Marketing Research and Analysis -II (Application Oriented)

Prof. Jogendra Kumar Nayak
Department of Management Studies
Indian Institute of Technology - Roorkee

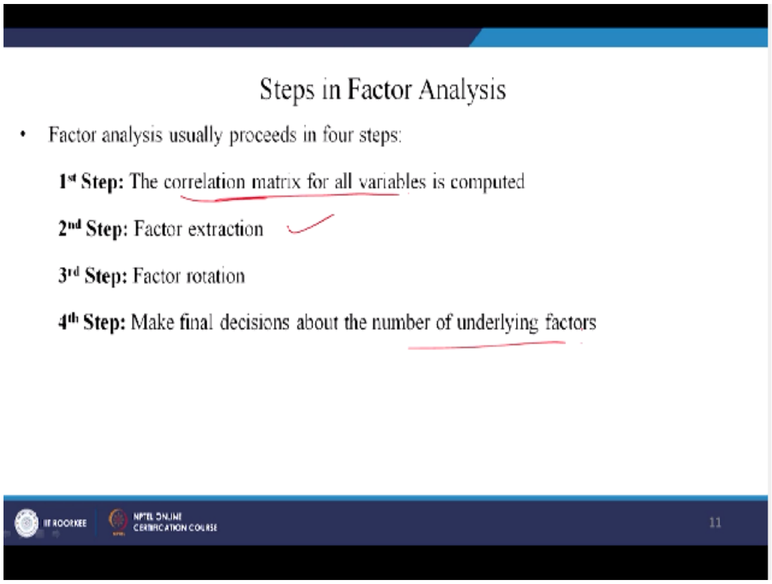
Lecture – 52
Factor Analysis in SPSS - II

Welcome friends to the class of marketing research and analysis. So we are continuing from the last class where we had left. In the last lecture, we had discussed about factor analysis. So factor analysis is a technique where which we have understood that it is a data reduction technique which is used highly all over, at least in the marketing space, to identify the customer segments, to understand the choice of customers, to make a right pricing system.

So in psychology also, you can see there are lot of utility of factor analysis. As I had given an example that for example, psychologist wants to break 60 to 70 traits of human beings. Traits means the characteristics, personal hereditary traits. Now it wants to, it cannot understand or explain 60 traits.

That is too much, right. So instead of having 60, if I can have 5 to 6 or 7 or 8 even, right, factors which can represent those 60 traits, then I think it is much useful. It is much better. So the whole purpose of doing factor analysis is to reduce. So that is why it is called data reduction or data summarization technique, okay. So we said how do we conduct the factor analysis? So there are 4 steps. The first step is to draw the correlation matrix.

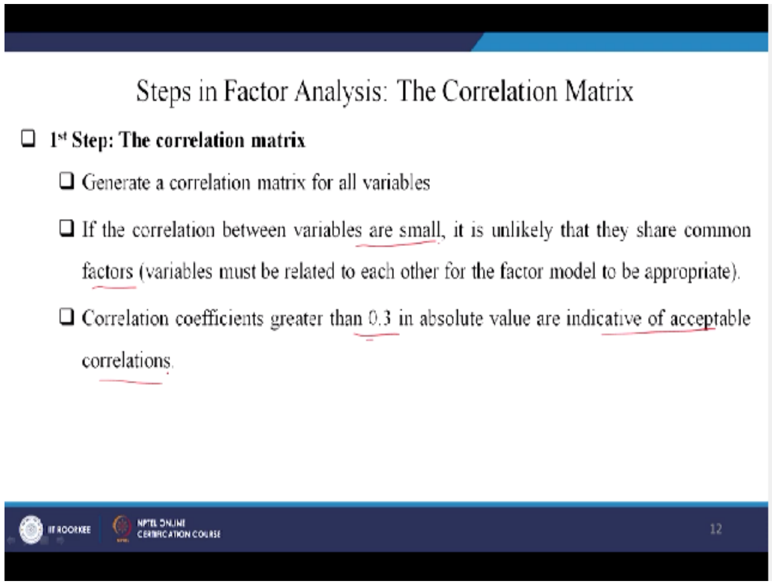
(Refer Slide Time: 01:50)



As we have understood by now that the heart and the spirit behind factor analysis is the correlation. There has to be correlation among the variables. If there is no correlation, then factor analysis is not the right tool. Then next is to extract the factors, right. After that, when suppose you do a factor extraction, you find that the factors are loaded into, the variables are loaded into only 1 and 1 factor mostly.

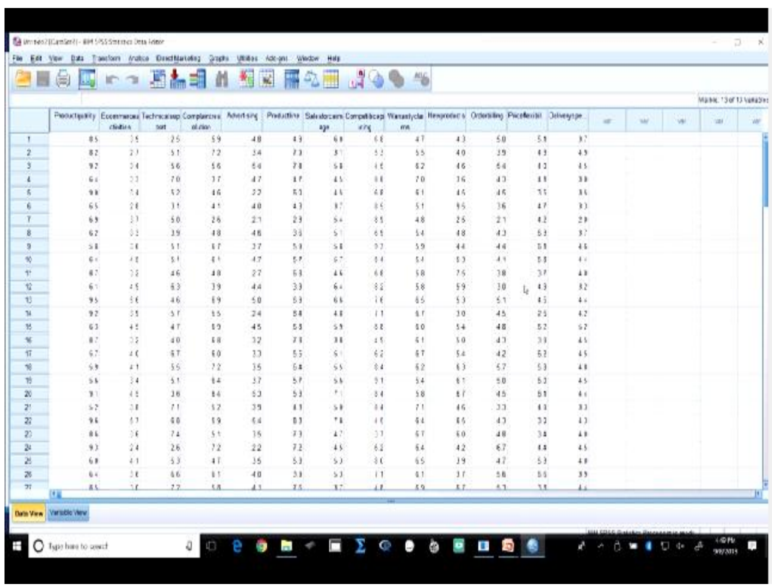
And they are not explaining the other factors well. So we need to make a structure little more simpler and easy. To do that, we do use a technique called factor rotation which I will explain, right. And finally we decide the number of underlying factors and we give names to them. So let us start.

