Marketing Research and Analysis -II (Application Oriented)

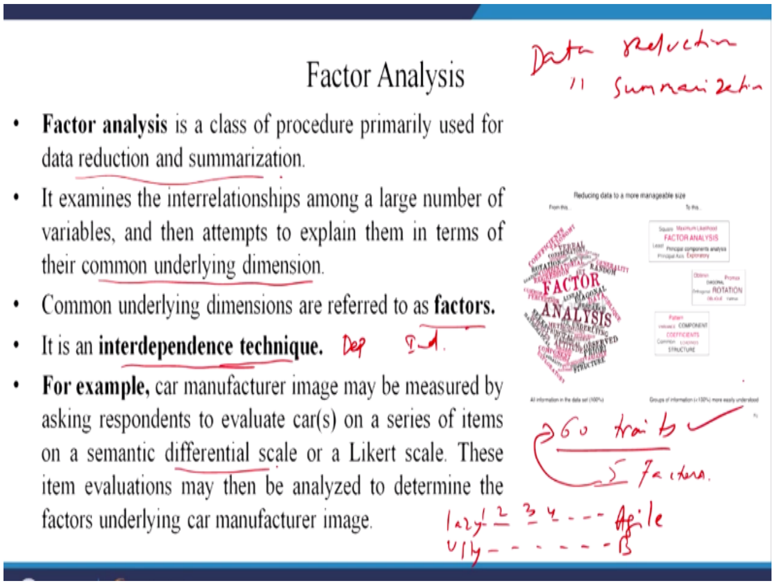
Prof.Jogendra Kumar Nayak
Department of Management Studies
Indian Institute of Technology - Roorkee

Lecture – 51
Factor Analysis in SPSS - I

Welcome friends to the course of marketing research and analysis, so today we will be starting with the new topic which I would also discuss earlier in my you know in the first course which I done on marketing research and analysis, there I would explained it theoretically the basic conceptualisation, the basic meaning of factor analysis which is a data summarisation, data reduction technique right.

But today, in this course the objective is that because we found lot of queries from people coming on that they need to understand it through how to use it so, that application part was lacking in the first session, first course, so that is what we have try to incorporate in this one, right, so today we will be starting with this factor analysis in SPSS. So, what is factor analysis basically?

(Refer Slide Time: 01:23)



Factor Analysis

- **Factor analysis** is a class of procedure primarily used for data reduction and summarization.
- It examines the interrelationships among a large number of variables, and then attempts to explain them in terms of their common underlying dimension.
- Common underlying dimensions are referred to as **factors.**
- It is an **interdependence technique.** Dep Ind.
- **For example,** car manufacturer image may be measured by asking respondents to evaluate car(s) on a series of items on a semantic differential scale or a Likert scale. These item evaluations may then be analyzed to determine the factors underlying car manufacturer image.

Data reduction
" Summarization

60 traits
5 factors.

lazy 1 2 3 4 - - - Agile
Ugly - - - - - - B

Reducing data to a more manageable size
From this...
To this...
FACTOR
ANALYSIS
FACTOR ANALYSIS
ROTATION
COMPONENT
COEFFICIENTS
STRUCTURE
All information in the data set (100%)
Groups of information (<100%) more easily understood

Factor analysis is a data reduction technique, sometimes it is called a data reduction or data summarisation, right so, data summarisation technique, so this basically as it says is a class of procedure used for data's reduction summarisation, so it has a lot of utility, so we will see what is the utility, right. It examines the interrelationships among a large number of variables, so suppose you have got a large number of variables let us say during a study, some psychological study you want to understand the personality or the traits of people, right.

So, when you are trying to understand the traits and personality of people, there could be may be 50, 60 such traits which are there and they are and you are observing them but if you start understanding or trying to explain each all these 60 may be 60 traits, then it will becomes very cumbersome, very complex, right, so in such a condition, there is a necessity of reducing this you know, number of variables to a few known factors which can explain it more easily, okay.

So, it is just to make a structure little more simpler that is the meaning, right, so as it says, it examines the relationships among a large number of variables and then attempts to explain them in terms of their common underlying dimension, so what is this common underlying dimension; the common underlying dimension is you have to understand that the whole philosophy of factor analysis is basically it is dependent on the correlation among variables.

