Cluster analysis in marketing research: Review and suggestions for application

Girish Punj
University of Connecticut

GIRISH PUNJ and DAVID W. STEWART*

Applications of cluster analysis to marketing problems are reviewed. Alternative methods of cluster analysis are presented and evaluated in terms of recent empirical work on their performance characteristics. A two-stage cluster analysis methodology is recommended: preliminary identification of clusters via Ward's minimum variance method or simple average linkage, followed by cluster refinement by an iterative partitioning procedure. Issues and problems related to the use and validation of cluster analytic methods are discussed.

Cluster Analysis in Marketing Research: Review and Suggestions for Application

Cluster analysis has become a common tool for the marketing researcher. Both the academic researcher and the marketing applications researcher rely on the technique for developing empirical groupings of persons, products, or occasions which may serve as the basis for further analysis. Despite its frequent use, little is known about the characteristics of available clustering methods or how clustering methods should be employed. One indication of this general lack of understanding of clustering methodology is the failure of numerous authors in the marketing literature to specify what clustering method is being used. Another such indicator is the tendency of some authors to differentiate among methods which actually differ only in name.

The use of cluster analysis has frequently been viewed with skepticism. Green, Frank, and Robinson (1967) and Frank and Green (1968) have discussed problems with determining the appropriate measure of similarity and the appropriate number of clusters. Inglis and Johnson (1970), Morrison (1967), Neidell (1970), and Shuchman (1967) have also expressed concern about the use of cluster analysis. More recently, Wells (1975) has expressed reservations about the use of cluster analysis unless very different, homogeneous groups can be identified. Such skepticism is probably justified in the light of the confusing array of names and methods of cluster analysis confronting the marketing researcher. As this confusion is resolved and as additional information about the performance characteristics of various clustering algorithms becomes available, such skepticism may disappear. Recent work on clustering algorithms affords a basis for establishing some general guidelines for the appropriate use of cluster analysis. It is useful to note that many of the problems associated with cluster analysis also plague multivariate statistics in general: choice of an appropriate metric, selection of variables, cross-validation, and external validation.

Two general sets of issues confront the marketing researcher seeking to use cluster analysis. One set of issues involves theoretical properties of particular algorithms. These issues are considered in the literature on cluster analysis (Anderberg 1973; Bailey 1974; Cormack 1971; Hartigan 1975), and are not addressed here. The second set of issues are more practical and pertain to the actual use of clustering procedures for data analysis. These issues are the foci of our article, in which we review applications of clustering methodology to marketing problems, provide a systematic treatment of the clustering options open to the marketing researcher, and use both theoretical and empirical findings to suggest which clustering options may be most useful for a particular research problem.

Cluster analysis has most frequently been employed as a classification tool. It has also been used by some researchers as a means of representing the structure of data via the construction of dendrograms (Bertin 1967; Hartigan 1967) or overlapping clusters (Arabie et al. 1981; Shepard and Arabie 1979). The latter applications

*Girish Punj is Assistant Professor of Marketing, School of Business Administration, University of Connecticut. David W. Stewart is Assistant Professor of Management, Owen Graduate School of Management, Vanderbilt University.

The authors acknowledge the support of the 1981 Dean's Fund of the Owen Graduate School of Management, Vanderbilt University, which facilitated the completion of this article.

Journal of Marketing Research
Vol. XX (May 1983), 134-48

are distinct from the use of cluster analysis for classification and represent an alternative to multidimensional scaling and factor analytic approaches to representing similarity data. Whereas classification is concerned with the identification of discrete categories (taxonomies), structural representation is concerned with the development of a faithful representation of relationships. Both uses of cluster analysis are legitimate, but the objectives of these applications are very different. The best clustering algorithm for accomplishing one of these objectives is not necessarily the best for the other objective. We restrict our treatment of cluster analysis to the more common of the two applications, classification.

Cluster analysis is a statistical method for classification. Unlike other statistical methods for classification, such as discriminant analysis and automatic interaction detection, it makes no prior assumptions about important differences within a population. Cluster analysis is a purely empirical method of classification and as such is primarily an inductive technique (Gerard 1957). Though some theorists have not been favorably disposed toward the use of cluster analysis, and criticism of the ad hoc nature of clustering solutions is common, classification is an important and frequently overlooked tool of science. Wolf (1926) has suggested that classification is both the first and last method employed by science. The essence of classification is that certain things are thought of as related in a certain way. Indeed, the final outcome of other methods of study may well be a new classification.

Kemeny (1959) and Kantor (1953), discussing the philosophy of science, point to the fundamental importance of classification. Wolf (1926) holds that verification of laws of science may occur only after classification has been completed. Thus, whether the classification exercise is completed explicitly or implicitly, it must occur. Cluster analysis provides one, empirically based, means for explicitly classifying objects. Such a tool is particularly relevant for the emerging discipline of marketing which is still wrestling with the problems of how best to classify consumers, products, media types, and usage occasions.

USES OF CLUSTER ANALYSIS IN MARKETING

The primary use of cluster analysis in marketing has been for market segmentation. Since the appearance of Smith's now-classic article (1956), market segmentation has become an important tool for both academic research and applied marketing. In a review of market segmentation research and methodology, Wind (1978) identifies both the impact of this most fundamental of marketing tools and some rather significant problem areas. Not the least of these problems is the plethora of methods that have been proposed for segmenting markets. This multiplication of techniques has served to confuse many marketers, shift discussions of researchers from more substantive issues to issues of method, and impede the development of meta-research directed at integrating market segmentation research. In concluding his review of the segmentation literature, Wind suggests that one important area of future research should be the "evaluation of the conditions under which various data analytical techniques are most appropriate" (1978, p. 334).

All segmentation research, regardless of the method used, is designed to identify groups of entities (people, markets, organizations) that share certain common characteristics (attitudes, purchase propensities, media habits, etc.). Stripped of the specific data employed and the details of the purposes of a particular study, segmentation research becomes a grouping task. Wind (1978) notes that researchers tend to select grouping methods largely on the basis of familiarity, availability, and cost rather than on the basis of the methods' characteristics and appropriateness. Wind attributes this practice to the lack of research on similarity measures, grouping (clustering) algorithms, and effects of various data transformations.

A second and equally important use of cluster analysis has been in seeking a better understanding of buyer behaviors by identifying homogeneous groups of buyers. Cluster analysis has been less frequently applied to this type of theory-building problem, possibly because of theorists' discomfort with a set of procedures which appear ad hoc. Nevertheless, there is clearly a need for better classification of relevant buyer characteristerics. Bettman (1979) has called for the development of taxonomies of both consumer choice task and individual difference characteristics. Cluster analysis is one means for developing such taxonomies. Examples of such use may be found in articles by Claxton, Fry, and Portis (1974), Kiel and Layton (1981), and Furse, Punj, and Stewart (1982).

