FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

July 28, 1975

CLUSTER ANALYSIS AND ITS APPLICATIONS IN MARKETING RESEARCH

Lawrence Sherman and Jagdish N. Sheth

#261

CLUSTER ANALYSIS AND ITS APPLICATIONS IN MARKETING RESEARCH

Lawrence Sherman and Jagdish N. Sheth*
University of Illinois

Introduction

Classification is the identification of an observation and its placement into a homogeneous group based on observed characteristics. When we are able to a priori specify the groups, multiple discriminant analysis (MDA) provides an analytical method to derive classification functions. See Anderson (1958) for an excellent discussion of classification procedures based on the linear discriminant function. However, many times in marketing research a priori specification of groups is impossible due to lack of formal theory and the researcher must choose for his analysis some of the heuristic, probabilistic, or combinational algorithms that have been proposed to deal with such situations. While it is virtually impossible to describe all clustering procedures in this paper, Frank and Green (1968), Anderberg (1973), Bijnen (1973), and Cormack (1971) provide a good starting place for a basic introduction to clustering multivariate observations.

An assumption underlying the use of cluster analysis is that homogeneous subgroups or clusters actually exist in the data. The basic problem in cluster analysis is to devise an algorithm that reduces the

---

* Lawrence Sherman was a doctoral student in the Department of Finance at the time of writing this chapter. Jagdish N. Sheth is I.B.A. Distinguished Professor and Research Professor in the Department of Business Administration at the University of Illinois.

sorting of an entity into g groups based on a profile of p attributes. When g is unknown, the number of possibilities of sorting n observations is

$$\sum_{j=1}^{g} f\ (j) \qquad (1)$$

where f is a Stirling number of the second kind and is defined by Abramowitz and Stegun (1964) as

$$f_n^{(g)} = \frac{1}{g!} \sum_{k=0}^{g} (-1)^{g-k} \binom{g}{k} k^n \qquad (2)$$

Complete enumeration of $f_n^{(g)}$ is impractical as a method of sorting each observation into groups. In fact, it would probably be difficult to differentiate the correct cluster from such a large number of clusters. Therefore, some hauristic or optimal rule must be designed which will make the task manageable and meaningful.

This paper discusses some of the problems and decisions in the application of clustering methods, reviews some recent marketing applications and concludes by stressing the problems implicit in cluster analysis.

Decisions in Cluster Analysis

Human judgment is the single most important factor in the generation of meaningful clustering results. Major decisions facing the analyst can be stated as:

(1) How do we select a similarity measure which will index a profile vector in order to make comparisons among entities?

(2) How do we compute the clusters?

(3) How do we determine the number of clusters in the data?

(4) How do we design the research strategy?

(5) Can the clusters by quantitatively and meaningfully justified?

It should be made clear that each decision must be made on the basis of sound criteria nurtured by the research problem.

The Similarity Matrix

To convert a profile vector of an observation into a similarity index, it is critical to know the type of measurement utilized in the research problems. The classical classification of scales are provided by Stevens (1951) and Torgerson (1958) and summarized below.

The Four Basic Scales

	No Natural Origin	Natural Origin
No Distance	Nominal	Ordinal
Distance	Interval	Ratio

Nominal scaled data refers to a numbering of the observations where measurement does not connotate properties of the observation. Ordinal data indicates a serial ordering of the entities such that the numbers are determined within a monotonic increasing or decreasing transformation. When numbers are assigned to reflect different amounts of a given property between objects, the data is said to be interval scaled. Ratio scaled data has the property of interval scaled data, plus a natural origin defined by the measurement. The measures can be further classified as continuous, discrete, or dichotomous--reflecting presence or absence of the phenomenon.

Making comparisons between entities depends on the similarity measure that is defined. Similarity measures are of two types---distance measures and association (proximity) measures. Selection of the similarity measure depends on the scale utilized in the data. Use of the distance measure requires the specification of a metric of measurement. The metric has the formal properties of

$$D(X,Y) = 0, \text{ if } X=Y$$

$$D(X,Y) \geq 0$$

$$D(X,Y) = D(Y,X)$$

$$D(X,Y) \leq D(X,Z) + D(Y,Z)$$

for X, Y, and Z in a metric space. The fourth property is the familiar triangular inequality.

In order to develop a similarity index between two entities based on the distance measure, the most generalized theorem is the Minkowski's constant λ, defined as:

$$d_{ij} = \left[\sum_{k=1}^{p} w_k \left(|X_{ik} - X_{jk}| \right) \lambda \right] 1/\lambda \qquad (3)$$

where we define

i, j	are the subscripts for entity i and j
d	represents the distance measure
X_{ik}	is the projection of entity X_i on orthogonal axis k
W_k	is 1 for unweighted distances
p	is the number of axis of the space
λ	is the metric of the space

