Marketing Research and Analysis -II (Application Oriented)
Prof. Jogendra Kumar Nayak
Department of Management Studies
Indian Institute of Technology - Roorkee

Lecture – 53
Exploratory Factor Analysis in SPSS

Welcome everyone to the class of marketing research and analysis. So in the last lecture, we had started with factor analysis, especially the exploratory factor analysis. So we had understood that factor analysis is a technique which is used for data reduction or summarization, correct. So why it is used? And it is also called as an interdependence technique. Because there is no dependent or independent variable in that as such.

So we say it is an interdependence technique, okay. And I also mentioned that the heart of factor analysis is correlation. That means we try to find out the correlation among the variables and the stronger the correlations are within a construct, it is better, okay. Or within a factor, it is better, right. So we assume that all the items or all the variables within the factor are strongly correlated, okay.

So we understood that in the last lecture that we first will, suppose we have a data set. Then we try to find out what are the different kinds of factors that are emerging. So that is based on several factors. For example, several ways you can select factors. One of them was on the basis of the variance explained.

That means we said if a study is explaining sufficient amount of variance, then we can select the right number of size of factors, okay. What does it mean? For example, as I said in the last lecture that your variance explanation should be at least 60%, right. So let me show you on the data set. It is better if I show you because then you will realize it better. So, okay, so this was the data set.

(Refer Slide Time: 02:10)

So we had if you remember several variables, product, quality, peak hours, activity, technical, support. So these are different variables a company is interested to see how they are influencing, right, the customers. So we went for a dimension reduction or a factor analysis. And this is the exploratory factor and we took all the variables, okay. So we will take all the variables in here, okay.

So now in descriptives, we had learnt that we have to find out for the KMO and Bartlett Test. What is the KMO? The KMO is a measure of sampling adequacy and the Bartlett Test says it should be significant. That means if it is significant, that the null hypothesis says that there is no correlation among the variables is rejected and we will say that at least there is some correlation among the variables which is an important assumption of the factor analysis.

So we did this, right. We said, we can extract on the basis of eigenvalue and fixed number of factors. So eigenvalue we said if you remember across the factor, how do the variables explain in a particular factor, the contribution of the different variables in a particular factor. So we take the loading square, the sum of the squared loadings across the particular factor and that value is called the eigenvalue.

So eigenvalue should be greater than 1, right. Or we could also use the fixed number of factors in which we said how do you determine the fixed number of factors. The fixed number of factors

could be determined on the basis of past knowledge or sometimes we can do it on basis of the amount of variance explained.

So let me show you, for example, or if you remember, if the amount of variance explained is sufficiently high. Let us say, you have 13 variables in this case and you have got, let us say, 4 or 5 factors. These 5 factors explained about 85%. Then you can say well, since it is explaining quite a good amount of variance, what I can do is, I can increase a little further.

So if I, in certain cases, in such cases, the variance explained is minimum should be 60%, right, in the social sense, right. So if it is 70% or 80%, you have a better chance. You can still reduce the number of factors. That means suppose 6 factors were explaining 85%, now I do not want 6 factors to take. I want only 4 factors from my theoretical understanding.

Then I will take 4 factors, that will reduce my variance explained, there is no doubt about it, right. So as the number of factors will reduce, so the amount of variance will also reduce. But if 4 factors now will explain, let us say, 70% of the study, of the variance in the study, then it is good enough, okay. So that is how you find the number of factors, okay.

(Refer Slide Time: 05:08)

Now I also explained about rotation, if you remember. And I said why do we rotate, if you remember? Sometimes to get a simple structure, we get a simple structure, we want to rotate the

variables so that the variables are placed in such a manner that they are not loaded too much into one factor and very much less on to the other. But they are more or less equitably distributed, okay.

So the most popular is the, among the orthogonal rotation is the varimax rotation. Now what is orthogonal and what is the oblique rotation? So I explained that orthogonal rotation is a rotation in which we say that the factors are at 90 degree to each other. That means what? If something is at 90 degree, mathematically you can understand that they would never meet. So they are not correlated, right. Uncorrelated.

But on the other condition, other hand, if you have a, let us say, oblique rotation, so it is not 90 degree. But the problem, the problem is that although in real life, in social life, hardly you find a relationship which is uncorrelated or not correlated. But the other side is when you take an oblique rotation which is more favorably saying that there is some correlation, but it is very difficult to analyze mathematically.