(Refer Slide Time: 02:34)



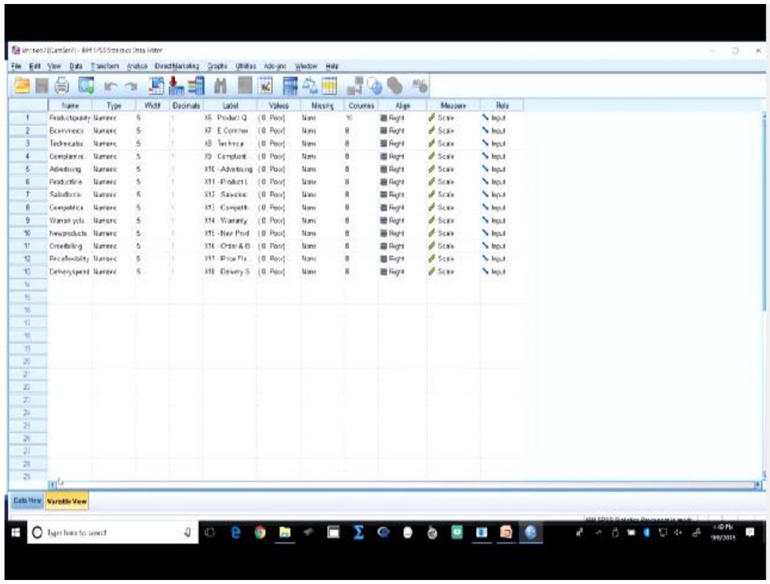
So first step is to generate a correlation matrix for all the variables, right.

(Refer Slide Time: 02:43)



So for example, let us say, I will just show you, this is a data set I have brought. So this data set has several variables.

(Refer Slide Time: 02:47)

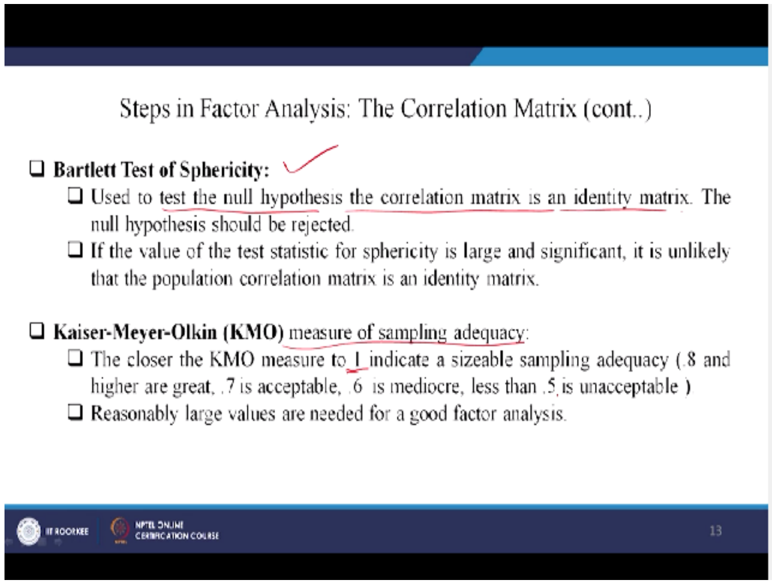


Let us say, the product quality, complaint, advertising, competition, flexibility, sales force image, the warranty claims, new products, order billing. It is of a company who is selling products in the market and it wants to know how consumers are feeling towards, what impression do they have? So there are several variables almost 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13; 13 variables. So these 13 variables, some of them might be correlated with each other.

Some of them might not be, right. So instead of explaining 13, if we can explain 3 or 4, it is much simpler. So in such a condition, we will go for this factor analysis, right. So the first thing is to generate a correlation matrix. So first you draw a correlation matrix. If the correlation between the variables are small, it is unlikely that they share common factors, right.

Variables must be related to each other for the factor model to be appropriate. There must be some correlation, okay. The correlation coefficients greater than 0.3 or 30% in absolute value are indicative of acceptable correlations. Very poor correlations are not acceptable, okay. That means below 0.3 are not acceptable.

(Refer Slide Time: 04:01)



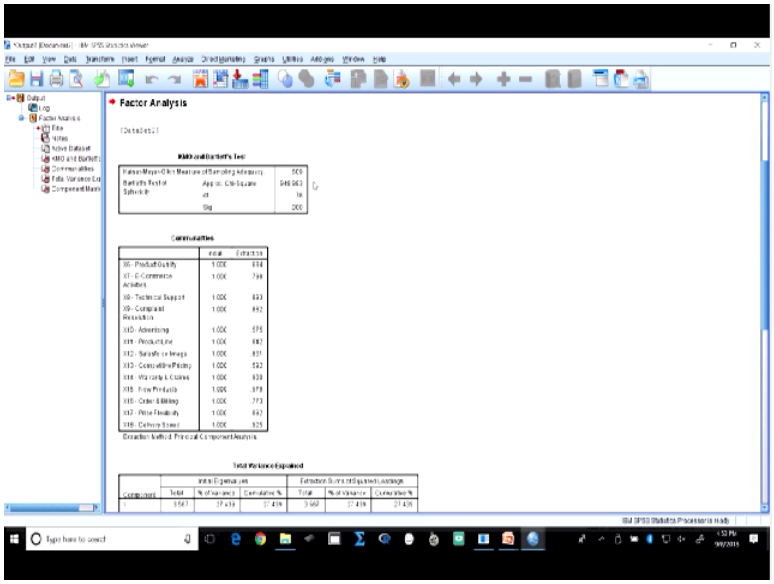
Second is and this continues. We measure something called the Bartlett Test of Sphericity, right. So what is this Bartlett Test of Sphericity, I will do it side by side with you. So Bartlett Test of Sphericity is a technique which tells, it is used to test the null hypothesis. Now what is it saying. Many people are not aware of it because the SPSS does not explain this. So you should be knowing.

So it says, it test the null hypothesis that the correlation matrix is an identity matrix, right. And it means what? The null hypothesis that the very sum correlation among the variables, right. So there is some correlation among the variables. Now suppose you write the null hypothesis in the alternate. How would you write? The null hypothesis is, let us say, we will say, no correlation should exist and the alternate is, there is correlation exist.

