INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

NPTEL

NPTEL ONLINE CERTIFICATION COURSE

Marketing Research

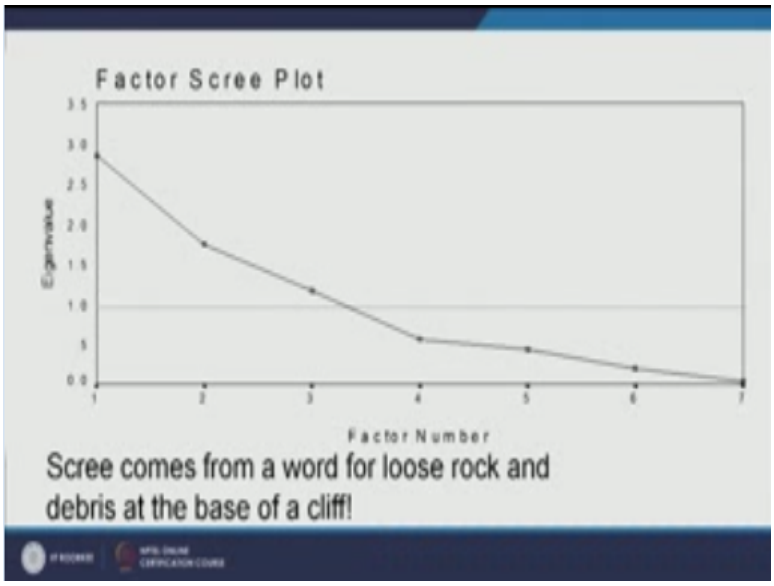
Lec -26
Factor Analysis

Dr. Jogendra Kumar Nayak
Department of Management Students
Indian Institute of Technology Roorkee

Welcome everyone to the class of marketing research and analysis in a previous section we are discussing about a factor analysis and we had started with the exploratory factor analysis right so we said there are two types of factor analysis basically largely speaking the exploratory factor analysis and the confirmatory factor analysis so we discussed the factor analysis is used when researcher wants to bring down a large number of variables into a few meaningful once.

So it is basically data reduction technique so data summary data reduction techniques which is used to bring in that you know better explanation into the process right so what accounts how do you how do we decided so we spoke about we discussed about how do you decide the number of factors right, so we said teat basically on 2 things one I called as creep low test which is something similar to like this so if you look at the bends at each bend.

(Refer Slide Time: 01:18)

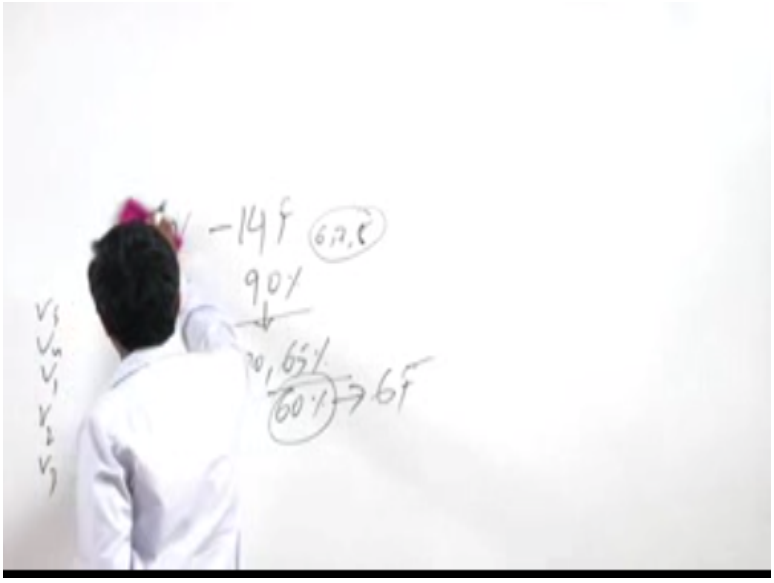


So each bend there is a factor right so let us say so this the 1st factor 2nd factor 3rd factor, 4th factor, 5, 6 and 7 but we can see that most of the slope of this point up to 4 the slope is better after than the slope is slowly decreasing so the researcher a to take a call this slope is basically the variants explained right.

So the 1st factor that explains the highest amount of variance the 2nd factor explains the second highest amount of variance it goes on right so the researcher as to decide okay at what cut of that does he want to stop the process for example let us say by utilizing 2 techniques the screen plot test and the carouser you know the Eigen value method we can decided up on the number of factors or the basically today the software's are built which are which you conduct that we do that exercise for you.

But what is the suppose let us say in case there are so this is possibility that suppose there are the study where 90% of variance has been explained right through the factor analysis and it has given you out so out of the 100 variables let us say 12 factors or 14 factors now 14 factors explaining let us say relate something like this 14 factors 100 variables where there okay.

(Refer Slide Time: 02:47)



Which came down to 14 factors right 100 variables where explaining 100% 14 variables are explaining let us say 90% right let us say 90% so the 10% is remaining with the rest if the items okay now the question is suppose I feel that or researcher feels that 14 is also to large in number so what you can do is you can fix the number of factors to let us say 6 or 7 or 8 or whatever is your number.

You can decide but when you reduce from 14 to 6, 7 or 8 what will happen is that the amount of variance explained by the factors we will also reduce with it right now may be 6 factors will explain only 70% or let us say 65% anything right so the cut of value for the amount of variance explained in a social science study is at least 60% so we say if through the factor analysis you can determine at least 60% of the variance in the study.

Then at adequate number of factors should be taken into account okay so if 66 factors if 6 factors 100 variables I am explaining the 00 variables but only 60% I am expanding it is still okay so that decision has to be taken by the researcher okay so second we went through this right now when I am talking about the factor analysis the inbuilt logic behind the factor analysis is the correlation right.

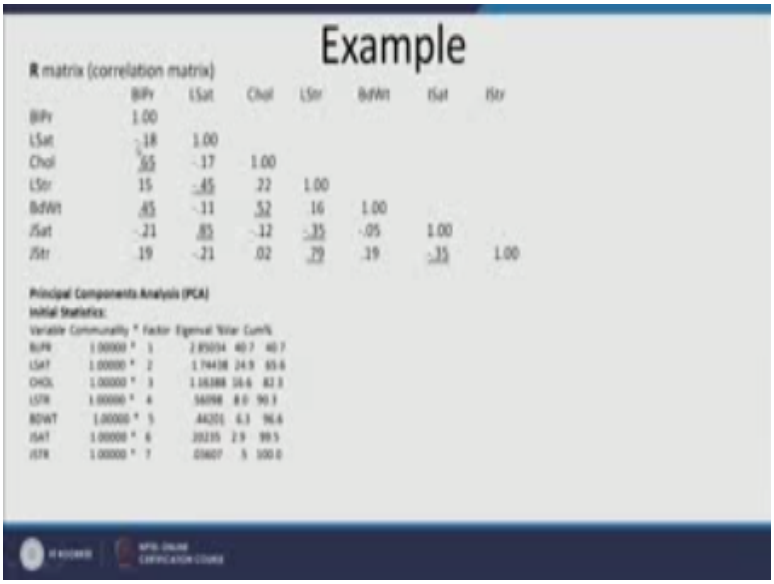
So why are saying the correlation because we are saying that the there as to be a relationship among the variables right so if there are there is a constant this is a factor the items within the factor the variables within the factor are some where related that means there is a correlation

among the variables right but 2 different factors should not be correlated the correlation between different factors two different f1 and f2.

For example should be poor why or the variables let us say that means what I am trying to I say let us say V_1 V_2 V_3 let us say V_4 V_5 let us say if I am doing this what I am saying is if factor 1 factor 2 suppose v5 let say is you know is explaining in factor 1 and also in factor 2 we are saying that the variable should be loading good into 1 factor and not into the other right this is one other basic assumptions that we make and it is desirable also because if it is loading on two different factors the same variable then we cannot theoretically say okay whether this factor whether this variable will actually be it should be considered into which factor right.

