**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**

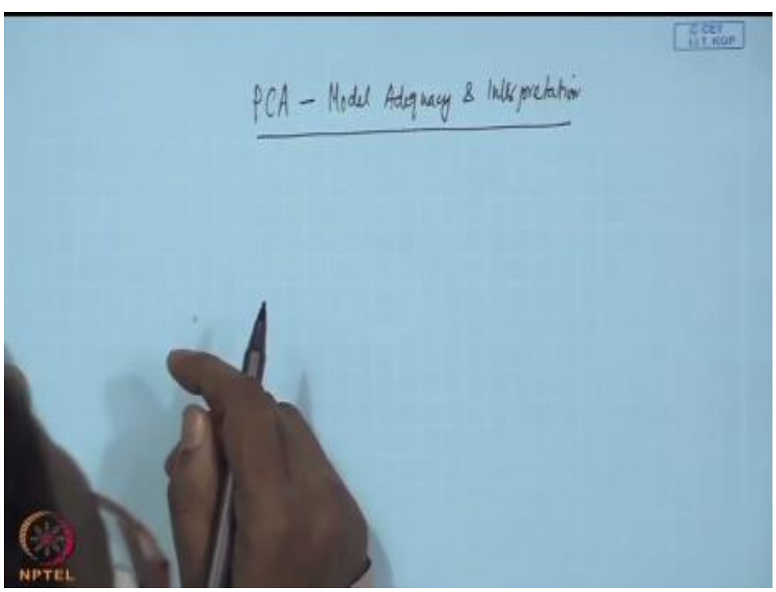
Lecture – 31

Topic

Principal Component Analysis (PCA)- Model Adequacy & Interpretation

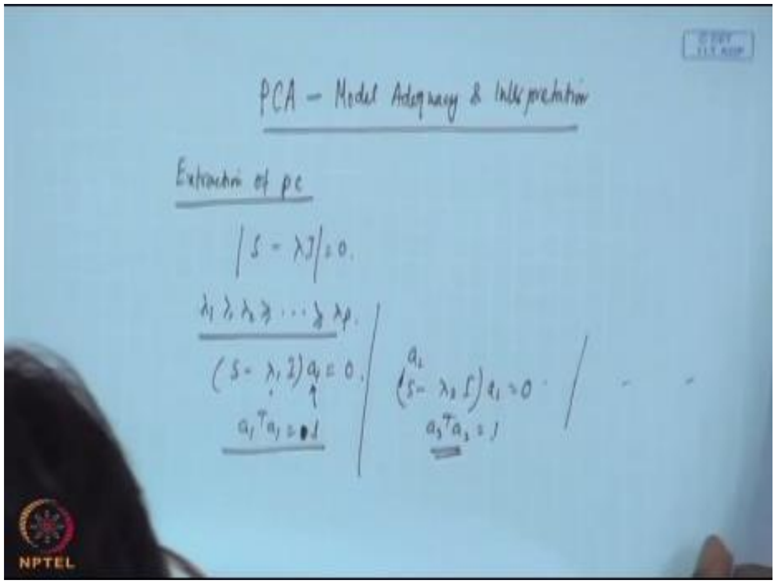
Good morning we will continue with PCA.

(Refer Slide Time: 00:25)



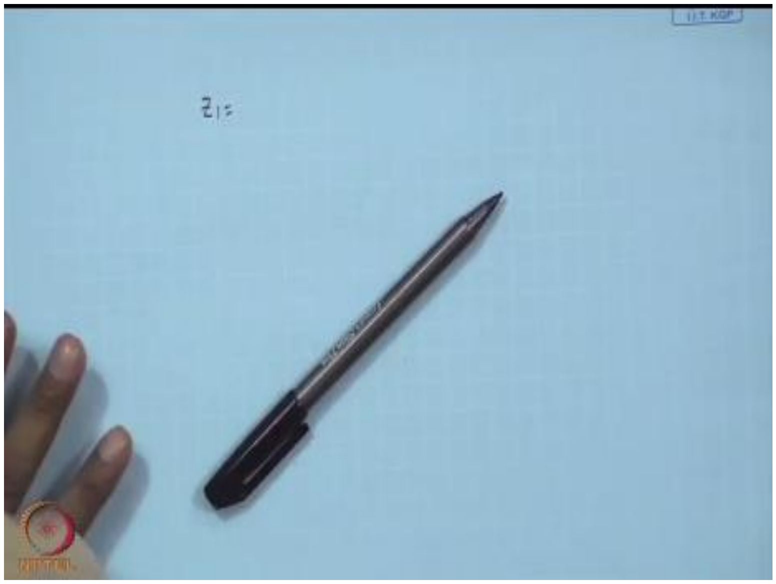
Today we will discuss PCA model adequacy and interpretation so last class we have discussed up to a extraction of PCA.

(Refer Slide Time: 00:43)



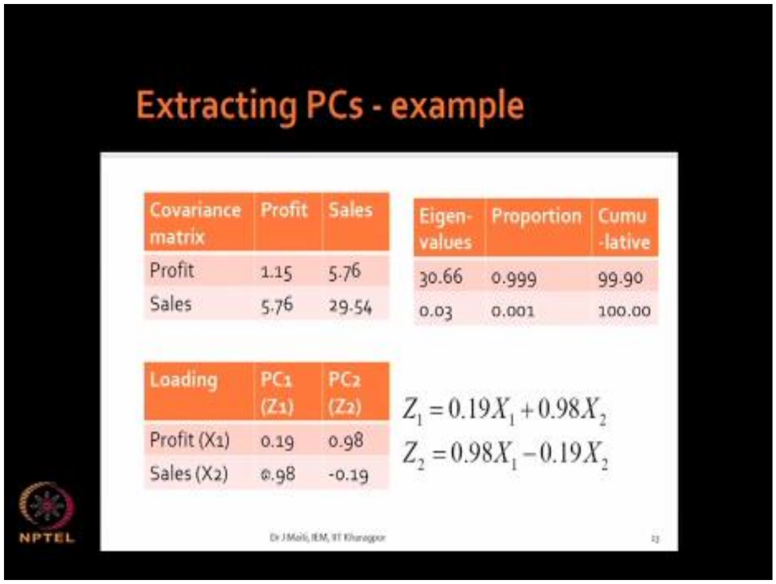
Extraction of principle component and we have seen that first we have used $S - \lambda$ determinant equal to 0 then we got the Eigen values and then using individual suppose of the for the first one a 1 first Eigen value with the first Eigen vector we have found out and subject to $a1^T$ a1 equal to 1 so similarly you can find out a 2 using $S^T \lambda$ 2 i a2 equal to0 and $a2^T$ a2 equal to 1 in this manner you we will be able to find out the Eigen vectors once you estimate the Eigen values and Eigenvectors you can write down the equations also.

(Refer Slide Time: 02:02)



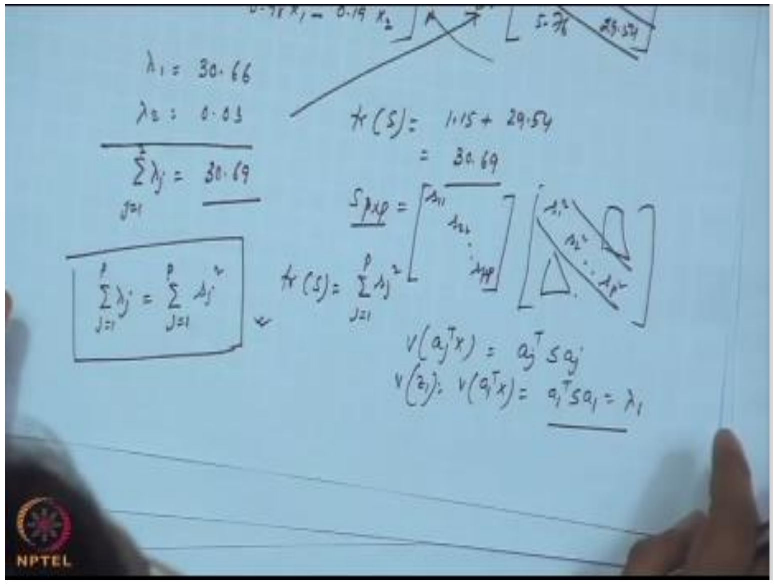
For example what the example we have considered.