It is a very complex process. Because the number of correlations that can occur will be very high. So in fact even the softwares are not developed to such an extent which can explain it properly. So most of the time what we do is, we consider and also theoretically, there is a logic that we fear of 2 factors, these 2 factors would be separate as much as possible.

So to believe that, we say well they are in a orthogonal rotation, will take an orthogonal rotation. That means they are not related. And taking that we take the varimax rotation, right. I said if you remember, you should sort by size and suppress the small coefficients. Now when you suppress the small coefficients, what do you mean?

I mean that the factor loadings which I would take should be at least 0.3 that is 30%, right. Because if it is, factor loading is what? The correlation between the variable and the factor, okay. So if it is 0.3 or 30%, that means the R square is at least 10%, right. So we take that as a magic number or a cut-off value, right. And we will say run, okay.

(Refer Slide Time: 07:39)
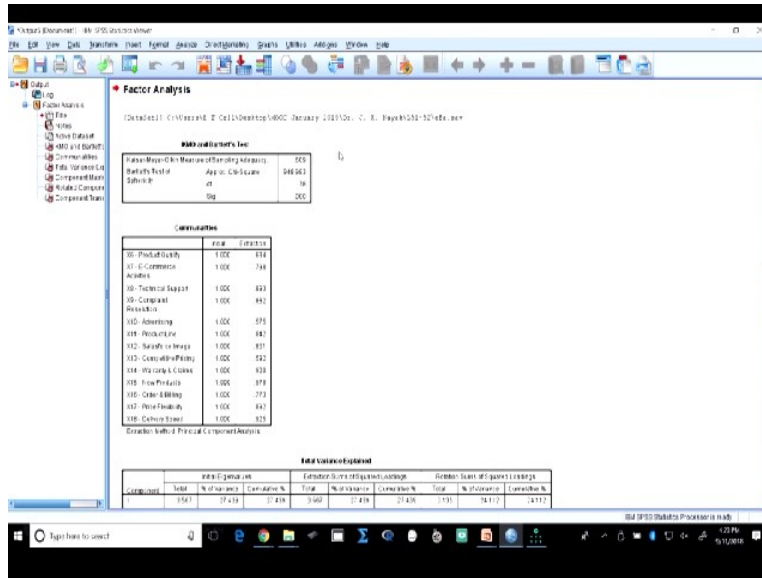
Now let us see the output. So KMO is good, is above 0.5. This is significant.

(Refer Slide Time: 07:45)
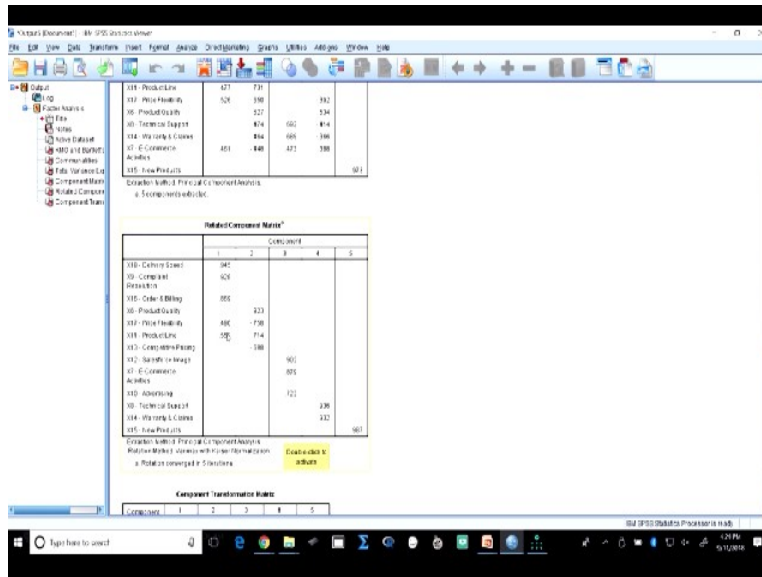
Now coming to our, and I had explained you about the role of communality also, right. So now look at the rotated component matrix. So if you look at this matrix, slightly let us go up.

