INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

NPTEL

NPTEL ONLINE CERTIFICATION COURSE

Marketing Research

Lec -25
Factor Analysis

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

Good morning everyone welcome to the class of marketing research and analysis till now have covered different aspects of marketing research the different tools which are unitized in conducting a marketing research study well this studies might not be this tools and techniques need to be used for marketing research they can be largely utilized for other research propose also right today what we are going to cover is another very important technique which is basically called as interdependence techniques.

(Refer Slide Time: 01:04)

So it is called as a interdependence technique okay now why it is called interdependence techniques basically the reason being very simple that in this case we do not have any dependent

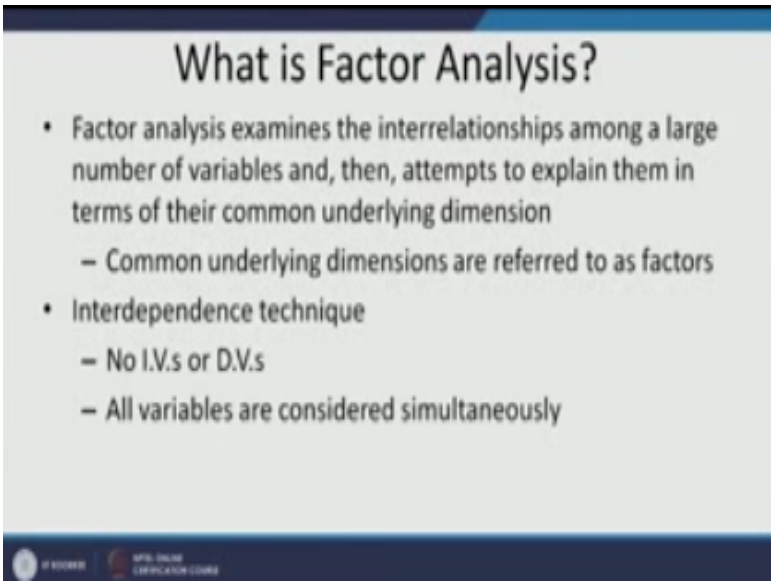
or independent variable right we do not have dependent or independent variable okay so what is the use of this technique why do we use it first of all and where it is us let me tell you this technique that I am going to describe or explain is one such technique which has been which is largely unitized heavily utilized and sometimes it is I can also say miss utilized people use researchers use it for different purposes we have understanding the very basic reason of they are doing it right.

So what is the static if we are taking about this technique is basically utilized to you know bring down large amount of data sets to a fewer meaningful once right that means what I am try to say for example let us say a company wants to know how do people by a certain product or what variables impact the you know the consumers right.

So supposes it has taken 100 variables let us assume right 100 variables now trying to analysis 100 variables and coming out with the meaningful you know explanation is a tough job because it is too tough to analysis 100 variables cross may be 500 or 1000 participants or 10000 participants whatever it is right so in such a case we need to have technique which can bring down this data to a fewer once to a very less number.

So that become simpler for the researcher to analyze and interpret and understand interpret that okay so this technique that we are talking about is basically called the factor analysis okay so factor analysis is nothing but a data summarization and a data reduction technique right it basically helps you in data summarizing the data and reducing the data okay, so let us see the factor analysis.

(Refer Slide Time: 03:20)



So it says it basically examines the interrelationships among a large number of variables as I said right there are 100 variables and you need to find out some meaningful you know meaningful meaning out of it so in such a condition if this 100 could be reduced to let us say only 6 or 7 or 10 maximum the we would assume okay it is much simpler to explain this 10 rather than the 100 okay.

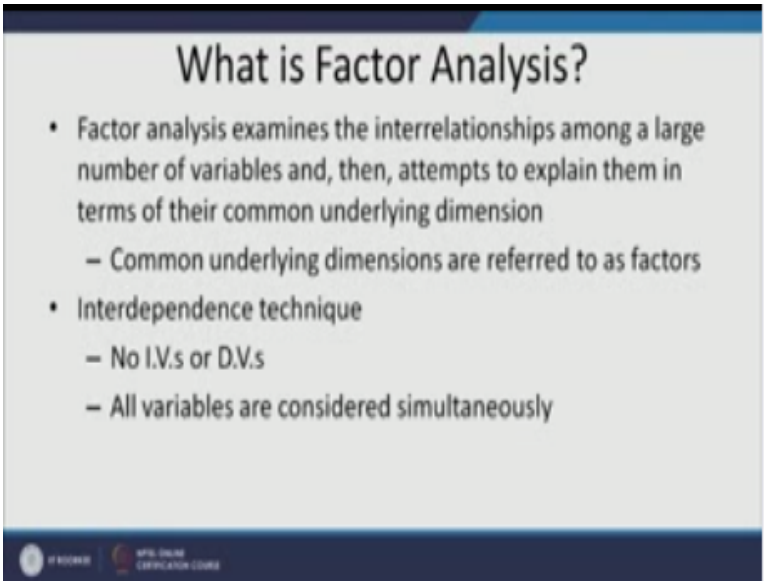
So what it does is basically it attempts to explain this 100 variables on bases of some common underlying dimension now what is this common underlying dimension you can understand is like some similarity some you know groups that could be formed right for example let us say there are the you know students who are who can be you know good in studies who can be good in sports how can be good in let us say culture.