(Refer Slide Time: 02:06)



Here that the loadings which from profit and sales are two variables interrelated variables we have extracted also two components PC1 PC2 and these are all Z_1 Z_2 and the loadings that basically these component loading the Eigen vectors 0.19 and0.98 for your first one.

(Refer Slide Time: 02:34)

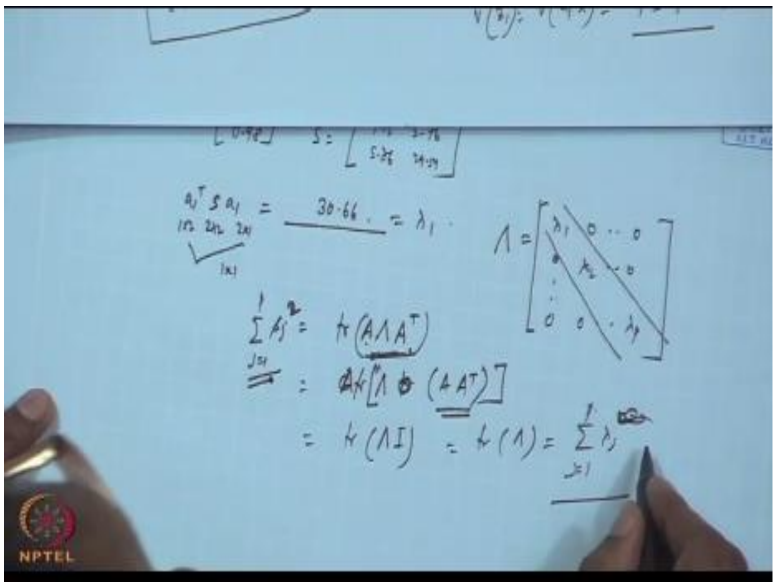


That principle component 10.19 x 1 + 0.98 x 2 and similarly for Z2 it is 0.98x 1 - 0.19 x 2 you see ultimately what is happening here just reverse only one sign reversal is there so what I ask you that you please compute in detail using these matrix S matrix like 1.15 5.76 and 5.76 29.54 you compute in detail and come to up to this level and you will be able to find out why what is the reason of this type of relationships now under adequacy test what you want to discuss first you have identified λ.

1 I think it is 30.66 probably and λ 2is 0.03 okay then the total is sum total of λ_j j equal to 1 to 2 here this equal to 30.69 now we have extracted this from this S matrix now what is the stress of S stress of S is the sum total of the diagonal elements this is 1.15 into 29.54 is not this it is 60 30.69 this is 30.69 now if there are if S is p x p matrix then you will be getting like this suppose we are using sample S_{11} S_{22} like this S_{pp} other way you can write $S1^2$ S^2 like this Sp^2 that you will be getting.

And are definitely off diagonal values are also there but stress is this so I can say stress of S is nothing but sum total j equal to one to p Sj^2 so these can be proved that your sum total of λj j equal to 1 to p equal to sum total of Sj^2 sum total of that means variability if you extract the maximum number of Eigen values and what you will find out that sum total of this Eigen values will be sum total of the variance component of the original matrix this matrix what you have considered and you have already also found out another thing that suppose I say variance of $aj^T x$ this one also we found out that aj^T S a_j so now if I want to know what is the variance of Z 1 that will be variance $a1^T x$ which is $a1^T S$ a1 it should be λ_1 it will be l λ_1 how what is a_1.

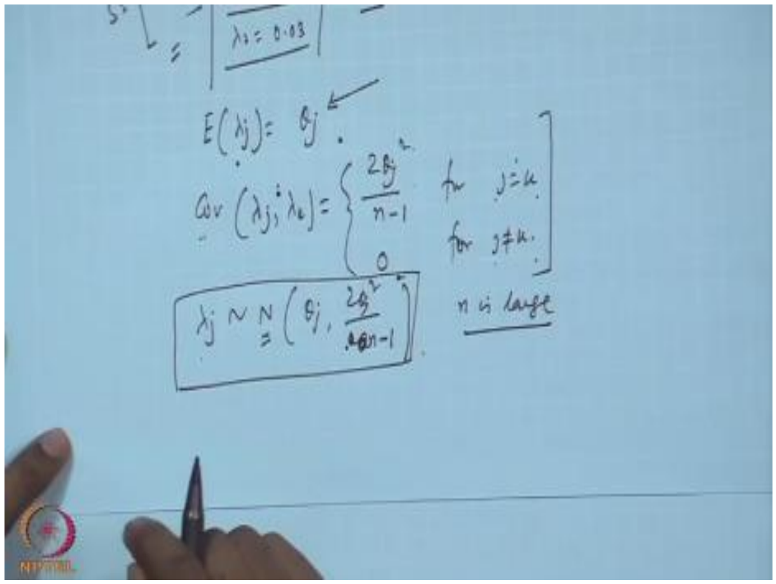
(Refer Slide Time: 06:56)



In this case a 1 is 0.19 0.98 S is 1.15 5.76 then 5.76 and your 29.54 so just you do this multiplication and find out that is $a1^T S$ a1 you find out because this one is $a1^T 1$x 2 2 x 2 x1 so resultant will be 1 cross1 and definitely this will be 30.66 which is this is a1 we can take a1 is λ a1 we can take lambda outside and that you can do also that is also possible basically what I am saying suppose if you create a diagonal matrix of λ 1 all 0 λ 2and λ p then what you can write basically is sum total of this one.

I think that is ja^2 j equal to 1 to p Sj^2 square that you can write stress of your λa λ I think A^T that you can be it can be written like this now what you will do basically is why am I using a is because you started with this and then λ is this transformation matrix so it will be λ^T you a sorry stress of this now A^T this is orthogonal I so stress of λ into i this will be stress of λ that means diagonal elements so sum total of j equal to 1 to pj^2 so this is what the decomposition λj λ j^2 j equal to 1 to p it is not square λj λ j if you go for singular manner decomposition that square will be there okay fine.

(Refer Slide Time: 09:48)

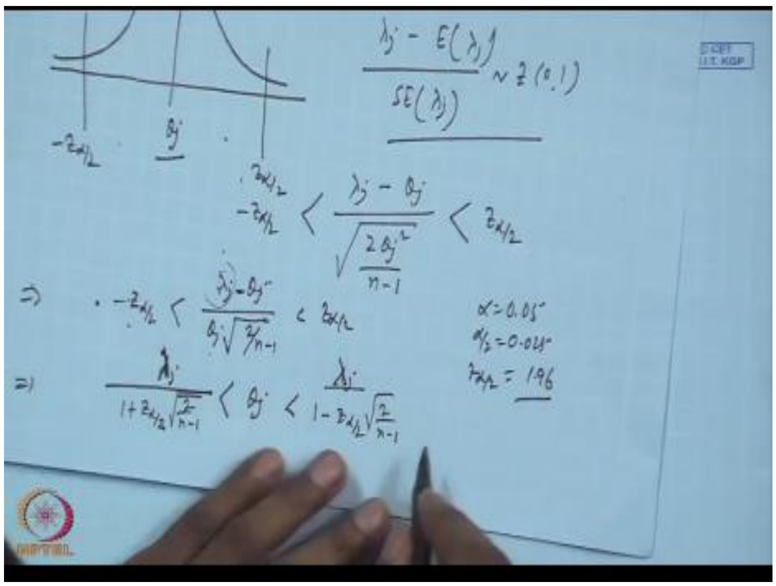


So this is the case so you now understand the relationship everything is wrong next is as I was telling you suppose you got $\lambda 1$ equal to 30.66 $\lambda 2$ equal to0.03 this are point estimates because we have taken example that 2 by 2 matrix from n equal to certain value I think in our case n equal12, two variable result suppose I go for next time as I told you earlier also point estimate if you want to find out then we are going for expected value of λ_j that is equal θ_j that means we are saying that in a population principle component analysis.