Cluster analysis has been employed in the development of potential new product opportunities. By clustering brands/products, competitive sets within the larger market structure can be determined. Thus, a firm can examine its current offerings vis-à-vis those of its competitors. The firm can determine the extent to which a current or potential product offering is uniquely positioned or is in a competitive set with other products (Srivastava, Leone, and Shocker 1981; Srivastava, Shocker, and Day 1978). Although cluster analysis has not been used frequently in such applications, largely because of the availability of other techniques such as multidimensional scaling, factor analysis, and discriminant analysis, it is not uncommon to find cluster analysis used as an adjunct to these other techniques. Cluster analysis has also been suggested as an alternative to factor analysis and discriminant analysis. In such applications it is important for the analyst to determine whether discrete categories of products are desirable or whether a representation of market structure is desirable. The latter may be more useful in many market structure applications, in which case cluster analysis would not be used as a classification technique and the analyst would face a different set of issues from those addressed here.

Cluster analysis has also been employed by several researchers in the problem of test market selection (Green, Frank, and Robinson 1967). Such applications are concerned with the identification of relatively homogeneous sets of test markets which may become interchangeable in test market studies. The identification of such homogeneous sets of test markets allows generalization of the results obtained in one test market to other test markets in the same cluster, thereby reducing the number of test markets required.

Finally, cluster analysis has been used as a general data reduction technique to develop aggregates of data which are more general and more easily managed than individual observations. For example, limits on the number of observations that can be used in multidimensional scaling programs often necessitate an initial clustering of observations. Homogeneous clusters then become the unit of analysis for the multidimensional scaling procedure. Fisher (1969) discussed the use of cluster analysis for data reduction from the perspective of econometrics and argued that cluster analysis is most appropriate whenever the data are too numerous or too detailed to be manageable. Such data simplification and aggregation are carried out for the convenience of the investigator rather than in the interest of theory building.

Table 1 is a brief description of some recent applications of cluster analysis to marketing problems. Although not a complete set of all applications of cluster analysis in marketing, it illustrates several points. First, the array of problems addressed by these studies is striking. Equally striking is the diversity of clustering methods employed. In constructing this table we had difficulty discerning the specific clustering algorithm used by the researchers. Cluster analysis methods were often identified by the name of the program used, e.g., BMDP2M, BCTRY, or Howard and Harris, rather than by the specific clustering algorithm used. Only by consulting a particular program's manual could we identify the method actually employed. For one of the studies cited we could not find any information on the clustering method used.

The lack of specificity about the method of clustering employed in these studies is illustrative of the problems associated with the use of cluster analysis. The lack of detailed reporting suggests either an ignorance of or lack of concern for the important parameters of the clustering method used. Failure to provide specific information about the method also tends to inhibit replication and provides little guidance for other researchers who might seek an appropriate method of cluster analysis. Use of specific program names rather than the more general algorithm name impedes interstudy comparisons.

This situation suggests a need for a sound review of clustering methodology for the marketing researcher. Previous reviews on this subject appeared prior to the publication of much of the research on the performance characteristics of clustering algorithms. Sherman and Sheth (1977) discuss selected similarity measures and clustering algorithms. Though they mention some empirical work on the characteristics of these measures and algorithms, their report is primarily a catalog of techniques and some marketing applications. Relatively little guidance is provided the researcher seeking to discover the characteristics and limitations of various grouping procedures. Indeed, the Sherman and Sheth report may mislead some readers because its categorization of clustering algorithms suggests substantive differences among identical algorithms which differ only in name. Frank and Green (1968) also provide an introduction and review of clustering methodology but make no specific recommendations to guide the user of the methodology. After reviewing the problems and issues facing the user of cluster analytic procedures, we offer clarification of the similarities and differences among various clustering algorithms and some suggestions about their use.

PROBLEMS IN USING CLUSTER ANALYSIS

Unlike other data analytic methods, cluster analysis is a set of methodologies that has developed outside a single dominant discipline. Factor analysis and various scaling methods were developed within the discipline of psychology and one would look to that discipline for guidance in the use of these methods. Regression, though used in a variety of disciplines, has tended to be the special province of econometricians, who have developed a large body of literature on the technique. In contrast, no single discipline has developed and retained clustering methodology. Rather, numerous disciplines (econometrics, psychology, biology, and engineering) have independently approached the clustering problem. Often working in parallel, researchers in these disciplines have arrived at similar solutions but have given them different names. For example, Blashfield (1978) reviewed the literature on hierarchical clustering methods and found as many as seven different names for the same technique. This diversity of names for identical techniques has tended to prevent comparisons of algorithms across disciplines. It has also served to confuse the data analyst by implying a much greater number of available clustering methods than actually exists.

Also confronting the potential user of cluster analysis is the problem of cluster definition. There are currently no clear guidelines for determining the boundaries of clusters or deciding when observations should be included in one cluster or another. Cattell (1978) has suggested that clusters are "fuzzy" constructs. The criterion for admission to a cluster is rather arbitrary. There are no well-established rules for the definition of a cluster. The preferred definition of a cluster seems to vary with the discipline and purpose of the researcher.

Clusters have most frequently been defined by relatively contiguous points in space (Stewart 1981). Cormack (1971) suggested that clusters should exhibit two properties, external isolation and internal cohesion. External isolation requires that objects in one cluster be separated from objects in another cluster by fairly empty

Table 1
SOME RECENT APPLICATIONS OF CLUSTER ANALYSIS IN MARKETING

Application	*Purpose of research*	*Nature of data*	*Clustering method used*
Anderson, Cox, and Fulcher (1976)	To identify the determinant attributes in bank selection decisions and use them for segmenting commercial bank customers	Determinant attribute scores on several bank selection variables	Iterative partitioning—MIKCA (McRae 1973)
Bass, Pessemier, and Tigert (1969)	To identify market segments with respect to media exposure	Attribute scores on several media exposure variables	Average linkage cluster analysis (Sneath and Sokal 1973)
Calantone and Sawyer (1978)	To examine the stability of market segments in the retail banking market	Attribute scores on several bank selection variables	*K*-means (Howard and Harris 1966)
Claxton, Fry, and Portis (1974)	To classify furniture and appliance buyers in terms of their information search behavior	Attribute scores on several prepurchase activity measures	Complete linkage cluster analysis (Johnson 1967, Lance and Williams 1967a)
Day and Heeler (1971)	To classify stores into similar strata	Factor scores on several store attributes	(1) Complete linkage analysis (2) Iterative partitioning (Rubin 1965)
Green, Frank, and Robinson (1967)	To identify matched cities for test marketing	Factor scores on several city characteristics	Average linkage cluster analysis (Sneath and Sokal 1973)
Greeno, Sommers, and Kernan (1973)	To identify market segments with respect of personality variables and implicit behavior patterns	Q sorts on 38 product items	Ward's minimum variance method (Ward 1963)
Kernan (1968)	To identify groups of people along several personality and decision behavior characteristics	Scores on several personality and decision traits	Ward's minimum variance method (Ward 1963)
Kernan and Bruce (1972)	To create relatively homogeneous configuration of census traits	Characteristics of census traits	Ward's minimum variance method (Ward 1963)
Kiel and Layton (1981)	To develop consumer taxonomies of search behavior by Australian new car buyers	Factor scores derived from several search variables	Average linkage cluster analysis (Sneath and Sokal 1973)
Landon (1974)	To identify groups of people using purchase intention and self-concept variables	Scores on self-image and purchase intention variables	Iterative partitioning (BCTRY, Tryon and Bailey 1970)
Lessig and Tollefson (1971)	To identify similar groups of consumers along several buyer behavior variables	Scores on several buyer behavior variables	Ward's minimum variance method (Ward 1963)
Montgomery and Silk (1971)	To identify opinion leadership and consumer interest segments	Scores on several interest and opinion leadership variables	Complete linkage cluster analysis (Johnson 1967)
Moriarty and Venkatesan (1978)	To segment educational institutions in terms of benefits sought when purchasing financial-aid MIS	Importance scores on financial-aid management services	*K*-means (Howard and Harris 1966)
Morrison and Sherman (1972)	To determine how various individuals interpret sex appeal in advertising	Ratings of advertisements by respondents	Iterative Partitioning (Friedman and Rubin 1967)
Myers and Nicosia (1968)	To develop a consumer typology using attribute data	Scores of supermarket image variables	Iterative partitioning (BCTRY; Tryon and Bailey 1970)
Sethi (1971)	To classify world markets	Macrolevel data on countries	Iterative partitioning (BCTRY; Tryon and Bailey 1970)
Sexton (1974)	To identify homogeneous groups of families using product and brand usage data	Brand and product usage rate data	Type not specified
Schaninger, Lessig, and Panton (1980)	To identify segments of consumers on the basis of product usage variables	Scores on several product usage variables	*K*-means (Howard and Harris 1966)
Green and Carmone (1968)	To identify similar computers (strata in the computer market)	Performance measures for different computer models	*K*-means (Howard and Harris 1966)