(Refer Slide Time: 07:58)

So if you look at this matrix, it looks little more complicated. There are lot of cross loadings. That means 1 variable is loading on to several factors. Factor 1, factor 2, factor 3, factor 4, factor 5. But if you look at this now, when we have rotated, now they are quite neat. The variables are loaded into 1 and only 1 factor. Only in few cases, there is a cross loading. For example, in this case if you see, there were 3 cross loadings or sometimes, this case is even 4.

But here if you see when we had done maximum there is a cross loading of 2 in to 2 factors which is not very, okay, it is fine, right. So what we do in here, now you see, you have got a factor. So factor 1, factor 2, factor 3, factor 4, factor 5, right. So all these 5 factors, now suppose you see, now I have got a 0.987 but it is a single variable, right.

(Refer Slide Time: 08:54)
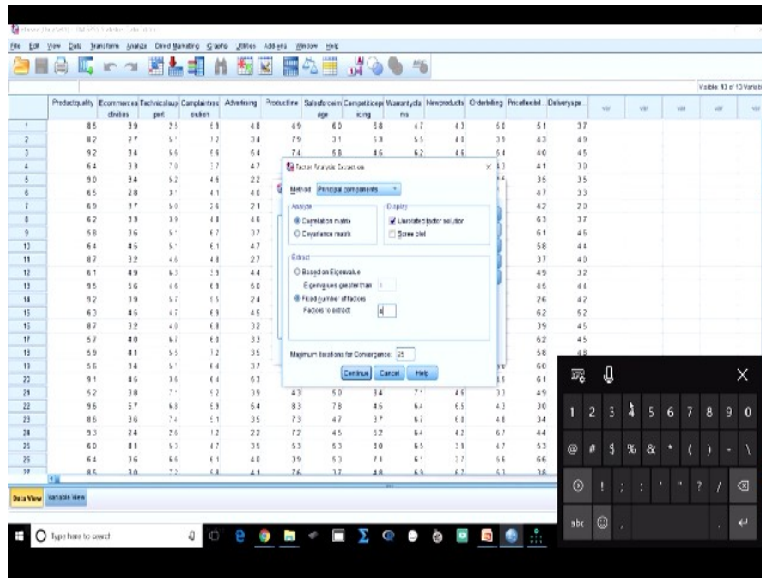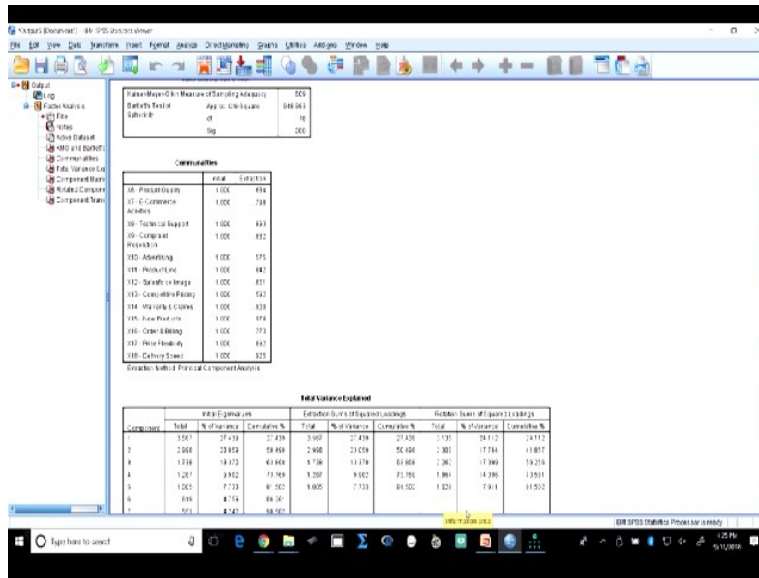
So if I want, I can run this in a slightly different way. What I can do is, I can go to analyze, dimension reduction again, and now this time, you see since I had, let us go back to the output. Look at the amount of variance explained here.