That means how 2 variables are correlated, what is the degree of correlation among them that is what actually, is the core of factor analysis, okay, although it looks it is like a regression, it is very similar to a regression only, right, so the heart of the basic algorithm is on the correlation, right. The common underlying dimensions are referred to as factors, so as I said suppose you have 60 traits right, of psychology and this 60 traits are very difficult to understand.

Because if you on go on explaining each one of them, it is very tough, complex, so let us say I do a factor analysis and the objective is to create some few factors out of this and let us say, I generate let us say 5 factors, okay, so that means what has happen; the 60 traits have been condensed, has been compressed to 5 factors, okay and this 5 factors are representative of the 60 traits, okay.

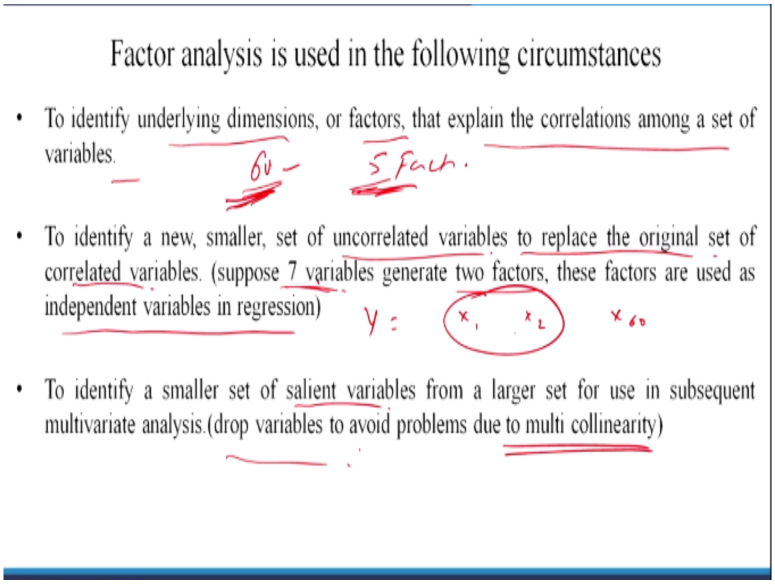
It as an interdependence technique, why it is interdependence technique; because the concept of dependence and dependent variable and independent variable does not exist here, right, so we are not talking about any cause and effect relationship, there is no you know, there is no outcome and there is no predictor, right, so that is why it is called an interdependence technique, all the variables are correlated to each other.

Example; a car manufacturers image may be measured by asking respondents to evaluate cars on a series of items on a semantic differential scale, I hope you understand about the semantic differential scale; a semantic differential scale is a scale in which you have bipolar adjectives, right. So, for example, a lazy and you have somebody, opposite of lazy is agile, you can say agile right.

So, there are 7, you know space, you know 7 numbers in between, so this when you; this semantic differential is a scale which is similar to a Likert scale but the only difference is that the Likert scale is an itemised scale where every item is being explained has to be explain, right. For example, when you say 1, 2, 3, 4, 5, 6, 7, what 1 means, what does 2 stand for, what is 3 stand for, what is 4, so the only difference between semantic differential and Likert is this that in the Likert scale, you are explaining what every number means, right.

How do you describe it and in semantic differential, we give 2 bipolar opposites, right let us say, ugly beautiful, lazy agile right or active, beautiful and there are 7 points, right and where does the person; the respondent want to fix its observation or its opinion he can fix it here, right, so this item evaluations may then be analysed to determine the factors underlying a car manufacturers image.

(Refer Slide Time: 05:59)



Factor analysis is used in the following circumstances

- To identify underlying dimensions, or factors, that explain the correlations among a set of variables.
- To identify a new, smaller, set of uncorrelated variables to replace the original set of correlated variables. (suppose 7 variables generate two factors, these factors are used as independent variables in regression)
- To identify a smaller set of salient variables from a larger set for use in subsequent multivariate analysis.(drop variables to avoid problems due to multi collinearity)

So, I just told you the better example is maybe we can take this 60 traits of psychology, okay, factor analysis is used in with circumstances let us see, first is to identifying the underlying dimensions or factors that explain the correlation among a set of variables, so as I said the heart and spirit is the correlation, right, so then once you that means what 60 traits came down to let us say 5 factors, okay.