space. Internal cohesion requires that objects within the same cluster be similar to each other. Everitt (1974) offered a similar concept which he defines as a natural cluster. The requirement of external isolation does not provide for overlapping clusters. Although a few algorithms have been developed for identifying overlapping clusters (Jardine and Sibson 1971; Peay 1975; Shepard 1974), these methods are primarily concerned with the representation of structure rather than classification. Applications of these methods have been few and are not reviewed here.

In the absence of a generally accepted or definitive definition of a cluster, various algorithms have been developed which offer particular operational definitions. Differences among clustering algorithms are frequently related to how the concept of a cluster is operationalized. Thus, to develop a set of recommendations for the application of cluster analysis, we must first develop a recognition of the clustering algorithms available to the marketing researcher and an understanding of the performance of these methods in relation to one another.

CLUSTERING ALGORITHMS

Table 2 provides a description of the more common clustering algorithms in use, the various alternative names by which the algorithms are known, and a brief discussion of how clusters are formed by each of these methods. Table 2 shows clearly that there are relatively few clustering methods from which to choose, far fewer than one might suspect from a reading of the literature on cluster analysis. Four primary hierarchical methods are available, single linkage, complete linkage, average linkage, and Ward's minimum variance method. Although there are several variations of the average linkage method, only one, simple average linkage, is widely used. In addition, two variants of the average method, the centroid and median methods, have very undesirable properties (Aldenderfer 1977; Sneath and Sokal 1973) which recommend against their use. The weighted average linkage method has been shown to produce results very similar to those produced by the simple average method (Blashfield 1977).

There is more variety among the nonhierarchical methods, though all work on similar principles. These iterative partitioning methods begin by dividing observations into some predetermined number of clusters. Observations are then reassigned to clusters until some decision rule terminates the process. These methods may differ with respect to the starting partition, the type of reassignment process, the decision rule used for terminating clustering, and the frequency with which cluster centroids are updated during the reassignment process. The initial partition may be random or based on some prior information or intuition. One method (MIKCA) uses several different random starting partitions to ensure an efficient solution. Two types of reassignment are generally employed, *K*-means and hill-climbing. These methods are briefly discussed in Table 2 as are the termination decision rules used with each method. Cluster centroids may be updated after each membership move or only after a complete pass through the entire data set.

Not included in Table 2 are two methods frequently used for cluster analysis: *Q* factor analysis and automatic interaction detection (AID) (Morgan and Sonquist 1963). *Q* factor analysis is not included because Stewart (1981) in the marketing literature and Cattell (1978) in the psychology literature have forcefully argued that factor analysis is inappropriate as a method for identifying clusters. Skinner (1979) discusses some relationships between factor analysis and cluster analysis. AID is not included because it operates on a rather different principle than the clustering procedures. AID requires the prior specification of independent and dependent variables and seeks to identify sets of nominal independent variables which group observations in a manner that minimizes the variance of the dependent variable within each group. Cluster analysis procedures require no such *a priori* specification of independent and dependent variables.

These clustering algorithms exist in various forms but most have been programmed. Several software programs are currently available for cluster analysis. They differ in their comprehensiveness and ease of use. Table 3 briefly describes several of the more common clustering software programs, identifies the types of clustering methods available within each program, and cites the original source of the program. Selecting an appropriate cluster analytic method or software package requires some knowledge of the performance characteristics of the various methods.

EMPIRICAL COMPARISONS OF CLUSTERING METHODS

One method for evaluating clustering methods that has been used with increasing frequency involves comparing the results of different clustering methods applied to the same data sets. If the underlying characteristics of these data sets are known, one can assess the degree to which each clustering method produces results consistent with these known characteristics. For example, if a data set consists of a known mixture of groups, or subpopulations, the efficacy of a cluster solution can be evaluated by its success in discriminating among these subpopulations. This mixture model approach to the evaluation of clustering algorithms has recently been employed by several researchers. Table 4 summarizes the findings of 12 such studies.

The number of clustering algorithms, distance measures, and types of data that might be incorporated in a mixture model study is so large as to preclude any one comprehensive study of the relative efficacy of clustering methods. We can look across the studies in Table 4, however, and begin to draw some conclusions about clustering methods. Three procedures seem to warrant special consideration. Ward's minimum variance method, average linkage, and several variants of the iterative partitioning method appear to outperform all other methods Ward's method appears to outperform the average linkage method except in the presence of outliers. *K*-means appears to outperform both Ward's method and the average linkage method if a nonrandom starting point is specified. If a random starting point is used *K*-means may be markedly inferior to other methods, but results on this issue are not consistent. Nevertheless, the *K*-