When λ=2, the similarity metric is the familiar Euclidean distance. Conceptually each entity can be viewed as a point in p-dimensional Euclidean

space. The closer the distance the more similar the entities; the farther the distance the more dissimilar the entities. This is the most often used distance measure in cluster analysis. However, Attneave (1950) has proposed the City-Block metric to deal with certain perceptual situations which has been used by Johnson and Wall (1969) in a clustering solution. Although not developed in this paper, and often forgotten in marketing applications, the selection of λ in the metric measurement model imposes a structure on the data. When scale of measurement among entities are different and contain no intrinsic information, W_k can be used to scale variable k (i.e. $W_k = 1/S_k^2$). With highly correlated variables the variable configuration and the orthogonal axis of the metric space do not correspond. Green and Rao (1969) discuss a method of dealing with redundancy in the data by computing distances in principal component space. This is equivalent to the Mahalanobis generalized distance (Morrison 1967):

$$D^2 = (X_i - \bar{X})^1 W^{-1} (X_i - \bar{X}) \qquad (4)$$

where

X_i is the ith observation vector

$\bar{X}$ is $\frac{1}{n} \sum_{i=1}^{n} x_{ij}$ and

W is $\sum_{j=1}^{g} \sum_{i=1}^{n} (X_{ij} - \bar{X}) (X_{ij} - \bar{X}_j)$

When individual differences in perceptions are expected in cluster analysis, a "modified" Euclidean distance measure

$$d_{jki} = \sqrt{\left[\sum_{k=1}^{p} (w_i a_{jm} - w_{im} a_{km})^2 \right]^{1/2}} \qquad (5)$$

where distance, d_{ijk}, is measured in a p-dimensional space between attribute j and k for entity i. The weight w_i is given to axis m by entity i and a_{jm} is the projection of attribute j on axis m. This measure is discussed by Horan (1969), and Carroll and Chang (1970) for the study of individual differences in multidimensional scaling. Bloxom (1974) suggests that it is a special case of the ACOVS procedure.

Proximity measures or measures of association depend mainly on the level of measurement of the data. When data is interval scaled, the cross-products matrix V, the variance-covariance matrix C, and the correlation matrix R has been mainly used in marketing research. Given a data matrix X

$$V = X'X \tag{6}$$

$$C = \frac{1}{n-1}\ (V-n\bar{X}'\bar{X}) \tag{7}$$

$$R = S^{-1}CS^{-1} \text{ where } S = (\text{Diag } C)^{1/2} \tag{8}$$

C implies that level of measurement is unimportant since it is subtracted out. When R is used, scale of measurement and level of measurement are assumed unimportant since R is scaled by the standard deviations of each variable. Fleiss and Zubin (1936) argue against the use of standardized data because they contend that scaling should be done on the clusters and not on the data matrix X. Implicit in the use of V, C, and R is a linear structure; Lehman (1974) has applied a non-linear correlation measure to marketing data in examining methods of grouping. Numerous other association measures have been proposed for specific purposes.

When data are in binary form, the matching coefficient is a useful method of computing measures of association. Given a contingency table and

		1	0	Total
i	1	a	b	a + b
	0	c	d	c + d
		a + c	b + d	a + b + c + d

allowing 1 to indicate presence and 0 to indicate absence, the Rogers-Tanimoto coefficient

$$\frac{a + d}{a + b + 2\,(b + c)} \qquad (9)$$

has been the most referenced, but by no means the only matching type coefficient. Cormack (1971) provides a discussion of various similarity measures, based on coefficients of association. When data consists of mixed scales, the choice of a meaningful similarity measure becomes troublesome. Perhaps the best advice is that the cluster analyst use foresight in the collection of his data to avoid mixed scale transformations. To facilitate the selection of a similarity measure, Table 1 lists selected formulae and appropriate references.

Please Insert Table 1 about here

With nonmetric data, Green (17) suggests that multidimensional scaling be carried out to bring out the metric qualities of the data and cluster analysis be performed on the configuration by computing distances in the derived space. While only experimental results are reported, the method appears promising for marketing data.

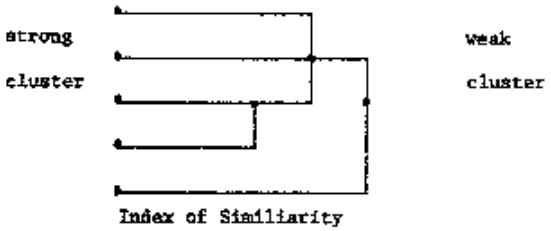
The Clustering Algorithm

The second major decision facing the cluster analyst is the choice of the clustering algorithm. Selection of the clustering algorithm must be made on the basis of anticipated properties of the clusters. Only recently has mathematical analysis been applied to provide a theoretical basis for clustering. Two basic methods of generating clusters of entities exist-hierarchical cluster analysis and non-hierarchical cluster analysis. In this section, references and a classification table of selected clustering methods are given. For the interested reader's benefit, it should be noted that most papers describe the author's prescribed method. Though this list is long, it is incomplete, but forms a basic core of readings from which additional references can be easily obtained.

Please insert Table 2 about here

The term hierarchical refers to the method of cluster analysis that starts with a strong cluster (i.e., each entity is a separate cluster) and on the basis of a similarity matrix S tries to achieve a weak clustering subject to an objective criterion specified by the clustering method. If the method starts with a weak cluster and tries to achieve a strong cluster, the method is known as agglomerative (divisive).

An Illustrative Hierarchical Clustering Tree



Hierarchical methods can be used to cluster observations or variables. Jollife (1973) reports several methods that may be used to discard redundant variables in principal components analysis. Since the similiarity measure contains $n(n+1)/2$ elements, hierarchical methods have usually been applied to samples with less than 400 observations. Johnson (1967) has programmed two methods of cluster analysis, quite similar to the method of single linkage and complete linkage discussed in Sokal and Sneath (1963) which are monotonically invariant under scaling. The similarity measure (S_{ij}) is derived from utilization of the Ultra-metric inequality

$$d(x,z) \leq d(x,y) + d(y,z) \qquad (10)$$

which is then minimized

$$d([x,y],z) = \min[d(x,z),\ d(y,z)] \qquad (11)$$

and is referred to as the minimum (single linkage) method. When

$$d([x,y],z) = \max[d(x,z),d(y,z)] \qquad (12)$$

is maximized, it is the maximum (complete linkage) method. Marketing and psychological applications have used these two methods with ordinal data when the use of a distance measure (usually Euclidean) has been untenable. Hubert (1974) has generalized the complete linkage and single linkage methods through graph theory. His approach offers the capability of overlapping clusters and asymmetric similarity measures. It would appear that the subjective decisions in clustering will diminish as the graph theoretical approach gives clustering methods the badly needed mathematical foundations for the derivation of clusters. In our opinion, the graph theory (Hubert 1973, 1974) and tree structures (Hartigan 1967) offer many advantages in the establishment of a mathematical foundation for clustering.