So in a study for this, what do we want as a researcher, that the Bartlett Test of Sphericity should always be rejected, should be significant. Now let us go back. And let us, we will explain this one, the KMO. The KMO is a measure of sampling adequacy. Now what does it says? the closer the KMO measure to 1, right. It ranges between 0 and 1.

It indicates the sizeable sampling adequacy. 0.8 and higher are great, 0.7 is acceptable, 0.6 is mediocre, less than 0.5 is not acceptable. So this talks about whether sampling adequacy is prevailing or it is not prevailing. Reasonably large values are needed for a good factor analysis, okay. So, okay, let me go to the slide and show you there. This is the one. So let me run this factor analysis with you.

(Refer Slide Time: 05:40)



So you see, I am going to dimension reduction, factor, okay. Now I am taking all these variables, okay. So we have taken all the variables. Now if you go to descriptives, you will see here the KMO. So this is what we are talking about, right. So let us check whether the KMO, what is the KMO and Bartlett Test, right, the value. So the KMO is 0.609. So we said above 0.8 is very good, above 0.7 is good, above 0.6 is acceptable, above 0.5 is acceptable in fact.

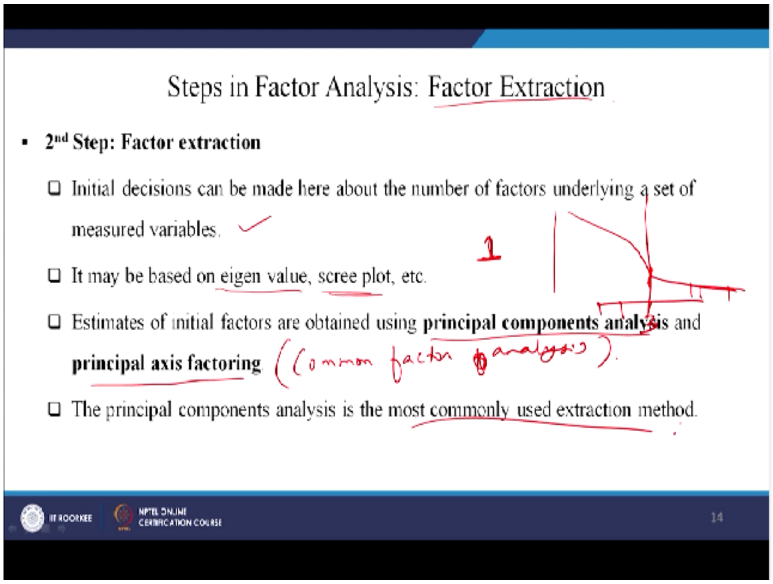
And this we are having 0.609. That means the sampling adequacy is okay. That means what is sampling adequacy? The number of variables to the number of respondents, the ratio which earlier I had said should be preferably 1:20, right. So 1 variable:20. So let me see, let me show you how many variables we have and how many respondents in this case. You see, here we have only 100 respondents.

And we have 13 variables. So what should be or ideally the sample size? Now you have 13 variables. So each variable measured by 20 makes 260 respondents should have been there. Sample size should have been at least 220, 260. If not 260, if you take 1:10 also which is the minimum. Then also it is 130. So we are still falling short. But anyway, the other factors also which decides.

So overall we will say it is okay, right. Because all the studies might not be having large samples, might not be possible. At least, as in this case. This is a case of a B to B case, business to business. So business to customers for example, retail malls, somebody is standing outside and collecting data is very simple or comparatively simpler to a case where you are collecting data from an industry which is very difficult to get, right.

So in those cases even 1:5 sometimes is accepted. So now look at this Bartlett Test. Now it is 0.000. That means what? The null hypothesis is rejected. So what was the null hypothesis? That there is no correlation. So that means now we can say the null hypothesis, the alternative is accepted which says that there is some correlationship among the variables. So which is a important requirement of factor analysis, okay.

(Refer Slide Time: 08:17)



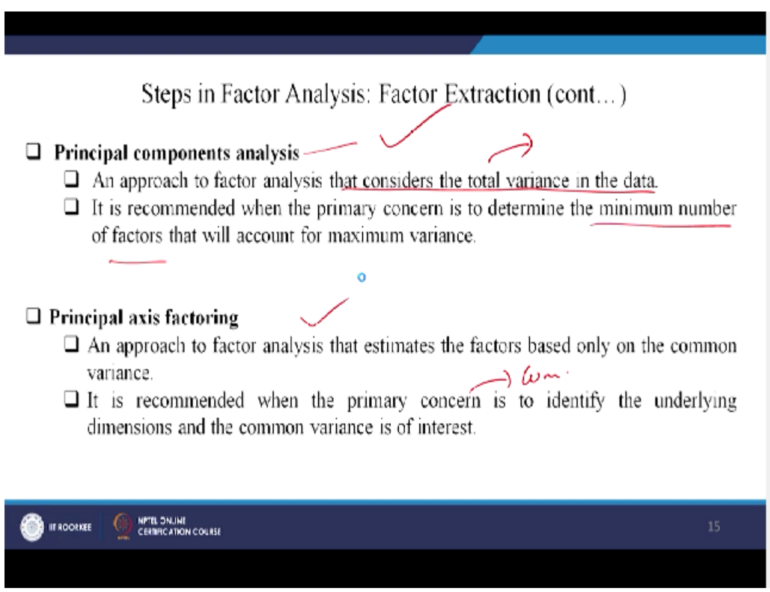
Now once we have done that, the second step. We will come to the second step which is the factor extraction. So the initial decisions can be made here about the number of factors underlying a set of measured variables, right. That means the researcher can decide that how many factors should get generated. But this requires lot of expertise, right. You should not be doing it blindly that I will say 3 or 4 or 5 without a logic.

So there has to be some theory or some reason behind it why you are saying. So extraction of factors is possible, that means you can force the computer or the software to give you the number of factors. But then you should know why you are doing it, okay. Sometimes it is based on eigenvalue. Eigenvalue greater than 1 is considered to be acceptable, okay. Scree plot also I had shown you how it changes.

So from the scree plot, we can say that after what point the changes are not much. So in one case, we saw that there are 3 number of factors which we selected, right. So suppose it falls like this and then like this. So up to this much, there is a significant change. Here there is no change, right. So we will say this is the number of, whatever the number of factors here, so 3 or 4 or whatever, right.

Estimates of initial factors are obtained using principal component analysis and principal axis factoring. Now this is very interesting and very important. What does it mean? And you must have heard the use of principal component analysis very frequently, right. But this you might not have heard much. But if I say use the other word, common factor analysis you might have heard, okay. So the principal component analysis is the most commonly used extraction method. Now what is this method, we will see, okay.

(Refer Slide Time: 10:16)

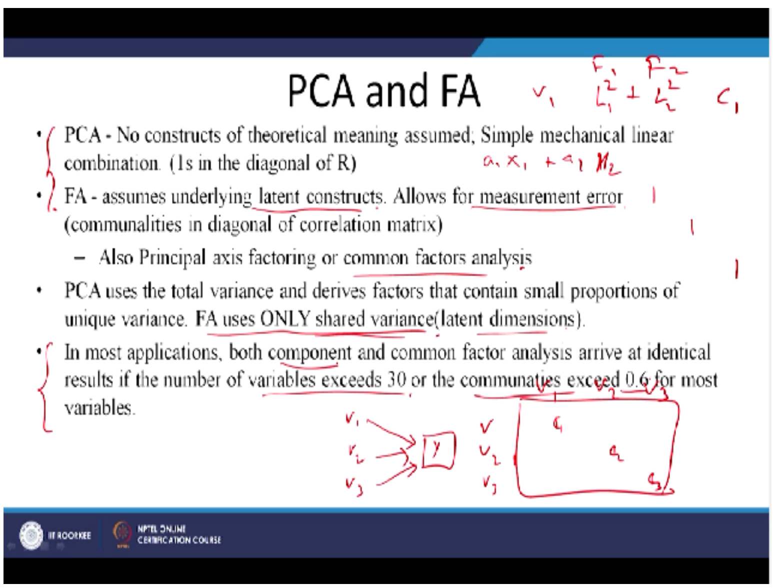


So what is this principal component analysis? An approach to factor analysis that considers the total variance in the data. So the variance in the data could be shared variance, unique variance, right. So we are saying that when you consider the entire variance in the data and then you extract the factors from it, then this method is called the principal component analysis. This is the basic difference you need to keep in mind, right.

It is recommended when the primary concern is to determine the minimum number of factors that will account for maximum variance, okay. So minimum number of, so there is a case of parsimony here also. The minimum number of factors explaining the maximum variance, okay. What is the principal axis factoring, the other method? So an approach to factor analysis that estimates the factors based only on the common variance, right.

So it is recommended when the primary concern is to identify the underlying dimensions and the common variance is of interest. So there can be a common variance which if there is a common variance, and this was taking total variance and this takes the common variance, right. So the two methods are slightly different from each other, right. So this method is the most popular method. This also is used because of its simplicity sometimes, okay.

(Refer Slide Time: 11:43)



So let us see, let us continue with that. So the principal component analysis, no constructs of theoretical meaning is assumed. Simple mechanical linear combination, if you remember in the first one, we had said, right, it is the addition of the, let us say, the like in multiple regression combination x1, so a1x1+a2x2. So it is like a linear combination, right. Factor analysis assumes underlying latent constructs.

The difference between PCA and FA, you see, is this. Now PCA for example, you see, what was happening. Now let us say, variable 1, variable 2, variable 3, right. So these variables are effecting, let us say, y. Now in the principal component analysis, what is happening? So this is all affecting y, correct. But in the case of factor analysis, we are assuming that there are latent constructs.

Now I hope you understand what is the latent construct. A latent construct is something which can directly not be measured. It can only be measured with the help of several other variables. Example, what do you mean by that? Let us take a case of satisfaction. Now how satisfied as a person you are? Now how do you, can I measure satisfaction directly? Not possible, right. So I will measure satisfaction may be through several other variables.

For example, how satisfied are you in your home. How satisfied are you in your office. How satisfied do you feel when you meet your family members, right. How satisfied you are generally when you are alone. So when I am, these are called latent constructs because they are hidden. So you want to find out through some way by asking other questions, right. Some related questions, okay.

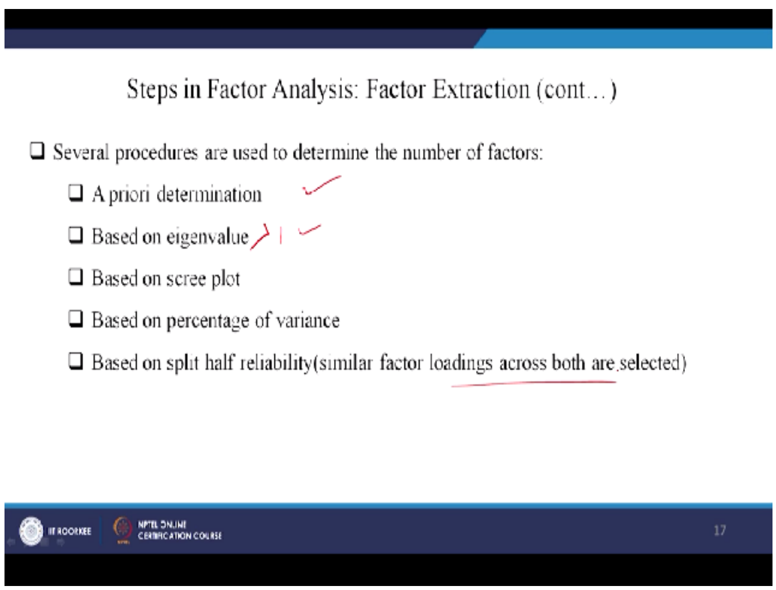
So it assumes underlying latent constructs and allows for measurement error, communalities in diagonal of correlation. That means what? When you do a correlation matrix generally, the diagonals if you see, there is one unity. There is a unity in the diagonals, okay. But in this case, we are saying because we are allowing for measurement error, the communalities when will be placed in the diagonal.