Table 2
CLUSTERING METHODS

Primary name	*Alternative names*	*Method of forming clusters*
Hierarchical methods		
Single linkage cluster analysis	Minimum method (Johnson 1967); linkage analysis (McQuitty 1967); nearest neighbor cluster analysis (Lance and Williams 1967a); connectiveness method (Johnson 1967)	An observation is joined to a cluster if it has a certain level of similarity with at least *one* of the members of that cluster. Connections between clusters are based on links between single entities.
Complete linkage cluster analysis	Maximum method (Johnson 1967); rank order typal analysis (McQuitty 1967); furthest neighbor cluster analysis (Lance and Williams 1967a); diameter method (Johnson 1967)	An observation is joined to a cluster if it has a certain level of similarity with *all* current members of the cluster.
Average linkage cluster analysis	Simple average linkage analysis (Sneath and Sokal 1973); weighted average method (McQuitty 1967), centroid method (Gower 1967); median method (Lance and Williams 1967a)	These are actually four similar methods. In all four methods an observation is joined to a cluster if it has a certain *average* level of similarity with all current members of the clusters These methods differ in the manner in which the average level of similarity is defined The weighted average method and median method provide for an *a priori* weighting of the averages based on the number of entities desired in each cluster. The centroid method provides for an initial computation of the centroid of each cluster. Average similarity is based on this centroid Only the simple average linkage procedure has been widely used.
Minimum variance cluster analysis	Minimum variance method; Ward's method; error sum of squares method (Ward 1963); HGROUP (Veldman 1967)	The minimum variance method is designed to generate clusters in such a way as to minimize the within-cluster variance. Unlike other hierarchical clustering methods, Ward's method optimizes an objective statistic: it seeks to minimize tr $\mathbf{W}$, where $\mathbf{W}$ is the pooled within-clusters sum of squares and cross-products matrix Ward's method is somewhat similar to the average method in that variance is a function of deviations from the mean. Some authors have included Ward's method as a special case of the average method (Bailey 1974). It is an average linkage method because it does not seek to minimize distance between one member of the cluster and the entity, or all members of the cluster and the entity as in single linkage and complete linkage, respectively, but minimizes the average distance within the cluster
Iterative partitioning methods (nonhierarchical methods)		These methods begin with the partition of observations into a specified number of clusters This partition may be on a random or nonrandom basis Observations are then reassigned to clusters until some stopping criterion is reached Methods differ in the nature of the reassignment and stopping rules
K-means		Cases are reassigned by moving them to the cluster whose centroid is closest to that case Reassignment continues until every case is assigned to the cluster with the nearest centroid. Such a procedure implicitly minimizes the variance within each cluster, tr $\mathbf{W}$
Hill-climbing methods		Cases are not reassigned to the cluster with the nearest centroid but are moved from one cluster to another if a particular statistical criterion is obtained Reassignment continues until optimization occurs The objective function to be optimized may be selected from one of four options, tr $\mathbf{W}$, tr $[(\mathbf{W}^{-1}\mathbf{B}]$, $\lvert\mathbf{W}\rvert$, and the largest eigenvalue of $[(\mathbf{W}^{-1}\mathbf{B}]$, where $\mathbf{W}$ is the pooled within-cluster covariance matrix and $\mathbf{B}$ is the between cluster covariance matrix
Combined K-means and hill-climbing methods		Uses a combination of K-means and hill-climbing methods

Table 3
COMMON CLUSTERING PACKAGES/PROGRAMS

Name of package/program	*Where available?/authors*	*Clustering methods*[a]	*Comments*
ANDERBERG	In the appendices of book entitled *Cluster Analysis for Applications* by M R Anderberg (1973)	S, C, A, W, K, H, KH	1 No missing value treatment 2 Only binary data type with octal coding scheme 3 User manual not available
BCTRY	D Bailey and R C Tryon, Tryon-Bailey Associates, Inc , c/o Mr. Peter Lenz, 2222 S E Nehalem St , Portland, OR	K	1 No MANOVA statistics are optimized 2 Initial partition has to be user specified 3 Factor analysis of variables may be performed as well
BMDP	W J Dixon (ed), Health Sciences Computing Facility, School of Medicine UCLA, Los Angeles, CA	S, C, A, K	1 Single and complete linkage available for clustering 2 Binary data not permissible 3 Continuous type similarity measure 4 Method for clustering cases and variables simultaneously is available 5 User cannot supply only similarity matrix for cases
CLUS	H Friedman and J Rubin (1967 *JASA* article), IBM SHARE system	K, H	1 Fixed number of clusters 2 Expensive to use
CLUSTAN	D Wishart, Computer Centre, University College of London, 19 Gordon St , London, WC1H OAH, Great Britain	S, C, A, W, K, H	1 High versatility (38 s/dis measures) 2 Initial partition for iterative partitioning methods has to be user specified 3 Binary and continuous data types 4 Binary data in 3 coding schemes 5 Variable transformations not available 6 Permits overlapping clusters
HARTIGAN	In the appendices of book entitled *Clustering Algorithms* by J J Hartigan (1975)	S, A, K	1 Fixed number of clusters for iterative partitioning methods 2 No variable transformations available 3 No user manual available
HGROUP	D J Veldman (1967), *FORTRAN Programming for the Behavioral Sciences*	W	1 Part of the University of Texas EDSTAT statistics package
HICLUS	S C Johnson (based on 1967 *Psychometrika* article), Bell Telephone Labs, Murray Hill, NJ	S C	1 No user manual available 2 No missing value treatment 3 No standardization of variables 4 No transformation of variables 5 User must supply similarity matrix (hence is versatile in some sense)
HOWD (Howard-Harris)	Britton Harris, F J Carmone, Jr , University of Pennsylvania, Philadelphia, PA	K	1 No user manual available 2 No MANOVA statistics 3 Number of clusters fixed
ISODATA	Daviel Wolf, SRI, 333 Ravenswood Avenue, Menlo Park, CA	K	1 No user manual available 2 No MANOVA statistics optimized
MIKCA	D J McRae, Coordinator, Testing & Computer Applications, Jackson Public Schools, Jackson, MI	K, H, KH	1 No user manual available 2 4 MANOVA statistics optimized 3 3 different distance measures
NT-SYS	F James Rohlf, John Kishapugh, David Kirk, Dept of Ecology and Evolution, SUNY at Stony Brook, Stony Brook, NY	S, C, A	1 Permits overlapping clusters 2 Alphanumeric coding scheme for binary data 3 Moderately versatile
OSIRIS	Institute of Survey Research, University of Michigan, Ann Arbor, MI	C	
SAS	James H Goodnight, SAS Inst Inc , P O Box 10066, Raleigh, NC	C	1 Continuous similarity measure

[a]S = single linkage
C = complete linkage
A = average linkage
W = Ward's minimum variance method
K = *K*-means
H = hill climbing
KH = joint *K*-means, hill climbing