If you do there you will be getting if you know everything about the population you will be getting the exact value of Eigen value for that a particular population this is from sample this is from population so this one is covariance of λ_j λ_k this will be 2 θj^2 by n - 1 for j equal k and 0 for j not equal to k this is the development so what I mean to that is that is two lambda component λ_j and λ_k they R covariance when k j equal to k this is nothing but the variable component j not equal to k.

It is the covariance component so covariance is 0 and variance is $2\theta j^2$ like this we can assume that it will follow a normal distribution with theta j and $2\ \theta\ j^2$ by 2 n – 1 when you that mean you know this solution now this distribution is common that is definitely a question but it is complicated question so there are several researchers statisticians develop this so we are sticking to this λ_j can be approximate to normal distribution provided definitely n is large is large.

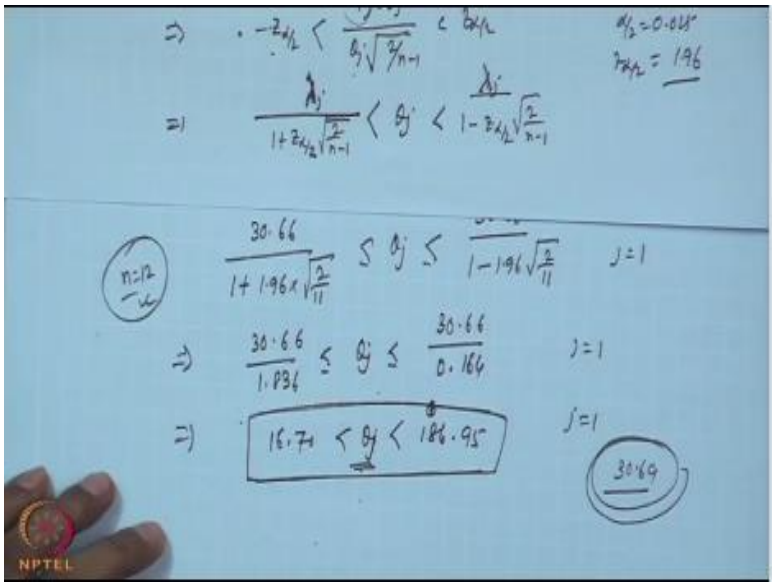
(Refer Slide Time: 12:52)



Now when this is the situation you can find out the confidence interval for θj so what we will do basically you want to find out this - Z α by 2 + Z α by2 we know that α_j - expected value of λ_j by standard error of λj this follows Z0 one the basic statistics that concept so that means λ_j - θ_j now what is the standard error of λ j now what is the value here that variance part $2\ \theta j^2$ n - 1 so that means standard error if you write down here it will be $2\ \theta j^2$ by n – 1 now these quantity will be this side - Z α by 2 this side Z α by 2.

So ultimately this one can be written like this Z α by 2 less than λ_j - θ_j by θ_j will come out root 2 by n -1 less than Z α by 2 you do now mathematical manipulation then ultimately what you will get you want something like this so θ_j will come out here then this θ_j is there so this θ_j + this quantity will be added and then as it is within this 1 by something will come final form will be like this one + Z α by 2 by n -1 then 1 by Z α by 2by n -1 but we have this λj component so it will be λ_j it will be λ_j into λ_j into λ_j you can write if our α value 0.05 then α by 2 is 0.025 then Z α by 2 this will be 1.96.

(Refer Slide Time: 16:00)

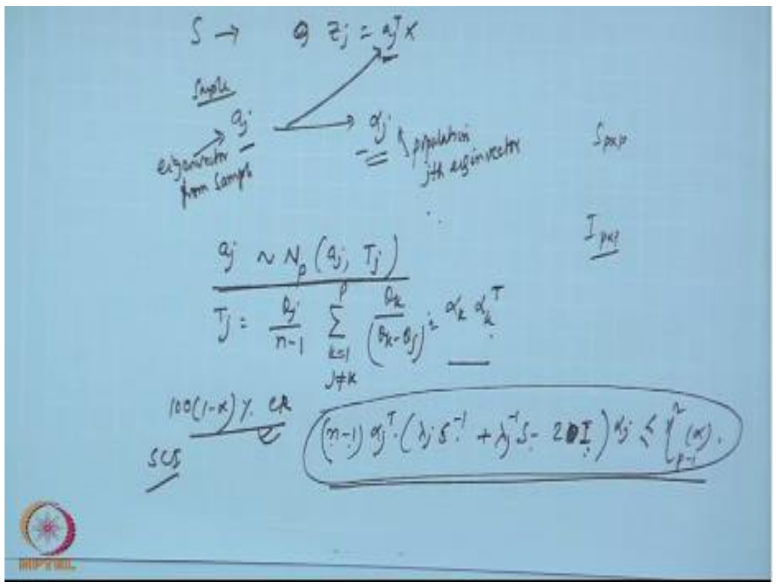


Now we got 30 our λ value 30.66 by 1 + 1.196 into2 by 11 n is 12 less than equal to also you can write less than equal to 30.66 divided by 1 - 1.962 by eleven so ultimately this quantity will yield 30.66 by 1.836 less than equal to θ_j less than equal to 30.66by this one - 0.164 and this leads to16.70 less than theta 1because we are considering j equal to 1 so if I write θ_j also j equal to 1 here then this value is 186.95 but do you agree with this there is an original see your original sample data if you see 3. 30.69 this is the total variability.

Now this is because of many things your number variables is number of observation is very small this is huge difference so I am sticking to that first one point estimate one only that point estimate will it is all practical purpose is it will solve the what you want to do but see this much variability for the population PCA is there If i go by this simple sample piece then this trial error method we are saying that this is possible when n is large we are assuming normal distribution our case n is very small value.

So we cannot say that this is correct but this is the way you will be finding out the confidence interval for the population Eigen value.

(Refer Slide Time: 19:09)



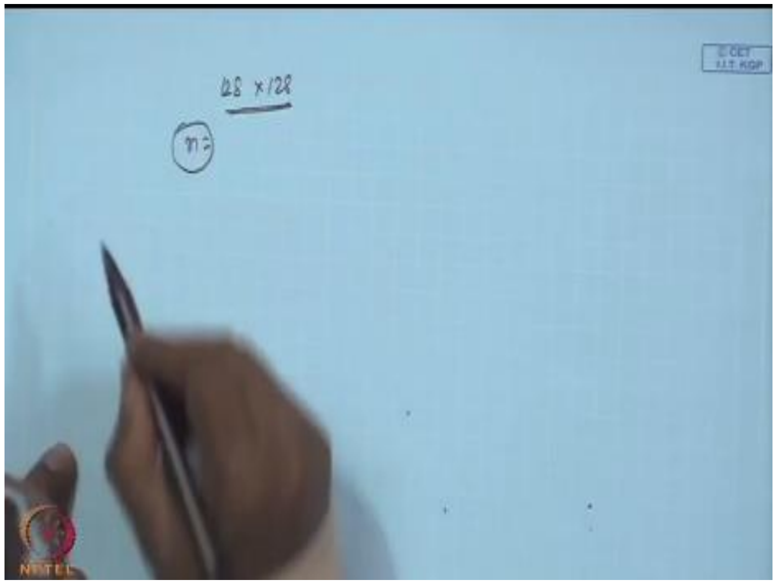
Now another one is there, similarly using S we found out that a j what is this Z j equal to $a_j^T xa$ now this a you have a compound root the value of a $_j$ but this core is sample a j corresponding in population in this λ_j is there this is the Eigenvector for from sample this is population Eigen vector Z so similarly there is definitely the distribution possible but that distribution again really they are all competitive distribution what is that the Eigen vector what you compute these follows multivariate normal multivariate normal with a_j and t_j where t_j is θ_j by n - 1 sum total of k equal to 1 top.

Then j not equal to k θk by $\theta k - \theta j^2$ λ k λk^T not λ that is αk αk^T this α is what this 1 so as we are using here α k k equal to 1 to p this type of mixed equation will come just for the sake of clarification request all of you to check this that k and j term the notation if there is just check this one so ultimately as we are getting as are getting that in p a_j t_j that is the multivariate normal you will be getting 100 into 1 - α percent confidence region if it is p variable case you will be getting the confidence region that confidence region takes this form n -1 α j^T λ_j S inverse + λ_j inverse S - 2 λa 2 i 2 into λ_j that follows chi square p - one α this is the confidence region so I repeat that one n - 1 α_j^T λ_j S inverse + λ_j inverse S -2 i this i is basically if S is p x p.