Cultural activities so now everything that is related to somewhat related to even culture would be brought under one group that is called culture right and everything the students does basically may be his GP is scores GPA or his some other examination score or something right all these could be brought under another category called let us say academics okay and similarly suppose he as done anything in sports, in yoga in anything right relative health and mental health and spiritual health.

So in that we would say this can be brought under the category of let us say the sports okay so making those bring in those let us say large number of variables to 3 now form example is what

is the bases and intention of factor analysis has I said it is a interdependent technique and there are no independent.

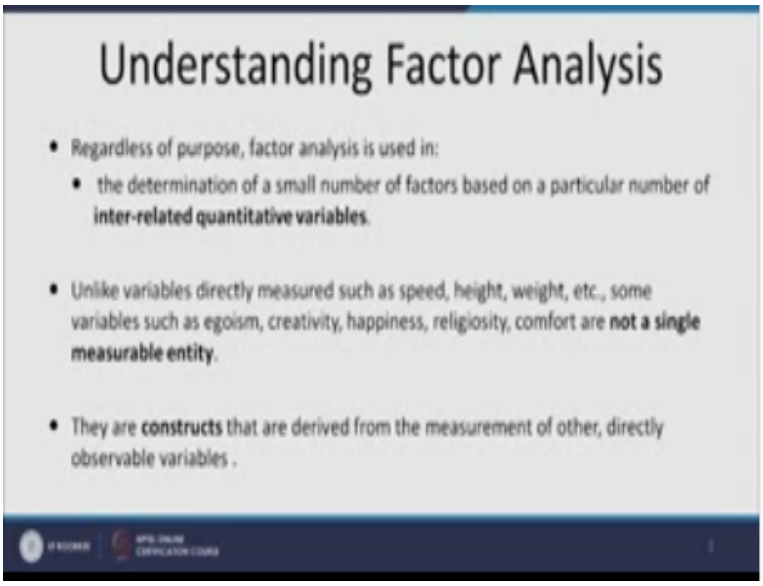
(Refer Slide Time: 05:06)



Or dependent variables earlier when we did regression analysis we said it is a causal model it is a cause and effect there is causal right, so in that we had a Y and X Y was the dependent and X was the independent variable so we said the whatever the change in why will happen is because of the change in X okay so that was something which was related there was a relation between the two variables but here we are not doing anything of that kind we are not doing relationship of dependent and independent.

However let m add to the you know understanding of the listener out here that the interpretations or the, the results that you derive from factor analysis can add the end maybe or can be utilized as a dependent and independent variable later on that means you can create dependent variables out of these out of this data summarization okay, I will explain that later so what it is saying is the.

(Refer Slide Time: 06:13)



Determines a small number of factors based on a particular number of inter related quantitative variables so first a fall please remember when you conduct a factor analysis factor analysis is to bring down all the variables together so that they can be you know some meaningful pattern can be brought out of it, here we are not taking any string variables or non quantitative variables right we even would like to avoid we are not doing any non metric right variables or let say categorical nominal data.

All kind of variables right suppose you are not interested in taking any demographic variables as such we are not interested because if you want to take a something if you want to do a factor analysis, factor is basically will be done and on continuous data that is data which is collected on a maybe an interval scale okay, interval scale so this is one thing another thing is that if you want if you want to do a factor analysis on let say a non metric data for that you have something called a Boolean algebra.

