
UNIT 14 CLUSTER ANALYSIS AND MULTI-DIMENSIONAL SCALING

Objectives

After going through this unit on Cluster Analysis, & Multidimensional Scaling you should be able to understand the following

- Importance of Cluster Analysis in today's market conditions
- Inputs for cluster analysis
- Different approaches to cluster analysis
- Nature of Multi-Dimensional Scaling (MDS)
- Applications of MDS

Structure

- 14.1 Introduction
- 14.2 Inputs for Cluster Analysis
- 14.3 Different Approaches to Cluster Analysis
- 14.4 Factors to be considered while using Cluster Analysis Techniques
- 14.5 Multi-dimensional Scaling
- 14.6 Applications of Multi-Dimensional Scaling
- 14.7 Summary
- 14.8 Self-Assessment Exercises
- 14.9 Further Readings

14.1 INTRODUCTION

Cluster analysis is a technique that is used in order to segment a market. The objective is to find out a group of customers in the market place that are homogeneous i.e., they share some characteristics so that they can be classified into one group. The cluster/group so found out should be large enough so that the company can develop it profitably, as the ultimate objective of a company is to serve the customer and earn profits. The group of customers that the company hopes to serve should be large enough for a company so that it is an economically viable proposition for the company. This is also true for the customer as customer would not be willing to pay beyond a certain price for a particular product (price of course is a function of positioning of product, cost of production etc.).

As an example, let us consider the Watch Industry. There could be many ways in which the Watch Industry could be segmented which are as follows

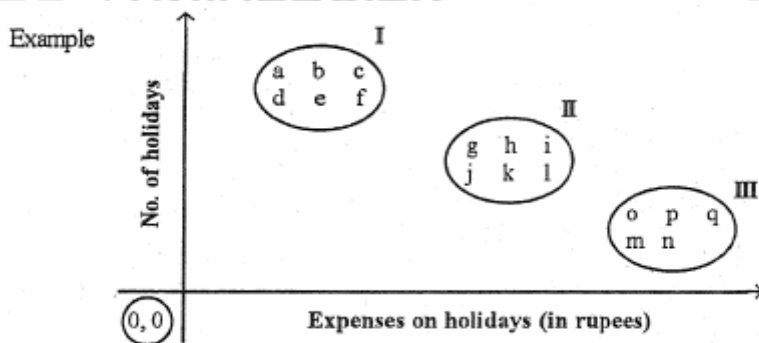
- a) Gender (Male/Female)
- b) Technology (Digital/Analog)
- c) Design Features
- d) Occasion of Use (Formal/Casual/Party)
- e) Price (Low/Medium/High/Jewellery)

Some of the above segmentation factors are demographic (price, gender) whereas some are psychographic factors (occasion to use.)

This, therefore, presents a problem to the market researcher/company, as to how to identify combination of factors that can be used to segment the market place. It is not always



possible to segment a market on the basis of one single factor. Thus, a combination of factors must be used to segment the market place. And this is where Cluster Analysis technique specifically deals with how objects (people, places, products) should be assigned to groups, so that there should be similarity within the groups; and as much difference between the groups, as possible.



In the above example, cluster analysis has identified three distinct clusters (groups) as follows:

Group I: People who spend less money but have more number of vacation days. (adventure seekers, students)'

Group II : People who spend average amount of money and have average number of vacation days (1" class train travel, 3-star hotels etc.)

Group III: People, who have very less vacation days, but spend heavily in these few days (five-star luxury, air travel, foreign trips etc.).

These grouping would help a travel and tourism company in the following ways:

- The travel and tourism company can study the various clusters of customers that have emerged, and decide which customer group it would like to serve, depending upon the company's own resources and capabilities, the volume of business in each cluster group that will generate sufficient business for the company's own survival.
- Once a cluster has been selected by the company, it can tailor various tour programmes for its cluster of customers.
- The company can keep profile of its customers, and identify any new emerging group of cluster.
- The company can decide to serve either only one group of customers (Niche marketing), or serve all groups of customers, at one and the same time, but having a range of travel and tour programmes.