Then i is also p x p that is the identity matrix then this quantity will follow this distribution chi square p - 1 with α you can check this also confidence region now as it is confidence is here you may be interested to know the simultaneous confidence interval then it will be kept to simplify it further but there are complex but I am giving you basically what you have applied in your PCA that is from sample.

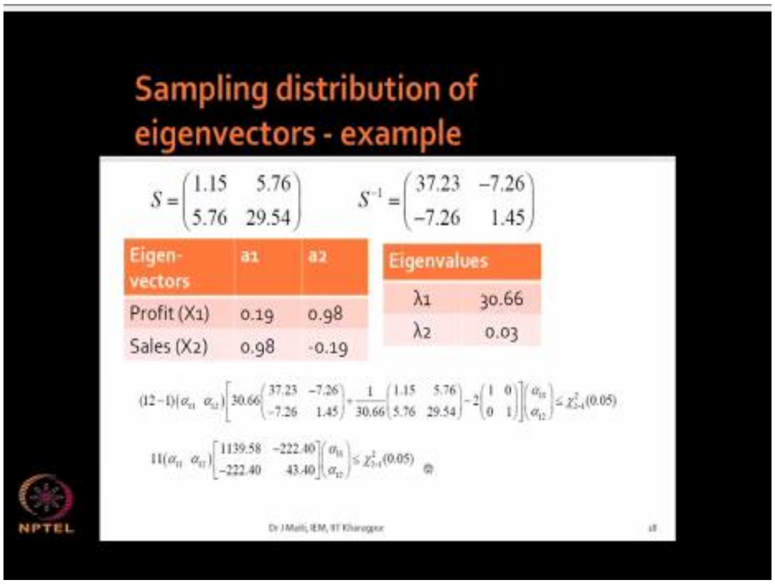
Now what is the guarantee that suppose in the texture direction what you are trying to find out that you are actually going for the clue what is your sample size it is a quite large yes very large very large so we are I think you can still check the particular the confidence interval for λj Eigen values that means the dispersion must be checked and if it is very big I think it will be within the limit because variant is large okay so you mean of size 128 cross 128 then is it 128 into 128this much data now what is this 128.

(Refer Slide Time: 24:08)



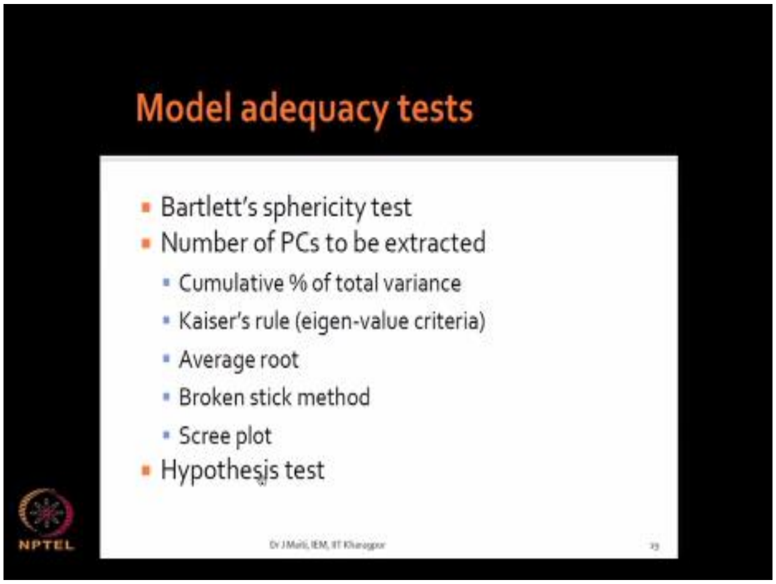
I understand but 128 this itself is a matrix 128 cross128 what about n what is n total observation 128 into 128 so these many observations are there it is quite larger so I think that as you are doing it may be we will be finding papers where this dispersion measures the Eigen values that is not considered that means uncertain part is not considered so you can propose the adequacy will be much better you can propose that and may be different sample size how the adequacy that your fitness will change that can be also a good addition to theoretical contribution now let us see this.

(Refer Slide Time: 25:05)



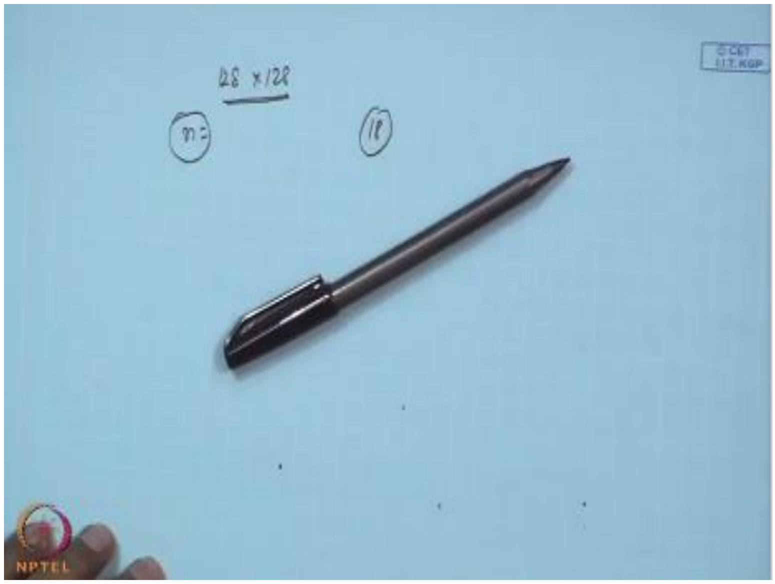
What that example is now see what happen ultimately λ_1 is this and ultimately this is what our confidence region is you have seen in multivariate normal distribution to the exponents are that same thing is basically when I am talking about that a j is multivariate normal for two variable case the ellipse is formed ellipse is formed so you find out this confidentially form your Eigen vector okay.

(Refer Slide Time: 25:44)



Then I will show you the that model adequacy test in terms of Bartlett's sphericity test then we will go for what are the different criteria that can be used to find out that what are the number of PC's that can be retained finally we will show that some hypothesis test what is this what is happening suppose there are p x variables you are extracting PC's m PC's you are extracting let m is much less than p for example I am taking 20 variables and then finally compacting into two dimensional using PCA so that means you are.

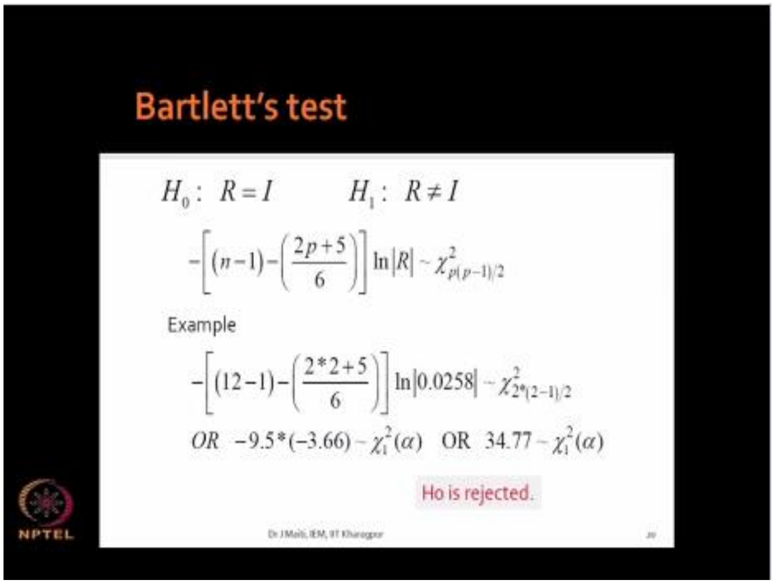
(Refer Slide Time: 26:45)



You are excluding 18 dimensions you are removing this if your data is very highly correlated then only it is possible that means the 18 dimensions are not required they are too much highly correlated the Eigen values will be 0 for all the remaining 18 only for 2 Eigen values will be significant others will be almost negligible so that 18 PCs what you are not considering are not the considering are not there they give you something wrong signal that if you go by subject in criteria of eliminating some of the some of the principle components it may give you some wrong results.

