

Marketing Research and Analysis-II
Prof. Jogendra Kumar Nayak
Department of Management Studies
Indian Institute of Technology – Roorkee

Lecture - 59
Cluster Analysis in Practice-I

Welcome everyone to the class of marketing research and analysis. So till now we have covered almost all the, you know, different techniques involved in marketing research. So we have talked about majorly hypothesis testing and techniques like regression, analysis of variance and you know this discriminant analysis, logistic regression so SEM structural equation modeling path analysis and factor analysis.

So today we will be doing one more technique which is called the cluster analysis. So what is cluster analysis by the way and how, what is it similar to and how is it different from other techniques. So to understand that let us first understand that cluster analysis is a technique very close to the factor analysis right. So as factor analysis you had done and you know that in factor analysis.

The basic objective of factor analysis is to do data summarization and data reduction right. Similarly cluster analysis helps you to create groups among similar objects right. So the only difference between the cluster and the factor is that, that in factor analysis you are you know the basis of grouping was the variables that means on basis of the various variables you are creating a data reduction technique.

And thus few of the, you know, a large number of the variables were being grouped in a few ones right, but here the objective in cluster analysis is not to group the variables, but rather the objective is to group the cases means the respondents. So can you know that means the objective is to see can the group or create groups out of a large number of respondent into a few numbers.

So that we can say well a particular group let us say one particular group has a particular kind of behavior and another group has another kind of behavior. So accordingly maybe a marketer can use that knowledge for his or own benefit okay. So let us see let us get into it now.

(Refer Slide Time: 02:35)

Cluster analysis

- ❑ Cluster analysis is a group of multivariate techniques whose primary purpose is to group objects (e.g., respondents, products, or other entities) based on the characteristics they possess.
- ❑ It is a means of grouping records based upon attributes that make them similar.
- ❑ If plotted geometrically, the objects within the clusters will be close together, while the distance between clusters will be farther apart. (homogeneous within and heterogeneous between)
- ❑ Cluster analysis is also called classification analysis or numerical taxonomy.



So cluster analysis is a group of multivariate technique right whose primary purpose is to group objects based on the characteristics they possess right. So it is a basically an interdependence techniques again like factor analysis it is an interdependence technique. So there is no dependent and independent variable as such right. It is a means of grouping records based upon attributes that make them similar.

So on what basis on what attributes what are the attribute that make similarity creates a similarity among the respondent is what is of interest to the researcher. If plotted geometrically the objects within the clusters will be closed together while the distance between the clusters will be farther apart that means if there are 2 clusters let us say this is cluster 1 and cluster 2.

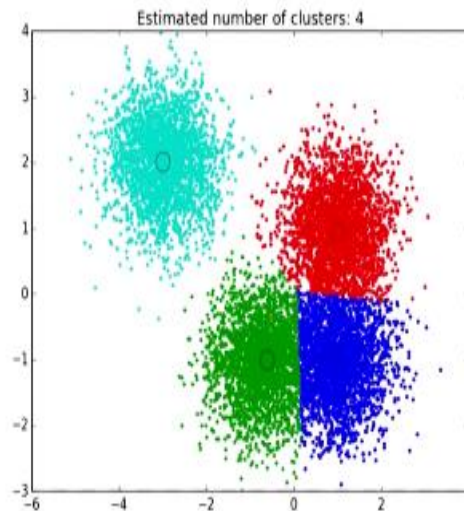
So what we are saying here is them, you know, the data points within the cluster are as homogenous as possible and 2 different clusters are as far as possible. So there is a clear cut distinction right. So that means it is homogenous within the cluster and heterogeneous between right. Cluster analysis is also called classification analysis or numerical taxonomy (03:50). Now if you have read in childhood when you are doing you know classes of biology and zoology you are studying about different species right.

How different animals can be grouped into different species or different species basically right. So we somewhere into like for example reptiles somewhere into Homo sapiens for example human being is a Homo sapiens, some was getting into fungi, somebody into algae

so there were different species right of plants and animals. So according to the characteristics we were classifying them. So how was it done?

This was done on the basis of with the help of such techniques like cluster analysis right.