(Refer Slide Time: 09:05)
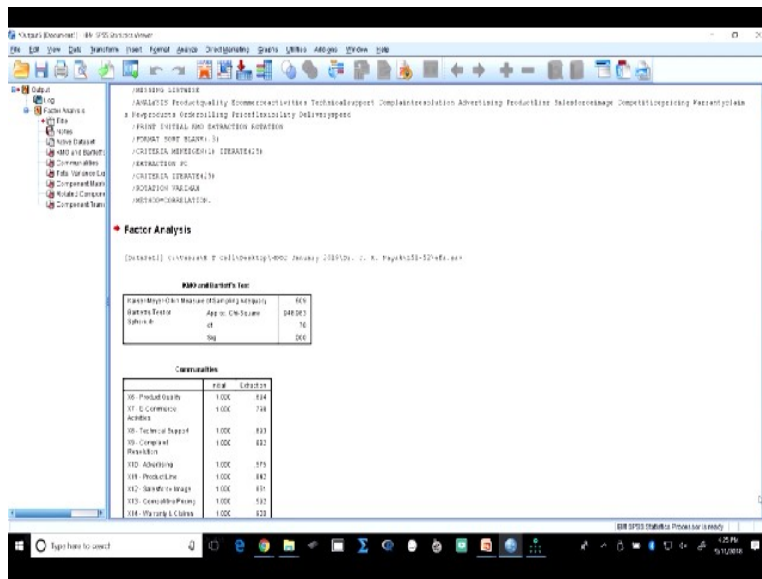
Now the amount of variance explained in this case is about 81% which is sufficiently high.

(Refer Slide Time: 09:09)

So instead of 5 factors, if I make it 1 factor, will it matter or will it make some difference. Let us see? So I will now extract only 4 factors. So what I am doing is, I am taking 4 factors. And I will run this again, okay.

(Refer Slide Time: 09:31)
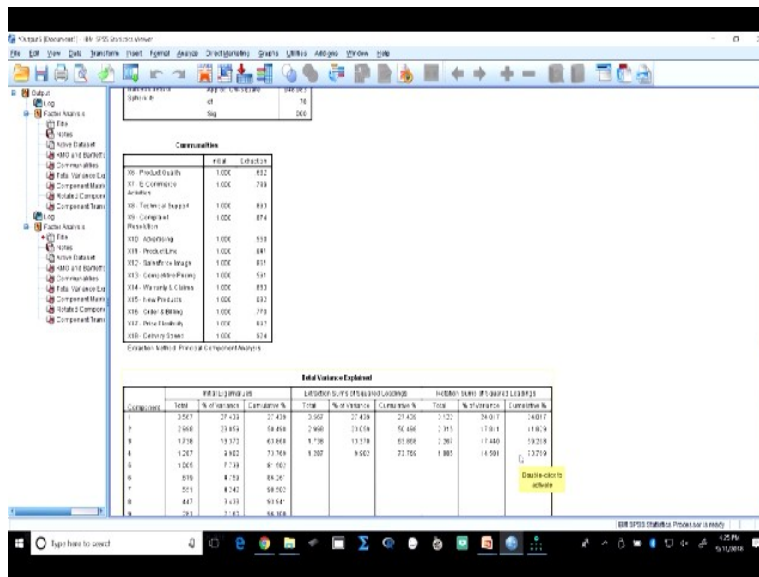
Okay, fine, everything remains the same. So now you see my variance has reduced from 81% to 73.7, 74, almost 74%, right. Has there been any change in the communalities? You see one of the communality has become very poor, new products, let us say, let us see that, okay.

(Refer Slide Time: 09:50)

And let us look at the rotated matrix now. So if you look at the rotated matrix, right. So the rotated, you see the new products is not visible. Because it is now less than 0.3. So it is much, it has a very low communality. It is not explaining to the factors. So it is automatically not visible to you. It is less than 0.3 that means, right, the loading which you have taken, putting those values into the SPSS.

Now look at it. Now we have got 4 factors. Are these 4 factors now explaining slightly better with 74% of the variance. If yes, then you should choose this model instead of the other one, right. So this is what, I am not forcing you to, asking you to do anything. I am just giving you an option. So you had 81% with 5 factors, but 1 factor was explaining only, there was only 1 variable in that.

So instead of having 1 variable, what we did is, we tried to see whether if we reduce it to 4 factors, will the variance decrease too much. If it is not too much and it is sufficiently explaining, then why not do it. And in this case, it happened, right. So after the analysis is done, now let us move to the PPT.