Nonhierarchical clustering methods have been developed to cluster n entities into g groups when g is unknown. MacQueen (1967), Friedman and Rubin (1967), and Ball and Hall (1965) have provided the early work in this area. Differences between the algorithms are generally in the generation of the initial configuration, in the criterion which is maximized or minimized to obtain the "best" partition, and in the method of determining the number of clusters that exist in the data. Recently McRae (1973) has developed a procedure (Mikca), which encompasses many of the concepts of non-hierarchical methods and will be discussed in some detail in this paper. Given a data matrix X and assuming g is known the total cross-products matrix can be decomposed as

$$T = X'X = W + B \qquad (13)$$

where: W is $\sum_{k=1}^{g} \sum_{i=1}^{m} (X_{ik} - \bar{X}_k)' (X_{ik} - \bar{X}_k)$

B is $\sum_{k=1}^{g} n_m \bar{X}'_k \bar{X}_k$

n_m is the number of observations in the mth cluster

g is the number of clusters

X_{ik} is the ith observation vector in the kth cluster.

The procedure generates g points in the space and on the basis of a choice of an objective criterion from among the following, develops clusters.

1. Minimize the $|W|$. Wilk's lambda is $\Lambda = |W|/|T|$. Since $|T|$ is fixed, minimizing $|W|$ results in small Λ which indicates large differences between groups.

2. Minimize trace W. Using this criteria minimizes

   Trace $W = \sum_{k=1}^{g} \sum_{i=1}^{n_m} (X_{ki} - \bar{X}_k)' (X_{ki} - \bar{X}_k)$ resulting in

   "minimum" variance partitions.

3. Maximize the trace of $W^{-1}B$. This is Hotelling's trace criteria which is max $\sum_{i=1}^{p} \lambda_i$ and is derived from the determinantal equation $|B-\lambda W| = 0$.

4. Maximize the largest root of $W^{-1}B$. This was proposed by S. N. Roy since when λ is large, large differences exist.

The cluster analyst specifies the number of g groups desired and g points are randomly dispersed within the space. On the basis of the criterion specified, an initial grouping of the entities is obtained, mean vectors are calculated and the procedure proceeds in an iterative manner until the selected criterion converges. Once the "final" form clusters are obtained, they may be described by a linear discriminant function using the derived clusters as the a priori specified groups. However, if the within-group variance-covariance matrices are not equal across groups, the linear discriminant function is not optimal in describing group separation. With the widespread availability of multiple discriminant analysis (MDA) procedures, a parametric clustering method independently proposed by Erbankh (1972), Mayer (1971), and Cassetti (1964) should see increasing use in non-hierarchical clustering applications in marketing. The procedure to implement this algorithm is

(1) Randomly divide the sample into g groups

(2) Run MDA using g groups

(3) Classify the groups on the basis of the linear discriminant function.

(4) Reclassify the groups on the basis of the Lachenbruch classification method to provide almost unbiased discriminant functions (Lachenbruch and Mickey, 1968).

(5) Switch misclassification entities into nearest discriminant group smallest distance from group centroids utilizing Mahalanobis D^2 statistics to form new predetermined groups.

(6) Repeat Steps 2 to 5 until no entities are misclassified.

(7) Repeat Steps 2 to 6 for g + k (k=1, 2, . . . = n-g) groups.

(8) Select the number of groups on the basis of Rao's F statistic for overall significance. If W^{-1} is singular, the generalized psuedo-inverse can be computed (Theil, 1971).

For the data analyst without a computer, there is no need to dispair. McQuitty (1967, 1970, 1971) provides a clustering technique based on hand computations. A discussion of the "quick" method of clustering and step by step directions for applications in marketing are given by Kamen (1970) and an extension using principal components analysis is given by Aaker (1971).

The Research Strategy

How can cluster analysis be used to aid in the interpretation of relationships latent in the data structure? This question hinges on the strategy employed. Due to the many implicit and explicit criteria that must be specified or assumed in clustering, the technique can not be blindly followed without a great deal of peril. Methods of cluster analysis enable the marketing researcher to work closely with his data. Roscoe, Sheth, and Howell (1974) have pointed out the need for inter-technique cross-validation in the search for invariant structure in marketing data. Since the selection of the similarity measure and the clustering algorithm imposes a given structure on the data, it is recommended that several clustering results be compared. Finally, Sokal and Rohlf (1962) utilized the cophenetic correlation coefficient as a measure of fit between a derived similarity measure from a hierarchical structure and an original similarity matrix which is the product-moment

correlation coefficient. Studies to date (Sammon, 1966; Sneath, 1965; and McQuitty, 1971), indicate that evaluating clustering procedures is not a minor problem since some methods produce dissimilar results and other methods produce comparable results.

A proper research strategy must encompass foresight in the collection of data, familiarity with clustering decisions, and a firm grasp of the research problem. Clustering can be used with factor analysis to produce clearer factorial structures, with discriminant analysis when a priori groupings are unknown, and with multiple regression when data structures are heterogenous and hypothesis testing is the objective. See Elton and Gruber (1970), for further discussion on this.

Applications of Cluster Analysis in Marketing

A widespread usage of cluster analysis in marketing research has not occurred despite the suitability to many marketing problems. This section of the paper reviews selected applications in marketing to illustrate the adaptability of the method to marketing problems. Particular problems described in these studies should form the basis for identifying marketing applications.

The subjectivity of the methods will be stressed and the potential problems—both methodological and theoretical--in the application of cluster analysis will be enumerated. It should, however, be noted that the subjective decisions in cluster analysis can often form the basis for imaginative application of the technique to marketing problems so long as one is aware of the decisions that must be made. To provide an intuitive perspective for the applied researcher, each review will discuss the purpose and nature of the research problem, the mechanics of the clustering

procedures, and the problem areas and decisions experienced by the researchers in their particular applications.

Test Market Selection

Orderly classification of multidimensional marketing phenomenon is a problem that remains unresolved in many marketing areas. Green, Frank, and Robinson (1967) approached the problem of test market selection through numerical methods of cluster analysis. They are among the pioneers in the application of cluster analysis to marketing. The research problem and purpose of the paper was to develop a method of matching representative test markets with larger product markets. Simultaneous consideration of a large number of market characteristics were considered by cluster analysis in an n-dimensional metric space. This paper is not only an important application, but is a clear explication of a research strategy by researchers aware of the decision and subjectivity problems inherent in cluster analysis.

The technique employed in the paper was to measure distances among test cities by the familiar Euclidean distance formula,

$$d_{jk} = \left[\sum_{i=1}^{n} (X_{ij} - X_{ik})^2 \right] 1/2 \qquad (14)$$

to identify a cluster in which cities within a cluster are more similar than cities between clusters. Similarity was defined by the distance in the metric space. As is often the case in the social sciences, market characteristics were measured in different scales and therefore properly normalized to have a mean of 0 and a standard deviation of 1. The number of cities to appear in each cluster was specified in advance by the prior desire of the researchers to have five cities in each cluster subject to a maximum cutoff distance that precludes clustering of distant cities (points) in the Euclidean space.

The problem of "weighting" market characteristics with highly correlated measures was addressed by the authors by first doing a principal components analysis on the data matrix and clustering on principal components scores for each city. Cluster analysis was performed on the component scores and the consequent distance measures (Mahalanobis generalized distance). "Implicit" weighting of correlated measures should be consciously considered and the analyst must decide which method of analysis is more appropriate. One method of weighting not dealt with in this paper, however, is "explicit" weighting schemes. Market characteristics more relevant to the research problem can be weighted by prior judgment rather than assigning either an equal weight of unity for each market characteristic or relying on some statistical criterion.

The heuristics of clustering methods may produce suboptimal clusters from a mathematical view, but when compared with the simultaneous assessment of multidimensional data by a market researcher the relevant question to ask seems to be: Does the method aid the assessment of multidimensional data? In this paper, the answer was a clear yes according to the authors. However, see Morrison (1967) for a discussion of alternative procedures for calculating distances.

International Marketing

Establishment of world marketing segments based on cultural, socio-economical, and political characteristics is of importance with the growth of international business. Sethi (1971) cluster analyzed 91 countries on the basis of 29 interval and ratio scaled variables. The objective of his paper was to establish homogeneous geographical markets segments.

The method of cluster analysis employed consisted of the V-analysis (variable x variable) and the O-analysis (object x object) which are subjects of the BC-TRY system discussed earlier in the book. Variable groups which have within-group similarity and between group differences are formed through V-analysis. The first step in V-analysis is to select k sets of n variable clusters that can reproduce the original matrix of intercorrelations among the variables. Each variable cluster dimension is defined by the collinear subset of variables defined by an index of proportionality, P^2, as

$$P^2_{xv} = (\Sigma r_{x_{ik}x_{jk}})^2 / \sum_{k=1}^{p} r^2_{x_j x_k} \sum_{k=1}^{p} r^2_{x_j z_k} \qquad (15)$$

where p is the index for the number of variables. Unlike the principal components analysis, this method factors common variance and not the total variance in the data and produces clusters of variables rather than linear combinations of variables. Since the cluster dimensions need not be orthogonal (uncorrelated), distortions in the distance measure based on them may occur.

Object analysis is obtained by assigning a variable cluster score to each object usually on the basis of a simple sum composite of the dimensions of a cluster from the V-analysis. Other methods using principal components scores are part of the approach. While V-analysis and O-analysis form one method of clustering variables or objects, disciples of this approach tend to treat the BC-TRY system as a unified method of data analysis. It must be emphasized that considerable subjectivity and unresolved mathematical problems still exist and like all clustering methods a heuristic defined by the program is maximized without regard to any statistical sampling theory.

In Sethi's paper four variable clusters termed: aggregate production and trade, personal consumption, international trade, and health and education were formed. Q-analysis produced eight country pattern types with differing profile descriptor patterns. Clustering results must be evaluated on the basis of the variables used to model the research problem. Contrary to many marketing problems, the variables selected to reflect comparative world markets did not appear to be theoretically selected. Unfortunately, they tended to be the usual United Nations type of census data which may or may not be relevant to the marketer of a specific industry. However, the paper does demonstrate at least one approach toward the development of international marketing segments and the cross-cultural analysis of world segments based on a profile of political, socio-economic, and demographic measures.