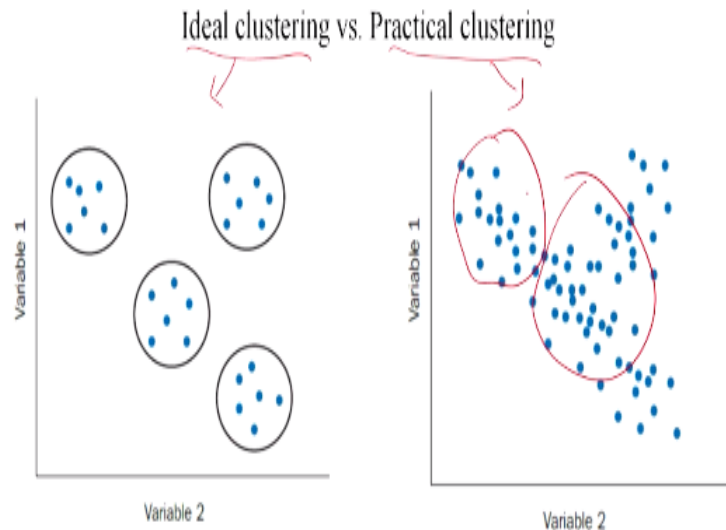
(Refer Slide Time: 04:31)



So you see for example this is just an example where it says if you can see there are estimated number of cluster is 4 all shown different colors. So this is a slightly greenish color sorry light blue, this is dark blue green and red. Now each one is showing depicting a particular cluster. Now if you can see this cluster is the farthest from the other clusters, but these 3 clusters are very close to each other right.

So this is an example to depict how clusters are formed. So you can say if you take it in like for example the species what I said so this could be humans and this could be like the fungi, algae and some other kind of plants or something right.

(Refer Slide Time: 05:18)



The ideal clustering is basically like this it seems to be, but nothing come so clear in real life right. So basically the data points are spread up in this manner in a real life. So practical this is ideal and this is practical. So when you are doing a clustering the data points if it is you know so clearly distinctly separated than nothing like it, but it does not happen in real life. So sometimes that is why you see some of the clusters will show some super imposition.

Some parts of it will try to you know jumble upon each other right. So this is why because this you see this data are so close so it might create a situation like this. So clusters are extremely close to each other.

(Refer Slide Time: 06:04)

Cluster analysis vs Factor analysis

Basis	Cluster analysis	Factor analysis
Technique	<u>Classification technique</u>	<u>Dimension reduction technique</u>
Basis	<u>Grouping is based on the distance (proximity)</u> ✓	<u>Grouping is based on patterns of variation (correlation)</u> ✓
Statistics	<u>Dendrogram, Cluster centroid and Agglomeration schedule.</u>	<u>Eigenvalue, Factor loading, Factor score and Commnality.</u>
	<i>Cases/Respondents</i>	<i>Variables</i>

So the difference between cluster and factor analysis. It is a classification technique this is a dimension reduction technique factor analysis. The basis is the grouping is based on the

distance. Now this is very important. Grouping is based on the patterns of variation using the correlation. So in the factor analysis the major objective was to you know create the groups of the variable on basis of the correlation among them.

But cluster analysis does not take the correlation it takes the distance between the variables okay. Statistics used is the dendrogram it is dendrogram is a diagrammatical representation cluster centroid and agglomeration schedule I will show you what they are this helps you to separate or create the clusters. Here we used if you remember Eigen value, factor loading which was nothing but the correlation with variable and the factor score and commonality.

So these are some of the things that differentiates the cluster and the factor and this is basically is done on cases or respondents this is done on the variables this is the most important simplest and the important you know difference.

(Refer Slide Time: 07:24)

Cluster analysis vs. Discriminant analysis

Basis	Cluster analysis	Discriminant analysis
Prior information	It <u>does not require a priori information</u> about the group or cluster membership for any of the objects.	It requires prior knowledge of the cluster or group membership for each object or case included, to develop the classification rule.
Technique	It is a <u>interdependence technique</u>	It is a <u>dependence technique</u>

Subjective measure

$$\frac{d}{Cate} = \frac{in d}{Cate.}$$

R₀/F how abstract, 2d.

Similarly cluster analysis versus discriminant. If you remember discriminant analysis was a technique which was used the dependent variable was a categorical variable right. So you wanted to take a decision whether somebody should be allowed into the class or not permitted into the class somebody should be given a bank loan or not a bank loan. So there we are using discriminant analysis right.

So what is the difference between these 2 why we are saying this? Now in case of a cluster analysis it does not require any prior information actually what happens is cluster analysis is subjective study right is a subjective measure basically. So any prior information about the

group or membership is not required, but in discriminant analysis prior knowledge of the cluster or group membership for each object or case included to develop the classification rule is required.

For example, when you are discriminating the groups for example there are 2 groups. So in order to discriminate this 2 groups you need very clear information about the groups okay. This is as I said this is an interdependence technique and this is a dependence technique right why because there is a dependent variable which is categorical and there is one or more independent variables which are continuous okay.