Table 4
EMPIRICAL COMPARISONS OF THE PERFORMANCE OF CLUSTERING ALGORITHMS

Reference	*Methods examined*	*Data sets employed*	*Coverage*[a]	*Criteria*	*Summary of results*
Cunningham and Ogilvie (1972)	Single, complete, average linkage with Euclidean distances and Ward's minimum variance technique	Normal mixtures	Complete	Measures of "stress" to compare input similarity/dissimilarity matrix with similarity relationship among entities portrayed by the clustering method	Average linkage outperformed other methods
Kuiper and Fisher (1975)	Single, complete, average, centroid, median linkage, all using Euclidean distances and Ward's minimum variance technique	Bivariate normal mixtures	Complete	Rand's statistic (Rand 1971)	Ward's technique consistently outperformed other methods
Blashfield (1976)	Single, complete, average linkage, all using Euclidean distance and Ward's minimum variance technique	Multinormal mixtures	Complete	Kappa (Cohen 1960)	Ward's technique demonstrated highest median accuracy
Mojena (1977)	Simple average, weighted average, median, centroid, complete linkage, all using Euclidean distances and Ward's minimum variance technique	Multivariate gamma distribution mixtures	Complete	Rand's statistic	Ward's method outperformed other methods
Blashfield (1977)	Eight iterative partitioning methods: Anderberg and CLUSTAN K-means methods, each with cluster statistics updated after each reassignment and only after a complete pass through the data, CLUS and MIKCA (both hill-climbing algorithms), each with optimization of tr **W** and **W**	Multinormal mixtures	Complete	Kappa	For 15 of the 20 data sets examined, a hill-climbing technique which optimized **W** performed best, i e , MIKCA or CLUS In two other cases a hill-climbing method which optimized tr **W** performed best, CLUS.
Milligan and Isaac (1978)	Single, complete average linkage, and Ward's minimum variance technique, all using Euclidean distances	Data sets differing in degree of error perturbation	Complete	Rand's statistic and kappa	Average linkage and Ward's technique superior to single and complete linkage
Mezzich (1978)	Single, complete linkage, and K-means, each with city-block and Euclidean distances and correlation coefficient, ISODATA, Friedman and Rubin method, Q factor analysis, multidimensional scaling with city-block and Euclidean metrics and correlation coefficients, NORMAP/NORMIX, average	Psychiatric ratings	Complete	Replicability, agreement with "expert" judges, goodness of fit between raw input dissimilarity matrix and matrix of 0's and 1's indicating entities clustered together	K-means procedure with Euclidean distances performed best followed by K-means procedure with the city-block metric, average linkage also performed well as did complete linkage with a correlation coefficient & city-block metric & ISODATA, the type of metric used (r, city-block, or Euclidean distance) had little impact on results

Table 4 (Continued)

Reference	Methods examined	Data sets employed	Coverage[a]	Criteria	Summary of results
	linkage with correlation coefficient				
Edelbrock (1979)	Single, complete, average, and centroid, each with correlation coefficients, Euclidean distances, and Ward's minimum variance technique	Multivariate normal mixtures, standardized & unstandardized	70, 80, 90, 95, 100%	Kappa	Ward's method and simple average were most accurate; performance of all algorithms deteriorated as coverage increased but this was less pronounced when the data were standardized or correlation coefficients were used The latter finding is suggested to result from the decreased extremity of outliers associated with standardization or use of the correlation coefficient.
Edelbrock and McLaughlin (1980)	Single, complete, average, each with correlation coefficients, Euclidean distances, one-way and two-way intraclass correlations, and Ward's minimum variance technique	Multivariate normal mixtures & multivariate gamma mixtures	40, 50, 60, 70, 80, 90, 95, 100%	Kappa and Rand's statistic	Ward's method and the average method using one-way intraclass correlations were most accurate, performance of all algorithms deteriorated as coverage increased
Blashfield and Morey (1980)	Ward's minimum variance technique, group average linkage, Q factor analysis, Lorr's nonhierarchical procedure, all using Pearson product moment correlations as the similarity measure	Multivariate normal mixtures	Varying levels	Kappa	Group average method best at higher levels of coverage; at lower levels of coverage Ward's method and group average performed similarly.
Milligan (1980)	Single, complete, group average, weighted average, centroid & median linkage, Ward's minimum variance technique, minimum average sum of squares, minimum total sum of squares, beta-flexible (Lance & Williams 1970a b), average link in the new cluster, MacQueen's method, Jancey's method, K-means with random starting point, K-means with derived starting point, all with Euclidean distances, Cattell's (1949) r_p, and Pearson r	Multivariate normal mixtures, standardized and varying in the number of underlying clusters and the pattern of distribution of points of the clusters Data sets ranged from error free to two levels of error perturbations of the distance measures, from containing no outliers to two levels of outlier conditions, and from no variables unrelated to the clusters to one or two randomly assigned dimensions unrelated to the underlying clusters	Complete	Rand's statistic, the point biserial correlation between the raw input dissimilarity matrix and a matrix of 0's and 1's indicating entities clustering together	K-means procedure with a derived point generally performed better than other methods across all conditions 1 Distance measure selection did not appear critical; methods generally robust across distance measures 2 Presence of random dimensions produced decrements in cluster recovery 3 Single linkage method strongly affected by error-perturbations; other hierarchical methods moderately so, nonhierarchical methods only slightly affected by perturbations 4 Complete linkage and Ward's method exhibited noticeable decrements in performance in the outlier conditions, single, group average, & centroid methods only slightly affected by presence of outliers, nonhierarchical methods generally unaffected by presence of outliers 5 Group average method best among hierarchical methods

Table 4 (Continued)

Reference	*Methods examined*	*Data sets employed*	*Coverage*[a]	*Criteria*	*Summary of results*
					used to derive starting point for K-means procedure 6. Nonhierarchical methods using random starting points performed poorly across all conditions.
Bayne, Beauchamp, Begovich, and Kane (1980)	Single, complete, centroid, simple average, weighted average, median linkage and Ward's minimum variance technique, and two new hierarchical methods, the variance and rank score methods, four hierarchical methods: Wolfe's NORMIX, K-means, two variants of the Friedman-Rubin procedure (trace **W** & $\|\mathbf{W}\|$). Euclidean distances served as similarity measure	Six parameterizations of two bivariate normal populations	Complete	Rand's statistic	K-means, trace **W**, and $\|\mathbf{W}\|$ provided the best recovery of cluster structure. NORMIX performed most poorly. Among hierarchical methods, Ward's technique, complete linkage, variance & rank score methods performed best. Variants of average linkage method also performed well but not as well as other methods. Single linkage performed poorly.

[a]The percentage of observations included in the cluster solution. With complete coverage, clustering continues until all observations have been assigned to a cluster. Ninety percent coverage could imply that the most extreme 10% of the observations were not included in any cluster

means procedure appears to be more robust than any of the hierarchical methods with respect to the presence of outliers, error perturbations of the distance measures, and the choice of a distance metric. It appears to be least affected by the presence of irrelevant attributes or dimensions in the data.

One conclusion in several of the studies is that the choice of a similarity/dissimilarity measure, or distance measure, does not appear to be critical. Despite the considerable attention given such measures (Green and Rao 1969; Morrison 1967; Sherman and Sheth 1977), the selection of a similarity measure appears to be less important for determining the outcome of a clustering solution than the selection of a clustering algorithm. Two cautions should be observed in taking this conclusion at face value, however. First, the number of studies of the relative import of distance measures for determining clustering solutions is small and many types of data have yet to be examined. There may be types of data for which the selection of a distance measure is critical to the clustering solution. Second, clustering algorithms which are sensitive to the presence of outliers (e.g., complete linkage specifically, and more generally all of the hierarchical methods of clustering) seem to produce better solutions when Pearson product moment or intraclass correlation coefficients are used. Such similarity measures tend to reduce the extremity of outliers in relation to Euclidean distance measures. This, in turn, reduces the influence of outliers on the final clustering solution. A similar effect is obtained if data are standardized prior to clustering.