Buyer Behavior and Personal Characteristics

Multidimensional relationships between consumer characteristics and buying behavior are known to be important in identification of market segments. Lessig and Tollefson (1971) desired to explore and demonstrate an approach to segment market identification and buyer behavior by assuming that consumers who exhibit similar buying behavior and personal characteristics are likely to have similar stimulus response functions.

Cluster analysis was performed on 20 buying behavior variables for 212 households. All behavior characteristics were given equal importance in the clustering procedure. This was a novel departure from most cluster analysis application and was achieved by dividing

squared distances $\left[(X_{ij} - X_{ik})^2\right]$ in the distance formula for each single characteristic by the number of dimensions for that characteristic. Average within clusters distance (AWCL) was used as a measure of cluster similarity.

Household personal characteristics were measured and related to each buying segment and for all households. To test the linear relationship between buying behavior and personal characteristics, canonical correlation analysis was used. A stepwise discriminant analysis was also performed for the prediction of buyer group membership on the basis of personal characteristics. An unbiased estimate of its predictive validity was conducted on a 28 household validation sample with rather poor predictive results.

This paper represents an excellent example of the complimentary, multistage use of multivariate methods. However, the poor classification results in the validation stage of the discriminant analysis do warrant some cautions that a researcher must be cognisant of, if a fruitful linkage can be made between different multivariate techniques. For example, it is not at all clear from the paper whether the poor predictive validation is due to small sample size or lack of homogeneity of the within-group dispersion matrix across the buyer segments.

Personality and Implicit Behavior Patterns

The applicability of cluster analysis to marketing studies relating personality and behavior patterns is demonstrated by Greeno, Sommers, and Kernan (1973). Self theoretical concepts of consumer behavior and personality trait theory were associated with the end result being a number of distinctive housewife types could be identified.

One hundred and ninety housewives between the ages of 30 to 45 years old were asked to sort 38 product items according to actual and idealized behavior. The set of 190 self-ratings were then cluster analyzed by Ward's hierarchical (1963) clustering algorithm. The similarity measure used was the Euclidean distance measure. Replicable stability was insured by conducting a separate analysis on two randomly split samples. Race and class structure were controlled in forming the samples. Six clusters were selected based on the information loss measure computed in the clustering procedure. Cluster naming proceeded on the basis of the cluster means and the rank order of the product array in the clusters. Tukey's test of mean differences and ANOVA procedures were used to evaluate differences in the clusters. Socio-economic and additional personality measures served as external measures to aid in the interpretation of the results.

Several methodological comments must be made at this point. Implicit in using Euclidean distance is the idea that the variables (products) were uncorrelated. This could result in the implicit differential weighting of products depending upon the choice of product configurations as discussed by Green, Robinson, and Frank (1967). Second, the relationship between "self" and "ideal" traits should have been first analyzed through other techniques such as the simultaneous factor analysis or the canonical correlation analysis. When size allows, the idea of sample splitting is a recommended procedure. The use of external measures was interestingly incorporated in the paper. Supportive validity of the results would have been achieved if actual usage rates of the consumer products were measured.

Market Experimentation

The results of an experimental approach to test the sales effect of three different price level changes in a new food product is reported by

Day and Heeler (1971). Their analysis is concerned with the construction of a randomized block experiment with five strata composed of three stores each being representative of a 58 store test market. A modified matching coefficient and a modified Euclidean distance measure were used to construct a 58 x 58 similarity measure. To reduce the redundancy of variables, factor analysis of the 12 store attributes explaining 77 per cent of the total variance was accomplished.

The subjective importance of each factor was weighed by λ_k (subjective importance assigned by experts) and the modified distance measure was calculated by

$$d_{ij} = \left(\sum_{k=1}^{n} [\lambda_k (X_{ik} - X_{jk})]^2 \right)^{1/2} \qquad (16)$$

This is equivalent to stretching the dimensionality of each factor by its subjective importance. The stores were iteratively reassigned to clusters by a hierarchical clustering method and by optimizing the average within cluster similarity subject to the constraint that five clusters be formed of three or more stores.

Representativeness and homogeneity of the strata were evaluated by reduced space analysis using non-metric scaling and principal components analysis. Representativeness was measured by

$$R = \sum_{i=1}^{n} (\sigma_{ij}/\sigma_{it})/n \qquad (17)$$

which compares dispersion across the 12 store attributes for each dimension i. This allows evaluation of the bias and dispersion produced by the clustering approach.

Several methodological problems arise in a study of this type. First, the sensitivity and the reliability of the weighting

scheme were not tested. The modified matching coefficient was developed to link ordinal-interval-ratio data and by using this measure on interval data the properties of the data is not fully exploited. Nevertheless, the comparison of the two similarity measures and the two clustering methods provided additional empirical support for their study. The representativeness achieved by reduced space analysis—whether metric or nonmetric--for randomized block experiments offers promise for further research.

The main disadvantage of the study is that the original objective of evaluating the effect of three price level changes on a new food product through a randomized block experimental approach was never discussed in the text of the paper.

Free Response Data Analysis

Green, Wind, and Jain (1973) suggest using a tandem reduced space and clustering approach in the analysis of free response marketing data. Free response marketing data is usually unstructured judgments expressing like-dislike or word association phrases. The purpose of this paper was to describe current limitations and methodological extensions in marketing of free response data analysis.

In the first example, the connotations of certain words for a new shampoo among 84 female respondents between the ages of 18-30 years of age were examined to find out the similarity between eight stimulus words and evoked word associations. An 8 x 19 word association frequency matrix was obtained in which the column entries were conditional responses to the raw stimuli. A hybrid version of Kruskal's M-D-Scale V scaling algorithm was applied to the word association matrix and five dimensions were required to obtain an "adequate" fit of the model to data. The 19

evoked stimulus words were positioned in the common reduced space.

Since the results were not easily interpreted and the configuration was nonunique, Euclidean interpoint distances were calculated in the five dimensional space for the 19 x 19 dissimilarity matrix. A hierarchical tree structure form of cluster analysis was then applied on the Euclidean dissimilarity matrix to determine the word association relationship between the stimulus phrases and evoked words. A second illustration in the study dealt with well-known women's home service magazines and the 107 respondents were media buyers for 41 different advertising agencies. Respondent protocols were analyzed to obtain frequency of evaluated type words and/or phrases. Further examples are illustrated and application areas listed.

The approach is interesting as one way of handling free response data and the use of cluster analysis provided a powerful way to aid in the interpretation of a multidimensional scaling solution. While the results are exploratory, little can be said about the stability, reliability, or feasibility of using the results in a marketing decision context. Graph theoretic clustering approaches proposed by Hubert (1973) seems to offer another structural approach that can be non-metric and capable of asymmetric clustering of free response data.

A Method of "Quick" Cluster Analysis

For the researcher without a technical background in multivariate analysis, the method of "quick" cluster analysis developed by McQuitty (1968, 1971) is elaborated and applied in market research by Kamen (1970). In this paper, quick clustering is viewed as a first approximation to the reality of a complex world. Emphasized is the research methodology and a solid understanding of the research area.

The method begins with a matrix of similarity coefficients, usually correlation coefficients. The clustering strategy is summarized in the following seven steps:

(1) The highest correlation in each column of the similarity matrix is identified.

(2) The highest element in the similarity matrix is selected as the nucleus of the first cluster.

(3) Any other object having its highest correlation with either one of the two entities in the first cluster is joined to that cluster.

(4) Excluding the already clustered objects the next highest correlation is selected.

(5) Repeat steps 2 to 3 for step 4.

(6) Repeat steps 4 to 5.

(7) Examine your results.

In the event of a tie, sum the correlations in each column with the highest sum having priority. One application of the quick clustering procedure related to consumer opinions of gasoline stations.

External criteria can often be used to validate and aid in the interpretation of the cluster analysis results which is suggested by Kamen. As a first approach, quick clustering has several definite advantages over the more complicated heuristic approaches. Directly working with the data enables the researcher to understand and to conceptualize his findings better. Unless one is familiar with the mechanics of the clustering methods, analytical results may be overinterpreted or even misinterpreted. In summary, this paper argues for simplicity in clustering rather than the complexity normally associated with a multivariate method. It should only be viewed as a first step, but the approach is worth the time and effort. Aaker (1971) has extended this approach by plotting points in a principal

component space and selecting clusters by a similar approach. This reduces the dimensionality of the problem. Computationally better methods of cluster analysis have been developed from which to make informed judgments. When weighing the advantages and costs of implementation, the market researcher with limited knowledge or interest in mathematical methods might well consider this approach.

Some Concluding Observations

Cluster analysis is an important addition to the family of multivariate techniques and to marketing methodology. This paper has attempted to summarize and introduce the concepts that are essential for proper application of the method to marketing research. While we reviewed some selected applications in this paper, they are not all inclusive of the substantive areas where the technique may be applied.

Subjective decisions in cluster analysis should be viewed as a challenge for innovation and not as an impediment to its use in providing understanding of complex multidimensional marketing problems when groups are not a priori known. The ability to deal with the major decisions in cluster analysis and awareness of the problem areas is a first step in the orderly classification of multivariate marketing phenomena.

Table 1. Some Proposed Methods of Similarity

	Formulae	References
I. Distance Measures		
A. Minkowski Metric	$\left[\sum_{i=1}^{i=k} w_{ij} \left\|X_{ij}-X_{ik}\right\|^{\lambda}\right]^{1/\lambda}$	Boyce (1969), Green and Carmone (1970)
B. "Canberra" Metric	$\sum_{i=1}^{n} \left\|X_{ij}-X_{ik}\right\| / \sum_{i=1}^{n} (X_{ij}+X_{ik})$	Lance and Williams (1966)
C. Angular Separation	$\sum_{k=1}^{j} X_{ij}-X_{jk} / \left[\sum_{k=1}^{j} X_{ik}^2 \sum_{i=1}^{p} X_{jk}^2\right]^{1/2}$	Gower (1967)
D. Squared Weighted Distance	$\sum_{j=1}^{k} (X_{1j}-X_{2j})^2 \Sigma x_{jk}^2$	Williams, et. al. (1964)
II. Association Measures		
A. Correlation Coefficient	$\sum_{i=1}^{n} (X_i-\bar{X})\ (Y_i-\bar{Y}) / \left(\left[\sum_{i=1}^{n} (X_i-\bar{X})^2\right]\left[\sum_{i=1}^{n} (Y_i-\bar{Y})^2\right]\right)^{1/2}$	Anderberg (1973)
B. Pattern Similarity Index	$1 - \Sigma d^2/2m$	Cattell (1949)
C. Polynomial Correlation	$\mathrm{Max}(R_{ij}, R_{ji})$	Lohmann (1974)
D. Index of Jaccard	$S = \frac{n_s}{n_s + n_d}$	Sneath (1957)
E. Smirnov Coefficient	$t = \frac{1}{\Sigma r_j} \sum_{jp} w_{jp}$	Sokal and Sneath (1963)