14.2 INPUTS FOR CLUSTER ANALYSIS

Since cluster analysis is used for the purpose of segmentation of the market place, for using the cluster analysis -technique, one has to go through the following procedure on a very broad level:

- Company identifies its own product capabilities, financial strength, distribution network (Who am I?)
- Company identifies the basis on which the customers are to be clustered e.g., one factor segmentation (age, income, gender, geographic location) or two-factor basis of segmentation of the market place (age and income, quality and price, nutrition value and price), or multi-factor basis of segmentation of the market place (Who am I going to serve?).

The company/market research agency has to decide which of the attributes are to be employed that' would generate natural groupings in the market place. The choice of variables that are



to be used to cluster the objects is one of the most critical decisions analysis make. The best way is to select those variables that make sense conceptually, rather than using any or all variables. Ideally, the variables should be selected within the framework of an explicitly stated theory that is used for supporting the classification.

- iii) Respondents who are part of the research work are given to fill up questionnaire. The questionnaire is typically Interval scaled (5pt, 7pt, 10pt) questionnaire.

The cluster analysis program that is run on the computer identifies the different groups (clusters) or respondent groups such that the members of any one group (cluster) are similar to each other, but different from the respondents in the other groups. The computer program for cluster analysis searches through the answers that the respondents give to the various questions in the questionnaire, and groups together those respondents who have given identical or almost similar answers of a certain combination of the questions. These respondents whose answers are similar are formed into one cluster. The procedure is then continued till more clusters are formed.

14.3 DIFFERENT APPROACHES TO CLUSTER ANALYSIS

There are various approaches adopted for cluster analysis

One extreme approach is to cluster respondents in terms of all available information about the respondent - although this is very thorough, but may lead to difficulty in interpretation and marked difference between groups i.e., the clusters or groups of respondents that are produced may be a hotch-potch, rather than clearly marked description of the target population.

Another extreme approach involves grouping respondents only on data that is directly representing the specific consumer need. Example - quality, price of a product. This has the advantage of clarity and accuracy, but may be quite limited in describing the target population.

The way one approaches cluster analysis depends upon the objective of study. Suppose the objective of the company is physical development of an existing product, then the consumers needs regarding the product are the most suitable variables for forming clusters.

Suppose an advertisement for a product is to be developed, then the attitude of the consumer towards the product gains more importance, as there are groups of people who have positive, negative or neutral attitude towards product/company. And the advertisement developed would be targeting specific clusters of consumers, depending upon the product/communication need.

Suppose a company wishes to develop a new product concept/product changes in the existing product offering, then also attitude of consumer towards product/company/product category gains more importance over other attributes when developing clusters,

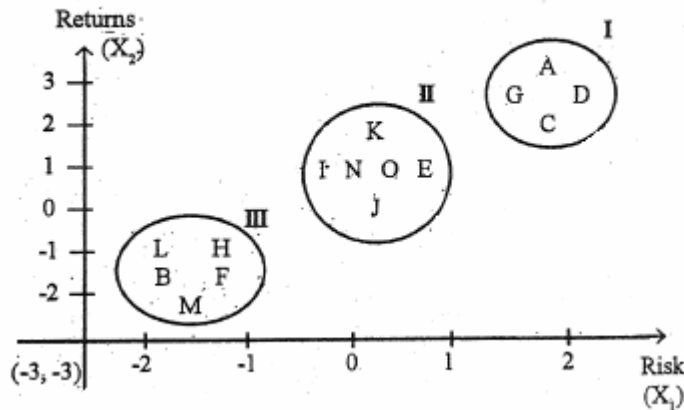
Therefore, depending upon the objective of study, various characteristics can be Used to describe the consumer, his attitudes, needs, demographic characteristics.

A general principle followed in cluster analysis is to omit behavioural and brand image data. One should concentrate on characteristics describing the consumer and the circumstances e.g., specific needs, demographic variables, general attitude and select any variables which would interact with each other to influence consumer behaviour. Once attributes are identified on the basis of which one is trying to form natural groupings of consumers; one has to scale the attributes, so that one has a basis to measure the similarity/differences between/within groups..



Example:

Suppose one wants to understand the consumers who operate in the finance market. With regard to the finance market, there are few outstanding aspects like Risk, Returns, Liquidity. One could use any two aspects at one time to form clusters of consumers or could use more than two attributes also.



There are 15 consumers, whose responses have been analyzed with respect to Risk and Returns in the financial market.