So to avoid such a case which of cross loading is as I show you I will discuss what we have to do okay, so now this is the correlation matrix.

(Refer Slide Time: 05:47)



Which tells about the different variables now this variables now suppose the first variable itself related with one right second is with the second one it is 0.18 – right okay now when it is a – value let say sometimes you find a very significant high value but it has a negative sign what does it means, it means that the variable is in the same factor there is no doubt about it because of it is high absolute value but you have to reverse core it now what is reversed score means now suppose the highest value.

In that particular variable was 6 that automatically suppose in this scale of 1 to 7 was let say 7 that should become 1 right so it has to be reverse course so that this negative value will be taking care of right this is one thing, so this is the initial statistics which says okay these are the variables and the commonality the Eigen value is given to you now let us look at how it is explaining the variance explained whether first factor if you see or the first you know is 40.7 then 24.9 it goes on right.

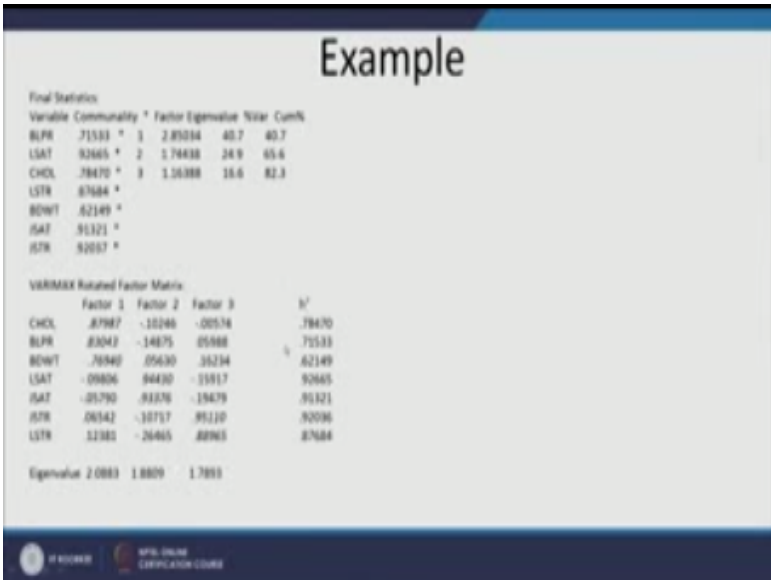
Now how much are be saying at what point should be stop the position is there right, now we say that whenever the variance explained by an item or a variable is less 10% that means when it will be a 105 variance will be less than 10% when the correlation of the variable will be let say above 0.3 is it not let us understand this what does it mean so if two variables if the r = r right r is the correlation.

(Refer Slide Time: 07:29)

If the r = 2 0 .3 and that means the variance which is the variance is equal to at least will say we will measure it through 0.3^2 right so we will say at least if we have a 0.32 that is 0.09 or 9% right so to get a variance of more than 10% at least the r value should be little higher right the loadings basically a loading has to be little higher than 3 so in the factor analysis the factor is loading that comes generally is those loadings are should be taken into account only which have a value of above at least 0.3 right

0.3 is also stays is not a very stringent value 0.3 let say 0.5 if you take it explains 25% 5.52 so to explain at least 50% of that into that factor a particular variable explain a 50% you need 0.7 that mans 7 7s are 49 72 okay, so one has to see that there has to be good amount of correlation right so this is basically.

(Refer Slide Time: 08:38)



What the you know the vales look like right so the factor 1 factor 2 factor 3 up to factor 7 and the Eigen value is explained Eigen value I said in the last session I explain that Eigen values are the squared loadings of the different variables on a particular factor right so if I take the variables loadings so the square the sum of the square across let say this is a f point for the movement just

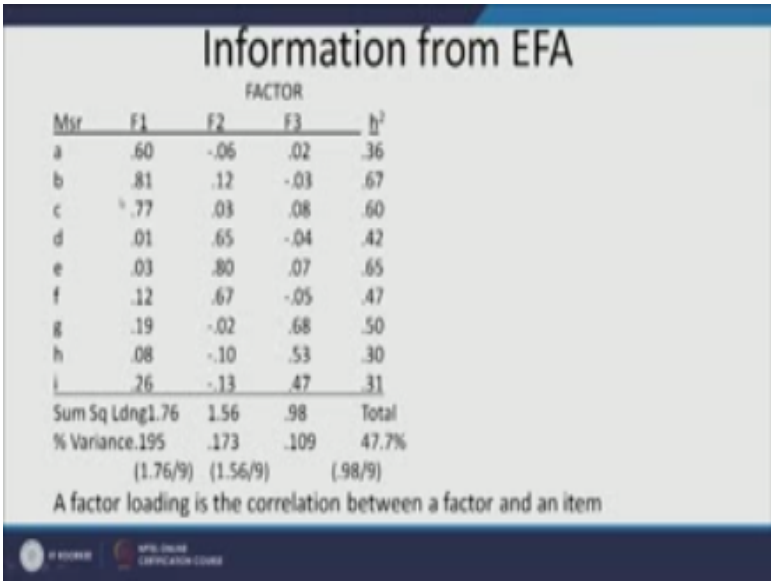
understand the sum is what we say is the Eigen values right, so now Eigen value and common relative also I had explained.

So commonality if suppose if the researcher finds that the commonality there is a cross loading or this is a problem is unable to understand what to do suppose in such conditions suppose a variable is cross loading find out the variable which has the lowest commonality right now the variable which ahs the because Eigen value has to be taken care of in the software itself or the when you decide the number of factors right through the latent square we said but suppose in your study you find there are some variables which give a very poor commonality of less than 0.5 let say 0.5 so if it is 0.5 less than 0.5.

Then it becomes an item for deletion so suppose you see cross loading that means a particular variable is cross loading into two factors or you see that there are certain variables are not being justifiably explained on basis of the theory in such conditions kindly look at the communality value and check okay what is the communality value and the item which has the lowest communality value is the item for deletion, right.

So you can remove that item and but please you have cross check it with the theory if it is theoretically highly justified to stay in the study then you should keep it, retain it but suppose it is not that strong theoretically then that particular variable should be an item for deletion because of its poor commonality value okay. Now this is how it looks like.

(Refer Slide Time: 10:55)



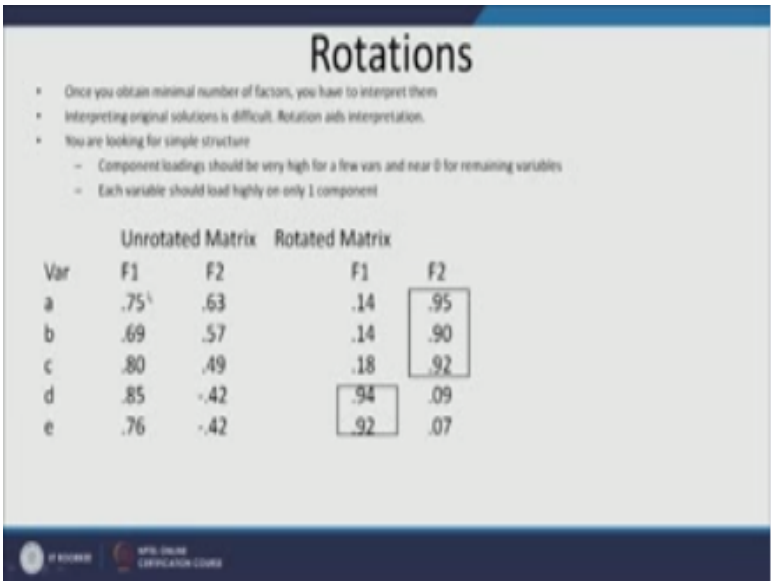
So F1, F2, F3 these are the, a, b, c, d, e, f, g, h, i all the different variables right, so coming up with three factors now you see a contributing 0.6 into the first factor 0.6 into F2 and 0.2 into F3 right, so that means it is a variable loaded with the first factor, b into first factor, c first factor again right, d is not in first factor but it is in the second factor because the highest loading is in the second factor, e second factor, third f second factor again.