(Refer Slide Time: 10:58)

Steps in Factor Analysis: Making Final Decisions

- **4ᵗʰ Step: Making final decisions**
  - ❏ The final decision about the number of factors to choose is the number of factors for the rotated solution that is **most interpretable**.
  - ❏ To identify factors, group variables that have large loadings for the same factor.
  - ❏ Interpret factors according to the meaning of the variables

So the final decisions, how do you make? The final decision about the number of factors to choose is the number of factors for the rotated solution that is most interpretable. So you can, either you take it from the, through the eigenvalue or the extracted number of factors, that is up to you. To identify the factors, group the variables that have largest loadings for the same factor. So we understood there also.

Interpret the factors according to the meaning of the variables. Now let us go back to the output file, right. So the output is here. Now let us see. Is there any relationship? That is what your job is now. So let us see, there are certain variables loading into factor 1. So what are they. Delivery speed, complaint resolution, order and billing, new products. New products is almost gone. So product quality.

Product quality is also not here. Why? It is in the second factor, right. Price flexibility, product line. So now you have to see whether there is any relationship among the variables which are loaded into factor 1. Similarly what is the relationship that is among the variables loaded into the factor 2. What is the relationship for example to the factor 3. Sales force image, E-commerce activities, advertising.

So most important is something linked to the promotional image, something, right. And this is, for example, if you see technical support and warranty and claims, more into the service part,

okay. So now you have to find the name. You have to suggest a name to the factor, okay. That is your job. So interpret the factors according to the meaning of the variables.

So as you saw the variables, you have now understood that there is some relationship among these variables. And on basis of this relationship, you have to give a name to this factor, okay. So now after you have understood this, now let us see what you can do with the factor analysis results, the exploratory factor.

(Refer Slide Time: 12:50)

We are talking about the exploratory factor analysis. Generally factor analysis when you say the word factor analysis, you mean the exploratory factor analysis. Because there is another type of factor analysis which is separately done, I mean, which is called as the confirmatory factor analysis which you will be doing later on, okay. Now once you have done with this exercise, then there are some options in your hand. What?

First of all, you can select a single surrogate measure. Now what does it mean? Choose a single item with a high loading to represent the factor. Now suppose in that factor, let us say, in the factor, there are 4 variables, right. So 1 of the variables you see let us say v_2 has a very high loading with this factor, right. So you can select this particular variable to represent this factor, that is also a, that is a possibility, right.

But it has own limitations, sometimes, why it has limitations? Because it is not good to say that this factor should only be representing, this variable should only be representing this factor and we should, we will completely overlook these 3. That is not a right way. That is, to me at least personally I think. The second way is that you create a summated scale. Now what is a summated scale?

Now the values, for example, v_1, v_2, v_3, v_4. There are 4 variables. Now these are the respondent 1, 2, 3, 4, 5. Let us say you have 300 respondents, okay. So the scores given by each respondent, whatever score they give, let us say, right, the values. So let us say, any value they give, let us say, 2, respondent 1 has given to v_2 as 3, 3, right. Then let us say 4, suppose. Then you create a new column which is, let us say, summated v, right.

Give a name or something and then, what is the average of this? 3+3+4+2=12, 12/4 is 3, right. Similarly for respondent 2. Suppose 3 4 3 2. Again it is 3. So you can find out, right. So take this, take finally this column and use it later on. We can use it. So what it says? form a composite from items loading on the same factor. That is what we did. Average all items that load on to the factor.
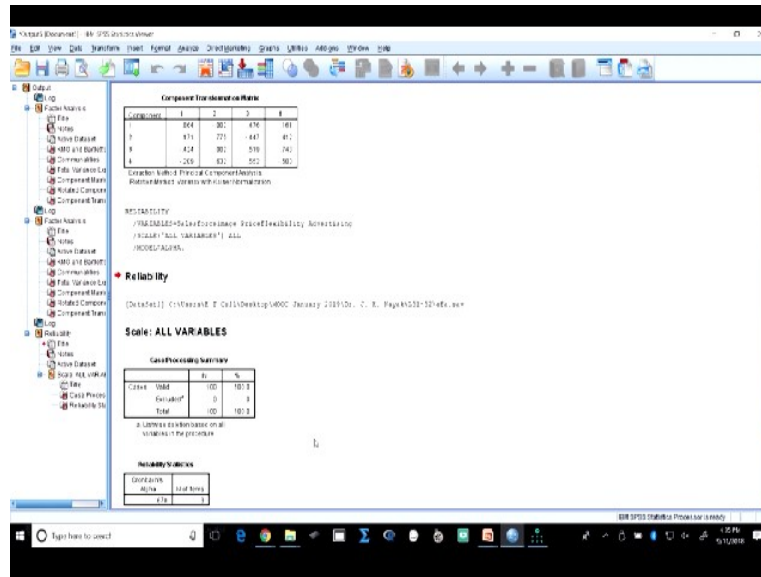
Calculate the alpha for the reliability, okay. Now this I have, I will have to explain you. What is this alpha and reliability, right? So we use a term called Cronbach's alpha. So it is an inter item, Cronbach's alpha, right. It is an inter item correlation. That means what is the correlationship among the variables within a certain factor, right. So we assume that within a certain factor, let us say, this is a factor 1.