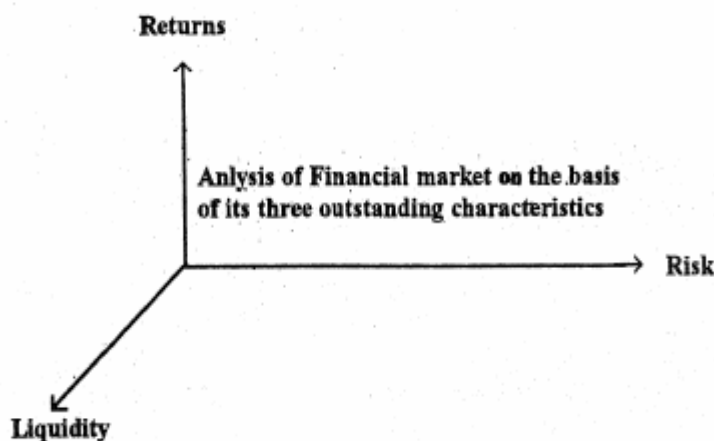
- Cluster I is a group of consumers who believe in High Risk High Returns instruments. e.g., Equity.
- Cluster II is a group of consumers who believe in Average Risk Average Returns (Mutual funds, 'Company fixed - deposits)
- Cluster III is a group of consumers who believe in Low Risk Low Returns (savings bank, fixed deposits).

An obvious way to measure similarity is to calculate the Euclidean distance between points. In the above figure, the distance between points A, and C would be calculated as

$$d(A,C) = \sqrt{(X_{c1} - X_{A1})^2 + (X_{c2} - X_{A2})^2}$$

And so on, one could calculate distance between various points. In three dimensions, the expression for distance between A and C would be

$$d(A,C) = \sqrt{(X_{c1} - X_{A1})^2 + (X_{c2} - X_{A2})^2 + (X_{c3} - X_{A3})^2}$$



		Distance			Risk				
		1	2	3	-	-	-	-	15
		A	B	C	-	-	-	-	O
Returns	1	A							
	2	B							
	3	C							
	-	-							
	-	-							
	15	O							

The distance between any two points is an inverse measure of similarity. Thus, larger the distance between two points, greater is the difference or dissimilarity between the two points. The closer the distance between two points, the more the similarity between them. Also more the number of respondents to be classified into clusters more will be the number of distance calculations. Hence for large sample sizes of respondents, computer would have to be used.

After calculating distances, one has to decide how to form clusters. There are various methods of doing it as follows:

i) Nodal Method

In this method one selects a point or points that will serve as nodes for forming clusters or focal points. The rest of the points are then allocated to each cluster on the basis of their distance from the focal point.

The above method is demonstrated as follows:

- a) Choose as nodes those objects (points) that have the greatest distance between them.
- b) Consider these two objects as focal nodes and allocate all the remaining objects to one cluster or the other based on their distance from the polar nodes.
- c) Split the resulting two clusters in the same way. Continue this process till the collection of points is split into its original members.

In the above example, the points 'D' and 'M' have the biggest distance between them, and they will be considered as nodes for the two clusters. Each of the remaining objects would then be allocated to each cluster on the basis of the shortest distance to either D or M. Thus the original cluster would be:

- Group1 : DACE GK
- Group2 : MBFHJLNO

In Group 1, the least similar (greatest distance) are D and K. They would thus be considered as new nodes, and the items in Group 1 would be allocated to each of the new clusters on the basis of their distance from these new nodes. The new groups would be:

- Group 1A : DACG
- Group1B : KE

Similarly, Group 2 would be divided (using B and J as new nodes) to yield

- Group 2A : BFHLM
- Group 2B : JINO



Thus, after every such stage; one has to find out if the sub-groups that have emerged can be further combined on the basis of some measure of similarity between the objects.

ii) Factor Analysis Method

Another way of developing clusters is by using the method called as Q-factor analysis. By this method one can determine which objects logically belong together: This is also called as inverse factor analysis.

iii) Linkage Method

There are three methods in this Single Linkage Method, Complete Linkage Method and Average Linkage method:

Using the previous example of analyzing financial markets, the method starts by finding out the points with shortest Euclidean distance. Therefore, initially the following groups emerge:

- C-D - Group 1
- A-G - Group 2
- N-O - Group 3
- FM - Group 4

In the next stage, depending upon the cut-off" distance, remaining objects are associated to the above clusters. The final clusters that emerge are

- Group 1 - AGDC
- Group 2 - NOKELT
- Group 3 - MFLBH

The complete linkage option starts out in the same manner by clustering the two closest points. However, the distance between two clusters is the longest distance from a point in the first cluster to a point in the second cluster,

In the average linkage option, the distance between the two clusters is the average distance from points in the first cluster to points in the second cluster.

Activity 1

Study the products and the consumers of your company (or the company you are familiar with) and try to develop cluster of customers depending on the different requirements of the customers.

.....

.....

.....

.....

.....

14.4 FACTORS TO BE CONSIDERED WHILE USING CLUSTER ANALYSIS TECHNIQUE

The following aspects should be kept in mind while using cluster analysis method:

- a) A number of cluster may emerge after doing analysis . However , their is limit to the number of cluster that a company can consider due to
 - Limitation of market potential within a cluster
 - Difference between clusters not sharply defined
- b) Cluster analysis provides a way of segmenting the market but these segments are



- c) The characteristics of a cluster may change over time, as the consumers economic status, education, lifestyle, etc., change over time then the company has to take a relook at the market place.
- d) The clusters that have been identified are used developing further marketing strategies in the areas " like 'product developments, advertising research, distribution strategies, pricing strategies etc,
- e) The most important assumption in cluster analysis is that the basic measure of similarity on which clustering is based is a valid measure of the similarity between objects. A second major assumption is that there is theoretical justification and basis for structuring objects into clusters. As with other multivariate techniques, there should be theory and logic underlying the cluster analysis.
- f) The major limitation of cluster analysis is the difficulty in evaluating the quality of the clusters. It is very, difficult to know exactly which clusters are very similar and which objects are dissimilar, and also difficult to select clustering criterion.

In conclusion, cluster analysis is a scientific method of help understand the consumer groups with 'their differing needs and perceptions.

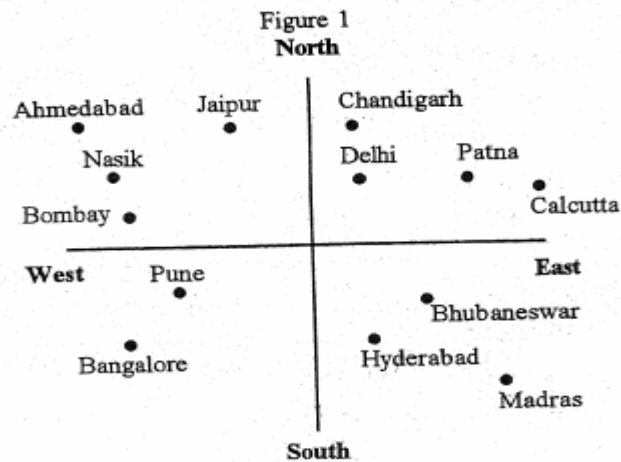
14.5 MULTI-DIMENSIONAL SCALING

Multi –dimensional scaling (MDS) technique or perceptual map or positioning map is a multi- variate analysis technique that is particularly suitable for measuring human perceptions and preference for products/brands. MDS is concerned with spetial representation of relationships among behavioural data i.e., buyer perceptions and preferences.

MDS Fundamentals

- a) Metric MDS
- b) Non-Metric MDS

Consider the following geographic location of various cities in our country:



The distance between any two points on the snap is given by the formula

$$d_{12} = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$$

The distance between various cities could be calculated and put in a tabular form as follow



Cities	Bombay	Pune	Nasik	Ahmedabad	Delhi
Bombay					
Pune					
Nasik					
Ahmedabad					
Delhi					

In short, it is a relatively simple matter to go from the map in figure I to the set of numerical distances in figure 2.

However, the converse or going from figure 2 to figure I is very difficult. Given different objects and their attributes, every consumer perceives the objects in a different manner.

So, one has to first understand the perceptions (distance, etc.) in the mind of the consumer regarding products, brands, and plot a graphical representation for the same. The task is to work backwards. That is, we would have to find out simultaneously, the

- Number of dimensions (can be more than two-dimension perceptual mapping)
- Configuration (or pattern) of points in that perceptual space, so that their computed inter-point distances most closely match the input data of figure 2. This is the problem of metric MDS. Whenever input data is interval-scaled or ratio-scaled, it is called as Metric MDS.