Now g is in the third factor right, h third and I third right, so if you see now this nine variables have been distributed across the three factors evenly 3,3,3 right this is the hypothetical example and how much of communality order variance explained by the individual variables across the factors now the total is for the first one a is 0.36, 36% right, so second one is 67 which is a substantially good value, right.

C is 60, 42, 65, 47, 50, 30 and 31 so in case you feel, in case sometimes in not in this study but in certain studies you feel that a variable is not because of a particular variable the amount of variance explained in the study is not being proper or the variables are not fitting properly into the right factors in such conditions if you want to avoid or delete the situation the particular variable kindly go through the communality.

The one which has the lowest communality is the one which is the right candidate for deletion okay, so I have already explain this now this is how a un-rotated factor analysis a factor matrix looks like.

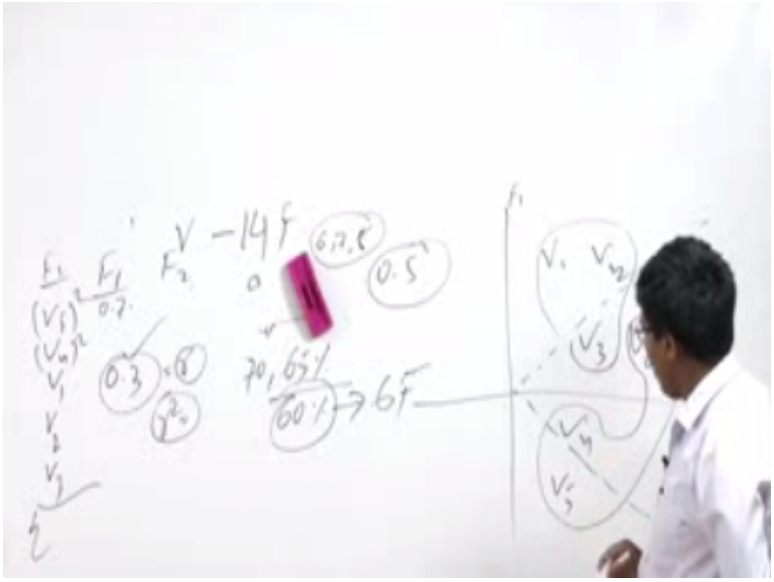
(Refer Slide Time: 12:48)



These are some five variables right, and into two factors so this two factors three all the five are let us say loaded here and here this two are loaded because if you see what is happened now 0.63, 0.57, 0.49, 0.42, 0.42 so these also do have a high value it is not key they have a poor value, but most of the items are loaded into the first factor only and this is the case when we do not rotate a factor this is why the rotation of a factor comes into play.

Now if you do not rotate a factor there is a possibility that most of the variables will load into the first or one or two factors right, to avoid that situation what we do is we rotate we rotate, so suppose this is v1, v2, let us say v3, v4, v5.

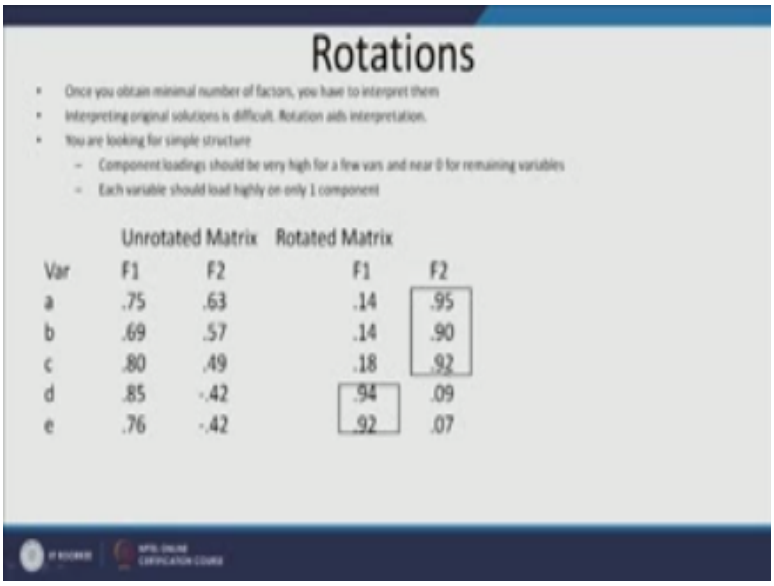
(Refer Slide Time: 13:42)



Let us say these are the v6 let us say like this, so if now I can say let us say this is my F1, this is my F2 okay, now that means what in F2 let us say these are ones which are coming let us say this two and maybe this two right, and F1 only these two are coming let us say. Now if I rotate this if I rotate this now what is happening is that possibly v1, v2 plus even v3 now comes into factor one right, and if v6 can come or otherwise v6 and v4 and 5 will come into the factor two.

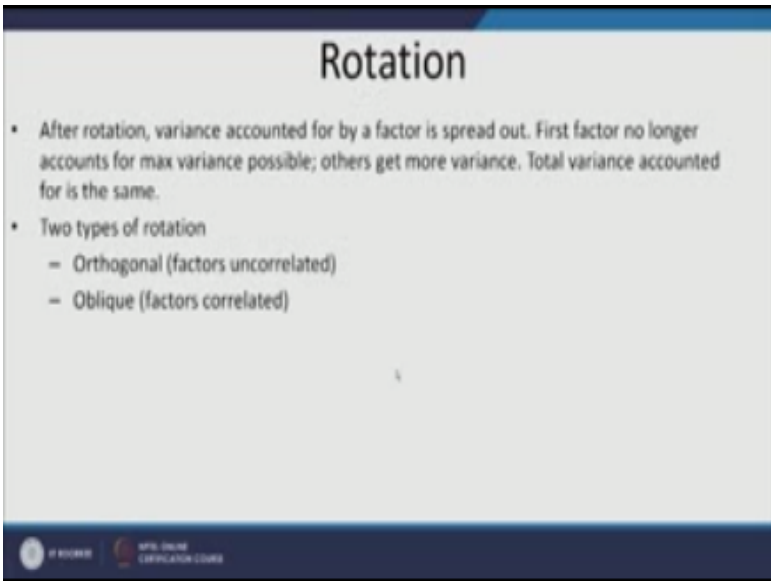
So this is an advantage that you get and the nothing changes actually when you do a rotation of the factor nothing chances as such the variance explained in the study everything remains the same, right.

(Refer Slide Time: 14:37)



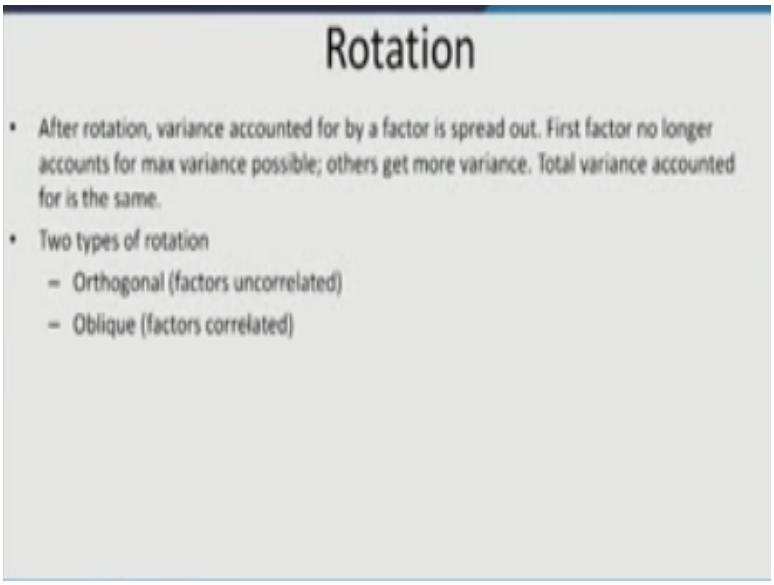
So it is you can see now 0.95, 0.92 in the second factor which was earlier not showing here right, and now and these two are only the coming the rotate in the first factor okay. So this clarity comes when you do a rotation of the factor.