Table 2. Some Methods of Cluster Analysis

I.	Hierarchical Analysis	
	1. Agglomerative	Johnson (1967), Ward (1963), Gruvaeus and Wainer (1972)
	2. Divisive	Edwards and Cavalli-Sforza (1965)
	3. Tree Structures	Hartigan (1967)
	4. Graph Theory	Hubert (1974)
II.	Non-Hierarchical Analysis	
	1. Minimum Variance Partitioning	Friedman and Rubin (1967) MacQueen (1967), McRae (1973)
	2. Discriminant Clustering	Urbankh (1972), Mayer (1971)
	3. Centroid Clustering	Ball and Hall (1965)
III.	Other Methods	
	1. Obverse Factor Analysis	Harman (1967)
	2. Key Cluster Analysis	Tryon and Bailey (1970)
	3. Pattern and Mixture Analysis	Wolfe (1971)
	4. Typal and Linkage Analysis	McQuitty (1967, 1968, 1970, 1971)

References

Aaker, D. "Visual Clustering Using Principal Component Analysis." In D. Aaker (ed.) Multivariate Methods in Marketing, Wadsworth Publishing Co., Belmont, California, 1971, 321-333.

Abramowitz, M. and Stegan, I. A. Handbook of Mathematical Functions with Formulae, Graphs, and Mathematical Tables, U. S. Government Printing Office, Washington, D. C., 1968.

Anderberg, M. R. Cluster Analysis for Applications, Academic Press, New York, 1973.

Anderson, T. W. An Introduction to Multivariate Statistical Analysis, Wiley, New York, 1958.

Attneave, F. "Dimensions of Similarity," American Journal of Psychology, 63 (1950), 516-556.

Ball, G. H. and Hall, D. J. "Isodata, a Novel Method of Data Analysis and Pattern Classification," Report Number RADC-TR-67-310, Stanford Research Institute, 1965.

Bass, F. M., Pessimier, E. A., and Tigert, D. J. "A Taxonomy of Magazine Readership Applied to Problems in Marketing Strategy and Media Selection." In R. L. Day and L. J. Parsons (eds.) Marketing Models Quantitative Applications, Haddon Craftsman, Inc., Scranton, Pennsylvania, 1970, 486-524.

Bijnen, E. J. Cluster Analysis, Tilburg University Press, The Netherlands, 1973.

Bloxom, B. B. "An Alternative Method of Fitting a Model of Individual Differences in Multidimensional Scaling," Psychometrika, 39 (1974), 365-367.

Boyce, A. J. "Mapping Diversity: A Comparative Study of Some Numerical Methods." In A. J. Cole (ed.) Numerical Taxonomy, Academic Press, New York, 1969.

Carroll, J. D. and Chang, J. J. "Analysis of Individual Differences in Multidimensional Scaling Via an N-Way Generalization of "Eckart-Young" Decomposition," Psychometrika, 35 (1970), 283-319.

Casetti, E. Classificatory and Regional Analysis by Discriminant Iterations, Report Number AD608093, Northwestern University, 1964.

Cattell, R. B. "Rp and Other Coefficients of Pattern Similarity," Psychometrika, 14 (1949), 279-298.

Cormack, R. M. "A Review of Classification," Journal of the Royal Statistical Society, Series A, 134 (1971), 237-242.

Day, G. S. and Heeler, R. M. "Using Cluster Analysis to Improve Marketing Experiments," Journal of Marketing Research, 8, August, 1971, 340-347.

Edwards, A. W. F. and Cavalli-Sforza, L. L. "A Method for Cluster Analysis," Biometrics, 21 (1965), 362-375.

Elton, E. H. and Gruber, M. J. "Homogeneous Groups and the Testing of Economic Hypothesis," Journal of Financial and Quantitative Analysis, January, 1970, 581-602.

Frank, R. E. and Green, P. E. "Numerical Taxonomy in Marketing Analysis: A Review Article," Journal of Marketing Research, Vol. V., February, 1968, 83-94.

Friedman, H. P. and Rubin, J. "On Some Invariant Criteria for Grouping Data," Journal of the American Statistical Association, 62 (1967), 1159-1178.

Frost, W. A. K. "The Development of a Technique for TV Program Assessment," In D. Aaker (ed.) Multivariate Methods in Marketing: Theory and Application, 1971, 334-350.

Gower, J. C. "Multivariate Analysis and Multidimensional Geometry," The Statistician, 17 (1967), 13-25.

Green, P. E. and Rao, V. R. "A Note on Proximity Measures and Cluster Analysis," Journal of Marketing Research, 6 (1969), 359-364.

Green, P. E., Wind Y. and Jain, A. K "Analysis of Free Response Data in Marketing Research," Journal of Marketing Research, 10, February, 1973, 45-52.

Green, P. E. and Carmone, F. J. Multidimensional Scaling and Related Techniques. Allen and Bacon, Rockleigh, New Jersey, 1970.

Green, P. E., R. E. Frank, and P. J. Robinson "Cluster Analysis In Test Market Selection," Management Science, 13, April, 1967, 387-400.

Greeno, D. W., M. S. Sommers, and J. B. Kernan "Personality and Implicit Behavior Patterns," Journal of Marketing Research, X, February, 1973, 63-69.

Gruvaeus, G. and Wainer, H. "Two Additions to Hierarchical Cluster Analysis," The British Journal of Mathematical and Statistical Psychology, 25 (1972), 200-206.

Harman, H. H. Modern Factor Analysis, (2nd ed.), University of Chicago Press, Chicago, Illinois, 1967.