So if you think of a correlation matrix, let us say, v1, v2, v3; v1, v2, v3. Usually what happens? So this is 1, this is 1, this is 1. But in this case, in the case of factor analysis, this is not that. Here we will place the communality values. Let us say, this is c1, this is c2, this is c3. I hope you have understood what is communality. In the last class, I had explained, last lecture. Communality is nothing but the square of the loadings of a variable across all the factors.

So it was v1 across F1, F2, let us say, 2 factors. So this L1 square+L2 square is my communality, okay. It is also called as common factor analysis. PCA uses the total variance, again I have said this and derives factor that contains small proportions of the unique variance. FA uses only shared variance, only uses the shared variance, okay. In most applications, both component; component means principal component and common factor analysis arrive at identical result.

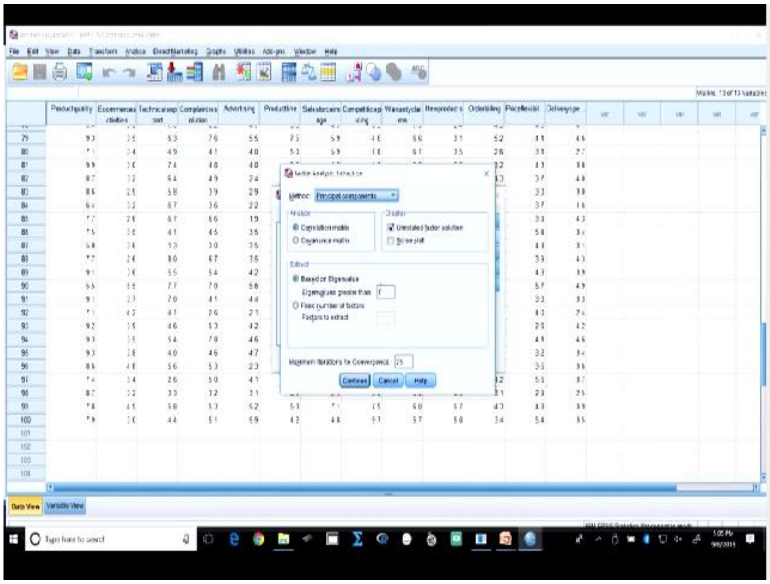
This is very important, please note. In most applications, both component and common factor analysis arrive at similar result. If the number of variables exceed 30 or the communalities for all the variables or most of the variables exceed 0.6. If this happens, then component analysis and common factor analysis, they converge towards each other and they give a very similar result. So the argument whether I should use the principal component analysis or a factor analysis, dies down there and is of inconsequential effect if you are having this kind of a situation, okay.

(Refer Slide Time: 15:53)



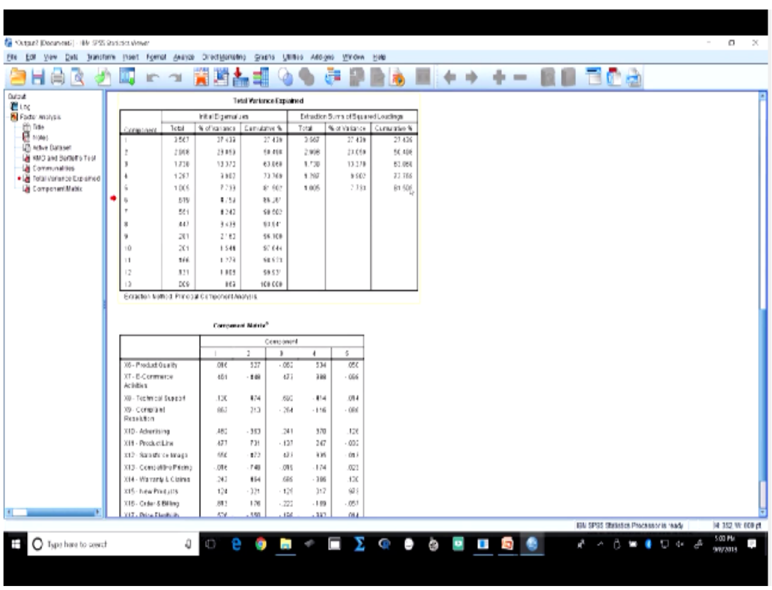
Next step is how do you determine the number of factors? So let us go to that, this one.

(Refer Slide Time: 15:57)



So let us first derive the number of factors, okay. So what I will do is, we will go to the dimension reduction factor. Now we will derive the factors. So I think the factors must have been derived.

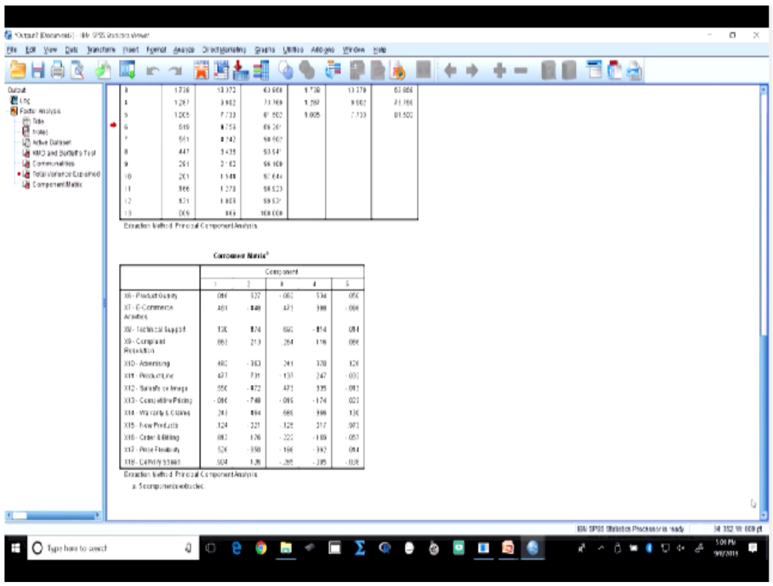
(Refer Slide Time: 16:14)



If you go to the output, now you see. Now you can see here how much of variance has been explained in this study. Now there are 5 components or 5 factors you can understand. So why it is called components? Because it is using the principal component analysis method, that is why it is called component, right. So 5 components are explaining 81.5% of the variance which is quite good, right.

And you can see here. The first variance is explaining how much? 27.43% of the variance. The second factor, 23. The third factor 13, fourth 9, fifth 7. So if you go till the end, you will achieve 100%. But then after 5, we are finding that it is significantly dropping down. So we will say up to 5, if we take 5 factors, then how much I will be explaining? If you take a cumulative of this, 81.5, okay.

(Refer Slide Time: 17:13)



Now we have done this component matrix. Now component matrix if you see, there are 5 factors have emerged, right. But it is becoming very difficult to analyze these 5 factors in this situation. So we will make some changes. But you can see here 5 factors have emerged. 13 variables, 5 factors, okay. So several procedures are used to determine the number of factors. How do you determine?

A priori determination. From begin inning only through some theoretical backup or understanding, you can decide, one. Second is based on eigenvalues. So at the moment, what we have done is, we have based it on eigenvalue. You can see. So if you see here, so how it has been selected, right. So extraction, if you go to extraction, there are 2, you can see here. So we have used the principal component analysis method, okay.

And unrotated factor solution. That means no rotation has been done. And eigenvalue greater than 1. if something, you please understand. This eigenvalue if it is less than 1, that means the variable is unable to explain for itself even, right. So minimum at least it should explain for itself. That means it should be at least 1. So you can have other fixed number of factors also which is I said based on your theoretical understanding.

But at the moment we have considered that by default is on the eigenvalue, right. So eigenvalue 1. So based on eigenvalue which is at least greater than 1. Scree plot also you can do from graphically. Based on percentage of variance. Now for example you see, this is the trade off. the percentage of variance. For example in this case, how much percentage of variance is being explained?

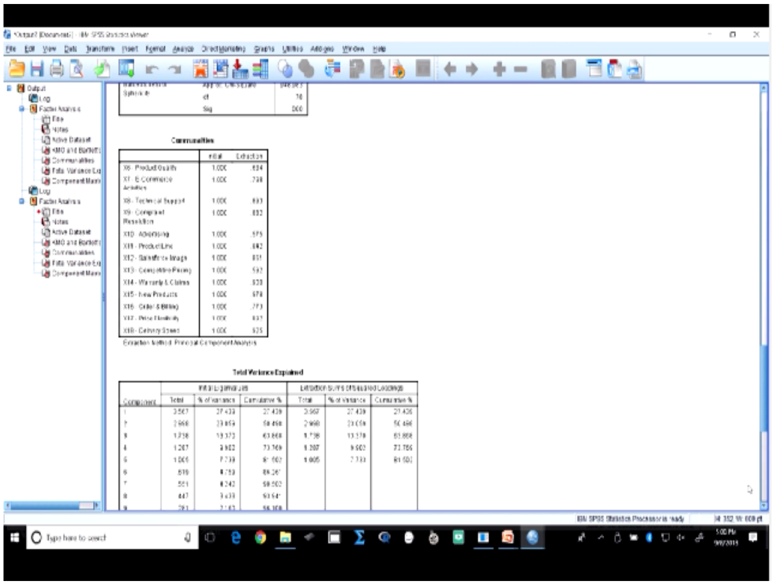
Now the percentage of variance being explained is, let us say, is how much? 81.5. Now 81.5 is a good amount of variance being explained in the study. Suppose I feel, suppose instead of 5 factors, there would have been 8 factors now and 8 factors explaining let us say 81%. And I would feel that 81 is quite high and 8 factors is also too large a number. So if I reduce it to 6 for example because of some theoretical understanding.

So if I reduce it to 6, the variance explanation will also go down. From 81, it might come down to 60 or 70 also, right, possibly, or even lower, I do not know. So you have to check. You have to do it and check it. So you suppose take first 6 variables which you feel and suppose they are explaining you 70% which is decent enough, not bad. Then you can say here, what you can do is?

You can extract. Instead of eigenvalue, you can say 6 here, right. So this is one thing that you need to understand. But here we are going on eigenvalue, right. The next thing is, so percentage of variance also you have understood. Split half reliability. So what you do is, you divide this sample into, if you have a large sample size, divide it into 2 parts, right. And you find out and you do 2 different factor analysis and look at the results on the 2 different sample groups.

And if you find that the factor loadings across both are similar, then also you can understand that whether the study is moving in the right direction or not, right. So these are some of the ways how you can extract the factors, okay. The next is factor rotation. So why you did, should do a factor rotation, I will explain. So let us go back to the data set. So here first let me show you, here you see. When I did this factor analysis, this matrix, right. There are 5 factors but they are very clumsy to understand, right. So let us rerun it.

(Refer Slide Time: 20:55)

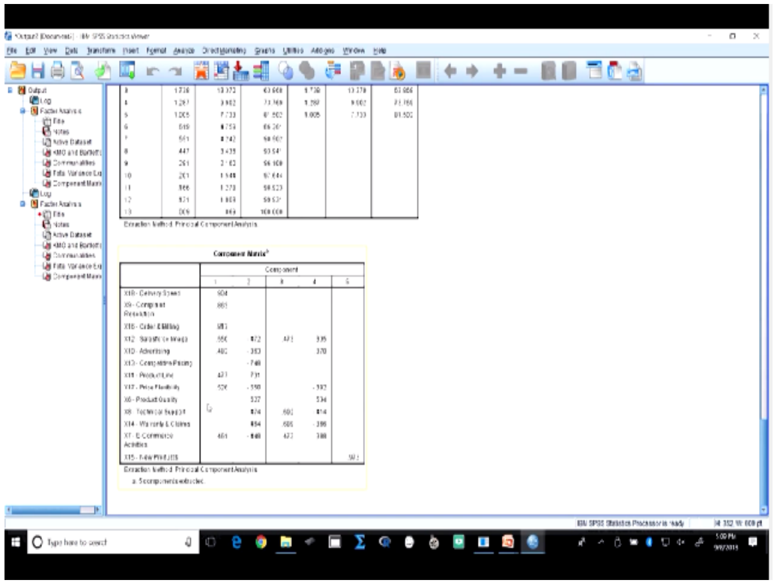


So what I am doing is, first I will clean it a bit. So what I will do is sort by size and suppress small coefficients. So here instead of having 0.31, what we will do is? We will take a minimum, we will set the factor loading, the correlation should be at least 0.3, we have said that. So we will take 0.3, so 0.3, okay, so continue.