(Refer Slide Time: 14:48)



So as I said there are two types of rotations the factor can be rotated orthogonally or there is an oblique rotation, orthogonal means is used when the factors are or the when we say that the factors are not correlated so you can see for example orthogonal factors are uncorrelated as it is written out here you can see this uncorrelated right.

(Refer Slide Time: 15:15)

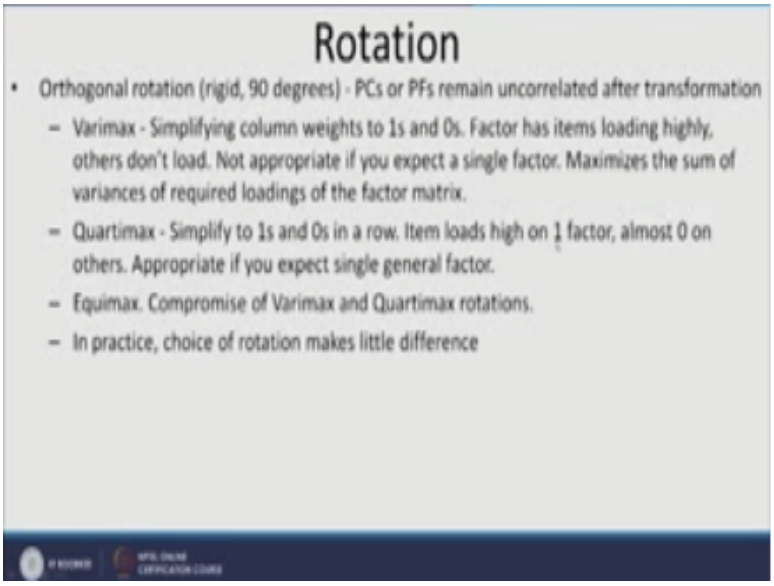


And oblique means factors are correlated although what happens in social science when we make a construct we say that each construct is sufficiently different from the other construct right there is a case of discriminate validity but the point is in social science whoever you may say there might some relationship between the factors for example now if I said trust satisfaction and let us say happiness are two different factors but easily our possible in the real life that satisfaction and actually happiness are not related it is difficult.

But the point is to understand or explain the construct well we assume that they are uncorrelated and that assumption what take in to the orthogonal rotation on the other case we have oblique rotation which says there is some correlation but although we say that and it is maybe practically true also but it is a complicated you know process of obtaining a oblique rotation right orthogonal just says they are uncorrelated and there is a 90^0 between them so it is uncorrelated so that becomes most simpler and it is more widely used okay.

But in case you feel while conducting a research that suppose there is a cross loading showing on two three variables on to two factors on two three factors then what you can do is you can go for orthogonal rotation and in case you do not find in any change in the cross loading patterns right that means the still the patterns of cross loading have been shown in the rotated factor also then in that case kindly try to see and use other rotational methods right in oblique also right that means there could be a correlation and may be if you use the oblique rotation method maybe the cross loading will disappear. So there are different methods of orthogonal rotation.

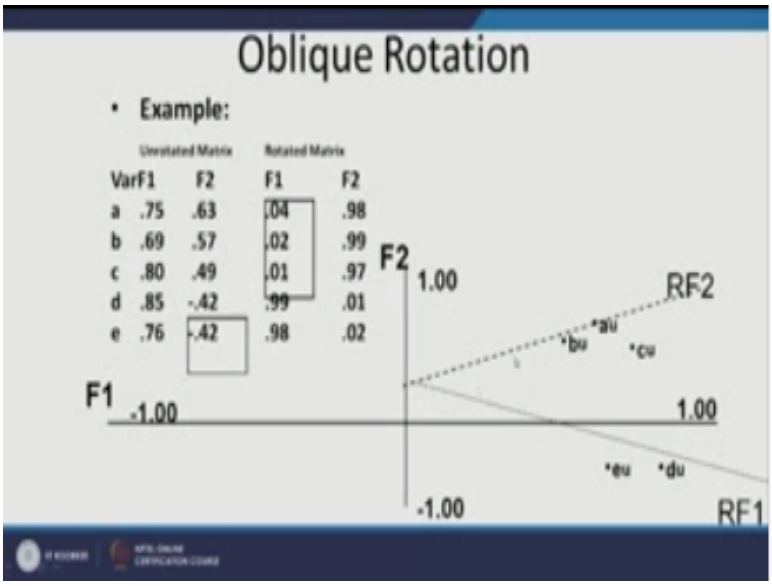
(Refer Slide Time: 17:18)



It is reject 90 0 as I said, so the uncorrelated right and out of this the varimax rotation if you can see is the one which is highly used right what it does is it maximize the sum of the variants of required loadings of the factor matrix right, it basically it is work in a column right the distribution of the weights across the columns other matrixes for example under orthogonal or the quartimax in the equimax which are seldom use which are not much utilizes because it works on basically the rolls right it simplifies to ones and 0 s in the row right item loads on one factor almost 0 on others.

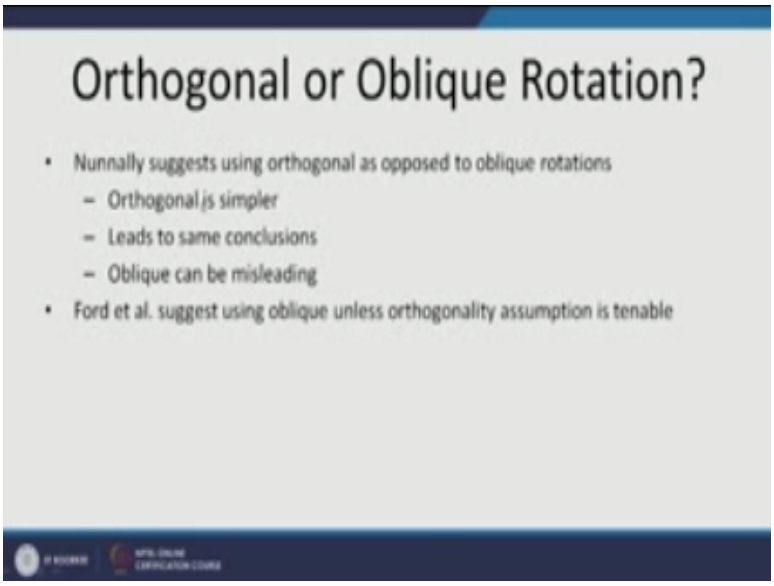
So this is generally less utilized in comparison to the varimax which is on a column basis right so and again as I said the oblique rotation is the one where it could be less or more than 900 that means there could be there is some correlation it is not 90 it is not you know this is not perpendicular to each other right. So these are some of the things that the researcher needs to understand, so it provides structure matrix of loadings and a pattern matrix of partial which to interpret now there is a question when you do a oblique rotation there are two things that will come now one is the structure matrix and a pattern matrix, now which one should you interpret we will see.

(Refer Slide Time: 18:43)



Now let us see a case now for example th9is is the orthogonal rotation which as we have done so if you see it is basically perpendicular as you can see and this two are at a 90 0 to each other okay 90 0 right on the other hand this is the oblique rotation where there is no you know the angle is not 90 so it is anything beyond less than 90 or more than 90 okay.