Or Boolean factor analysis which is not the part of our course and we are not doing it, so what is we saying basically so interrelated quantity variables first thing right, second it says if you see in social science what happens is to measure a particular you know particular concept let say we cannot may of times we are not able to measure it directly right, so what we do we measure it through indirectly through some other ways for example if I am interested in measuring let say honesty, honesty I may not be able to measure a honesty through a single item right because it is

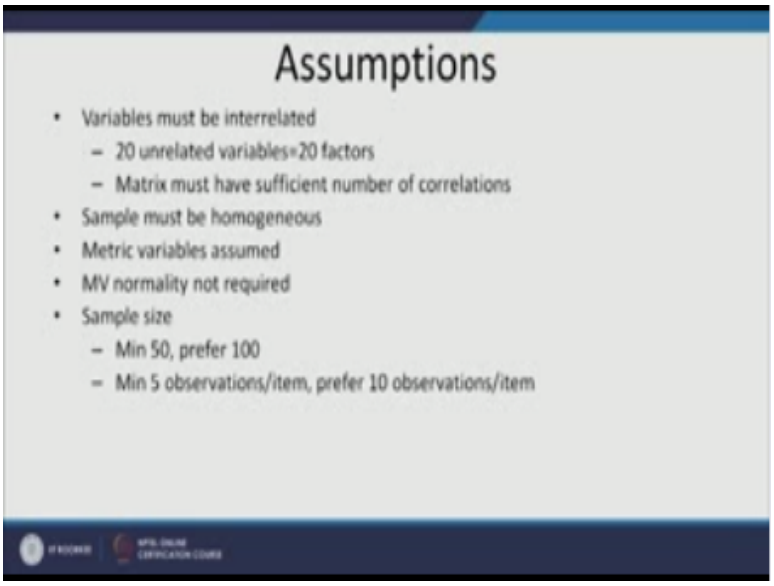
a it is kind of an abstract thought right, so in order to understand better what we do is we ask certain number of questions.

Or items to which are concerned about honesty okay so third thing they are saying is they construct that are derived from the measurement of other directly observable variables that means what are those observable variables now suppose in your question here suppose you have framed a survey instrument a question at in which you had 1 2 3 4 let say 10 questions okay, and this was related to honesty this was related to honesty this was related to trust this was related to satisfaction.

Maybe this was related to again honesty right this was related to satisfaction so now this these are actually a observable variables which are somewhere related to honesty that is why I have given the name honesty and when we bring this three together they come under the group honesty okay so why is it required for a marketer a marketer is requires it very largely because when we do the initially the study we take large number of variables in order to understand the respondents physiological profile but somewhere what so what happens is that in the context of doing the research we have taken large amount of variables and at the end we feel okay we have taken too much.

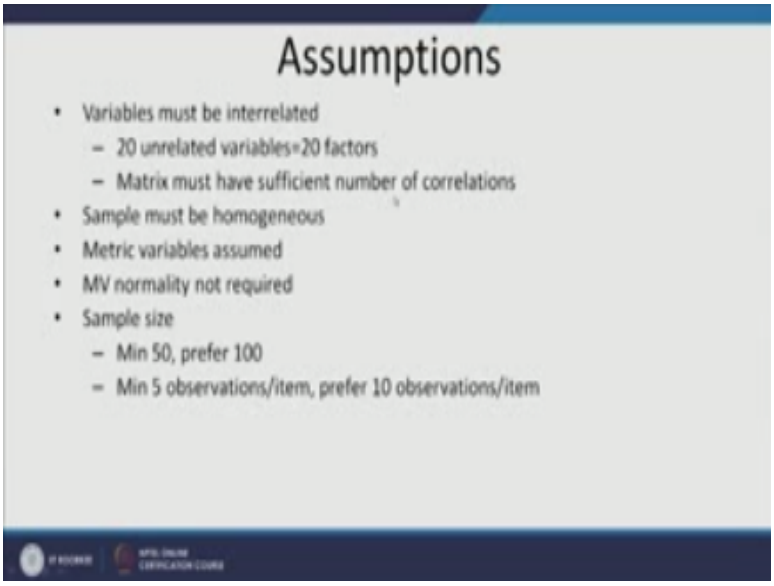
And by taking too much we are unable to actually reduce a or come to a proper inference okay what is a assumptions you have in factor analysis.

(Refer Slide Time: 09:47)



Basically the variables must be related that means what when you take a very when you take when you conduct a factor analysis there is assumption that the items within the factor there has to be some degree of correlation right there should be sufficient number of.