To identify a new smaller set of uncorrelated variables to replace the original set of correlated, now you see 60 variables where there all related to psychology, now we have got 5 factors, now each factor is separate from each other, when I will talk about validity, we will use the word called term called discriminant validity, right, convergent validity, discriminant validity, what is discriminant validity?

It says that one factor is sufficiently different from another factor, okay so when you have let us say to identify a new smaller set of uncorrelated variables, so these are the uncorrelated variables, to replace the original set of correlated ones, right, to so we are replacing this 60 with this 5, right, suppose 7 variables generate 2 factors, this factors are used as independent variables in regression.

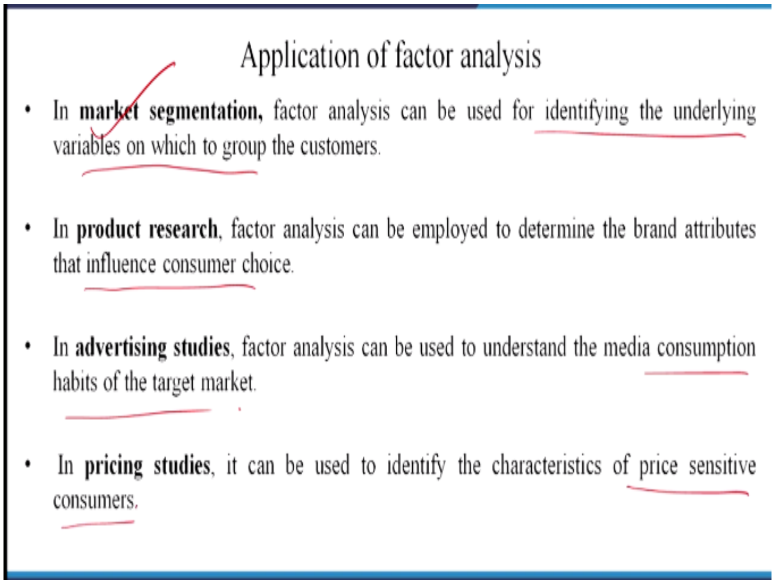
So, the whole objective is suppose, you want to use this factors let us say now or you wanted to use this variables as an independent variable in the regression equation, so y =; so you could have had now in this case up to x60, right, instead of having this x60, what we are doing is; we are using this 5 factors, right, the scores of this 5 factors, so x1, x2, x3, x4, x5, so this may be took; taken in as the you know the variables, right.

Or another example I have given is there are 7 variables, there are 2 factors have been generated and this 2 factors are nothing but they represent this 7 variables, okay, to identify a smaller set of salient variables; important variables from a large set for use in subsequent multivariate analysis, so you can drop variables, so it helps also to you know identified the variables which are highly similar, right.

And which can create the problem of multi collinearity and drop them, right, so you have understood by now when we did regression that multi collinearity is a very serious issue, right and what happens when there is a problem, when multi collinearity exists, if multi collinearity exist the entire you know, coefficients can become dissuaded and the science can become opposite of what they should be.

And similarly, the whole you know, the study can become insignificant, so it can pose serious trouble, right, so it helps you to you know drop variables and understand, by after understanding where there are too much of correlation or extremely strong correlations, okay.

(Refer Slide Time: 08:45)



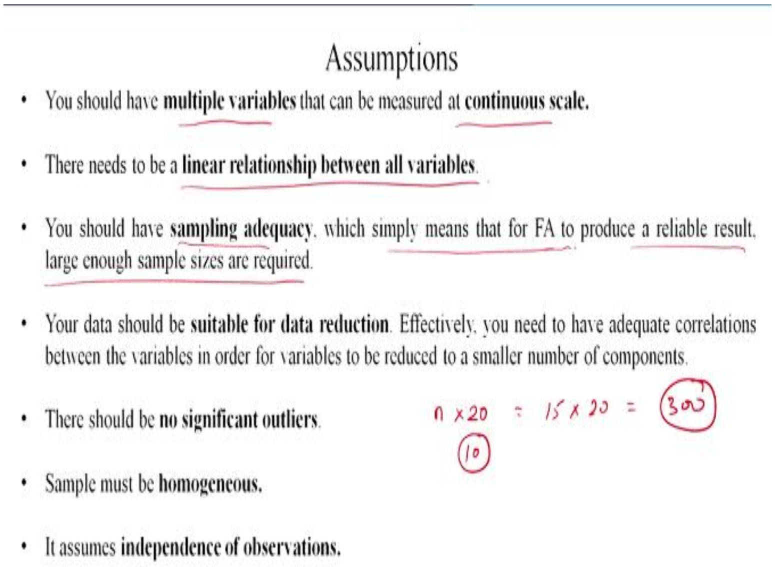
Application of factor analysis

- In **market segmentation,** factor analysis can be used for identifying the underlying variables on which to group the customers.
- In **product research**, factor analysis can be employed to determine the brand attributes that influence consumer choice.
- In **advertising studies**, factor analysis can be used to understand the media consumption habits of the target market.
- In **pricing studies**, it can be used to identify the characteristics of price sensitive consumers.

Some areas of application of factor analysis is market segmentation, so in market segmentation, we use it, can to identify the underlying variables on which to group the customers, how do you want to group the customer, let us say on the psychological traits you can understand and may be 2, 3 psychological groups you can form and say well this group needs this kind of a service or product and this group needs this kind of a product, so that is how you segment the market.

Although, classically we used the cluster analysis for market segmentation but factor analysis also can help you. Product research; factor analysis can be employed to determine the brand attributes that influence consumer choice, fine similarly, advertising studies; it is used to understand the media consumption habits of the target market, right and in pricing studies, it is used to identify the characteristics of price sensitive consumers.

(Refer Slide Time: 09:44)



Assumptions

- You should have **multiple variables** that can be measured at **continuous scale.**
- There needs to be a **linear relationship between all variables**.
- You should have **sampling adequacy**, which simply means that for FA to produce a reliable result, large enough sample sizes are required.
- Your data should be **suitable for data reduction**. Effectively, you need to have adequate correlations between the variables in order for variables to be reduced to a smaller number of components.
- There should be **no significant outliers**.
- Sample must be **homogeneous.**
- It assumes **independence of observations.**

$n \times 20 = 15 \times 20 = \boxed{300}$

$\boxed{10}$

So, there are several uses of factor analysis in the marketing sphere, right, now what are the assumptions; there some assumptions for example, you should have multiple variables which are measured at a continuous scale, so you can measured it in a let us say continuous scale, so for example in a Likert or any interval or ratio scale and then there needs to be more than a sufficient number of variables should be there, so that then only the point of you know, factor analysis is justified.

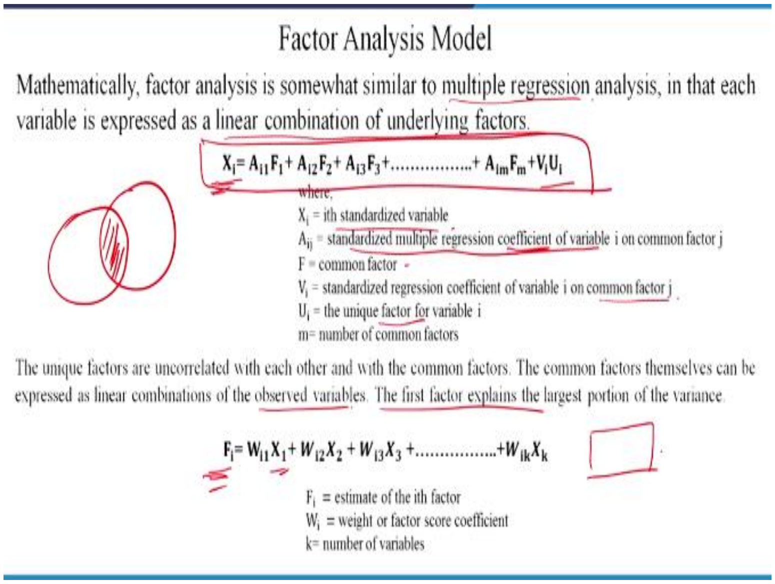
Otherwise, why should you do a factor analysis when only 3 or 4 variables are existing, they needs to be a linear relationship between all the variables that means, there needs to be a linear relationship among the variables which I have explained what is linearity mean, right, so every variable should be corrected linearly in a linear fashion, so that one change; the change in one you know reflects a similar corresponding change in the other, right.

Sampling adequacy which simply means that for factor analysis to produce a reliable result, large sample sizes are required now, I will tell you what are the sample sizes in initially, I have also said I think so for any study, there is a thumb rule that the number of variables, the number of variables into ideally the best is 20, so if you have let us say, 15 variables then you need 20, you know respondents for each of samples, so that comes to 300 sample size, right.

And this can go up to if you do not find it is a difficulty then the you know up to 10 is desirable, right but below 10, it is not desirable, right, so data should be suitable for data reduction, now your data should effectively lead to have adequate correlations between the variables now, this is very important as you will see this is (()) (11:34) which is done to show there are at least some correlation between the variables exist.

If there is no correlation then factor analysis is of no use, you should not be doing factor analysis at all, okay, outliers should not be present as much as possible, the sample must be homogeneous and interdependence of observation, again the same sample must not be repeated, right. So, how does mathematically a factor analysis look like?

(Refer Slide Time: 12:05)



Factor Analysis Model

Mathematically, factor analysis is somewhat similar to multiple regression analysis, in that each variable is expressed as a linear combination of underlying factors.

$$X_i = A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \dots\dots\dots\dots + A_{im}F_m + V_iU_i$$

where,
X_i = ith standardized variable
A_{ij} = standardized multiple regression coefficient of variable i on common factor j
F = common factor
V_i = standardized regression coefficient of variable i on common factor j
U_i = the unique factor for variable i
m= number of common factors

The unique factors are uncorrelated with each other and with the common factors. The common factors themselves can be expressed as linear combinations of the observed variables. The first factor explains the largest portion of the variance.

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \dots\dots\dots\dots + W_{ik}X_k$$

F_i = estimate of the ith factor
W_i = weight or factor score coefficient
k= number of variables

Like multiplication regression, factor analysis also is expressed similarly, here in that you see mathematically, factor analysis somewhat similar to multiple regression analysis in that each variable is expressed as a linear combination of the underlying factors, now let us see, so X, right =; X is the standardised variable, right, Xi is the standardised variable, A is my standardised multiple regression coefficient, okay F1 is my common factor.

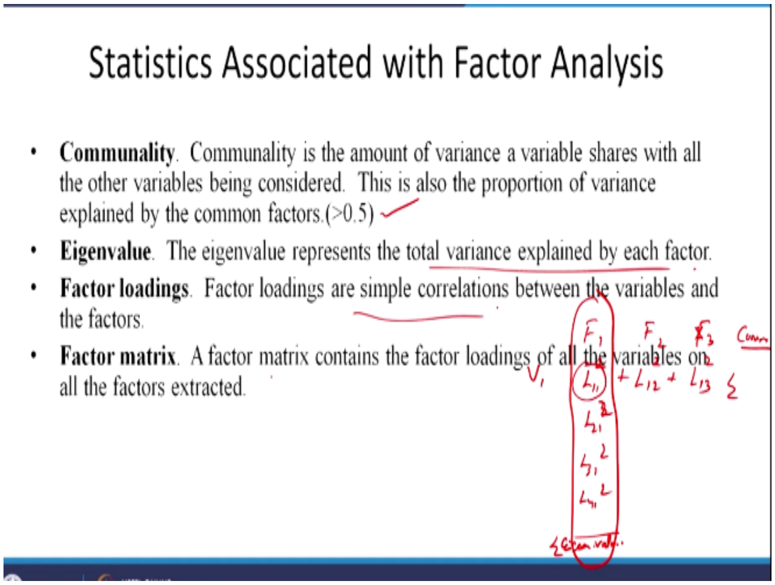
So, F1, F2 are my factors, right, A is my standardised multiple regression coefficient, so as you of by now you have understood the coefficients like you have the slope the coefficient; beta coefficients, right, so Vi is the standardisation regression coefficient of variable I on common factor j, right and Ui is the unique factor, now there are two terms; the unique factor and the common factor.

So, for example if you when like a shared variance, so something you know this is shared or common, right and some are unique, so it talks about the these are the unique factor for variable I and the standardised regression coefficient, anyway just understand this, if you want; if this is; it looks more like a regression equation; multiple regression equation, right. The unique factors are uncorrelated with each other and with the common factors.

The common factors themselves can be expressed as linear combinations of the observed variables, let us see the first factor. The first factor you should remember in factor analysis when you get several factors, the first factors explain the highest amount of variance, the second factor express the second highest amount of variance, the third factor express the third highest amount of variance, right.

So, as you go on, you see the first factor let us say the Fi is the F1, so estimate of the highest factor is equal to the weight of the factor score coefficient, weight or factor score coefficient, now factor score I will explain, in the next slide, maybe it is there, into the X is a variable, right, so this will tell us the; this will give us the individual factors variance or how much of variance in the study is this particular factor you know explaining, right.

(Refer Slide Time: 14:28)



Statistics Associated with Factor Analysis

- **Communality**. Communality is the amount of variance a variable shares with all the other variables being considered. This is also the proportion of variance explained by the common factors.(>0.5)
- **Eigenvalue**. The eigenvalue represents the total variance explained by each factor.
- **Factor loadings**. Factor loadings are simple correlations between the variables and the factors.
- **Factor matrix**. A factor matrix contains the factor loadings of all the variables on all the factors extracted.

So, there is some statistics associated with factor analysis, first is communality; what is this communality; is the amount of variance, a variable shares with all other variables being considered now, how does it look like, for example let me show you both eigenvalue and the communality at the same time. Eigen value represent the total variance explained by each factor, now let us say these are the factors; F1, F2, F3, right, so the loadings let us say this is L11, L12, L13, right.

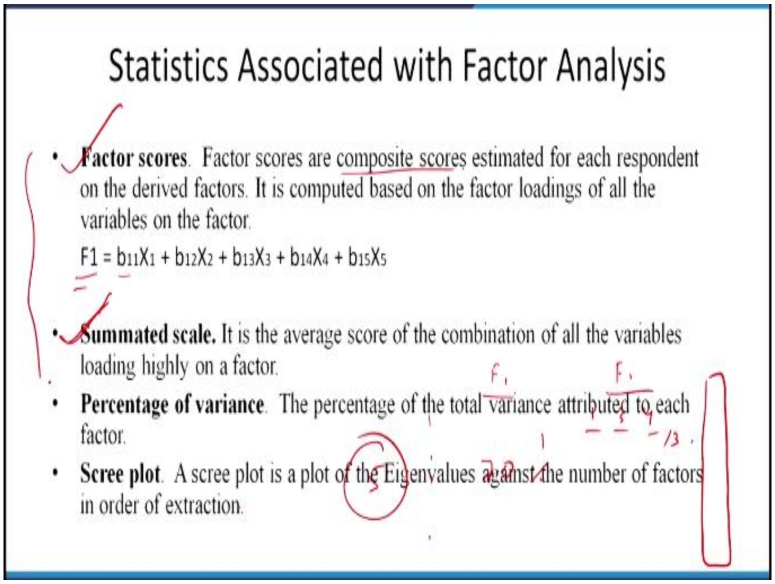
So, there are my variables this side and what I do is; I just take the square of these; right, the sum of the square of this and this summated you know value is what is called my communality, right, so this is my communality and when I take the across the variables as I take the summation of this on the vertical, so the summation here of let us say for factor one let us say, this is called my eigenvalue, right.

So, as you can see communalities amount of variance, a variable shares, right with all the other variables now, this variable let us say V1, how much it is sharing with the all the other variables through the factors let us say factor 1, factor 2, factor 3, let us say, how much it is sharing, this is also the proportion of variance explained by the common factors, right and the communality should be more than .5, should always be > .5, right.

Similarly, if you look at the eigenvalue, the eigenvalue represent the total variance explained by each factors, so F1 explains how much, now this is the square, so this let us say 21 square sorry, square, 31 square, 41 square, so the summation of this is called my eigenvalue, right, so this is very simple to understand and this tells me what; they explains the total variance of each explained by each factor, the eigenvalue represent the total variance explained by each factor, okay.

So, similarly for F2 and F3, we can find out, what are factor loadings; these are simple correlations, so these are the factor loadings, right, the simple correlations between the variables and the factors, so the factor and the variable, these are the loadings, so this loadings are nothing but simple correlations, right factor matrix. A factor matrix contains the factor loadings of all the variables on all the factors extracted.

(Refer Slide Time: 17:06)



Statistics Associated with Factor Analysis

- **Factor scores**. Factor scores are composite scores estimated for each respondent on the derived factors. It is computed based on the factor loadings of all the variables on the factor.

  $F1 = b_{11}X_1 + b_{12}X_2 + b_{13}X_3 + b_{14}X_4 + b_{15}X_5$

- **Summated scale.** It is the average score of the combination of all the variables loading highly on a factor.
- **Percentage of variance**. The percentage of the total variance attributed to each factor.
- **Scree plot**. A scree plot is a plot of the Eigenvalues against the number of factors in order of extraction.

So, if I do draw the entire chart here, so this will be my factor matrix, okay, factor score; now, what is this factor score and what is its use; we will see later on but just now to understand, factor scores are composite scores, right, estimated for a each respondent on the derived factors, it is computed based on the factor loadings of all the variables, so factor score for example here you see factor and score is b11X1 + b12X2, b13X3, it goes on.

So, this is another term which is similar which is called the summated scale, if you can see here, factor scores sometimes during factor analysis or after doing a factor analysis, we use the output of the factor analysis or the factors for you know makes it for to conducting some other studies for example, a regression study right, so there how what are the variables we will take; so we can take the factor scores as the variables, right, so we can take the factor scores.

Or you may take another term called a summated scale, so what is this summated scale and there was a difference between the factor score and summated scale, I will explain, so summated scale is the average score of the combination of all the variables loading highly on a factor, so that means what; when I am doing a factor score or I am doing a summated scale, there are 2 differences; there is one difference, sorry.

Thus factor score I may use it as an input for a multiple regression analysis or I may use the you know the variable after doing a summated scale analysis now, what is the difference but? The factor score you remember is derived by using all the variables used in the study, right but summated scale only uses the respective variables which are loaded on to the factor, are you getting my point?

Suppose, there are 10 variables, so if you want to do a calculate a factor score, so all the 10 variables loadings are to be taken but suppose, I am doing a factor summated scale and factor 1 has got loaded with only 1, 3 and 4, the variable number 1, 3 and 4, then only the you know the each respondent score will be summated; summed up that means what; he has given the score for one; question number one or the variable 1, variable 3 and variable 4.

This will be taken summated and the average will be recorded, so this and this new column which we generate, right is called my summated scale, right, so this summated scale has a very large implication and it is sometimes considered better and easy then instead of taking the factor score because factor scores have their own complications also, they have the own difficulties because they change every time with a change in the respondents, so calculation becomes extremely complex okay.

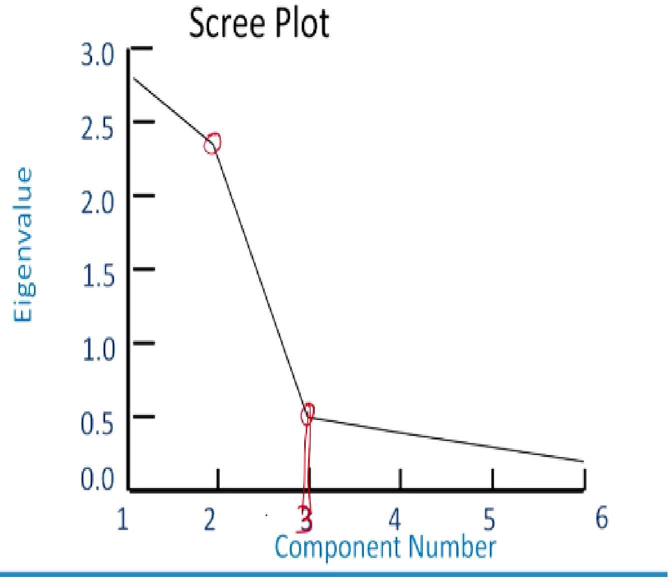
Percentage of variance; the percentage of the total variance attributed to each factor, now during a study when we will conduct a factor analysis, you will see that every factor contributes something to the variance, right variance means how much is the explained, how much you know of explanation is being done by this factor let us say, so there are let us say 5 factors, so 5 factors let us say are explaining us 70% of the study.

So, then we have to have a you know understanding is it is sufficient or not; sometimes when you are doing a factor analysis, there will be some loss of variance that means explanation power will be lost, there is no doubt about it but how much are we losing; are we losing too much of explanation power, if we are losing too much of explanation power then our study is not doing well, right.

So, whenever you draw factors, you should ensure that your factors are sufficiently explaining a large amount of the variance, right, then we can say well, although we have lost some explanation power but it is not much, right, so what is that value of variance that you can expect, at least you should be you know the study should be explaining about 60% minimum in a social science case we will say, the factors should be at least contributing 60% of the variance of the study, right that is minimum, right.

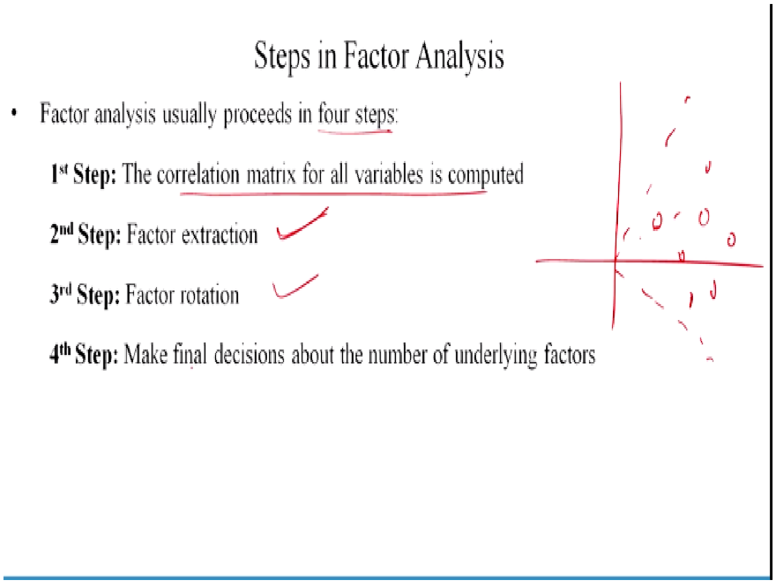
Scree plot; is a plot of the eigenvalues against the number of factors, so it helps to identify the number of factors.

(Refer Slide Time: 21:19)



This is the scree plot you can see, now how do you identify the number of factors to be considered, now you can take the points where it is changing, so 1, 2, right, so after this if you see there is a steady; steadiness, right, so however there is no more change here, so we can say well up to this much, so there are 3 factors to be generated, right, so the scree plot is a diagrammatic representation that is all.

(Refer Slide Time: 21:45)



Steps in Factor Analysis

- Factor analysis usually proceeds in four steps:

  1st **Step:** The correlation matrix for all variables is computed

  2nd **Step:** Factor extraction

  3rd **Step:** Factor rotation

  4th **Step:** Make final decisions about the number of underlying factors

Now, how do you conduct the factor analysis; so you have understood the importance of it, the steps also to some extent, so let us go to the importance of it and where it is used, now let us understand the steps. The first step is; there are 4 steps, the first step is the correlation matrix for all the variables need to be computed, right. First draw a correlation matrix to see that there is some correlation among the variables that exists.

If correlation among the variables is not existing, is very weak then factor analysis is not the right tool, okay, then how to extract the factor, third; that means you are extracting the factor at this stage, third; you are rotating the factors, sometimes we have seen, we will see that while doing it that after you extract the factor, the factors are not properly explaining all the variables, right, most of the variables are maybe are loaded onto one factor only, right.

So, there is a problem, so we will have a problem there, so how do we come out of that problem, then there is a technique called rotation which I will explain, so when you rotate the variables let us say, if this is my axis, so let us say, if I rotate the axis, then what will happen; the variables will be size are in a much better form, so this will help in better explanation of the variables, okay.

So and finally is to make final decisions about the number of the underlying factors, how many factors you want and what are the name, you know you have to give a name to them, right, so these are some of the steps in the factor analysis, so what we can do is; we will continue in this the same you know lecture in the next class but try to understand that factor analysis is a very, very important technique which needs to be very clearly understood.

There is a lot of mis-utilisation also here, people try to you know do anything and everything with you know and run a factor analysis, you should be very clear that factor analysis also needs a theoretical, has needs a very sound theoretical background, right, so when you are doing a factor analysis, when you are creating the factors, it is not possible that there is no you know theoretical explanation behind it, it has to have a theoretical explanation, right.

Because if the variables are correlated, we should know why they are correlated and whether they should be correlated or not that has to be there, right, so sometimes people can you know, why I am saying this is because people can give you; respondents can give you a scores without checking or you know not in a very serious way, so there wrong correlations can an generate and wrong factors can generate.

Because the wrong factors means in the sense, the factors may encompass wrong variables, right, so for that reason you need to be very careful that a theoretical understanding also has to be there, right, anyway we will continue this in the next class, right, next lecture where I will explain the details about the factor analysis, thank you so much.