So assume it is a house. It is a particular residence or house. So we can say that the members, let us say v_1, v_3, v_6, right. And here v_2, v_4, v_7, let us say. So we can say that the members within this house or this factor, they should be having strong correlationship, right. Similarly, the variables within this factor, they should be strongly correlated among themselves. But should they be related to each other? No.

That is why most of the time you say that we need an orthogonal rotation because we feel that

the factors are uncorrelated to each other, okay. So but there has to be a strong correlation. So this strong correlation if that is there, then that means what? We will be measuring what we are intending to measure, right. That means the value of what we want to measure that will be more or less repeatedly becoming the same.

That means it will be repeating the same result again and again. So most of the variance will be explained. So here we, in this condition, when we say this, right, so we find this a reliability. So reliability, how to measure, we will see. So that is a, as I said the Cronbach's alpha has to be measured, okay. So now you name the scale or construct. Now let me show you how to do the, let us say, measure, right.

(Refer Slide Time: 17:09)

So take one of this. For example, 12, 17 and 10. Let us remember this, okay. So we go to this file. So how do you measure reliability? Go to scale, reliability, right. So 12, we will take 12. Then it was I think 17, okay and 10, okay. So now we want to see the alpha. So you can see there are several things out here. So we do not want to do anything. You can even check. Suppose you have large number of variables, 7 or 8 variables.

And because of some variables, let us say the entire reliability is coming down. So you can even use this, scale if item deleted. So what it will do is, it will reduce, it will deduct those variables which are contributing to the poor reliability of the study, okay. So that is a very nice thing. But I

do not want anything at this moment. I just want to see the alpha.

(Refer Slide Time: 18:09)

Now you see, the Cronbach's alpha in this case, the number of items and the Cronbach's alpha is saying 0.67. So 0.67 is as, earlier it has been said that this reliability value should be 0.7 or more. But researchers like for example, if you go to Anderson's multivariate analysis. In his book, he has said if it is even above 0.6, then it is fine. Because in social science, the reliability measures do not, are not, will not be as high as you find in basic science, right, like engineering and medicine.

So in this case, we will try to see whether it is sufficiently above 0.6 or around 0.6 or not, okay. So this is the second method and this is the, one of the most preferred method now a days. The third method is also good but this is a preferred method, right. So now remember the third method is the factor score method. Now factor scores I had shown you I think that if you go back.

(Refer Slide Time: 19:05)
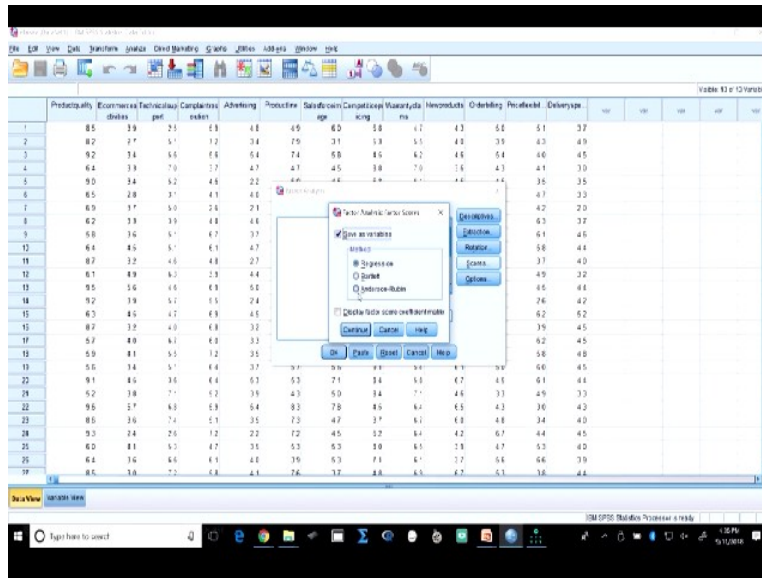
So you can do one thing. You go to the exploratory factor analysis here, right and see scores. So what you can do is, there is a method to locate this, you can save the scores, right, factor scores.