(Refer Slide Time: 10:06)

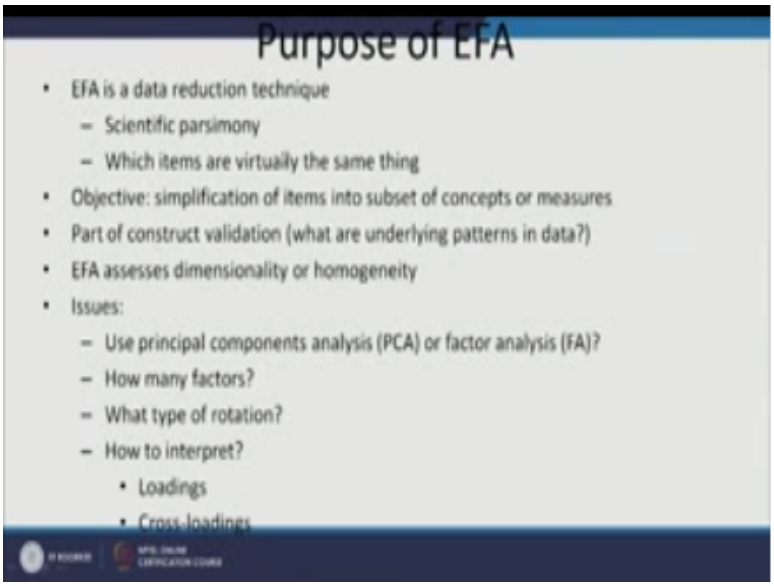


Sufficient number of correlations there is a even a test for which we do the battle test of sterility which we will test, we will test to we want it to be significant we will see that when I show you and most of the factor analysis studies when you see some software you know output also you see that battle test it says if it is significant that means there is some correlation among the variables that is the meaning of it, right. The variables are assumed to be metric as I said multi variant normality is not a condition that is important, right.

If it is not there it does not make of much of a difference in your study right, now what is the sample size to conduct a factor analysis the sample size should be around 100 at least right, allow 50 is there but 50 is a very small number right, that is the minimum amount but if you have anything less than 100 it is so wise to conduct a factor analysis and if you have anything above 100 it is a ideal number, right.

How I think I have explained okay, what should be the criteria of understanding the number of respondents or number of cases it is one variable or one item that you have taken in the study multiplied with an average of 10 respondents, so if you have 20 variables in your study that means your average should be respondents size should be at least 200, right. So minimum it says is 5 that is basically in a b to b sector where data is very difficult to obtain, so 5 is the minimum and maximum is up to 20 we say, 20 is the very ideal number so if you have 20 variables 400 that means, okay. But in between is 10 right, now what is the purpose as I said.

(Refer Slide Time: 11:54)



Is a data reduction technique right, so it is objective is to simplify the items into subsets of concepts or measures right, so it simplifies into creating subsets okay. It helps in validating the construct, the construct is the factor on say for example there, right so it helps to even validate so we will check how validation is also done with through discriminate validity convergent validity basically construct validity, there is a process, right. So issues, now what the two methods basically students are always they are interested to know okay.

(Refer Slide Time: 12:33)

What is the method very famously we have the principle component analysis, now what is this principle component analysis right, and against what I am talking about the two techniques which are used to derive factors in a study so one is the principle component analysis the other being the common factor analysis, right common factor analysis or just factorization also people say that, right.

Now what is this difference between the principle component and the common factor, the difference between the principle component and the common factor being 1 the point is here the total variance is taken into account to derive the factors, right so all the complete all variance is taken where basically we talk about complete variance means the unique variance right, and the you know the error variance right, so we have unique variance specific variance and error variance, share variance basically.

The unique specific or shared variance right, or common variance shared or common right, so this three variances together make the 100% so in during the principle component analysis there is no difference, they do not create any difference between the three and the total variance is taken into whole, but in the common factor analysis only those variance, those data are taken which share this common the variance commonly.

So like if you look at like a Venn's diagram so suppose this is the common area right, so this is the common variance so in a common factor analysis we talk about this variance right, and we are less bothered about the others, right. But there is a problem with you know the most utilized

is the principle component analysis and we highly sell them talk about the you know common factor very less, principle component is mostly utilized.

So the question is how many factors are the right, you know how many should the researcher derive suppose your 100 variables so how many factors should I derive out of this 100 variables 5,7,10 how many what is the right number we do not know, then comes a question when you create the factor right, sometimes what happens there are terms which I will be used slowly the factors sometimes you know the variables are loaded in to only one factor many a times that means when you create a factor analysis right.

You will see that most of the factors they are loaded most of the variables not factor variable are loaded on to the first factor. Now what does it mean now suppose let us say like this suppose I have ten factors ten variable sorry v3 v4 v5 v20 okay this is factor 1 factor 2 factor 3 okay three factors are there it might be possible that was the first six variable are loaded in to the first factor only right.

And only two of this are only one is loaded in to the second factor and two again the third factor so because of this kind of problems what happens although if we are purpose is only data redaction then no issues right but suppose you want to have a better pattern because it is so happens that it looks very strange that six variables are loaded in to the first factor and the other are not getting sometime it might not be even two only one so it looks very hard in those case what we do is we use something called a factor rotation so why by rotating the factor the distribution of the variables is made much better across the factor okay we will see that.

