it will verify the fact that we are trying to basically divide the total effects into two groups, one of the factors which are orthogonal to each other technically. And, the elements which are taken in each group in factors would have basically a maximum dependence structure amongst themselves.

So, with this I will close this 40th lecture and continue the discussion further in the factor analysis do the problem. And, I know the that the progress has been slow, please bear with me because the number of types things we are going to cover is huge and we will try to basically keep it as and many many as possible. Because only doing the problems would not give you the idea that why we are doing it, like the concept of eigenvalues, eigenvectors or orthogonalities they are coming up time and again in multivariate analysis.

So, I am repeating them timing and again please bear with me and I am sure you will be able to understand. And, for any queries please put up your queries on the forums so, we will definitely take our time and answer all of them accordingly.

Thank you very much and have a nice day.