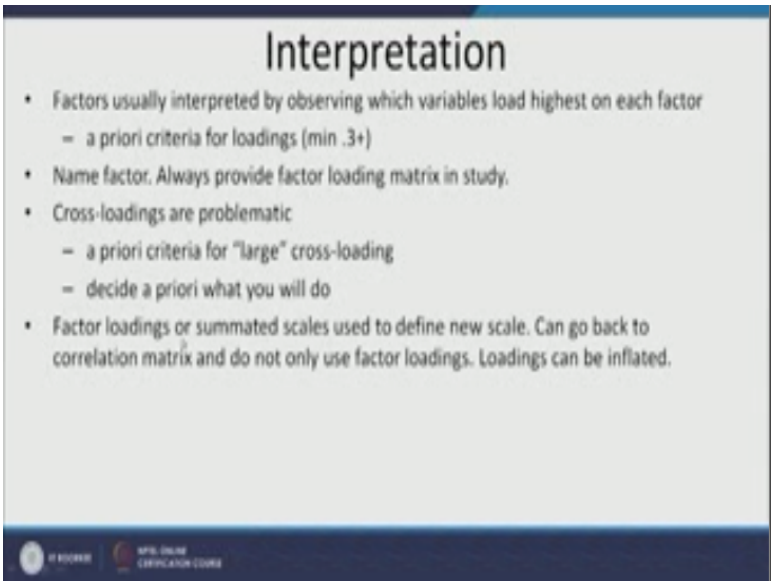
(Refer Slide Time: 19:10)



So which one now should you use so many researchers have said Nunnally being a very popular researcher, who has explain that orthogonal is more simpler although it might not be very practice to you know I have many a times used seen that oblique rotation gives better results but orthogonal is still simpler to understand and get a better explanation okay and oblique can be actually misleading sometimes yes it is possible.

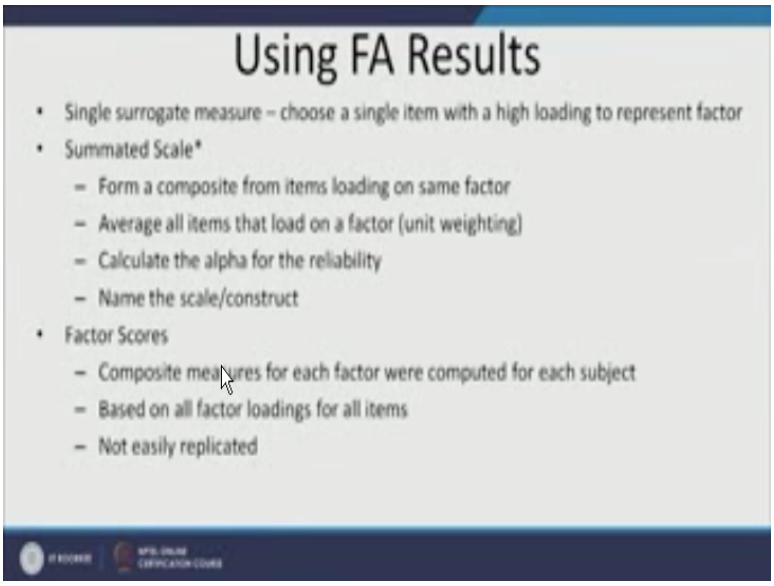
So many thing we have out of this some all things we have discussed already so cross loading are problematic so you need to understand what you do want to do as I told you if there is a cross loading that means one factor then is one variable loading in two factors then you need to somewhere first it is an item for delusion or else you can see whether theoretically that variable should go into which factor so that is completely the researchers you know propagating or he could just delete this by utilizing the different factor with rotational methods okay.

(Refer Slide Time: 20:23)



And finally two things which are important as a factor analysis loading for the some matrix scales which I have earlier explained in the last session also right okay.

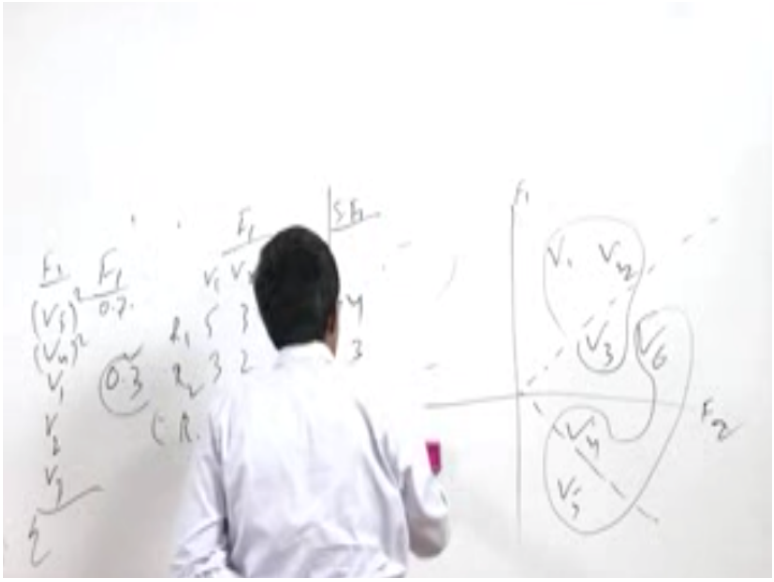
(Refer Slide Time: 20:47)



So how do you do this I am skipping this slides I am getting into directly the one oaky from here we will talk about the using FA results this is very important till now we have understood how to create the factors okay and how to make the factors basically right on what basis you make the factors now the question comes we said it is an interdependence techniques.

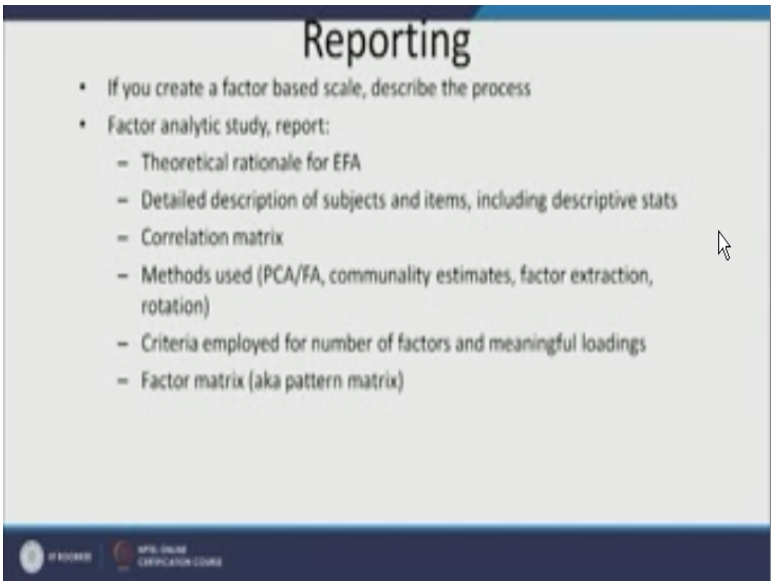
So the question is can I use this factors how will I use the factors what is the basis or what the methodology of using so in the last session also I told you two things one is called the factor score right now the other is called as summated scale right summated scale was the nothing but the average of the variables.

(Refer Slide Time: 21:35)



Across the particular respondents right for a particular factor that means what, what do I mean by this when I am saying I am saying that let us say let us say when I wanted to do summated scale let say this is factor 1 which has got V1 V2 V3 V4 and I said 5 3 4 4 so if I want to take a summated scale of factor 1 summated scale of factor 1 it will be like for example in case of respondent 1 it will be 5 3 8 4 2 6 4 so respondent 2 could be let say 3 2 3 4 7 9 12 2 so it goes on okay.

(Refer Slide Time: 22:25)



And this the value is highly efficient and highly utilizable important value to us okay so once you have this right once you get this factor summated scale and the factor scores then you can use this factor scores for your other study in the regression and other things right so I think now I have explained you enough to understand the explorative factor analysis the rational beyond the explorative factor analysis which is a correlation.