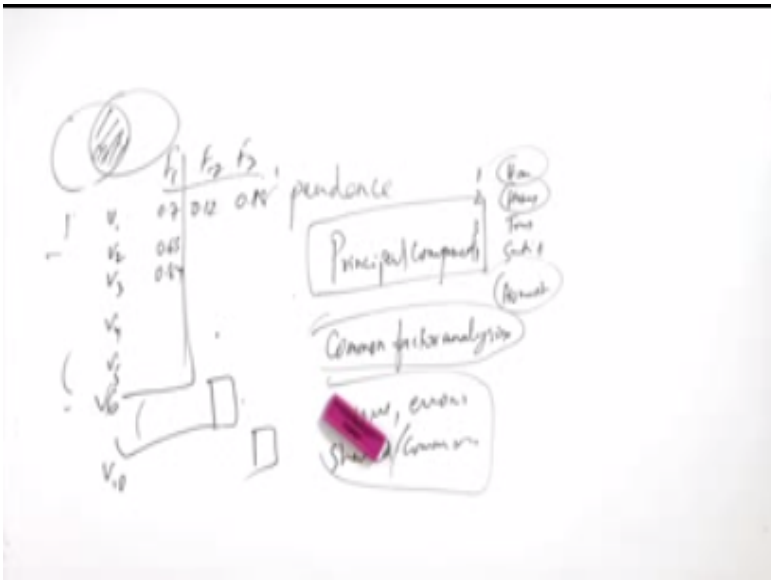
And finally how to intrepid now the tooth one thing is very important to understand loadings, now what is a loading? Loading is basically every variable loads in to the factor it has certain value let us say 0.7 let us say 0.65 okay 0.84 now what does it mean? It means that in simple terms if you want to understand that these loadings are nothing but the correlation of the variable with the factor.

So you will see that sometimes this is 0.7 this could be maybe 0.12 this could be the remaining or something like this right let us say 0.18 okay. Now similarly what happen is that means if a variable is loaded very high on to one factor it generally should be loaded less in to the other

factors okay that means it is very unique thing it is only for this factor it should not spreading across to other factors.

But we do face a problem in certain cases what are the problems, the problems are that sometimes we see that some variables show a high correlation between two factors factor one and factor two factor one and factor three factor two and factor three so this is something a problem right this problem is called a problem of cross loading now what should you do? So what should you do in this cross loading we will we see that.

(Refer Slide Time: 18:09)

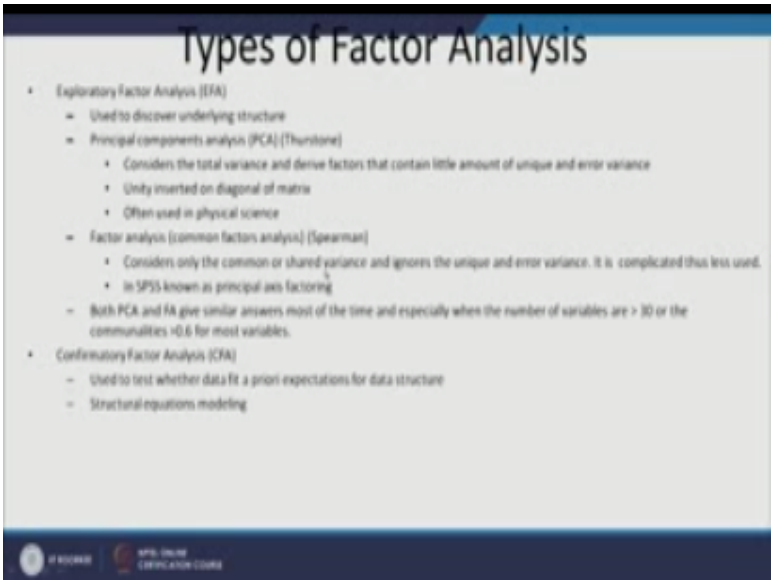


So if you look at now the as I said I started with I did not say one thing that factor analysis when I say there are two types of factor analysis the first is call the EFA and then there is another one called CFA right EFA being the exploratory factor analysis as a name suggested you are exploring so you are exploring the variables to come out with certain number of factors that you are not knowing at the beginning.