For example, as I said pass fail right and this is on hours of study you know IQ of the student etc where is it applied.

(Refer Slide Time: 09:02)

Application of Cluster analysis in Different Fields

- ❑ **Field of psychiatry** - where the characterization of patients on the basis of clusters of symptoms can be useful in the identification of an appropriate form of therapy.
- ❑ **Biology** - used to find groups of genes that have similar functions.
- ❑ **Information Retrieval** - World Wide Web consists of billions of Web pages. Clustering can be used to group these search results into small number of clusters, each of which captures a particular aspect of the query.
- ❑ **Climate** - Understanding the Earth's climate requires finding patterns in the atmosphere and ocean. Cluster analysis has been applied to find patterns in the atmospheric pressure of polar regions and areas of the ocean that have a significant impact on land climate.

In the field of psychiatry where the characterization of patients on the basis of clusters of symptoms can be useful in the identification of an appropriate form of therapy. See in when you talk about the mental health for example there are different kinds of patients everyone cannot be treated similar right or same. So in such a condition it helps us to decide at what level or what level of mental condition is a patient lying.

So that helps us to give different kinds of medicines right. Biology for example to group the genes that have similar functions. Some genes are responsible for your happiness; some genes are responsible for somebody is being aggressive right. So what are the genes that are responsible for these so that helps you to identify you know through a cluster analysis. You

know the application of cluster analysis has improved like anything today right.

In Worldwide Web consist of billions of web pages for example you see. Clustering can be used to group this results into small number of clusters each of which captures a particular aspect of the query. For example, somebody is interested to know about you know what is there under the sea. So anything linked to sea and all that can be cluster into one group something about the sky, the universe the milky way.

So that is another group or sometimes you can find out if sea and reptiles could be taken together maybe that is a cluster. So such kind of clusters can be formed right. Understanding the earth's climate requires finding patterns in the atmosphere and ocean. Cluster analysis has been applied to find patterns in the atmospheric pressure of polar regions in areas of the ocean that have a significant impact on land climate. So such kind of applications are very much there right.

(Refer Slide Time: 11:02)

Purposes of Cluster Analysis in Marketing Research

- ❑ Segmenting the market:
 - For example, consumers may be clustered on the basis of benefits sought from the purchase of a product/service.
- ❑ Understanding buyer behaviors:
 - Cluster analysis can be used to identify homogeneous groups of buyers. Then the buying behavior of each group may be examined separately, as can happen for example with different types of car buyers.
- ❑ Identifying new product opportunities:
 - By clustering brands and products, competitive sets within the market can be determined.
- ❑ Selecting test markets:
 - By grouping cities into homogeneous clusters, it is possible to select comparable cities to test various marketing strategies.
- ❑ Reducing data:
 - Cluster analysis can be used as a general data reduction tool to develop clusters of data that are more manageable than individual observations.

Hand
Spirit
Team

1	2
-	-
-	-
-	-

Similarly cluster analysis has been largely and largely used in marketing right. First of all it is used for segmenting the market. Now on what basis do you segment what is segmentation to divide the markets into several parts so that each part has got some measurable value okay. Example consumers may be clustered on the basis of benefit sought from the purchase of a product or service.

For example, you see sometimes some customers are divided in such patterns for example some are called to be innovators. One who try the product for the very first time and they

want to test it initially, some people are those who want to check the product after only it has been used. So what is the kind of pattern or behavior of these people right. So what factors effect them?

So these kind of benefit sought and things help in segmentation which can be done through cluster analysis. Understanding buyer behavior cluster analysis can be used to identify homogenous group of buyers then the buying behavior of each group may be examined separately. That means suppose I have created 2 clusters let us say 1 and 2. So cluster 1 has got certain behavioral characteristics.

For example, we have taken characteristics like happiness you know sportsmanship right. For example, team person now what is the score that the cluster 1 people are getting what is the score the cluster 2 people are getting. So accordingly we can see which cluster is a better cluster for us. Identifying new product opportunities by clustering brands and products competitive sets within the market can be determined obviously.

So when you can create different cluster obviously then you can target which cluster you like more and would be more beneficial for the company. Sometimes test markets are selected on basis of clustering by grouping the cities into homogenous clusters. So it is possible to select comparable cities to test various marketing strategies. So what you can do is you can create a cluster or cities so any city that falls in a particular cluster are similar right.