The same input data in the format of figure 2 could be transformed into a rank-order. Suppose we take the smallest distance between two cities and give it the Rank 1. Then we could arrange the remaining distances between cities in ascending order and give it ranks from 2 onwards up to Rank 13 (e.g., largest distance could be between cities Calcutta and Bombay on the map. We therefore find with this method.

- Number of dimensions
- Configuration (pattern) of points in that dimensionality, so that the ranks on their computed inter-point distances most closely matched the ranks of the input data. This is the problem of non-metric MDS. Whenever input data is in the form of nominal or ordinal form, it is called as non-metric MDS.

Whichever method for MDS one selects (metric; and non-metric MDS), it does not change the relative distances of the points.

Attribute Space - Objective Space and Perceived Space.

Any product or service can be visualized as composed of both objective and perceived attributes or 'dimensions'. A company has an objective description of its product in physical or chemical terms. Such an objective attribute 'space' in which various brands are viewed as points positioned in space, will usually not agree with the buyers' perception of the products. The buyers' perception about a product, service, company is a sum total of not only the company which is selling the product, but also other factors in the market like competition, competitor's products Advertising, after-sales service etc.

Also note that the dimensions of perceived space need not be the same as the objective space. As an example, for many consumer goods that are highly individualistic items like tea, coffee, toothpastes, perfumes, shampoos, -where strong individual preferences play a key role in buying decisions, the objective space of the company and perceived space of the consumer may be widely different. But in the case of industrial product, the perceived configuration of brands may agree rather closely with an objectively constructed configuration in which



measurements of such characteristics as speed, reliability, efficiency can be made rather straightforwardly.

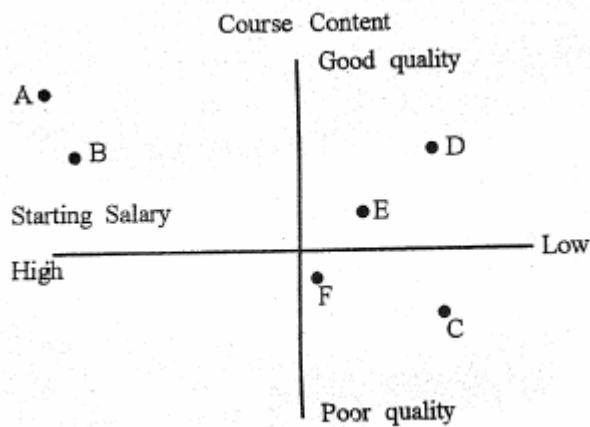
Ultimately, the view of the consumer is the most relevant one, as this is what would make the consumer go out and buy a particular product/brand.

Example of Multi-Dimensional Scaling

A study was conducted in a leading business school (A) among the first year MBA students to find out how other, business schools were 'similar' or 'different' from the 'A' business school. Various attributes are used to evaluate a business school - faculty, faculty/student ratio, starting salaries of the MBA students, physical facilities like library, computers, and so on. But which of the attributes do students 'actually' use in making a competitive comparison between the business schools?

Let us consider six leading business schools in a place as A, B, C, D, E, F. Around hundred students were given a questionnaire in which the students were asked to respond to the questions which indicated various attributes of a management school, and students had to mark schools that were similar/dissimilar on the different counts.

From the information that was gathered from the questionnaire and subjected to NOS software program, what emerged was that course content and starting salary are the most important criteria by which, students judge a business school.



The multi-dimensional scaling technique is therefore very useful in plotting the perceived perception of the consumer about different brands/products/services. And this analysis could be used by the marketing manager of a company for developing further marketing strategies.

Activity 2

Consider the following products and try to develop a perceptual map of the various brands in that particular product category. Before attempting such an exercise, the student should understand the product category in its totality i.e., target consumer, pricing strategy, advertisement strategy, distribution, usage and attitude for the product, and then try to develop a positioning map, (NOS) for the same: Once a product has been studied, the study should try to identify two dimensions along which 'brands could be clubbed as similar/ dissimilar.

- 1) Toothpaste (Colgate Gel, Colgate Total, Pepsodent 2-in-1, Babool, Close-up etc.)
- 2) Shampoo (Kyle Herbal, Dabur Vatika, Sunsilk, Lakme, Clinic etc.).

.....

.....

.....

.....

.....



14.6 APPLICATIONS OF MULTI-DIMENSIONAL SCALING

Some of the typical marketing applications that emerge from the MDS technique are

1) **Market Segmentation**

Market segmentation is the technique of trying to identify groups of consumers who exhibit commonality of perception of products and preferences. One can use MDS techniques to identify present perceptions of products by consumers, and use it to modify the company's product, package, advertising, additional features, so that the product offering of the company moves more and more closer to the 'ideal' requirement of the consumer.

2) **Advertisement Evaluation**

The MDS technique could be used at the stage of advertisement pre-testing. Once an advertisement has been developed, it could be tested for similarity/dissimilarity with other advertisements in the same product category. As the ultimate objective of an advertisement is to communicate with the target consumer effectively, and this is possible only if the advertisement is distinct in its message from the other competing advertisements,

3) **Product Re-positioning Studies**

If a company is interested in re-positioning its product/service (in the mind of the consumer), the first and foremost activity to be done is to assess the current perception of the product in the mind of the consumer. The classic re-positioning case is that of Cadbury chocolates, which kept on assessing its positioning platform, and successfully moved Chocolates from a product perceived as one for children, to a product which could be consumed by a person of any age, at any time, of the day, and for varied occasions.

4) **New Product Development**

MDS technique shows us the various perceived perceptions of the different brands. Spaces/ Gaps in the product perceptions could be used to develop new offerings for the target consumer.

5) **Test Marketing**

MDS technique can be used to identify cities that have similar demographic characteristics, and one could then identify a city which could represent a national character, and use that city for test marketing.

One can thus observe that MDS is a very useful technique to help understand the market place and develop strategies for the future.

Advantage of MDS

The advantage of MDS methods is not in the measurement of physical distances, but rather "psychological distances", also called as 'dissimilarities'. In MDS, we assume that every individual has a 'mental map' of products, people, places, events, companies, and individuals keep on evaluating their external environment on a continuous basis.

We also assume that the respondent is able to provide either numerical measure of his or her perceived degree of similarity/dissimilarity between pairs of objects, or can rank pairs of objects (ordinal scale of measurement) in terms of similarity/dissimilarity to each other.

We can then make use of methodology of MDS to construct a physical map in one or more dimensional whose inter-point distances (or ranks of distances) are most consistent with input data.



Now-a-days a number of software programmes are available for conducting MDS analysis. These programmes provide for a variety of input data. Some of the widely used softwares include MDPREF, MDSCAL SM, INDSCAL, PREFMAM, PROFIT, KUST.

14.7 SUMMARY

In this Unit we have discussed two of the techniques of multi-variate analysis viz., 'Cluster Analysis' and Multi-Dimensional Scaling. Cluster Analysis is mainly used for market segmentation. Some of its other applications include test market selection and to establish grouping of products within a product line. Various inputs required for -cluster analysis have been discussed in the Unit. Further various application to cluster analysis have been explained. Finally some of the important consideration which should be kept in mind while using cluster analysis have been highlighted. In this Unit we also discussed the Multi-dimensional Scaling technique or perceptual map or positioning map is a multi-variate technique that is particularly suitable for measuring human perceptions and preferences for products/brands. Its applications include market segmentation, positioning studies, test marketing etc.

14.8 SELF-ASSESSMENT EXERCISES

- 1) What is cluster analysis? What are its possible applications?
- 2) State how the following techniques differ from each other (a) Cluster Analysis (b) Discriminant Analysis.
- 3) An airlines marketing manager wants to segment his customer. How can cluster analysis be used for this purpose?
- 4) Give few examples of marketing situations where cluster analysis can be used.
- 5) Discuss with the help of examples the areas where Multi-dimentional Scaling can be applied for marketing.

14.9 FURTHER READINGS

Green, Tull and Albaum "Research for Mcarkeint Decision", PHI. N. Delhi.

Majumdar, R. "Marketing Research: Teat, Applications and Case Studies Wiley, New. Delhi.

Beri, G.C. "Marketing Research", Tata McGraw Hill Publishing Company Limited, New Delhi. Parasurama,-A. "Marketing Research", Addison Wesley Publishing Company Inc., U.S.A.