So after the study you after conducting the exploratory factor analysis you can come to you can get a knowledge well, five or six factors are coming maybe out of this 100 variables or ten right whatever. But in the case of it confirmatory factor that is a different story where already there is a theory behind it and already the factors you are only conforming the whether the factors are ideally or adequately explaining the you know the research study or not that means what in such

cases the researcher already knows what are the factors and how they are related he is only going to test them okay, cross checking basically you can say.

(Refer Slide Time: 19:24)



So what it does basically principle component analysis explains consider the total variants and derives factors contain little amount of unique and other variants. So it is takes the total variants right and often you within physical science on the other hand the factor analysis of the common factor analysis considers only the common or shade variants which I have drawn there right.

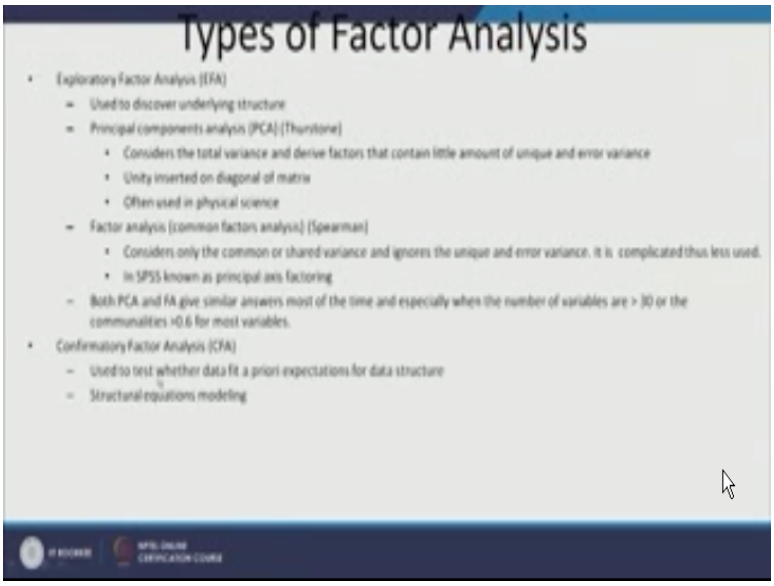
And ignores the unique and the error variants right unique or specific variants what we say it is complicated and thus less utilized that is why most of the time in the anybody ask you, you can always know you can always say what is the principle component analysis the principle component analysis covers the total entire variants and the common factor release on the other hand only takes the share variance okay so identify the share variance when there is large number of data said is difficult right.

So that is why it is and the beauty is both factor and both the you know principle component and the common factor analysis give us a similar result once you are if you see once you have the when the number of variables or the items in your case or greater than 30 right if your number of items that you are studying is more than 30 then the result that you drive from principle component analysis and a factor analysis common factor analysis is more or less the same right.

(Refer Slide Time: 20:57)

And one more thing if you have a commonality I will explain what is commonality, commonality is nothing but the shared variances the shared variances basically the shared valance that means commonality is the variables contribution to each you know factor right so the square of this value the sum of the square of this value is here is called commonality okay.

(Refer Slide Time: 21:36)

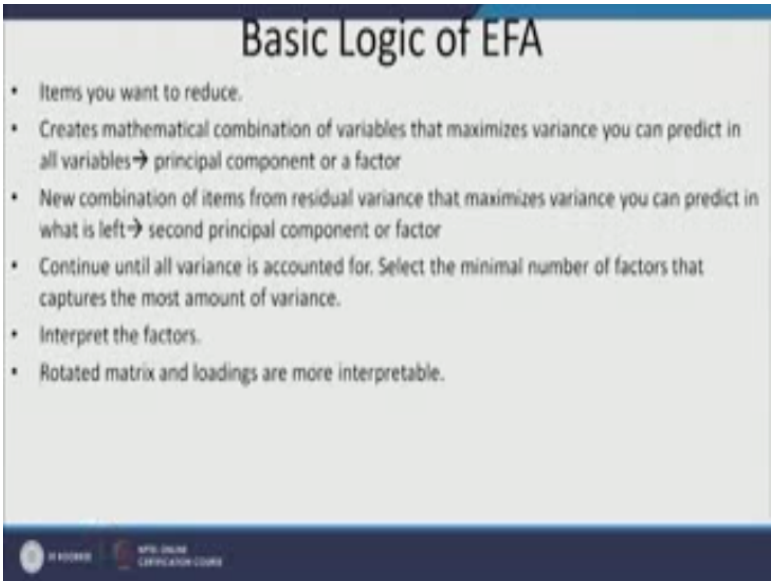


So this commonality if it is above .6 that means in almost all the cases if it is .6 then we see and if fa does not make a difference okay and as I said conformity factor analysis is used to test

whether data fit a prier expectations right that mains already the researcher has in mind a particular theory for example if he is let say two constructors a and b a and b have let say there are three variables V1 V2 V3 V4 V5 V6 right so if we there is a clear cut relationship which they are understanding right.