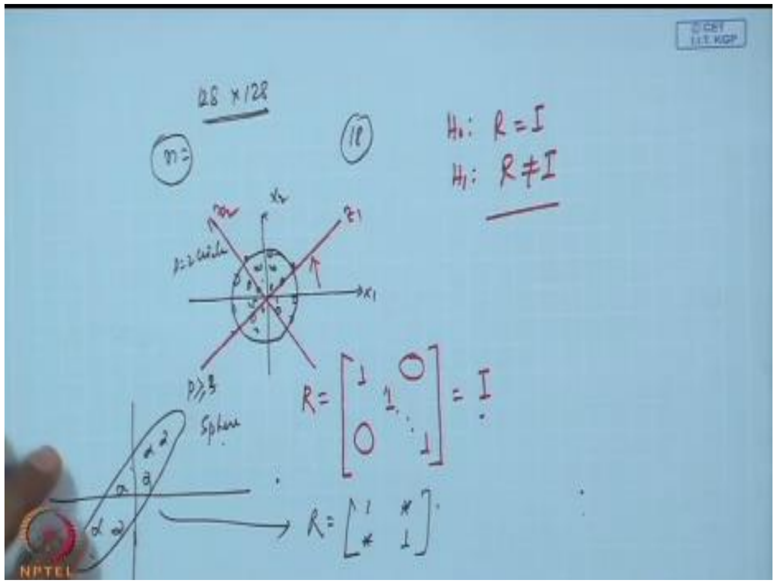
So finally that is why what we want to say that some hypothesis test that we are removing at some subset of principle components we want to see that whether those sets are contributing to the extraction or not but to the explaining the variability in two sense that is not extraction I mean to say that explaining the variability of the x or whether the discarded components variability is very limited or small what is Bartlett’s sphericity test.

(Refer Slide Time: 28:17)



Bartlett’s sphericity test is interesting for example.

(Refer Slide Time: 28:21)

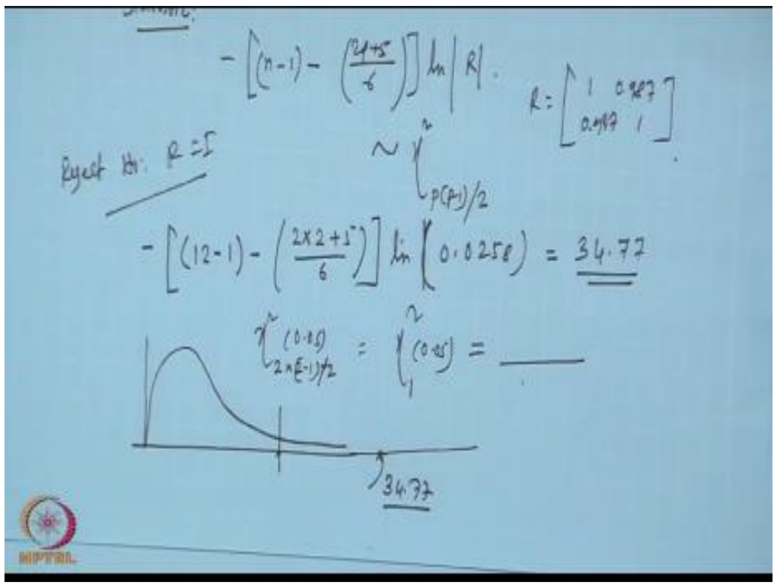


You see your data scattered data x1 and x this is your data this is random data if you increase this is a random data sheet I mean there is more relation between x_1 and x_2 so it will resemble a circle if we go for p greater than 2 that means p > than equal to 3 or what will happen it will become a sphere so for this is for p equal to 2 it is a circle for p greater than 3 it will be a sphere greater than equal to 3 it will be sphere now when you find out the circle or sphere that means the variables are basically random in nature in variables are un correlated variables scattered randomly without any systematic component.

This means variables are un correlated if each of the variables are un correlated with others what will happen how many PC component you can extract can I go for principle component analysis for example this is my case suppose if I rotate this suppose this Z one this side it is Z 2 you are getting now if I take the correlation matrix between these variables with no correlation between each other what will happen you will get only diagonal elements one and off diagonal elements all will be 0 so if true sense the variables are not correlated what is the meaning of going for principle components analysis.

No need of principle analysis in that case R will be I that is what happens in that Bartlett's test it is like this h0 that R is i and correlation co efficient is that matrix is a identity matrix now instead of this figure if you get a figure like this high correlation is there so in that case your this R it cannot become this cannot become 0 this will be having substantial help so Bartlett says then alternative hypothesis is R not equal to i and he developed.

(Refer Slide Time: 31:35)

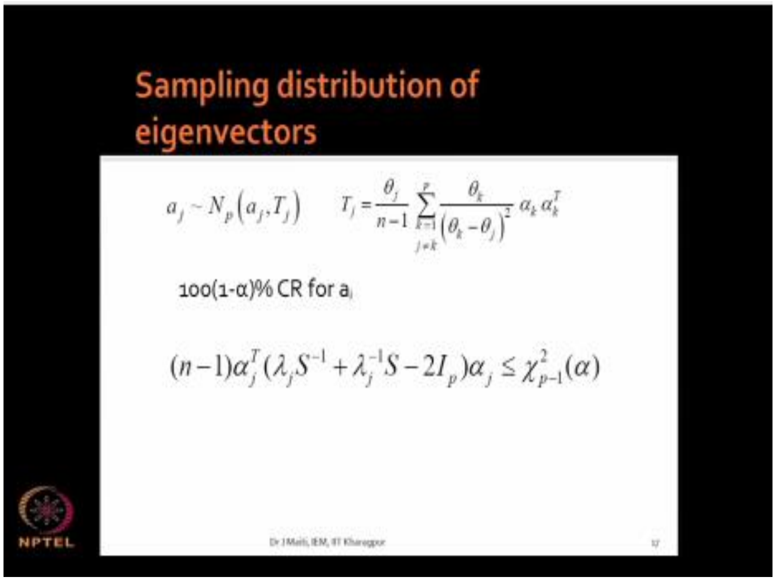


The statistics value also what is the statistics to be tested the statistics to be tested is here - n -1 - 2p + 5 by 6 log base a determinant of r this is the statistic now these statistics follows χ distribution with p into p - 1 by 2 sorry now the example we have taken 3 for 12 observation then we have two variables then our log R the determinant part the determinant if you say that the determinant that log R that determinant of R this one will be computed like this 0 2 5 8 the determinant of R.

So this quantity this quantity will become some value 9.5 into 3. that is almost 34.77 this quantity becomes 34.44 okay now what is your χ that p means your2 into 2 - 1 divided by 2 let α is 0.05 so this value will be ki^{2} square 0.05 if you go if you see the table I think this value may be very low value not this much 34 you will not get this so ultimately what will happen your α values for this one is this value and this 34 will be coming somewhere here so for our example reject ho R equal to i.

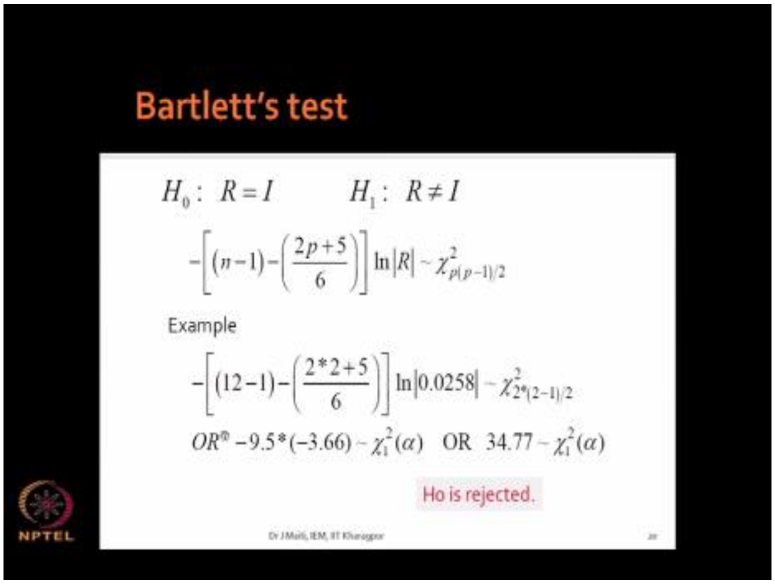
This is obvious if you see the R matrix here this R matrix is 1 0.987, 0.9871 this is r matrix I think similar may be 9871 or something like this.