And we are not changing other things. So we will say okay. Now if you go into this, nothing has changed if you can see, right. And you see, these are the communalities. Is there anything that is below .5, in this case? No. So 0.5 is the minimum benchmark. If it is less than 0.5 and your model is not very good.

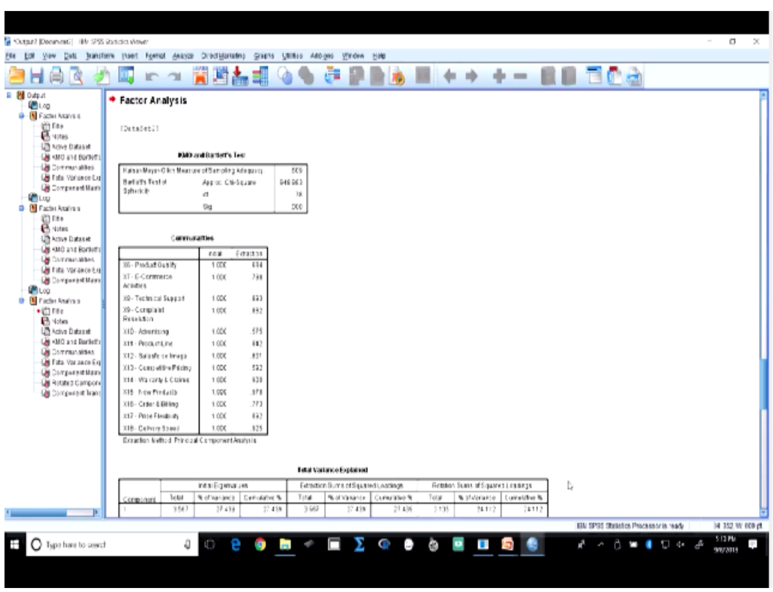
Then you may start thinking of dropping the one which has the communality less than 0.5, okay. But do it one by one. Do not, if there are the 2 variables below 0.5, do not delete it at once. First delete the lowest value, run it and if things change for better, it is good. If they do not, then still go further, then next go for the next, the lowest one again. Now look at the component matrix.

(Refer Slide Time: 22:11)



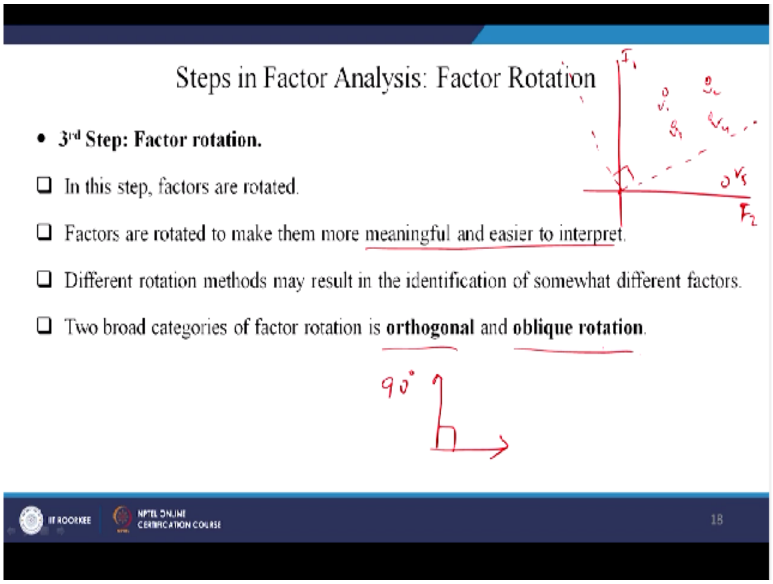
Now if you see, now it looks little cleaner than the earlier one but still there is a problem. I see that most of the variables are loaded and there is a lot of cross loadings. Can you see? For example, sales force image is cross loaded on 4 factors. Some of them are loaded on to 3 factors. Some are loaded on to 2 factors. So cross loading is again a very dangerous thing, right. Cross loading is very dangerous. So we should try to avoid cross loadings and if cross loadings do not go, then that variable may be as an item for deletion, okay.

(Refer Slide Time: 22:39)



So now we will go for extraction method, right, rotation method. So by rotation, we may think of improving our model. So let us go to rotation. So the most popular rotation is varimax rotation. Now first let me run it and then I will go to the; before that, I will show you the first the theory.

(Refer Slide Time: 23:02)

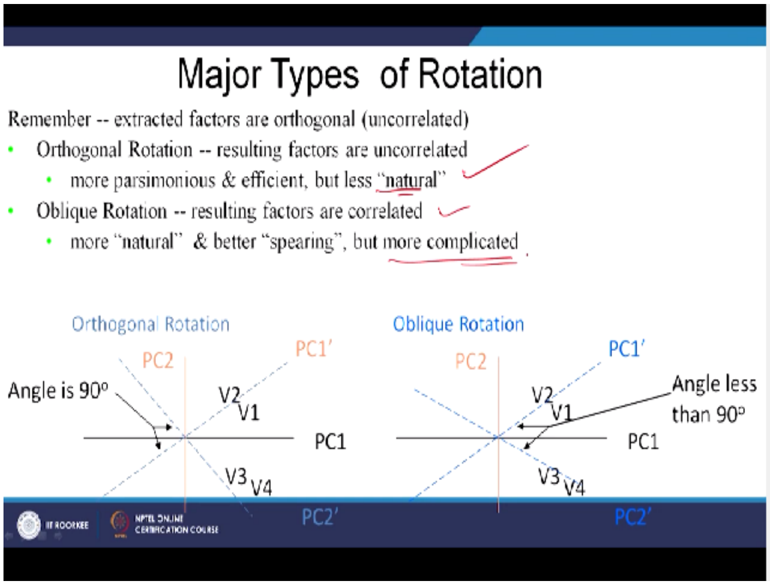


In this step, factors are rotated, right. So as we just did the varimax rotation. Factors are rotated to make them more meaningful and easier to interpret. So as I said suppose, let us say, the variables are, the data points are like this, okay, right. So we can see, suppose this is, let us say, factor 1. This is factor 2. So most of the variables are, these are the variables, right. So v1, v2, v3, v4, v5.