One characteristic of data appears to have a marked decremental effect on the performance of all clustering methods—the presence of one or more spurious attributes or dimensions. A variable that is not related to the final clustering solution, i.e., does not differentiate among clusters in some manner, causes a serious deterioration of the performance of all clustering methods, though this problem is least severe with the K-means procedure and is probably less serious for other iterative partitioning methods as well. This finding indicates the need for careful selection of variables for use in clustering and the need to avoid "shotgun" approaches where everything known about the observations is used as the basis for clustering. Clearly one cannot know in advance what variables may differentiate among a set of as yet unidentified clusters. Nevertheless, it is not unreasonable for a researcher to have some rational or theoretical basis for selecting the variables used in a cluster analysis.

A final conclusion can be drawn from the empirical findings on the performance of clustering algorithms: as a clustering algorithm includes more and more observations, its performance tends to deteriorate, particularly at high levels of coverage, 90% and above. This effect is probably the result of outliers beginning to come into the solution. Clustering all observations may not be a good practice. Rather the identification and elimination of outliers or the use of a decision rule to stop clustering short of the inclusion of all observations is probably advantageous. Suggestions for identifying outliers are provided hereafter. The *K*-means procedure has shown less decrement in performance as coverage increases than have the hierarchical methods.

Though a reasonable amount of evidence suggests that iterative partitioning methods are superior to hierarchical methods, particularly if nonrandom starting points are used, it is not yet clear which of the iterative partitioning methods are superior. *K*-means procedures and tr **W** and |**W**| hill-climbing procedures all appear to perform well. Some evidence (Blashfield 1977) suggests that hill-climbing methods which minimize |**W**| have an advantage over other iterative partitioning methods.

RECOMMENDATIONS FOR USING CLUSTER ANALYSIS

It should be clear from the preceding discussion that the research analyst must make several decisions which affect the structure of a cluster solution. These decisions can be grouped in the following broad categories:

1. *Data transformation issues*
   - A. What measure of similarity/dissimilarity should be used?
   - B. Should the data be standardized? How should nonequivalence of metrics among variables be addressed?
   - C. How should interdependencies in the data be addressed?
2. *Solution issues*
   - A. How many clusters should be obtained?
   - B. What clustering algorithm should be used?
   - C. Should all cases be included in a cluster analysis or should some subset be ignored?
3. *Validity issues*
   - A. Is the cluster solution different from what might be expected by chance?
   - B. Is the cluster solution reliable or stable across samples?
   - C. Are the clusters related to variables other than those used to derive them? Are the clusters useful?
4. *Variable selection issues*
   - A. What is the best set of variables for generating a cluster analytic solution?

Often these decisions are not independent of one another because the choice of a means for addressing one of these issues may constrain the options available for addressing other issues. For example, choosing to use a Pearson product moment correlation coefficient also determines that the data will be standardized because standardization is implicit in the computation of the correlation coefficient. Thus, it is not possible to offer recommendations for the resolution of any one of these issues without an explicit understanding of the interactions among these decisions.

DATA TRANSFORMATION ISSUES

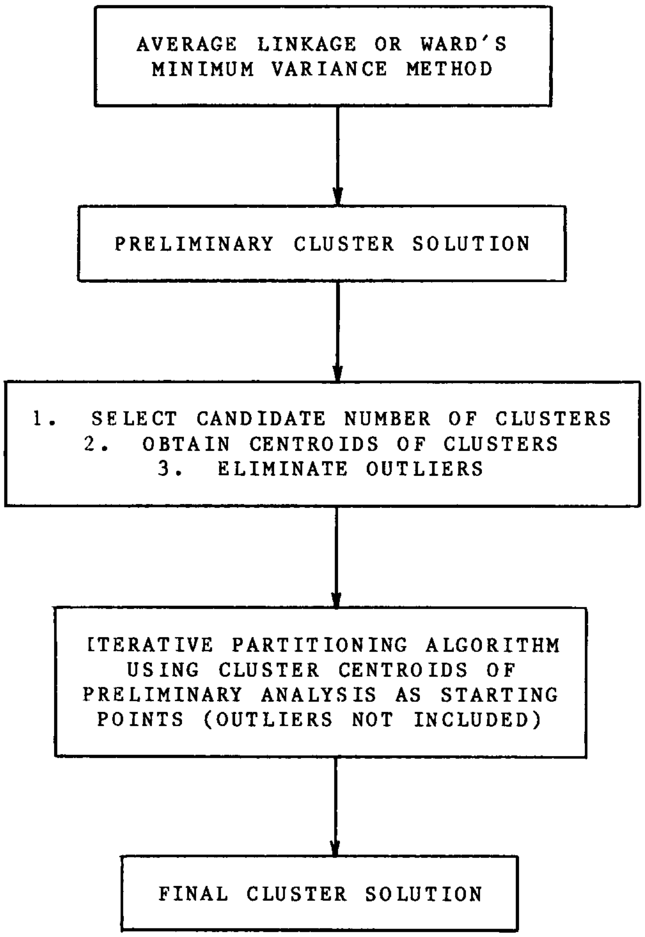
Although issues related to the choice of a similarity/dissimilarity measure have received considerable attention (Green and Rao 1969; Morrison 1967) the results of the empirical studies cited above suggest that the choice is not crucial to the final clustering solution. The same appears to be true of the standardization issue. To the extent that a particular measure of similarity or standardization reduces the extremity of outliers, the performance of some algorithms which are sensitive to outliers may be improved. Otherwise the selection of a similarity measure or the standardization of data prior to clustering appears to have minimal effect. We do not suggest that the choice of a similarity measure should be indiscriminant; the measure should be appropriate for the type of data being considered. Rather, the choice of a correlation coefficient, a Euclidean distance, or a city-block metric does not seem to produce much difference in the final outcome of a clustering exercise involving data for which each of the similarity measures is appropriate.

Some measures of similarity/dissimilarity explicitly correct for interdependencies. Other measures do not consider interdependencies. Interdependencies among variables may exist by design or, more often, are the unexpected result of the research design. Careful selection of variables may reduce unwanted interdependencies but the problem is likely to remain even in the best of circumstances. Bailey (1974) provides an illustration of the problem, the effect of which is to weight more heavily certain dimensions along which clustering will be carried out. When this is desirable for some theoretical or practical purpose, correcting for interdependencies is inappropriate. When the researcher desires that all dimensions or attributes be given equal weight in the clustering process, it is necessary to correct for interdependencies. This can be achieved by selecting a similarity measure which corrects for interdependencies, Mahalanobis D^2 or partial correlations. Correction may also be achieved by completing a preliminary principal components analysis with orthogonal rotation. Component scores may then be used as input for the computation of a similarity or distance measure. Skinner (1979) gives an example of this latter approach.

SOLUTION ISSUES

The selection of the clustering algorithm and solution characteristics appears to be critical to the successful use of cluster analysis. Empirical studies of the performance of clustering algorithms suggest that one of the iterative partitioning methods is preferable to the hierarchical methods. This holds, however, only when a nonrandom starting point can be specified. In addition, iterative partitioning methods require prior specification of the number of clusters desired. Hierarchical methods require no such specification Thus, the user is confronted with determining both an initial starting point and the number of clusters in order to use the methods that have dem-

Figure 1
TWO-STAGE CLUSTERING



onstrated superior performance. Information for determining starting points in the form of *a priori* descriptions of expected clusters may be available. In the absence of such information a means for obtaining starting points and an estimate of the number of clusters is required. A two-stage procedure may be employed to cope with this problem.