So if I want to select let us say my target segment falls into cluster 1 let us say and in cluster 1 there are 5 cities so I can select any one of them and use it in my test market okay. Finally reducing data for example as you are doing in factor analysis. You can do it in cluster analysis too.

(Refer Slide Time: 13:40)

How does Cluster Analysis work?

- The primary objective of cluster analysis is to define the structure of the data by placing the most similar observations into groups.
- To accomplish this task, we must address three basic questions:
 - How do we measure similarity?
 - How do we form clusters?
 - How many groups do we form?

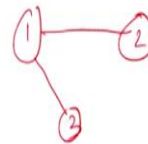
How does it work that is most important? The primary objective is to define the structure of the data by placing the most similar observations into groups right. So it is a matter of similarity or dissimilarity right. So how similar this observation are? To accomplish this task we should know how do we measure similarity, how do you measure right, how do you form clusters and how many groups should you form right.

So how many clusters how do we measure the similarity and how do we then understand these each clusters okay.

(Refer Slide Time: 14:13)

Measuring Similarity

- Similarity represents the degree of correspondence among objects across all of the characteristics used in the analysis.
- It is a set of rules that serve as criteria for grouping or separating items.
- ✓ - Correlational measures.
 - Less frequently used, where large values of r 's do indicate similarity.
- ✓ - Distance Measures.
 - Most often used as a measure of similarity, with higher values representing greater dissimilarity (distance between cases), not similarity.



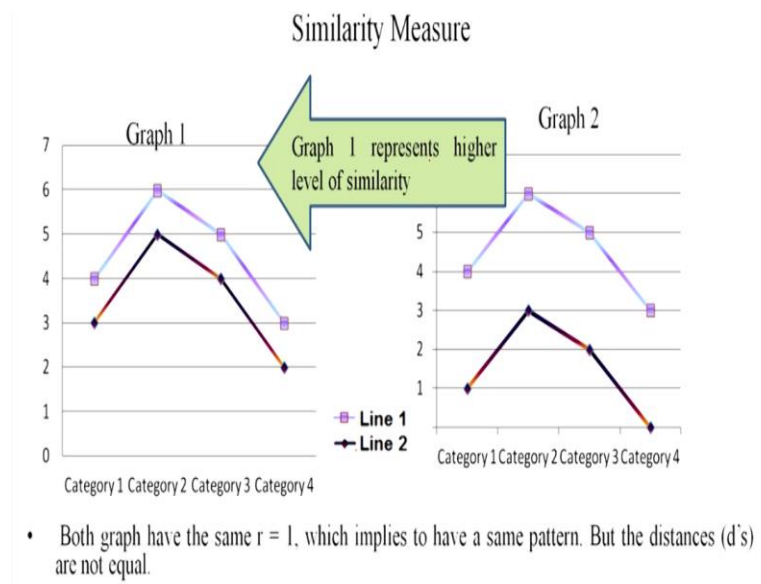
So similarity represents the degree of correspondence among objects across all the characteristics used. Please remember this is what I was trying to say. There are 2 ways you can understand the similarity right. One through the correlation measures which you are

doing in factor analysis. So in the factor analysis we said the variables which are very close to each other or have a high correlation would fall into one factor.

But in the case of cluster analysis we are not taking the correlation measures right. We are often related with the distance measure most often used as a measure of similarity with higher values representing greater dissimilarity distance between cases not similarity that means what if the distance between 2 clusters let us say cluster 1, cluster 2 and let us say this is cluster 3. Now this is 1, 2, 3.

So higher the distance that means farther the 2 clusters that means in a layman terms you can understand that cluster 1 and cluster 2 are more different are much more separate to each other then in comparison to cluster 1 and cluster 3. So this is a closer pair and this is much farther from each other.

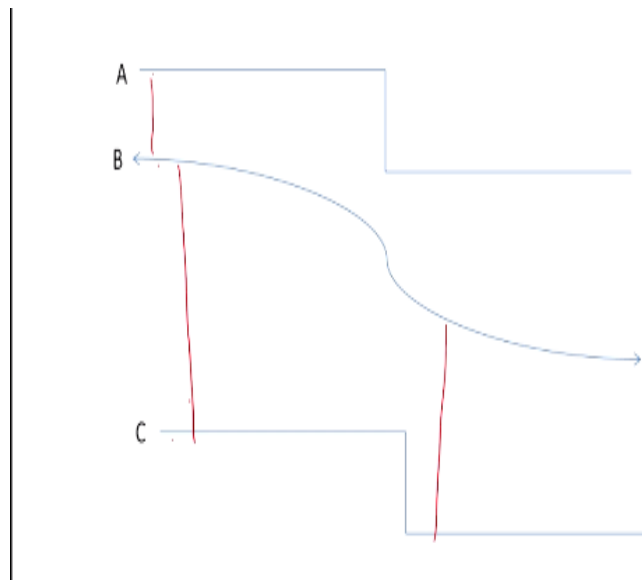
(Refer Slide Time: 14:29)



Look at this similarity measure now both graph if you can see if you look at this right the pattern of the data. So it rises then it falls then it falls more rises, falls, falls more right. So same thing is happening in both the cases. So this is if you check it in terms of correlation if you see both are having a correlation of 1 right. So both graph has the same $r=1$ the correlation=1, but if you look at the distance the distance between the points.

Now you can very easily see that here in this case there is a much farther distance that means the data points are at a much farther from each other right. So that is what it says graph 1 represents higher level of similarity right.

(Refer Slide Time: 16:15)



So look at this for example now this is 2 lines so this is A and C. If you look at the pattern of the data if I put in a data you know they are very similar to each other, but if I now put in a third point right let us say this one the B line and now you can see that the pattern of the data is different. Now if I measure they say this point and this point and this point and this point you see now the distance is becoming different.

At this point it is very less, but at this point it is very high or this point it is very high. So now A and B and B and C are different from each other right. So the distance is different from each other.

(Refer Slide Time: 17:00)

Distance Measures

Several distance measures are available, each with specific characteristics.

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

- Euclidean distance:** It is referred to as straight line distance.
- Squared Euclidean distance:** It is the sum of the squared differences without taking the square root.
- City-block (Manhattan) distance:** It uses the sum of the absolute differences of the variables.
- Chebyshev distance:** It is the maximum of the absolute difference in the clustering variables' values.
- Mahalanobis distance (D^2):** It is a generalized distance measure that accounts for the correlations among variables in a way that weights each variables equally.

So what basically we are doing in a cluster analysis is to measure the distance. Basically we

want to say how far is one cluster from the other or how far is one variable from the other or how far is one data point from the other. This is what we want to see not the correlation not the trend. Correlation is basically the trend the pattern. So we are not talking about the pattern we are talking about the actual distance right and we measure here the Euclidean distance.

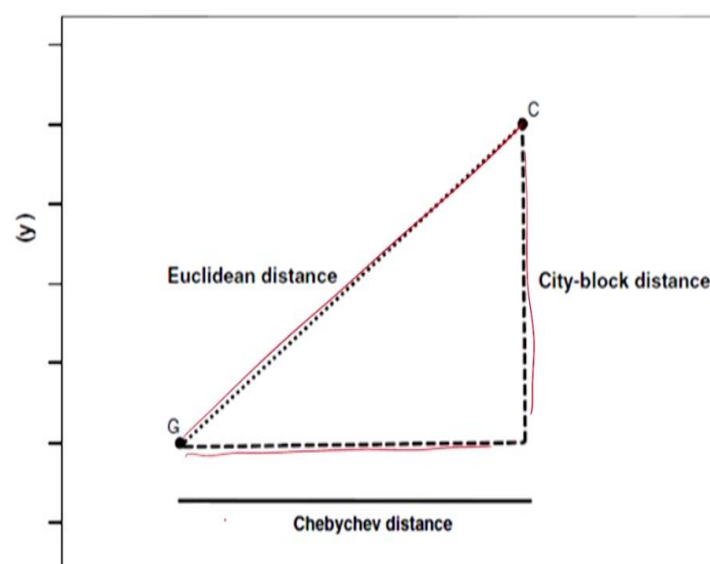
The Euclidean distance is referred to as the straight line right and there are some distance measures for example Euclidean distance which is a straight line distance square Euclidean distance is a sum of the squared differences sum of the square differences without taking the square root right. So for example x we say equation of straight line is

$$= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Now if you just remove the square root then this is the square Euclidean distance. City block distance it uses the sum of the absolute differences of the variables right. I will show you in the next slide. Chebychev is the maximum absolute difference in the clustering variables values and Mahalanobis. This is used when there is a high degree of correlation among the variables in order to avoid that situation.

Whenever you have high correlation among the variable at that condition the Mahalanobis distance is a very right measure to be used right. It basically what it does is it standardizes. It is a generalized distance measure that accounts for the correlations among variables in a way that weights each variably equally then what it basically does it standardizes the variables okay the values.

(Refer Slide Time: 18:49)



Now this is what we are talking about the Euclidean distance, the city-block distance. Now this city-block distance is as you can see you can read it. It uses the sum of the absolute differences of the variables. Now this right+ this right is the city-block distance. The chebychev distance is this part. So there are difference distance measurers okay.

(Refer Slide Time: 19:10)

Example

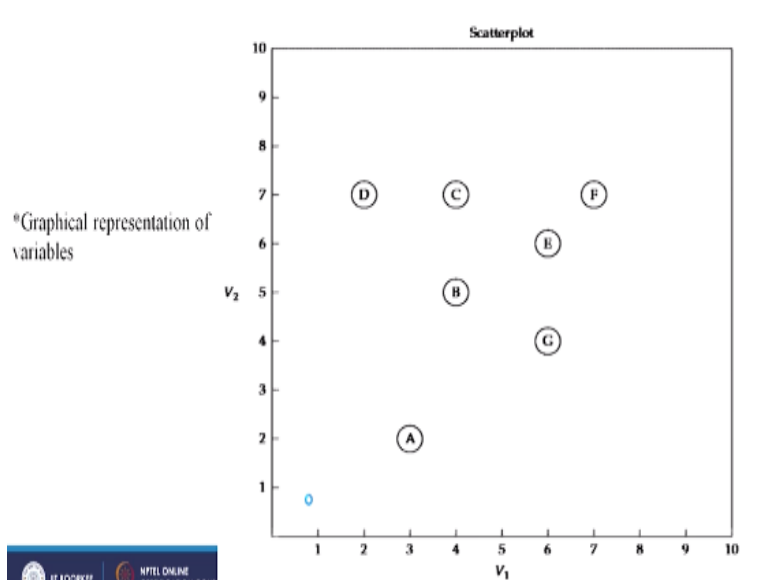
- Suppose a marketing researcher wishes to determine tourism market segments in a wellness destination based on patterns of tourists' emotions (joy or surprise) towards the destination. A small sample of seven respondents is selected as a pilot test of how cluster analysis is applied. Two measures of emotions- V_1 (joy) and V_2 (surprise)- were measured for each respondents on 0-10 scale.

Data Values

Clustering Variable	Respondents						
	A	B	C	D	E	F	G
V_1 ✓	3	4	4	2	6	7	6
V_2 ✓	2	5	7	7	6	7	4

Now let us take this example a market researcher wishes to determine tourism segments in a wellness destination based on patterns of tourist emotions. Now on basis of the emotion he wants to create segments towards the destination. A small sample of seven respondents is selected as a pilot test right and cluster analysis is applied. So two measures of emotions V_1 and V_2 right and seven respondents are there and the data are given to you right. The emotions were measured on each respondent on a 0 to 10 scale okay.

(Refer Slide Time: 19:47)



Now when we represent the same data V1 and V2 you see so the first case was 3, 2 for example so A is 3, 2 right. So B is 4, 5 right so we have plotted it right. So but from here we cannot say anything much okay.

(Refer Slide Time: 20:06)

How do we measure similarity? ✓

Proximity Matrix of Euclidean Distance Between Observations

Observation	Observations						
	A	B	C	D	E	F	G
A	---						
B	*3.162	---					
C	5.099	2.000	---				
D	5.099	2.828	2.000	---			
E	5.000	2.236	2.236	4.123	---		
F	6.403	3.606	3.000	5.000	1.414	---	
G	3.606	2.236	3.606	5.000	2.000	3.162	---

$$d_{\text{euclidean}}(AB) = \sqrt{(V_{1(A)} - V_{1(B)})^2 + (V_{2(A)} - V_{2(B)})^2}$$

$$= \sqrt{(3-4)^2 + (2-5)^2} = 3.162 \quad \checkmark$$

Clustering Variable	Data Values						
	A	B	C	D	E	F	G
V ₁	3	4	4	2	6	7	6
V ₂	2	5	7	7	6	7	4

Now what we done is how do we measure the similarity that was one question we had in our mind how do we measure the similarity. So to do that this is the formula we use. So for example for A and B. So I have put that same table here I have copied it here see A and B.

Now the distance is how do I measure? Now

$$= \sqrt{(V_{1(A)} - V_{1(B)})^2 + (V_{2(A)} - V_{2(B)})^2}$$

$$= \sqrt{(3 - 4)^2 + (2 - 5)^2}$$

$$= 3.162$$

So this is 3.162 similarly for A and C you can do A and D, A and E, A and F, A and G you can do everything right so 3.162 we have found for B and A or A and B same thing right.

Now let us say A and C so

$$= \sqrt{(3 - 4)^2 + (2 - 7)^2}$$

$$= \sqrt{26}$$

$$= 5.099$$

so similarly you can do it for the rest right. So every value has to be written.

(Refer Slide Time: 21:14)

How do we form clusters?

□ Simple rule:

- Identify the two most similar (closest) observations not already in the same cluster and combine them.
- We apply this rule repeatedly to generate a number of cluster solutions, starting with each observation as its own "cluster" and then combining two clusters at a time until all observations are in a single cluster.
- This process is termed as hierarchical procedure because it moves in a stepwise fashion to form an entire range of cluster solutions.
- It is also an agglomerative method because clusters are formed by combining existing clusters

How do we form the cluster identify the two most similar observation not already in the same cluster and combine them? So what do you do is first once you have this values after that you start doing your clustering. So first find the similar clusters and try to combine them right and you continuously do it till you reach the single cluster This process is termed as hierarchical procedure hierarchical because it is in order right.

Because it moves in a stepwise fashion to form an entire range of cluster solutions. It is also an agglomerative method because clusters are formed by combining existing cluster so that is why it is an agglomeration right. So it could be agglomerative or divisive (21:58) we will see right.

(Refer Slide Time: 22:00)

How do we form clusters?					
AGGLOMERATIVE PROCESS			CLUSTER SOLUTION		
Step	Minimum Distance Unclustered Observations	Observation Pair	Cluster Membership	Number of Clusters	Overall Similarity Measure (Average Within-Cluster Distance)
	Initial Solution		(A)(B)(C)(D)(E)(F)(G)	7	0
1	1.414	E-F	(A)(B)(C)(D)(E-F)(G)	6	1.414
2	2.000	E-G	(A)(B)(C)(D)(E-F-G)	5	2.192 ✓
3	2.000	C-D	(A)(B)(C-D)(E-F-G)	4	2.144
4	2.000	B-C	(A)(B-C-D)(E-F-G)	3	2.234
5	2.236	B-E	(A)(B-C-D-E-F-G)	2	2.896
6	3.162	A-B	(A-B-C-D-E-F-G)	1	3.420

□ In steps 1,2,3 and 4, the OSM does not change substantially, which indicates that we are forming other clusters with essentially the same heterogeneity of the existing clusters.

□ When we get to step 5, we see a large increase. This indicates that joining clusters (B-C-D) and (E-F-G) resulted a single cluster that was markedly less homogenous.

So how do you do it? So we have found out this values the initial solution now this is the pair

we have noted that 1.414 so E and F so E and G is this C and D, B and C, B and E, A and B so these are the different values right. Now this is the first initial solution A, B, C, D, E, F, G there is no combination there is no clusters have been clubbed till now. Now first we will start how do we do it. Now we will first club the closest one.

So this is if you see it is put in a order right So E and F first we are clubbing E and F right and then how many clusters are left 1, 2, 3, 4, 5, 6 right. So this values remains the same then next is A, B, C, D, E, F and G so E, F and G we have clubbed. So how do you do here? Now you see now this value you must be thinking how this value has come. Now to do that what you do is E and F how much 1.414 E and G 2 so let us add these two right $1.414+2+$ F and G right.

So F and G you can measure it here so F and G for example so

$$= \sqrt{(7-6)^2 + (7-4)^2}$$

$$= \sqrt{10}$$

3 point something you see right 3.162. So now this+ this let us say this 3.162 kind of a value. So if you add it here and divide it by 3 then you will get this score. So rest you can do it for the others also. So now we have clubbed and you see each by clubbing each step.

Now you have come across to a point where finally you have got one cluster. So in steps 1,2, 3 and 4 the overall similarity measure does not change substantially right 1, 2, 3 and 4 which indicates that we are forming other clusters which essentially with essentially the same heterogeneity of the existing cluster okay. When we get to step 5 we see a large increase step 5 right so there is a large increase okay.

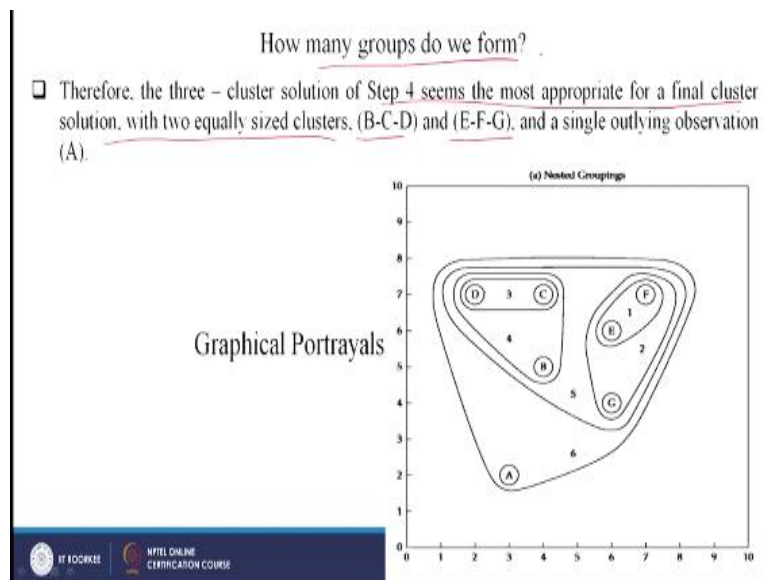
This indicates that joining clusters B, C, D B, C and D this one right B, C, D and E, F, G right resulted in a single cluster that was markedly less homogenous. Okay you just understand that this is how the clusters are formed okay. Therefore, how many groups do we form the question comes. Therefore, the 3 cluster solution of step 4 seems the most appropriate right. Let us see for example here if you look at this values now where is a change maximum.

Now here there is hardly any change there is a change yes there is a big change, but there is hardly any change slight change. Now if you go from the bottom right let us say you go from the bottom. So there is some change here there is change also, but there is slight change and

this is almost no change rather it is increasing. So how many clusters have to be formed we can think accordingly.

So if you see step 4 seems the most appropriate for a final cluster solution with 2 equally sized clusters B, C, D and E, F, G and a single outlying observation A. So this is what we are talking about B, C, D E, F, G and a cluster A. So on the fourth step in the 4 step this part right we are talking about this part.

(Refer Slide Time: 25:44)

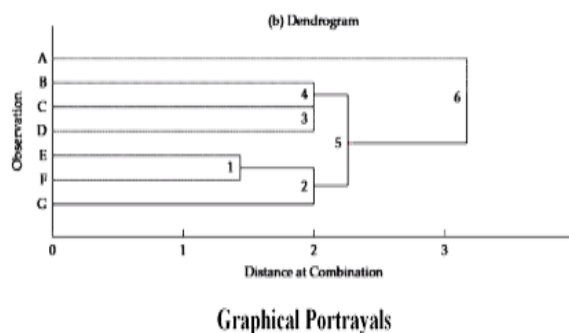


So this is how it looks like B, C, D, E, F, G and A is alone right.

(Refer Slide Time: 25:52)

□ **Dendrogram**

- It is the graphical representation (tree graph) of the results of a hierarchical procedure.
- Starting with each object as a separate cluster, the dendrogram shows graphically how the clusters are combined at each step of the procedure until all are contained in a single cluster



What is the dendrogram it is a graphical representation a tree graph of the results in the hierarchical procedure. So how it has been clubbed you can see here. So starting with each

object as a separate cluster the dendrogram shows graphically how the clusters are combined at each step of the procedure until all are contained in a single cluster. So first you add one then you go on adding to the next so slowly, slowly you go on adding till you get the sixth cluster right.

So let us say this is 1 then with 1 and 2 is combined then after that 3 and 4 are combined then 3 and 4 are combined with 1 and 2 and then this total 5 is combined with 6 so it goes like this. **(Refer Slide Time: 26:36)**

Deriving Clusters

- There are number of different methods that can be used to carry out a cluster analysis which may be classified as follows:
 - ✓ Hierarchical Cluster Analysis ✓
 - ✓ Nonhierarchical Cluster Analysis ✓ (k-means)
 - Combination of both ✓

SSS

How do you derive the clusters there are number of different methods that can be used to carry out a cluster analysis to understand how many clusters should be made? One is the hierarchical cluster analysis the other is a non hierarchical cluster analysis and third is a combination of both. So hierarchical cluster analysis I will just brief you and then we will wind up in the next lecture we will carry on from here.

In the hierarchical cluster analysis as you have seen it helps you to indicate the number of clusters that need to be formed okay. And in the non hierarchical cluster which is basically if you see any test it is shown as a k means cluster right is used to identify and analyze the behavior of each cluster okay. So when these 2 methods are used together A and B this and this are used together it is a combination of both.

It gives a very appropriate (27:26) solution for any researcher to derive the number of clusters and understand them very critically right. So what we will do is we will continue with this lecture in the next this you know topic in the next topic and we will also see how it is being

done in the SPSS right. So thanks for today.