(Refer Slide Time: 19:17)
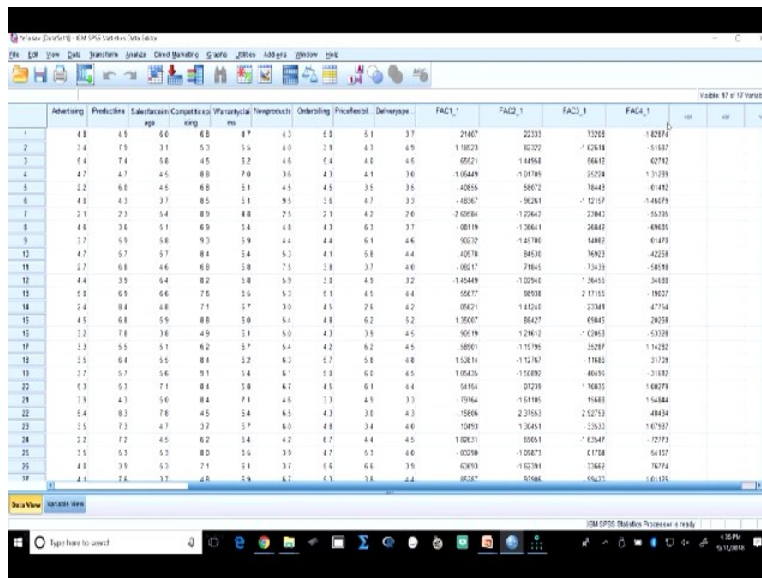
So if you go back to the data set now again, you will find that 5 or 4, I do not know how many, as the number of factors, the same number will be the factor scores. So 4 columns have been inserted into the data set. So these 4 columns are nothing but called factor scores. They are the scores which are representing each factor. Now this factor scores, you can use it for your study in some other studies, may be in some regression or some other study where you want, insert it as an independent variable, right.

But also remember, the same thing could have been done had you taken the summated scale. So because in summated scale also there were 4 factors. So you only did, what you did was? You took the variables only linked to that particular factors. So let us say, factor 1, let us go back to the slide. So in this case, the difference between the summated and factor score is that, while calculating factor score, the software or using, is taking, considering all the variables, right, all the variables.

But in summated scale, the difference is, while calculating summated scale, we are using only those variables that are linked to that particular factor, that is the difference. So if you are using factor 1, let us say, v_1, v_2, v_3 are the 3 factors, then only these 3 will be considered. And v_4, v_5, v_6, v_7, v_8, v_9 are not to be considered.

But while calculating the factor score, all the variables are considered to calculate the factor score, right. And obviously the weightage is more to the ones which are having highest connection with it. So these 2 methods but most of it, now it is, if you see, mostly the summated scale is used, right.

(Refer Slide Time: 21:00)

Now coming to this concept of, we just spoke about validity and reliability. What is this validity? And what is reliability? So validity as it says, what does it mean? The soundness or appropriateness of a test or instrument in measuring what it is designed to measure, right. Who

said? Vincent 1999. He says that whether an instrument is actually measuring what it is intended to measure or it is not measuring.

So if it is doing the work, then it is called a valid instrument, right. On the other hand, he says, there is another term which is important for any construct or factor which needs to be measured, is that. The degree to which a test or measure produces the same scores when applied in same or other, different other circumstances, right. Then we will say it is reliability. Suppose if you are going to measure your weight, for example.

So you measure your weight in 3 different instruments, right. For example, a weighing machine or 3 different types of weighing machines. Then if they are, if it is a weighing machine, then we will say it is a valid instrument, correct. But then suppose let us say, let us do one thing. If you go to weighing machine and check your weight first, first time. Then you again check your weight and you again check your weight.