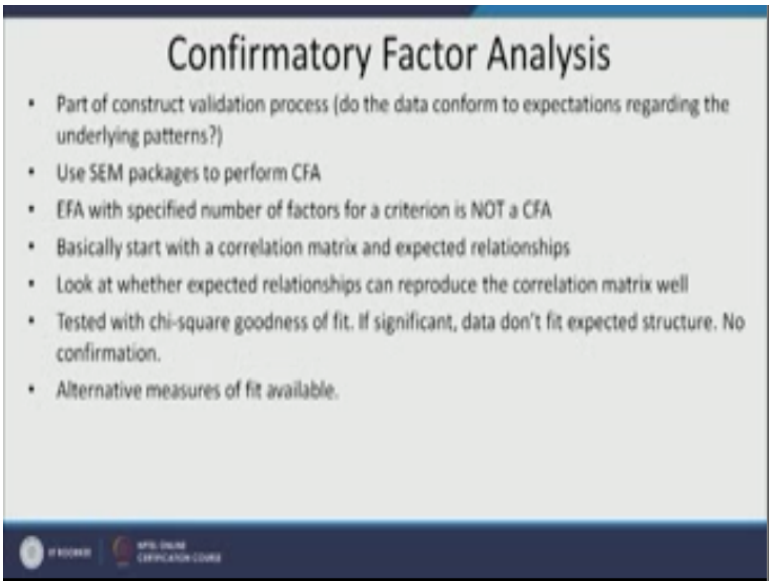
You know the correlation value right which is the basic underline backward you know the structure skeleton of this factor analysis is a regression model but an it uses a correlation value also to explain itself so the question is this is how we generated several factors out of a large number of possible variables right so now the question comes you can always go through this, this are later on right but the point is now what happens is there is a different kinds of factor analysis also which is very important right.

(Refer Slide Time: 23:34)

And which we need to understand now what this factor analysis is so in this factor analysis instead of exploring we will try to make it confirmatory study now what is the confirmatory study now we have been saying we as a word suggests confirm as I said right a and b so I know that say there are three variables V1 V2 V3 there are let say three variable here let say what I am doing is.

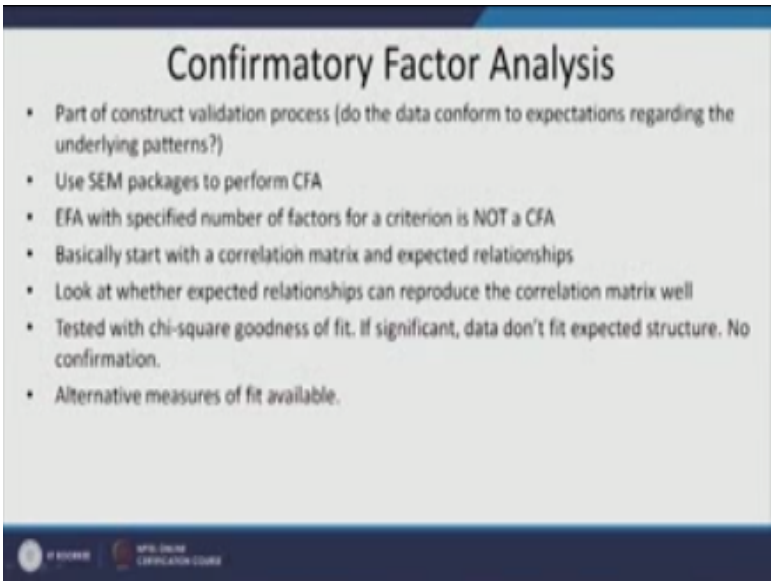
I am just changing like this six variables in total let say V6 okay and V3 V5 and V4 comes into let say now how did I know this, this has come out of theory right and there is a this is the basically co variance model that means both the variables are the constraints also variable right so if you have the constraints the summated scale of the making it as summated scale it becomes a typical variable right.

(Refer Slide Time: 24:43)



So the question is if I have this condition and I want to check the effect or I want to even test the hypothesis or something I can do it I can do it with the help of the confirmatory factor analysis now let us see what is the confirmatory analysis so it uses something called as a structurally questioning modeling now to understand hypothesis in order to test hypothesis we use something a confirmatory factor analysis method a special case of confirmatory factor analysis which is a SCM right.

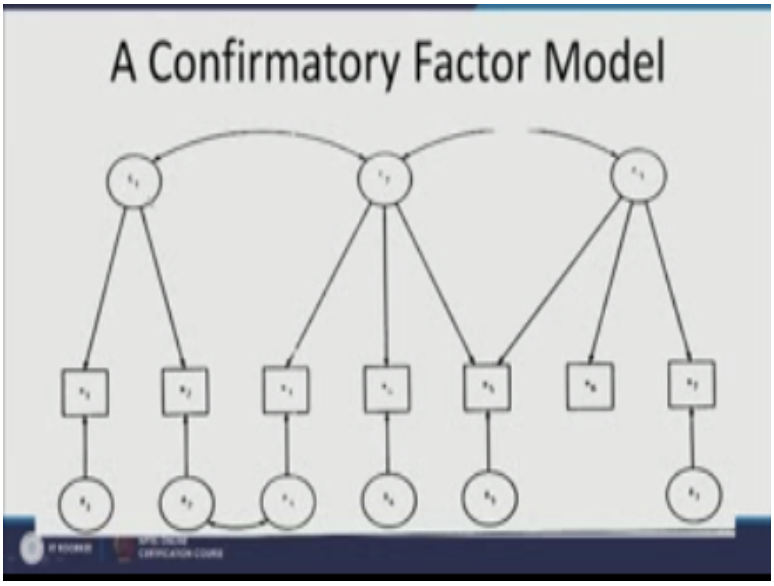
(Refer Slide Time: 25:04)



So what happens is it starts with you know it is like a construct validation process, so the construct needs to be validated and how do you validate the construct, now this construct is first of all are correct thing or not correct first you have to check that, to check that we have very different construct validation techniques for example we use convergent validity, discriminative validity, nomological validity right of this convergent.

And the discriminative are the popular one. SEM is used right, so it basically starts with the correlation matrix, as I said that the co variance is the correlation. In the earlier topic we discussed but a correlation cannot be a co variance why because correlation is the co variance basically. So that has to be remembered right.

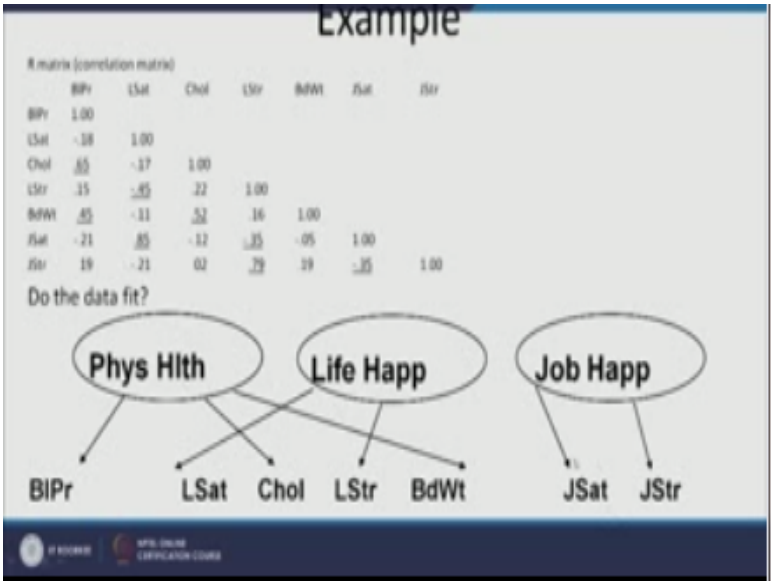
(Refer Slide Time: 25:55)



Now this how the confirmatory factor model looks like, so there is clear evidence x1 x2, x3 x4, x5 x6 x7, now we know that these relationships are very clear the researchers are aware of it and these are the basically the error terms, which is taken into the account, those errors are ignored, but in the case of the confirmatory factor analysis the errors are taken into the account and we also feel that there is a possible relationship among the errors.