They are mostly loaded into the first factor, right. And because they are close to the factor 1. But suppose we tilt it a bit, we tilt it. What we will do? We will tilt it a bit. Turn it 90 degree. Keeping this 90 degree, so this is 90. So what is happening now? We can see that some of the variables like for example, v1 and v3 will go to factor 1 but v4, v5 may come down to the factor 2, right.

So factor rotation only tells this work. It makes it more easier to interpret, right. So different rotation methods are there and they result in the identification of the different factors. So 2 broad categories are orthogonal and oblique. Orthogonal is when it is 90 degree. So when it is 90 degree, means what? The variables are uncorrelated. So 1 variable is going this side, the other variable is going this side. They are not correlated at all. But oblique rotation is something where you see.

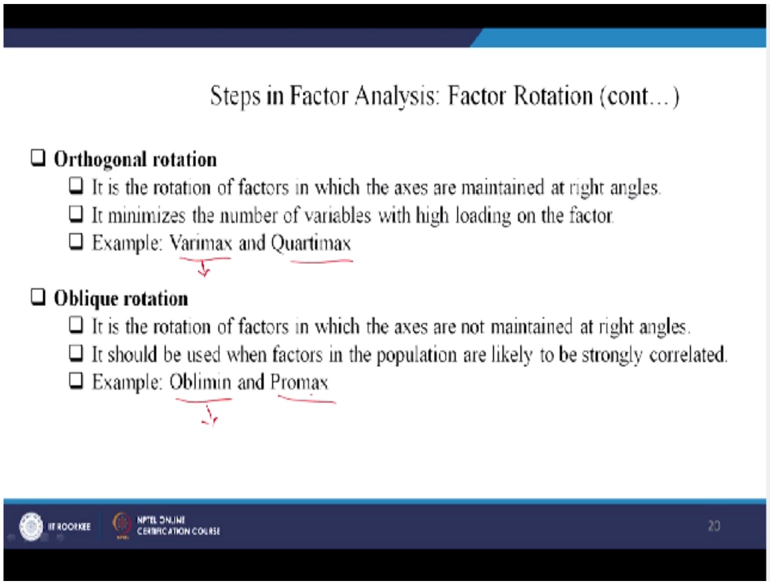
(Refer Slide Time: 24:29)



I have brought it. So major types of rotation. Orthogonal, the resulting factors are uncorrelated, right. So there is a 90 degree. It is more parsimonious and efficient but it is less natural. In a social science, just imagine which is on human relationships, how do you feel 2 factors will not be correlated. There will be some correlation. But it is easy to interpret and easy to conduct, right.

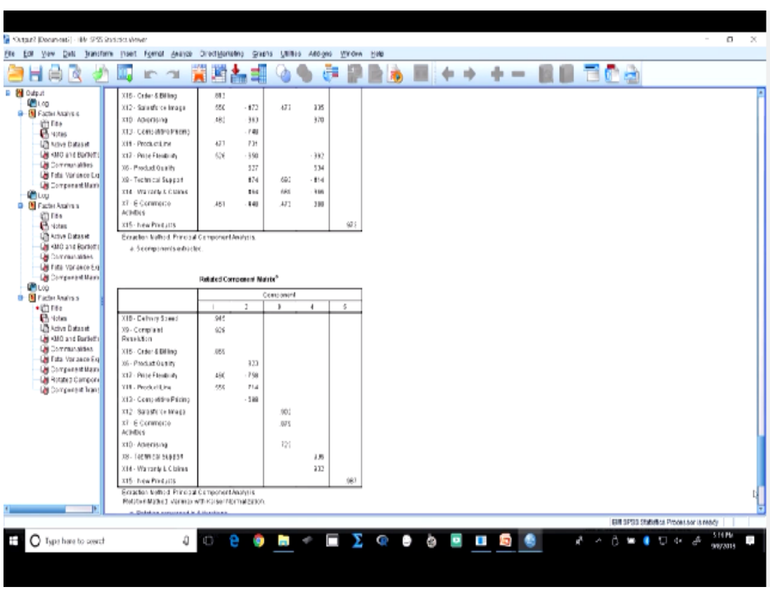
On the other hand, oblique rotation is where resulting factors are saying it is correlated. That means the factors are correlated. But for a mathematical understanding, under theoretical understanding, we will say that they should not be correlated, right. But in truth, correlation between the factors will surely exist. So you see the angle is less than 90 degree in this case, right. So that means what? At some point of time, they will converge somewhere, right. But it is a very complicated technique, right. Anyway.

(Refer Slide Time: 25:26)



So orthogonal rotation, there are basically varimax and quartimax but varimax is most likely utilized. And among oblique, it is oblimin and promax. There are 2 methods. And generally what we have seen in factor analysis. We have used, you may try using; according to your logic, you may use the oblique rotation also. Does not matter, right. And you will see if there is any change for the better, theoretically the explanation is getting better. It is much good, right. So let us take, let us go back to the data set, the output now.

(Refer Slide Time: 25:58)



Now we had done some rotation, correct. And the variance does not change, nothing changes, right. So this was the component matrix. And now let us look at the rotated component matrix. So you can see there is a C change. There is a big change, right. Now for example, you see, so now most of the variables were loading into here, right.

Here and here and there was lot of cross loadings. Can you see? Now this cross loadings are somewhere gone, disappeared. Now you see, it is very clear, the first factor has got 1, 2, 3, 4 and 5, okay. Although there is a cross loading in this case, right. But does not matter. It is much better than the earlier one. Factor 2 has got 1, 2, 3, 4. Factor 3 has got 1, 2, 3, 4. This is something similar, right.

And last factor has got only 1. So anyway, so we have understood by now that when we do a factor rotation, the factors are equitably or distributed in a much better way and their explanation power improves, okay. Well, what I will do is, we will continue from here. Then we will go into the next things, right, about rotation and then how to check for, after rotation, we have to make a final decision and then we have to check for reliability and validity.

And how to utilize the outcomes or the results that you derive from factor analysis for other studies, right. So we will do that in the next lecture. So today, I think you are clear with what is factor analysis and how to conduct a factor analysis. So step by step I have tried to explain you. So we will proceed similarly in the future in the next lecture and we will finish there, right. Thank you so much.