In the first step one of the hierarchical methods which has demonstrated superior performance, average linkage or Ward's minimum variance method, may be used to obtain a first approximation of a solution. By examining the results of this preliminary analysis, one can determine both a candidate number of clusters and a starting point for the iterative partitioning analysis. In addition, this preliminary analysis can be used for examining the order of clustering of various observations and the distances between individual observations and clusters. This provides an opportunity for the identification of outliers which may be eliminated from further analysis. The remaining cases may then be submitted to an iterative partitioning analysis for refinement of the clusters.

Similar two-stage clustering approaches have been suggested by Hartigan (1975) and Milligan (1980). Figure 1 is a schematic representation of the procedure. Only four cluster analytic software packages provide average linkage or Ward's method and an iterative partitioning algorithm: BMDP, CLUSTAN, and the Anderberg and Hartigan series.

VALIDITY ISSUES

Even after careful analysis of a data set and the determination of a final cluster solution, the researcher has no assurance of having arrived at a meaningful and useful set of clusters. A cluster solution will be reached even when there are no natural groupings in the data. This problem is similar to that encountered with a variety of other procedures ranging from factor analysis to regression analysis. Some test or set of tests must be applied to determine whether the solution differs significantly from a random solution. Milligan and Mahajan (1980) and Milligan (1981) reviewed several such methods for testing the quality of a clustering solution and found them wanting on a number of dimensions. A method suggested by Arnold (1979) appears to overcome the problems of other methods. Arnold (1979) proposed using a statistic first suggested by Friedman and Rubin (1967) as a test of the statistical significance of a cluster solution. The statistic is given by

$$C = \log(\max / |\mathbf{T}|/|\mathbf{W}|)$$

where $|\mathbf{T}|$ is the determinant of the total variance-covariance matrix and $|\mathbf{W}|$ is the determinant of the pooled within-groups variance-covariance matrix. A number of iterative partitioning methods seek maximization of the ratio of $|\mathbf{T}|$ to $|\mathbf{W}|$: MIKCA, CLUSTAN, and CLUS. For algorithms which do not optimize $|\mathbf{T}|/|\mathbf{W}|$ the test becomes even more conservative. Arnold generated distributions of the C statistic for 2, 4, and 8 group solutions, 10, 20, 50, 100, 200, 500, and 1000 entities, and 5, 10, 20 attributes and indicated values of C which allow rejection of the null hypotheses that the data arise from unimodal or uniform distributions. He presented data to support the use of the statistics and provided formulas for the derivation of other values.

As with other multivariate statistics, one must demonstrate the reliability and the external validity of a cluster solution as well as its statistical significance. Reliability may be established by cross-validation. External validation requires a demonstration that the clusters are useful in some larger sense. Numerous authors have recommended cross-validating cluster solutions (see, e.g., Sherman and Sheth 1977) and several methods of cross-validation have been proposed. One of the more frequently used methods involves dividing the sample in half and carrying out clustering on each half. Descriptive statistics of the two sets of clusters are compared to determine the degree to which similar clusters have been identified. The problem with such an approach is that no objective measure of reliability is obtained.

Several authors have recommended the use of discriminant analysis for cross-validation (Field and Schoenfeldt 1975; Nerviano and Gross 1973; Rogers and Linden 1973). The approach involves using cluster membership as the group membership variable in a discriminant analysis. After a cluster solution has been developed on one sample, discriminant functions are derived which are applied to a second sample. The degree to which the assignments made with the discriminant functions agree with assignments made by a cluster analysis of the second sample serves as an estimate of the stability of the cluster solution across samples. A coefficient of agreement, such as kappa, may be used to provide an objective measure of such stability. Using discriminant analysis for validating cluster analysis has several drawbacks. Discriminant coefficients may be poor estimates of population values and need to be cross-validated themselves This procedure is not cost-effective and the sample size available may be insufficient for cross-validating both the cluster analysis and a discriminant analysis.

McIntyre and Blashfield (1980) discussed an alternative approach to cross-validation which is recommended here. The procedure is relatively simple and easy to implement on a computer. Cluster analysis is first carried out on one half of the observations available for analysis. Once a statistically significant clustering solution has been identified, centroids describing the clusters are obtained. Objects in the holdout data set are then assigned to one of the identified clusters on the basis of the smallest Euclidean distance to a cluster centroid vector. The degree of agreement between the nearest-centroid assignments of the holdout sample and the results of a cluster analysis of the holdout sample is an indication of the stability of the solution. A coefficient of agreement, kappa, may be used as an objective measure of stability. If an acceptable level of stability is obtained the data sets may be combined to obtain a final solution.

The demonstration of the statistical significance and stability of a cluster solution is necessary before one can accept and use the classification developed by the methodology. The acceptance of a particular classification system, whether developed through cluster analysis or some other method, requires a further demonstration of utility, however. Classification is only useful if it assists in furthering an understanding of the phenomena of interest. Clusters, or classes, must have demonstrable implications for hypothesis generation, theory building, prediction, or management. The ultimate test of a set of clusters is its usefulness. Thus, the user of cluster analysis should provide a demonstration that clusters are related to variables other than those used to generate the solution. Ideally, only a small number of variables should be required to classify individuals. This classification should then have implications beyond the narrow set of classification variables. The task of classifications is not finished until these broader implications have been demonstrated.

VARIABLE SELECTION ISSUES

The findings of empirical studies of cluster methods suggest that attention to initial variable selection is crucial because even one or two irrelevant variables may distort an otherwise useful cluster solution. The basis for classification must be carefully identified to ensure that extraneous characteristics do not distort an otherwise useful cluster analysis. There should be some rationale for the selection of variables for cluster analysis. That rationale may grow out of an explicit theory or be based on a hypothesis. Clearly more attention needs to be paid to this critical issue. As a science develops, researchers must agree on those dimensions which are most relevant to classification for a particular purpose. Much debate in the science of marketing involves the issue of variable selection. Thus, it is not surprising that a variety of different classification systems have been developed for similar phenomena. Indeed, it is probably unrealistic to expect that a single classification system will emerge in any area of marketing in the foreseeable future. Rather, there are likely to be numerous competing systems. The development of diverse systems is healthy and has been observed in other sciences. Experience with rival systems and a comparison of their usefulness ultimately provide a basis for selection of one system over another. Cluster analysis has much to offer as an aid for developing classification systems. To the extent that classification is both the first and last step in scientific investigation, cluster analysis should have increasing application in marketing.

REFERENCES

Aldenderfer, M S. (1977), "A Consumer Report on Cluster Analysis Software· (2) Hierarchical Methods," working paper, Department of Anthropology, Pennsylvania State University

Anderberg, M. R. (1973), *Cluster Analysis for Applications* New York: Academic Press

Anderson, W. T., Jr , E. P Cox, III, and D. G Fulcher (1976), "Bank Selection Decisions and Market Segmentation," *Journal of Marketing*, 40, 40–45.