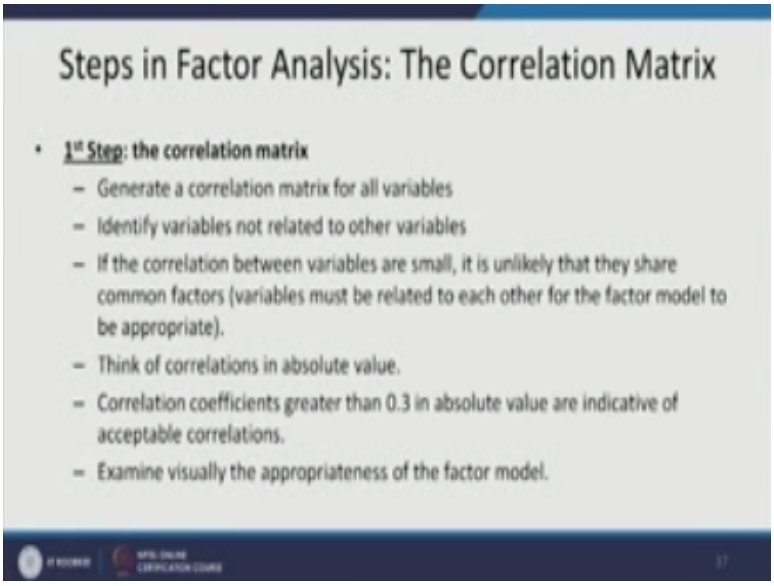
Why? There is nothing but the unexplained variances, so unexplained variances could have the relationship between the variances right. This is how this is the example right.

(Refer Slide Time: 26:44)



That this constructs have their own variables some are shared and some are unique okay. I would like for people let me just brief the confirmatory analysis and the explain infact what I have done is that, I have created the slide to tell you the steps, so here I just explain that part, there are few steps.

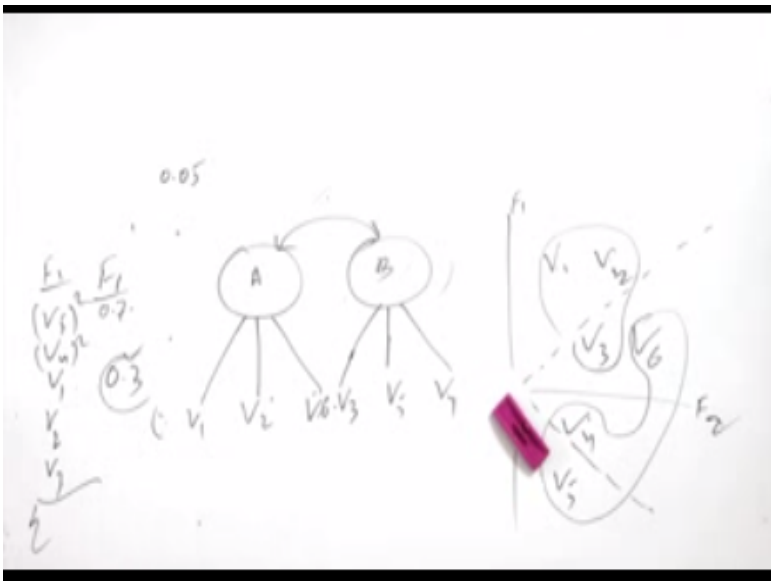
(Refer Slide Time: 27:19)



First step is to developed the correlation matrix any factor analysis.There is something called a Bartlett test of sphericity infact why I kept the slide I remember now because I thought you might I wanted to teaching the theory you should also remember the steps to be followed right, 1st is when you do a factor in SEM what you should do is basically the exploratory we are talking about. So what we dop basically is that we try to find the correlation matrix among them variables 2nd is to understand whether there is correlation existing at in the variables are not.

And that is done through a Bartlett test once which you have in the last session also, that there should be loadings that means, Bartlett test when you say.

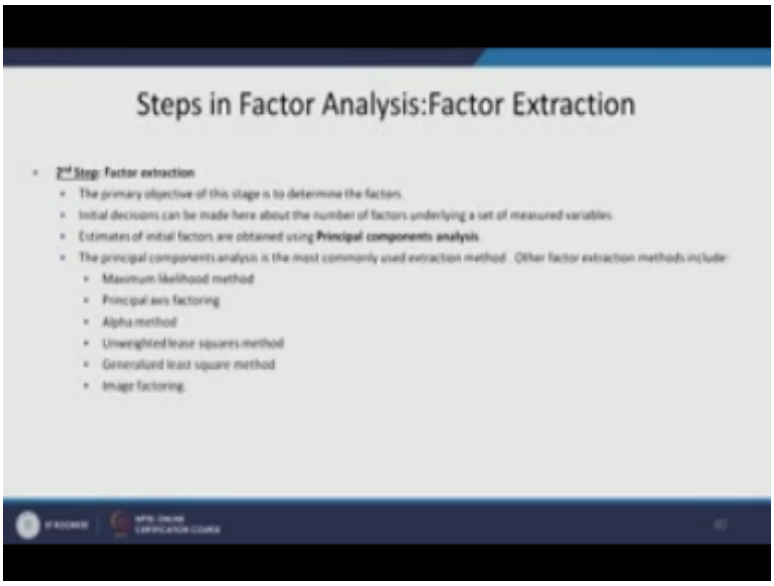
(Refer Slide Time: 28:09)



The value should be less than 0.05, that si if it is less than 0.05 that means the null hypothesis will be rejected; null hypothesis will be there is no relationship correlation between the variables. And there is something called the KMO is used, KMO value of 0.7b around we take, 0.5 is the cut of value, if we say that whether the number of the sample is using is justified is good enough to explain the factor or not, that can be known through the KMO study.

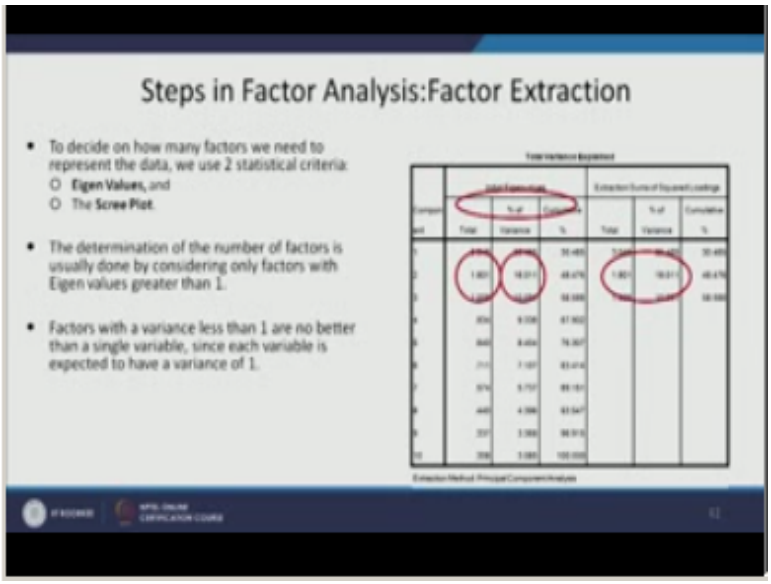
The KMO value you can see that, the KMO value basically tells that if it is about 0.5 that means there is the sampling adequacy that means the numbers of the samples used are good enough but a very high value like point 8 or point 7 is much better is a great right.

(Refer Slide Time: 29:06)



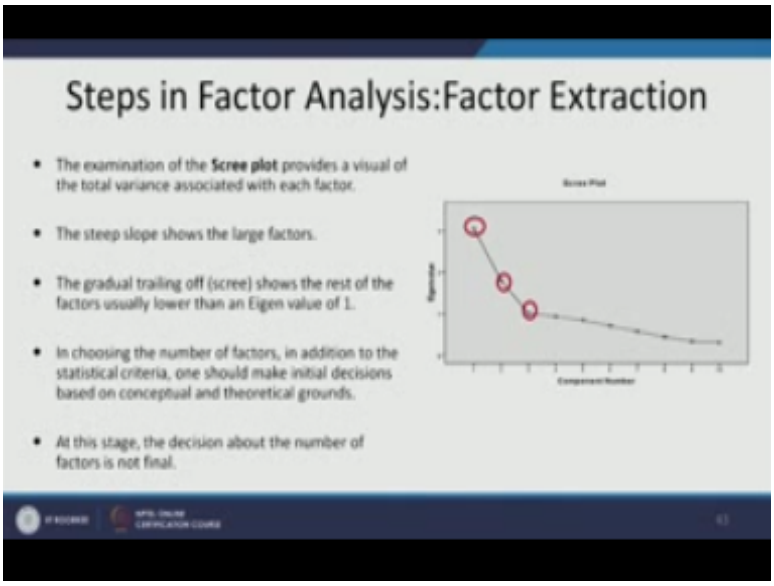
Third is the factor rotation extraction which I always said right I always have explained then we did you know.

(Refer Slide Time: 29:14)



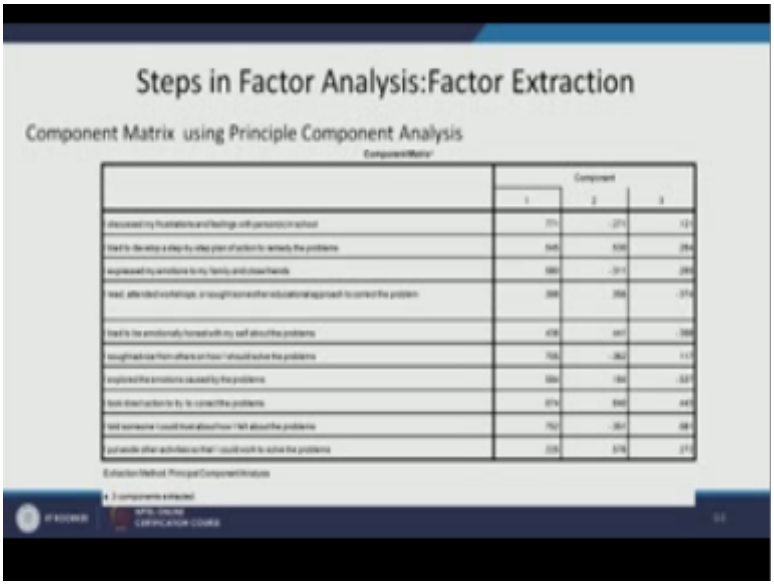
We explain the variance so this is how you understand the variance so the first 10 variables the three factors emerged and this is three factors have explained 58.56 okay.

(Refer Slide Time: 29:28)



And this is how also you find a graphical I explain that

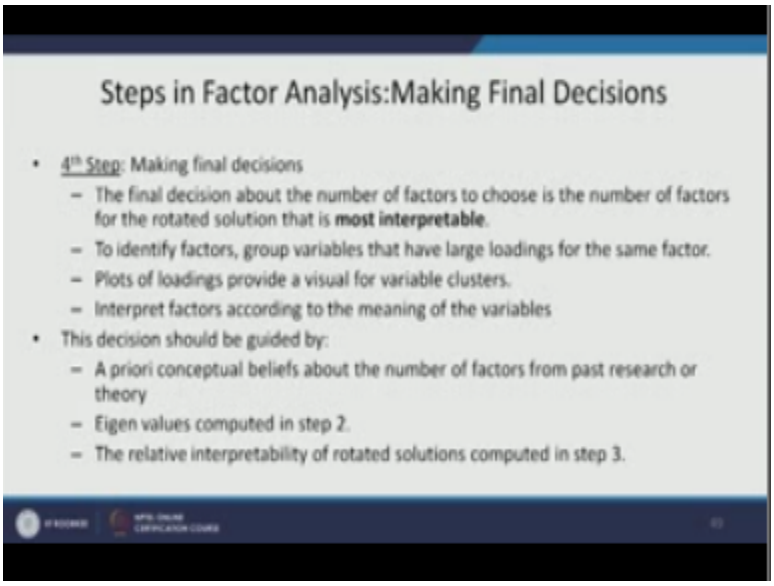
(Refer Slide Time: 29:32)



Okay this are the three factors you know this is a I think good to see one example this are the very for example you see I discussed my first exaction persons in schools I tried to develop step by step action so all this things all this are loaded into three different factors the first factor has got one, two, three right then how do you know this now you can just look at the values highest is here right in the three then this one is the second highest right but this is the sign of cross loading okay third is this one again fourth is this one again but again you see they are very close to each other right.

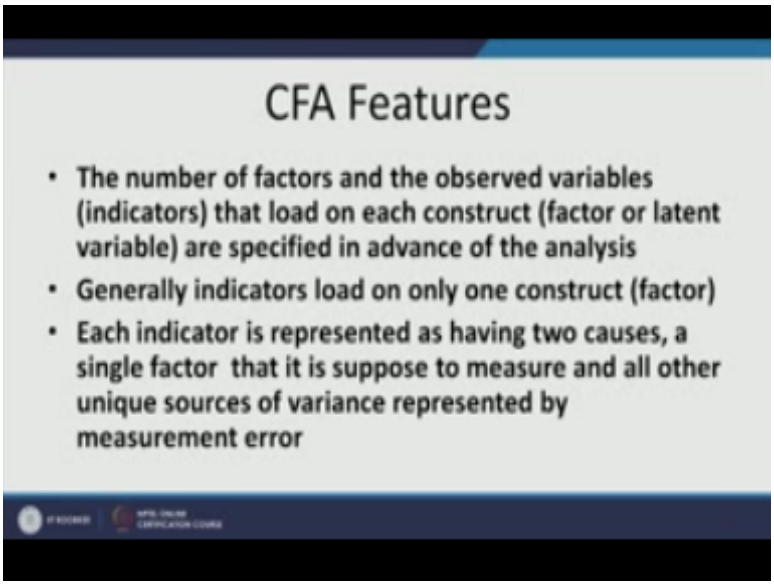
So this is how you decide which variable goes into which you know loading which factor so then you do the factor rotation and after factor rotation through orthogonal or whatever now the same result has been re distributed you can see you can compare it and you can find that the cross loading are minimized if you see in second for example now if you look at this there was a cross loading in the second case right now let us look into here now it is very clear this is the one it has now moved into the third you know factor okay so this is what it does

(Refer Slide Time: 30:49)



And fourth step is to make the you interrupt you name the factors right.

(Refer Slide Time: 30:53)



And you do it now this is what I had prepared but now what I will do is maybe I will just explained you today the exploded factor analysis only and I have just briefed what is the confirmatory and then again I showed you a table so this stable was to explain you the you know the how to conduct the factor analysis you can when you see this slides later on also you can go step by step and understand what exactly is to be done and how to utilize this factors later on for other purposes or other studies right which is of high importance okay.

So well this is for the session what I will do is in the next session may be I will carry on with the second path confirmatory factor analysis and structural equation modeling right so structural equation modeling is a special case of a confirmatory factor analyze only so it is the part of factor analysis so where we will try to test hypothesis right in a exploratory factors you were unable to do test hypothesis so in a or path analysis we will try to test a hypothesis also with the help of you know with the study right.

So that we will do in the next coming session and I think today what we have done is introduction or a beginning into the factor analysis and I am sure you are very clear with it and you can just go through and browse to the terms that I have explained and if there is something you can maybe we can talk about it so the point is please remember this theses are very important terms what a like for example I said submitted scale factor scores factor loading you know rotation orthogonal rotation oblige rotation.

Then all this values all this different concept and terms are very important and to one should be very clear with it and please remember one thing never ever always when you do a factor analysis it has to be on the variables and this variables are basically variables in a continuous scale right on a interval scale for that right so this is what we have for this session we will meet in the next session with confirmatory factor analyses thank you so much.

For Further Details Contact

Coordinator, Educational Technology Cell
Indian Institute of Technology Roorkee
Roorkee 247 667
E-Mail Etcellitrke@gmail.com etcell@itr.ernet.in
Website: www.itr.ac.in/centers/ETC. www.nptel.ac.in

Production Team
Sarath Koovery
Mohan Raj. S
Jithin. K
Pankaj saini
Graphics
Binoy. V.P

Camera
Arun. S

Online Editing
Arun.S
Video Editing
Arun.S

NPTEL Cooridinator
Prof B.K Gandhi

An Educational Technology Cell
IIT Roorkee Production