So and this is a covariance model so if, if they know already then it is a case of a confirmatory but suppose they would not have known then how it would have been it would have been something like this so all variables running into all other variables right so when you have all variables running into each other that is exploration case you do not know.

(Refer Slide Time: 22:46)



That is why exploring in other cases the confirmatory okay the basic logic it says when you it creates a mathematical combination of variables that maximize the variance, variance means as I explained earlier the way is explains the, the explained variance basically we were talking about whenever we say variance in this case we talking about explained variance in regression also if you remember.

We had talked about explained and unexplained variance right so more they explain variance the better there is researcher has conducted his study right his explanation is better so creates new mathematical combination variables that maximize the variances you can predict in all variables right.

A new combination of an items from residual variance that maximizes the variance and what is the means what in the first once it derives the first factor let say first factor will explain the highest amount of variance right let say the overall variance explained is 70 .7 then the first factor out of it may be explains 30 right and the residual 40 is divided among the other factors the second factor have explains the second highest variance.

The third factor explains the third highest variance goes on right continue until all variance is accounted for right all variances that explained variances so the minimum number of factors that captures the most amount of variance interpret the factors right so once you have got this factors right now the researcher needs to give a name to this factors.

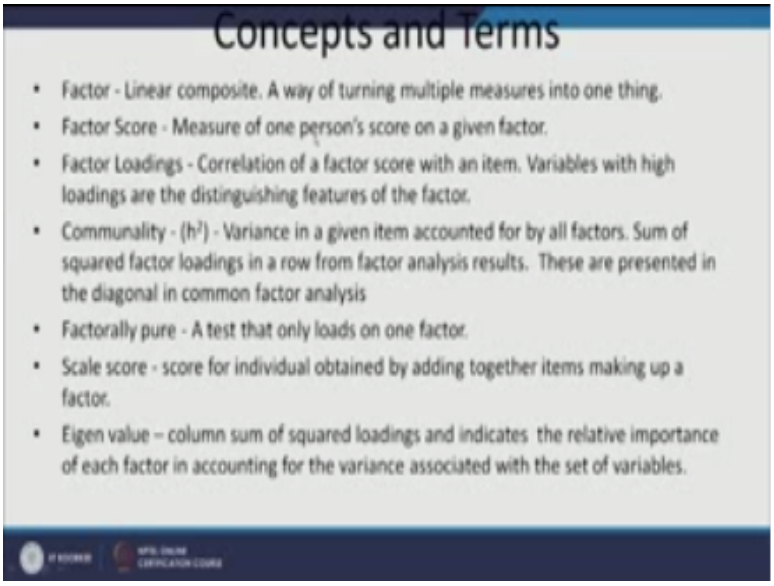
Now how will you give a name on what basis will you give a name the name will ebb given on basis of the similarity of the variables as I said there are time in the beginning the all the trades that are related with academia would be clapped into the group of academics all the trades are related to sports will be group under sports and the remaining right.

So this is basically what it has then interpret the factors, once you interpret the factors then some times as I said , now you have to rotate the factors, now rotating the factors I will explain again, there are two things, this is how it looks like. So understand it is like a car, it is like a car starring right, you are holding the staring, so if you can turn the axis.

So if I turn the axis let say, so this comes here and this automatically come here, perpendicularly through an orthogonal rotation or it might not be perpendicular this is called abler rotation. So if I rotate what is happening the variables will be better distributed, let say variables are like this right, so these variables distribution would be done in the better way and instead of falling into one factor only which happens in the rotational that will be distribution rotation okay.

So 3 things conception terms that you have to understand, so what is the factor it is the linear composite of the variables right, so multiplied with the weight, weight x independent variable, w1, x1 x w2x2 goes on all the variables together and factors score.

(Refer Slide Time: 26:30)



What is that person's opinion on the score on the given factor, what is the value or score which is given to the particular variable is called the factors score. Factors score are utilized heavily at the end of the study, these factors score can be utilized as the dependent variable is or independent variable or a ration study, we will see that. Factors loading I have already explained right=, community I have explained. What is the factorally pure?

Sometimes it test only loads only on one factor, so that means we have only single factor, it is good in some cases that means there are no other factors. So there is something called another term which is important for us to understand that is called scale score. Now what is the scale score? A scale score is basically nothing but it is the summated scale score, there are two scores that you can use.