(Refer Slide Time: 34:21)



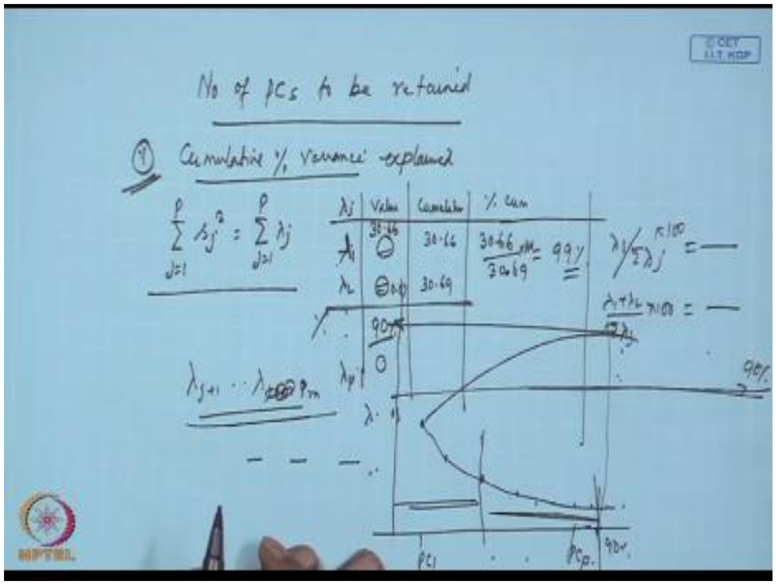
Somewhere it is there okay.

(Refer Slide Time: 34:57)



But very high value So when there is high value this Barlett's null hypothesis is rejected what we are saying you can go for principle component analysis now this is one of the test in fact this test can be done much before the principle component analysis first see the correlation matrix then this correlation go for Bartlett's test if you find that yes there is felicity no need of going back to principle component so much of effort why should you put then suppose Bartlett's test is rejected and you can go for several PC's now what is there how many numbers of how many numbers how many PCs you will give.

(Refer Slide Time: 35:17)



Keep number of PCs to be retained there are several criteria one of the criteria is cumulative percentage variance explained what we have seen we have seen that the variance component of the original data matrix that is equal to j equal to 1to p λ_j that we have proved so that is why what you do you do like this suppose your this is λ_j that is λ_1 λ_2 so like this there will be λ_p you find out the value of λ_1 how much some value like this you are getting then this is the value then you find out cumulative value for example in this case our first one is 30.66 second one is 0.03.

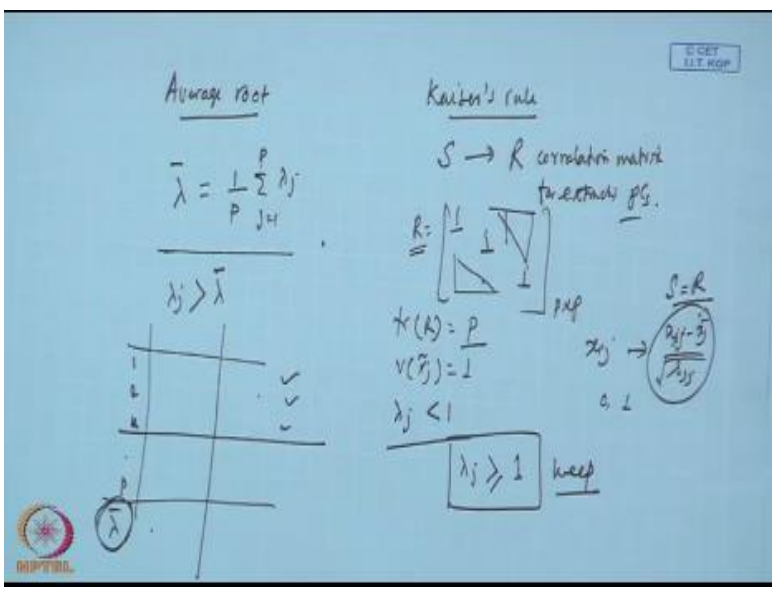
So this is 30.66. If I see the percentage percent cumulative 36.55 by 36.69this will be almost 99% 30.66 by total sum of λ that is the total sum of the λ because we have two only two from the example that is 2 so that is why this into 100 it will be almost 99% now the thing is that we just go on doing like this that λ one divided by sum of λ_j that $\lambda_1 + \lambda_2$ divided by sum of λ_j so this into 100 into 100 in similar way then you find out what is the percentage what percentage like this then you put a cut of flate cut off by 90%.

So those many λ components you keep that is cumulative percentage of total variance explained this is what the first criteria you can do is so in this criteria what will happen you may find out that you have to go for 95 90% of the total variability of x to be explained you may take some of the principle component for example j + 1 to may be j + j + m these many component or upto j + 1 to λ_p these many components these components we have taken or if I say no problem m components we have taken these values are almost equal very negligible.

For example if I plot like this is my λ value this side the PC value PC1 to PC p component you may find out first λ value is here second lambda value is here third λ value is here fourth λ value will be here like this you may get a figure like this actually suppose 90% is coming here this is your 90% because it will go for cumulative at this point it will extract here you make a percentage column also to cumulative percentage this will be 90% so see from contribution point if I see I think up to here it is there is difference but see all those values are not significantly different there is many screw contribution still there because I want this.

So in that case you may not be interested in 90% total variability explained you may be sacrifice little bit and then you come to some other total variability may be it is 80% then you go for 80% and then these many these many components you keep there are some other methods.

(Refer Slide Time: 40:27)



Just to do this one how nicely you will be able to do this another method is average root average root basically what will be your average root λ average one by p sum total j equal to one to p λ_j so you keep those many principle components value is greater than λ bar getting there is a I spoke about it so many components where p component is extracted and the λ bar then component 1 2 2like this p component you check suppose k components are having values which is greater than λ bar these many will be kept understood there is another method known as Kaisers rule Kaiser rule says that that you instead of .

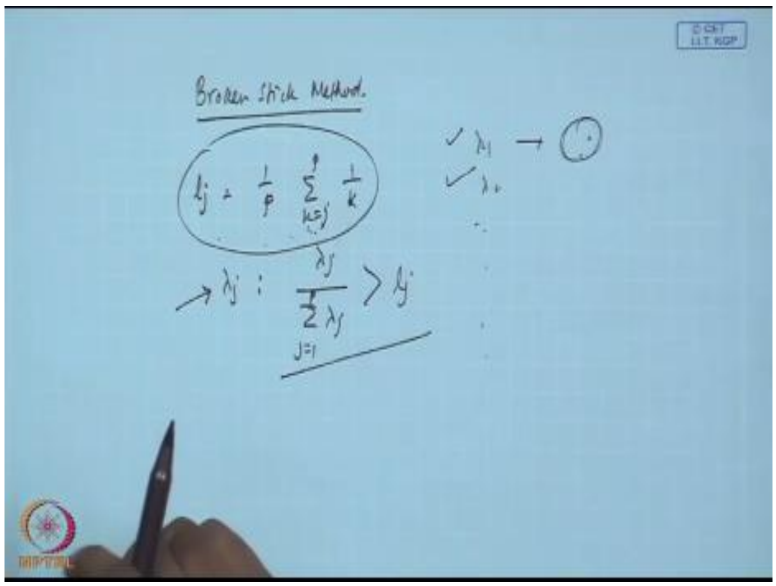
You use correlation matrix R matrix that is the correlation matrix for extracting PCs so your correlation matrix is like this all diagonal elements will be one end of diagonal values will be there and some values will be there then what is this stage of R if it is p x p matrix that will be p so if I use correlation matrix instead of covariance my total variability is number of variables correct now what is the each variable variability in terms of when he go for standardized variables.