And suppose, the deviation in the weights is not much. Let us say in first one you were, let us say, 80 kg. Second one, you were 79.5 kg. Third one, let us say, 79.8 kg. Then we can say it is more or less a reliable machine. But suppose you go to another machine and it gives you 80, let us say, then 78 and then 83. Then it is a not reliability machine. So that is what it says, right. So our measure, our construct, our scale or factors have to pass this test of validity and reliability, okay. Now this is the example you see.

(Refer Slide Time: 23:00)

Now target A, poor validity. So it should be hitting here, right. But it is hitting somewhere here. But it is a good reliability. Why? Because all the values are more or less giving the same result again and again. This case, poor validity, not hitting anywhere in the center and even reliability is not good. Because it is haphazard, right. Now look at this. This is a case of a good validity, right and also a good reliability, okay. So this is what we are understanding from the validity and reliability.

(Refer Slide Time: 23:30)

And I showed you how to find out the reliability, right. So when you talk about reliability, you can do it through a Cronbach's alpha which should be, the values should be above 0.6, so should be above 0.6. If it is above 0.6, we say it is decent enough, right. This also gets affected by the

sample size, okay. So the other thing is the validity. So now, for now, we will talk about, when I talk about validity, I will talk about face validity or content validity, face or content.

That means what? What we do here is, we try to take the opinion of experts and see whether the instrument that I am using, is this instrument good enough to find out what my objective of the study is. So to do that, we do a face validity. That means the instrument, we are checking the instrument through some experts and saying kindly check it for us and say whether I have used the right questions or the right instruments or not, right.

Then you have content validity, face validity. Then you have other validity, types of validity also. Nomological validity, discriminant validity, right, construct validity, convergent validity. So what is convergent validity? When the factors, as I said, factor 1 has got, let us say, v_1, v_2, v_3. If the items, the correlation among this items is high, let us say above 0.7 or more, then we say it is a strong correlationship, right.

That means we will say that there is a convergent validity. So convergent validity is also measured through the correlation among the variables, okay. Now this is what I will continue with the reliability and validity in the next section also when I will get into confirmatory factor analysis, okay. Because there it is more deeply explained. Now how do you report. Finally you have written everything.

You have done a factor analysis. Now you have to report. How do you report? So if you create a factor based scale, describe like this. So the report is, what is the theoretical rationale for EFA. Why did you do EFA? Suppose you say well I did a study to understand the suppliers behavior. So to understand the suppliers behavior, I needed to do a factor analysis. I had asked several questions, there were several items in my study.

And when I had too large study, number of variables, the size of, the number of variables were very high, so I needed a exploratory factor analysis to squeeze it or to reduce it to a few meaningful factors. Then detail description of the subjects. So what is the questions you asked and what are the factors that were generated, you need to explain it very clearly, okay. Then also

include the descriptive statistics.

So the description, like the mean, the number of occurrence, the maximum value, minimum value, the standard deviation, also like that. Show the correlation matrix, why? It will tell you which variables are correlated with other and in which way, okay. Then you say what did the, what was the method you used? Was it the principal component analysis, was it a common factor analysis, right.

So I had already explained the principal component analysis and factor analysis. So and then you can also write the communality estimates and the factor extraction if it is required. You can omit this part, right. And what was the kind of rotation you had used to bring in some sanity into the study. Suppose your study was not behaving properly or it was not giving a good result. Then you should do a rotation.

Otherwise, you need not do a rotation. Suppose you found that most of the variables are rolled into only 1 factor, then we will do a factor rotation. So which type of factor rotation did you do? For example, a varimax, or equimax, promax, whatever. And what was the criteria employed for the number of factors. For example, as I said, so we will not take any loading less than 0.3, right. So that is a cut-off value.

So you have to mention that. And the meaningful loading. So any loading which was above 0.5 was only considered, right. So 0.5, 0.3, whatever it is, right. So saying all this, then you have to write it in your research report. So then finally you will say after doing all this, I got 3 or 4 factors and this factors explained; one thing is not given, the variance.

During here, how much variance in the study was explained through this factor analysis. That also you have to mention. When you do this, right, then I have already explained how to do the, conducting the factor analysis in SPSS. So need not go further. So when you write this, then your factor analysis, exploratory factor analysis is over.

So in the next lecture, what I will do is, I will continue with a new technique which is called the

confirmatory factor analysis, right. So confirmatory factor analysis is also a very important tool. Now it is being used largely in various researchers. So what does it mean and how it is done, I will explain in the next lecture. Thank you so much.