Arabie, P., J. D. Carroll, W. DeSarbo, and J. Wind (1981), "Overlapping Clustering A New Method for Product Positioning," *Journal of Marketing Research*, 18 (August), 310–17

Arnold, S. J. (1979), "A Test for Clusters," *Journal of Marketing Research*, 16 (November), 545–51

Bailey, K. D. (1974), "Clusters Analysis," in *Sociological Methodology*, D. Heise, ed. San Francisco: Jossey-Bass.

Ball, G. H. (1970), *Classification Analysis*. Menlo Park, CA: Stanford Research Institute.

Bass, F. M , E. A. Pessemier, and D. J. Tigert (1969), "A Taxonomy of Magazine Readership Applied to Problems in Marketing Strategy and Media Selection," *Journal of Business*, 42, 337–63

Bayne, C. K., J. J Beauchamp, C L Begovich, and V E Kane (1980), "Monte Carlo Comparisons of Selected Clustering Procedures," *Pattern Recognition*, 12, 51–62.

Bertin, J. (1967), *Semiology Graphique*. Paris: Gauthier-Villars.
Bettman, J. R. (1979), *An Information Processing Theory of Consumer Choice*. Reading, MA: Addison-Wesley.
Blashfield, R. K. (1976), "Mixture Model Tests of Cluster Analysis: Accuracy of Four Agglomerative Hierarchical Methods," *Psychological Bulletin*, 83, 377–88.
——— (1977), "A Consumer Report on Cluster Analysis Software: (3) Iterative Partitioning Methods," working paper, Department of Psychology, Pennsylvania State University
——— (1978), "The Literature on Cluster Analysis," *Multivariate Behavioral Research*, 13, 271–95.
——— and L. C. Morey (1980), "A Comparison of Four Clustering Methods Using MMPI Monte Carlo Data," *Applied Psychology Measurement*, in press.
Calantone, R. J and A. G. Sawyer (1978), "Stability of Benefit Segments," *Journal of Marketing Research*, 15 (August), 395–404.
Cattell, R. B. (1949), "R_p and Other Coefficients of Pattern Similarity," *Psychometrika*, 14, 279–98.
——— (1978), *The Scientific Use of Factor Analysis in the Behavioral and Life Sciences*. New York: Plenum Press.
Claxton, J. D., J. N Fry, and B. Portis (1974), "A Taxonomy of Pre-purchase Information Gathering Patterns," *Journal of Consumer Research*, 1, 35–42.
Cohen, J (1960), "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, 20, 37–46.
Cormack, R. M. (1971), "A Review of Classification," *Journal of the Royal Statistical Society* (Series A), 134, 321–67.
Cunningham, K. M. and J. C. Ogilvie (1972), "Evaluation of Hierarchical Grouping Techniques: A Preliminary Study," *Computer Journal*, 15, 209–13.
Day, G. S. and R. M Heeler (1971), "Using Cluster Analysis to Improve Marketing Experiments," *Journal of Marketing Research*, 8 (August), 340–7.
Dixon, W. J. and M. B. Brown (1979), *BMDP: Biomedical Computer Programs, P Series, 1979*. Los Angeles: University of California Press.
Edelbrock, C. (1979), "Comparing the Accuracy of Hierarchical Clustering Algorithms: The Problem of Classifying Everybody," *Multivariate Behavioral Research*, 14, 367–84
——— and B. McLaughlin (1980), "Hierarchical Cluster Analysis Using Intraclass Correlations: A Mixture Model Study," *Multivariate Behavioral Research*, 15, 299–318.
Everitt, B. S. (1974), *Cluster Analysis*. London: Halsted Press.
Field, H S. and L. F. Schoenfeldt (1975), "Ward and Hook Revisited: A Two-Part Procedure for Overcoming a Deficiency in the Grouping of Two Persons," *Educational and Psychological Measurement*, 35, 171–3.
Fisher, W. D. (1969), *Clustering and Aggregation in Economics*. Baltimore, MD: Johns Hopkins Press.
Frank, R. E. and P. E. Green (1968), "Numerical Taxonomy in Marketing Analysis: A Review Article," *Journal of Marketing Research*, 5 (February), 83–98.
Friedman, H D and J. Rubin (1967), "On Some Invariant Criteria for Grouping Data," *Journal of the American Statistical Association*, 62, 1159–78.
Furse, D. H., G Punj, and D. W. Stewart (1982), "Individual Search Strategies in New Automobile Purchase," in *Advances in Consumer Research*, Vol. 9, Andrew Mitchell, ed., 379–84.
Gerard, R. W. (1957), "Units and Concepts of Biology," *Science*, 125, 429–33.
Gower, J. C. (1967), "A Comparison of Some Methods of Cluster Analysis," *Biometrics*, 23, 623–37.
Green, P. E. and F. J. Carmone (1968), "The Performance Structure of the Computer Market: A Multivariate Approach," *Economic and Business Bulletin*, 20, 1–11.
———, R. E. Frank, and P. J. Robinson (1967), "Cluster Analysis in Test Market Selection," *Management Science*, 13, B-387-400.
——— and V. R. Rao (1969), "A Note on Proximity Measures and Cluster Analysis," *Journal of Marketing Research*, 6 (August), 359–64.
Greeno, D. W., M. S. Sommers, and J. B. Kernan (1973), "Personality and Implicit Behavior Patterns," *Journal of Marketing Research*, 10 (February), 63–9.
Hartigan, J. A. (1967), "Representation of Similarity Matrices by Trees," *Journal of the American Statistical Association*, 62, 1140–58.
——— (1975), *Clustering Algorithms*. New York: John Wiley & Sons, Inc.
Howard, H and B. Harris (1966), "A Hierarchical Grouping Routine, IBM 360/65 FORTRAN IV Program," University of Pennsylvania Computer Center.
Inglis, J. and D. Johnson (1970), "Some Observations On, and Developments In, the Analysis of Multivariate Survey Data," *Journal of the Market Research Society*, 12, 75–8.
Jardine, H. and R. Sibson (1971), *Mathematical Taxonomy*. New York: John Wiley & Sons, Inc.
Johnson, S. C. (1967), "Hierarchical Clustering Schemes," *Psychometrika*, 32, 241–54.
Kantor, J. R. (1953), *The Logic of Modern Science*. Bloomington, IN: Principle Press.
Kemeny, J. G. (1959), *A Philosopher Looks at Science*. New York: Van Nostrand.
Kernan, J. B. (1968), "Choice Criteria, Decision Behavior, and Personality," *Journal of Marketing Research*, 5 (May), 155–69.
——— and G. D. Bruce (1972), "The Socioeconomic Structure of an Urban Area," *Journal of Marketing Research*, 9 (February), 15–18.
Kiel, G. C. and R. A. Layton (1981), "Dimensions of Consumer Information Seeking Behavior," *Journal of Marketing Research*, 18 (May), 233–9.
Kuiper, F. K. and L. A. Fisher (1975), "A Monte Carlo Comparison of Six Clustering Procedures," *Biometrics*, 31, 777–83.
Lance, G. N. and W. T. Williams (1967a), "A General Theory of Classificatory Sorting Strategies. I. