CLUSTER ANALYSIS

- **Cluster analysis** is a class of statistical techniques that can be applied to data that exhibit “natural” groupings. Cluster analysis sorts through the raw data and groups them into clusters. A **cluster** is a group of relatively homogeneous cases or observations. Objects in a cluster are similar to each other. They are also dissimilar to objects outside the cluster, particularly objects in other clusters.

- **Examples**
- The diagram below illustrates the results of a survey that studied drinkers' perceptions of spirits (alcohol). Each point represents the results from one respondent. The research indicates there are four clusters in this market. The axes represent two traits of the market. In more complex cluster analyses you may have more than that number.

- Cluster analysis, like factor analysis and multi-dimensional scaling, is an interdependence technique: it makes no distinction between dependent and independent variables. The entire set of interdependent relationships is examined. It is similar to multi-dimensional scaling in that both examine inter-object similarity by examining the complete set of interdependent relationships. The difference is that multi-dimensional scaling identifies underlying dimensions, while cluster analysis identifies clusters. Cluster analysis is the obverse of factor analysis. Whereas factor analysis reduces the number of variables by grouping them into a smaller set of factors, cluster analysis reduces the number of observations or cases by grouping them into a smaller set of clusters.

SIMILARITY MEASURES

- DISTANCE MEASURE WHICH IS KNON AS DISSIMILARITY MEASURE
- CORRELATION COEFFICIENT
- ASSOCIATION COEFFICIENT

DISTANCE MEASURE

- The widely used similiarity measure is the euclidean distance measure. Actually, any distance measure is known as dissimilarity measure.
- If the distance between a point(object)/partial cluster and another point (object)/partial cluster is the least among all such values, then they can be grouped together.

Correlation coefficient

- Correlation coefficient between variables can act as a measure in cluster analysis. But it is not widely used. The correlation coefficient is a kind of similiarity coefficient. So grouping of objects is done based on the maximum value of such coefficient example mark class

Association coefficient

- It is a measure defined in terms of attributes of the objects. If a particular attribute is present in an object, then the respective association coefficient is a kind of similarity coefficent. So grouping of objects is done based on the maximum value of such coefficient

Factor analysis

- **Factor analysis** is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called **factors**. In other words, it is possible, for example, that variations in three or four observed variables mainly reflect the variations in fewer unobserved variables. Factor analysis searches for such joint variations in response to unobserved latent variables. The observed variables are modeled as linear combinations of the potential factors, plus "error" terms. The information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset. Computationally this technique is equivalent to low rank approximation of the matrix of observed variables. Factor analysis originated in psychometrics, and is used in behavioral sciences, social sciences, marketing, product management, operations research, and other applied sciences that deal with large quantities of data.

Type of factor analysis

- CENTROID METHOD
- PRINCIPAL COMPONENTS METHOD
- VARIMAX METHOD OF FACTOR ROTATION

- **Exploratory factor analysis (EFA)** is used to identify complex interrelationships among items and group items that are part of unified concepts.[2] The researcher makes no "a priori" assumptions about relationships among factors.[2]
- **Confirmatory factor analysis (CFA)** is a more complex approach that tests the hypothesis that the items are associated with specific factors.[2] CFA uses structural equation modeling to test a measurement model whereby loading on the factors allows for evaluation of relationships between observed variables and unobserved variables.[2] Structural equation modeling approaches can accommodate measurement error, and are less restrictive than least-squares estimation.[2] Hypothesized models are tested against actual data, and the analysis would demonstrate loadings of observed variables on the latent variables (factors), as well as the correlation between the latent variables.[2]

Multidimensional scaling (MDS) is a series of techniques that helps the analyst to identify key dimensions underlying respondents' evaluations of objects. It is often used in Marketing to identify key dimensions underlying customer evaluations of products, services or companies.

example

The marketing manager of a two wheeleer company designed a questionnaire to study the customers feedback about its two wheeler and in turn he is keen in identifying the factors of his study. He has identified six variables which are:

1. Fuel efficiency x1
2. Life of the two wheeler x2
3. Handling convenience x3
4. Quality of original spares x4
5. Breakdown rate x5
6. Price x6

- After administering a aquestionnaire among respondents, the opinion of the customers can be obtained on the above variables. Let the range of the score for each of the above variables is assumed to be between 1 and 10. The score 1 means the lowest rating and 10 earns the highest rating. Let us assume that the application of factor analysis groups theses variables, as follows:
- X1,x2,x4 and x5 into factor -1
- X6 into factor 2
- X3 into factor -3
- If all the three factors are significant they they are retained for future analysis.

Multidimensional scaling

- Multidimensional scaling and conjoing analysis are the extensions of multivariate techniques for measuring human perceptions and prefernces. Multidimensional scaling deals with the judgements of respondents about the degress of similarity of pairs of stimuli on similarity basis or distance basis

- It is used to measure human perceptions and prefernce towards some stimuli like products, organizations, places, events, brands, etc, and position them in a perceptual space.
- Three basic steps
- Identification of the dimensions of perceptual space on which customers perceive objects
- Input the respondents judged values of perception prefernce of the objects into MDS package
- Positioning of the objects on the perceptual space. The output of the multidimensional scaling is a perceputal map which consists of location of the objects in it

Technology positioning

- Electric current
- Petroleum gas
- Gohar gas
- Solar energy
- Kerosene
- Fire wood
- Rice husk