Hartigan, J. A. "Representation of Similarity Matrices by Trees," Journal of the American Statistical Association, 62 (1967), 1140-1158.

Horan, C. B. "Multidimensional Scaling: Combining Observations When Individuals Have Different Perceptual Structures," Psychometrika, 34 (1969), 139-165.

Hubert, L. J. "Some Applications of Graph Theory to Clustering," Psychometrika, 39 (1974), 283-310.

Hubert, L. "Min and Max Hierarchical Clustering Using Asymmetric Similarity Matrices," Psychometrika, 38 (1973), 63-72.

Johnson, R. L. and Wall, D. B. "Cluster Analysis of Semantic Differential Data," Psychological Measurement, 29 (1969), 769-780.

Johnson, S. C. "Hierarchical Clustering Schemes," Psychometrika, 32 (1967) 241-254.

Johnson, R. M. "Market Segmentation: A Strategic Management Tool," Journal of Marketing Research, February, 1971, 13-20.

Jollife, I. T. "Discarding Variables in Principal Component Analysis II: Real Data," Applied Statistician, 22 (1973).

Kamen, J. M. "Quick Clustering," Journal of Marketing Research, 7, May, 1970, 199-204.

King, B. F. "Step-wise Clustering Procedures," Journal of the American Statistical Association, 62 (1967), 329-350.

Kruskal, J. F. "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis," Psychometrika, 29 (1964) 1-27.

Kruskal, J. B. and Shepard, R. N. "A Nonmetric Variety of Linear Factor Analysis," Psychometrika, 39, June, 1974, 123-159.

Lachenbruch, P. A. and Mickey, M. R. "Estimation of Error Rates in Discriminant Analysis," Technometrics, 10 (1968), 1-11.

Lance, G. N. and Williams, W. T. "Computer Programs for Hierarchical Polythetic Classification," The Computer Journal, 9 (1966), 60-64.

Lehman, D. "Some Alternatives to Linear Factor Analysis for Variable Grouping Applied to Buyer Behavior," Journal of Marketing Research, 10 (1974), 206-213.

Lehmann, D. "Some Alternatives to Linear Factor Analysis for Behavior Variables," Journal of Marketing Research, May, 1974, 206-213.

Lesaig, V. P. and Tollefson, J. D. "Market Segmentation Through Numerical Taxonomy," Journal of Marketing Research, 8, November, 1971, 480-487.

MacQueen, J. "Some Methods for Classification and Analysis of Multivariate Observations," In the Fifth Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 1967.

Mayer, L. S. "A Method of Cluster Analysis When There Exists Multiple Indicators of a Theoretical Concept," Biometrics, 27 (1971), 143-155.

McRae, D. J. Clustering Multivariate Observations, Unpublished Ph.D. Thesis, University of North Carolina, 1973.

McQuitty, L. L. and Clark, J. A. "Clusters From Iterative, Intercolumnar Correlational Analysis," Educational and Psychological Measurement, 28 (1968), 211-238.

McQuitty, L. L. "Group Based Pattern Analysis of the Single Individual," Multivariate Behavior Research, 1967, 529-536.

McQuitty, L. L. "A comparative Study of Some Selected Methods of Pattern Analysis," Educational and Psychological Measurement, 31 (1971), 607-626.

McQuitty, L. L. "Hierarchical Classification by Multiple Linkages," Educational and Psychological Measurement, 30 (1970), 3-10.

Morrison, D. G. "Measurement Problems in Cluster Analysis," Management Science, 13, August, 1967, 755-780.

Roscoe, A. M., Sheth, J. N. and Howell, W. "Intertechnique Cross Validation in Cluster Analysis," AMA Educators Conference, Portland, Oregon, 1974, (in press).

Sammon, J. W. "A Nonlinear Mapping for Data Structure Analysis," IEEE Computer Journal, 18 (1966), 401-409.

Sethi, S. P. "Comparative Cluster Analysis for World Markets," Journal of Marketing Research, 8, August, 1971, 348-354.

Shepard, R. N. "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function," Psychometrika, 27, 125-139.

Sneath, P. H. A. "A Comparison of Different Clustering Methods as Applied to Randomly Spaced Particles," Classification Society Bulletin, 1965, 1, No. 2, 2-7.

Sokal, R. R. and Sneath, P. H. A. Principles of Numerical Taxonomy, Freeman, London, 1963.

Stevens, S. S. Handbook of Experimental Psychology, Wiley, New York, 1951.

Theil, H. Principles of Econometrics, New York, Wiley, 1971

Tryon, R. C. and Bailey, D. E. Cluster Analysis, McGraw-Hill, New York, 1970.

Urbankh, V. Y. "A Discriminant Method of Clustering," Journal of Multivariate Analysis, 2 (1972), 249-260.

Ward, J. H. "Hierarchical Grouping to Optimize an Objective Function," Journal of the American Statistical Association, 58 (1963), 236-244.

Williams, W. T., Clifford, H. T. and Lance, G. N. "Group-Size Dependence: A Rational for Choice Between Numerical Classification," Computer Journal, 14 (1971) 157-162.

Williams, W. T., Dale, M. B., and MacNaughton-Smith, P. "An Objective Method of Weighting in Similarity Analysis," Nature, 201 (1964) 426.

Wolfe, J. N. "Pattern Clustering by Multivariate Mixture Analysis," Multivariate Behavioral Research, 1971, 5, No. 3, 329-350.

Zubin, J. A. "A Technique for Pattern Analysis," Psychological Bulletin, 33 (1936), 733.