One is the factor score, the factor score which comes eaxp0lantion which tells you about, a person, a responded or a case you know, how much values put on to the particular factor or how much importance, similarly we have something called the summated scale. Now summated scale will be largely utilized and it is the new development which is largely used.

Now what is hammered scale now let us say there is a factor 1 okay so factor 1 was nothing but combination of factor v1+v2+v3 let us say v4 now submitted scale says suppose this is respondent 1 respondent 2 it goes on right so whatever score he has given for variable let us say

in the scale of 1 to 7 may be right he has given 5 for this he has given let us say 3 for this he has given let us say 4 for this he has given again let us say 4 right.

So the submitted scale submitted value will be nothing but the average so it is divided by 4 so that is 4 for this respondent similarly for respondent 2 for respondent n 100, 200, whatever so you this submitted scale is highly utilized is a very important tool because later on you can use those factors as an independent variable or the dependent variable for a different kind of study for a caution effect study right so that is where it comes off great use right one more thing is when I am saying factor score I have submitted scale I have explain then there is also something called I Eigen value.

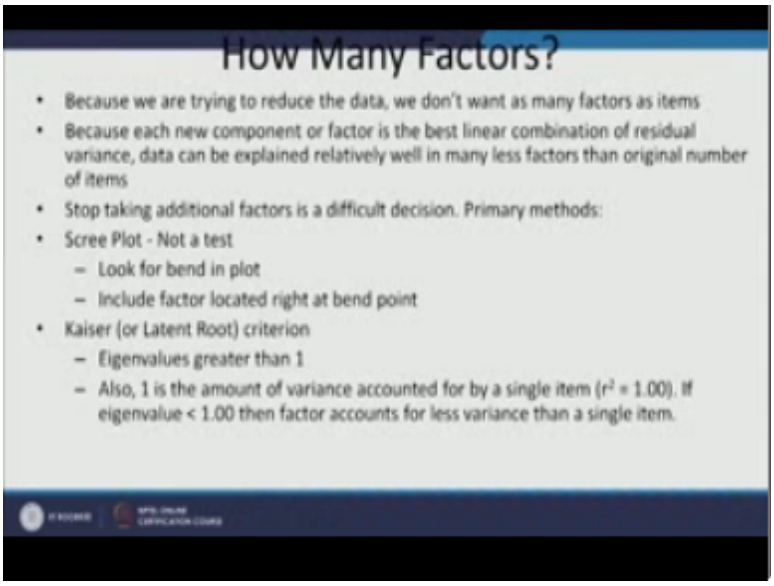
Now what is an Eigen value this is also very important for you to understand now as I said I explain the commonality right I explain the communality there is something also called an Eigen value now Eigen value is a vertical score right so it is the how you know variables are loading into particular factor so this squared the sum of the squared loadings across the factor this total is called the Eigen value right so the Eigen value is one of the ways which is used to extra factors right in a factor analysis study so Eigen values if it is ;less than 1 we generally omit right we avoid any factor analysis study which obtains the Eigen value of less than 1 because that means it is not explaining the item is not explaining itself right.

So as good as that so Eigen value above 1 is at least that means that the variables are factor is explaining itself as good as that okay so we will see how many factors how do you interpret how does the researcher understand how many factors to be taken right okay what I will do is we will just I will tell you brief you about the way of identifying the factors which I were just saying one of the method is through a graphical method which is called as plot method right like there is a twist you know the bend in the arm so what is the plot I will just show you a screen plot is basically nothing but you know

(Refer Slide Time: 31:04)

This is how the data plot changes for example let say this is something like suppose this is the data point so this is first second curve third curve fourth then it is may be so when you see such you know such kind of the arrangement of data then we say key well there are four curvature four points where there are curves the curve is bending right so we will say there are 4 factors so out of all the variables v1 to v end so we are saying there are 4 factors coming out so this is the method which is used graphically called plot test okay the second is through the later in root criterion Eigen values are Latin roots if you see you would not get confused is the same thing right so it says

(Refer Slide Time: 31:53)



Eigen value greater 1 is used I have just explained y1 right that means it explains itself at least so Eigen value is greater than 1 is taken as a criteria to generate the number of factors so one is the amount of variances accounted for by a single item one is the amount of variance accounted for single item so if Eigen value is less than 1 then factors account for less variance the factor is explaining less variances than the single item.

So I item is I factor and if the Eigen value is less than one that means it is not even explaining a single item well what we will do is we will continue this same session in the same thing same factor analysis in the next session now we will take a break here 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 Coordinator
Prof B.K Gandhi

An Educational Technology Cell
IIT Roorkee Production