So when you are using correlation matrix what actually you are doing instead of x_{ij} you are transforming this x_{ij} to x $_{ij}$ - x j bar by S_{jj} √ then if you compute x sorry S co variance matrix for this will be equal to R so standardized variable when you are taking for every standardized variable mean is 0 and standard deviation is 1 and variance is 1 so now in this case what will happen for every variable the variability of x_j that is 1 but standardization and giving that symbol so if any of the λ_j that is obtained from that correlation matrix value is less than one no need of keeping this because these component is not able to explain a single variable variability so that is why what it says that you keep those principle component whose λ_j λ value is greater than equal to 1 this is keep these many so there is another criteria i have one question regarding average root this one.

Speaker 2: Here we are arranging λ s in descending order.

Now when you that like when you extract your λ one is always greater then equal to λ_2 yes in any ways so we arranging Eigen values in descending order which is what it is obvious it is obvious because this λ one must be greater than the average λ it has to be it has to be so one component is sufficient according to that criteria no that is not the criteria you take all those components whose λ values is greater than λ bar not 1 1 not one it is not all the components have been Eigen value greater than the average Eigen value okay another criteria is broken stick method.

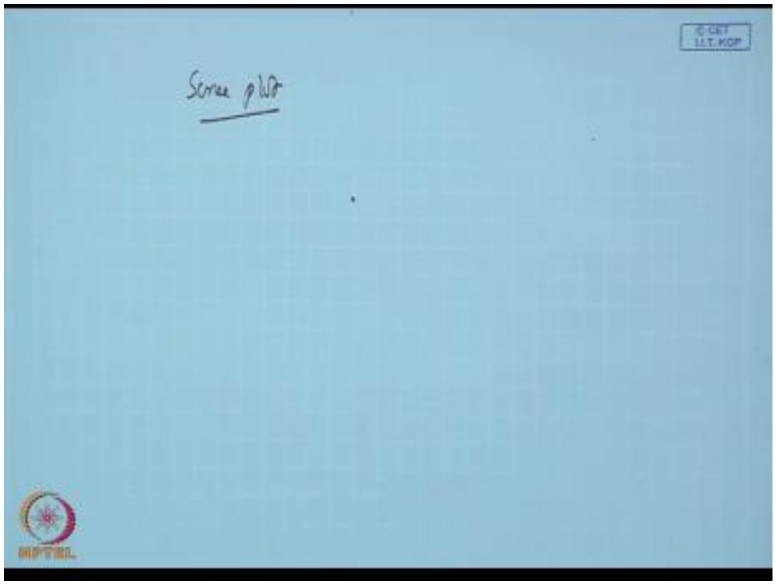
(Refer Slide Time: 45:23)



Here this combination is first to find out a quantity called l j which is 1 by p sum total k equal to j to p1 by k this quantity here it is just like a stick broken into several components that it is random randomly that all the components in any, any size you can get so that is why this type of 1 by k this things are coming now your λ_j the cumulative value for each λ_j you find out so that cumulative value is λ_j by sum total of λ_{jj} equal to 1 to p you keep those λ_j whose value is greater than l j so again the descending order is there in $\lambda_1\lambda_2$ like this.

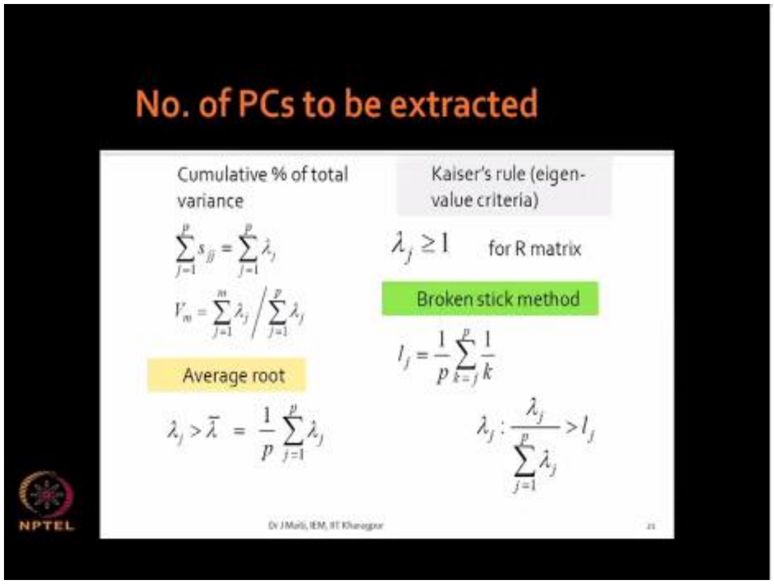
So you find out for the first one k equal to what found out this quantity is this quantity this quantity is less than the cumulative percentage explained here then keep this keep this when you are going forward at the end k is changing if it is first one is 1 second one is 2 so from 2 to p 1 by k that way you are taking a broken stick method so this is another criteria there is the popular criteria one more popular criteria is there that is now see the screen plot.

(Refer Slide Time: 47:13)

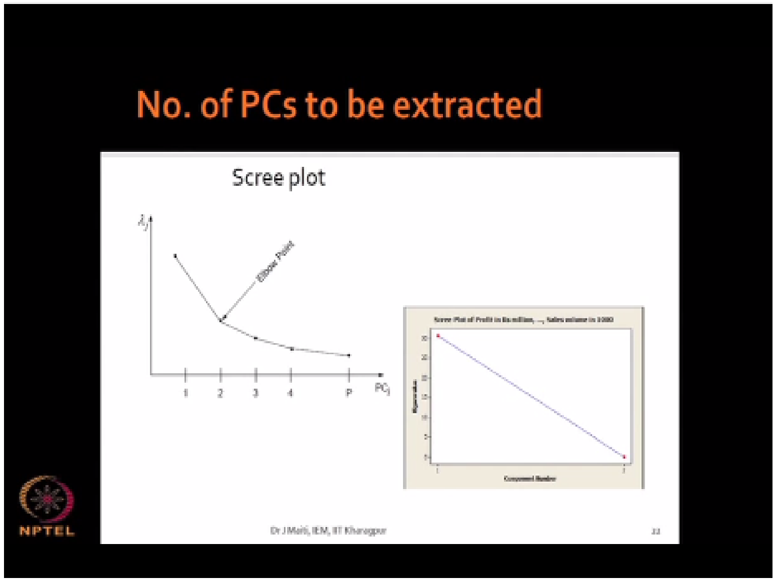


Now see the Scree plot here.

(Refer Slide Time: 47:15)



(Refer Slide Time: 47:15)



Scree is nothing but you just decide you keep the λ value this side the PC then put for every lambda values you put here and you will find out that ultimately it will create the carbon create in elbow type of shape elbow will be there elbow is the point if you take the normal posture the standard normal posture sitting posture in that case what will happen that whole arm will be and upper arm should make 90 degree angle perpendicular that is the thing so that means this elbow point you find out you find out the point where this type of 90 degree will not get here.

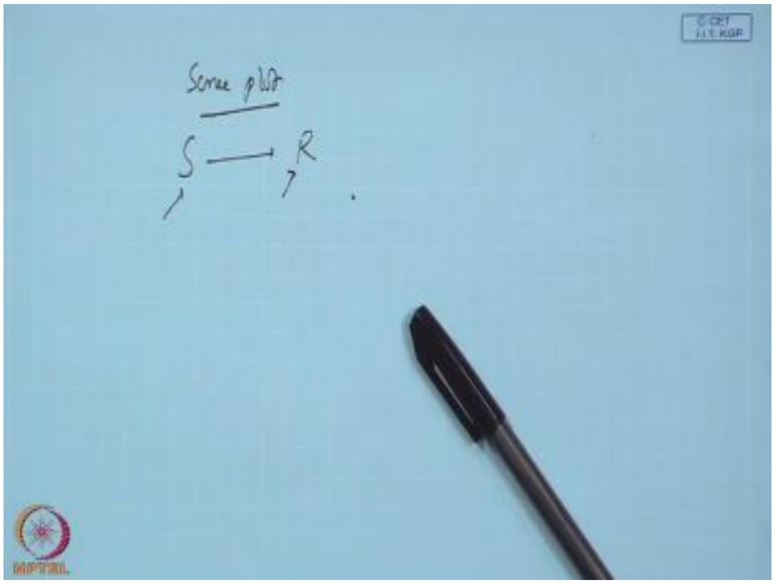
But if you get that this the best one so you are finding out where this elbow lies you take principle component number of component up to that elbow level this is the use alone the reason is if I say my elbow and then this is parallel all this is perpendicular parallel to horizontal my four arm that means other point here they are equally contributing there is no improvement in terms of addition of some other principle component then you have to add everyone all the things because they are parallel now in this in our case this is two variable case only.

So two principle component we have found out that only one component is enough we do not require more any question here up to this what we should study for all those terms which one you want to follow now how many so many things are there see ultimately what happens I can say that each are basically almost the same thing they are talking about one way or other thing it will difference now cumulative percentage we have to use that that is probably you cannot ignore that one because ultimately the original data sheet these are certain variability that co variance structure is there what you are doing using PCA that co variance structure you want to explain in some transformed dimensions in transformed variables.

So if I am not able to explain the majority of the variability present in the original data sheet then I think it is not a good model so cumulative percentage where that is the first one we have to look into and then you have to think what the total variability that you must explain the percentage of the total variability that you must explain you will find out that it is very difficult to explain in 50% of the variability now question is then if I require twenty variables to explain 90% variability now if I reduce it to five dimensions.

And able to explain 80% of variability then definitely you will go for the five dimensions with the deduction of 20 percent variability you are not able to explain the 25% variability at 90 to 80 that is 10% deduction is there correct that is possible second thing is then why all those things are there basically it goes for everyone then you see which one is favoring you but whatever you will do either broken stick or Kaiser Rule or your Scree plot finally you have to see that how much the total variability you are able to explain these are the Scree plot is basically mutual representation another important question is that whether we will go for S or R.

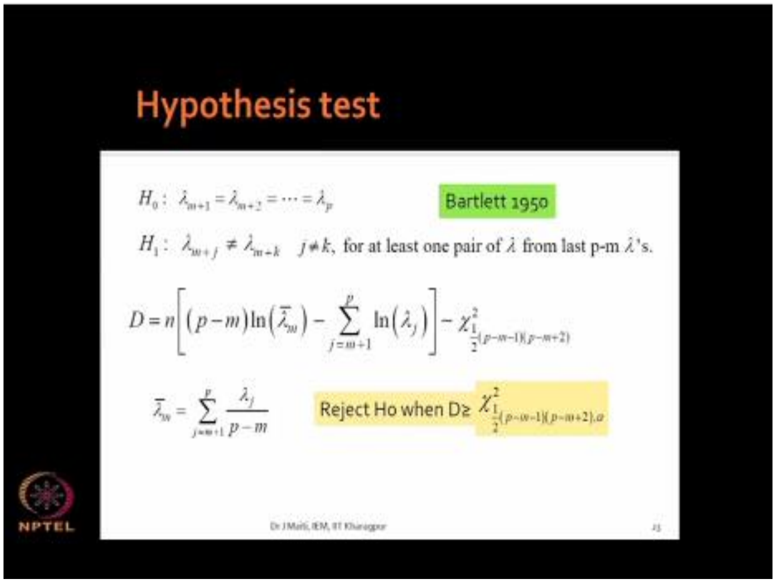
(Refer Slide Time: 51:52)



Which co variance which matrix you will explain see if you use S and if you use R that will get different type of different regions it is in the transformed this now it all depends on that what you want you want that suppose you are measuring the variables in different units the variability is wide under this case suppose your interest is there not on Scree the pattern of relationship co variance relationship then I can ask you to go for correlation matrix okay that mean whatever axis you give dimensions you give it is in terms of correlation matrix.

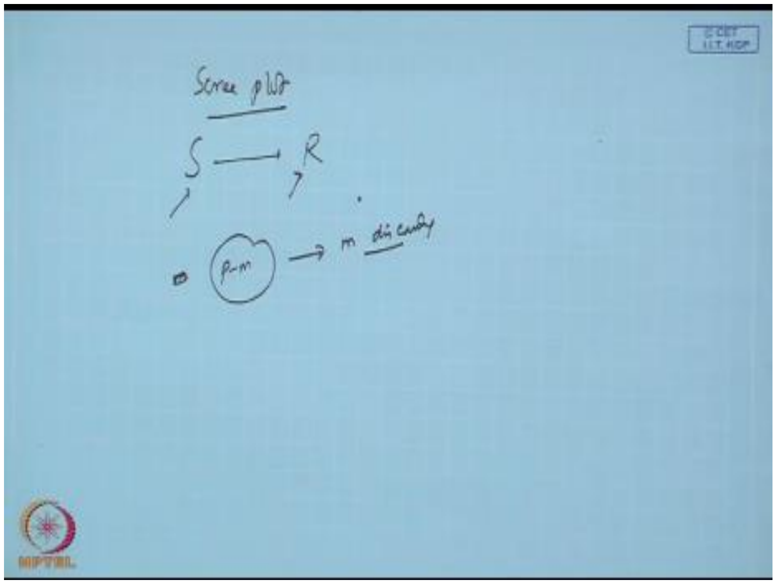
That vectors Eigen vectors will be created and there the pattern of relationship is more important not the strength if it is the strength is also equally important then you will go for co variance matrix because strength covariance is in the original domain not transforming when you are going for R you are going for the standardized variables so that you have to look into where you want to do.

(Refer Slide Time: 53:03)



Okay then final one is that as I told you the hypothesis test that Bartlett 1950 developed this one so suppose by these traditional method like Scree plot broken sticks and other things you are keeping out of the P components you are keeping m components.

(Refer Slide Time: 53:30)



Or let p - m components you are keeping that means m components you are discarding now my question is I want to test the m component collectively what this contributing in explaining the variability of the original data matrix significantly in that case our hypothesis is that λ m+ 1 equal to λ m + 2 equal to λm p means the Scree plot you make the Scree plot is horizontal that is what you are trying to test then you are saying they are not equal for at least one pair in the hypothesis the statistics is d equal to n p - m log λ m bar and where λ m bar is this one because the component you are discarding the average of that.

Then this quantity will follow high square distribution with this divisional freedom and then you know that chi square the table you have to follow find out this values then you go for a change accordingly you take your decision we should accept all λs whichever greater than λm bar greater than λ m bar but not that is straight forward here it is saying that if actual thing is that if they are equal then λ m bar will be individual λs.

So then using this but in sample you will not get exact symbol there will be little bit of difference so if I assume that λ suppose λ m is λj + one or m + 1 is greater than λ m bar and you will take this then you are going by point estimate what it is saying you just see that there is a difference or that difference is significant or not if that λm + 1 - λ m bar is significantly different then you keep otherwise you do not keep if say that can be discarded that is why the sample this distribution is very important and using this distribution.

You are first finding out the statistics then you are finding out the value of statistics then you know that what the chi square digit of freedom is then you go to table find out this value if you find that d value is greater than this then you reject null hypothesis h0 what is all the λ values here the discarded λ values are equal so these many tests are there and I think and we should nothing problematic these are all simple to calculate although very difficult to derive for example Bartlett's test and all this are difficult.

But very simple to calculate from the application point of view all those things when we come to this multivariate domain what will happen, ultimately that you will find out that a large number of statistics are used to test the same thing now what is the reason is all those statistics are based on certain assumptions now and there will there will be deviation from reality in raw data when you collect the data you will not find out that they are actually following the assumptions of the statistics of the models you are using.

So you will use several such statistics and see that all are favoring or some are majority favoring or not if majority favors to a 0 go for H0 otherwise do not go for H0 yes, Rahul any question no p sir so I think we will stop here and afternoon I will show you in that given data how to go for your multiple regression multivariate regression then principle component analysis and also we will see part very simple add data things will click business, things will be done even then once I show it will be easy for you to use the Software's thank you very much.

NPTEL Video Recording Team

NPTEL Web Editing Team

Technical Superintendents

Computer Technicians

A IIT Kharagpur Production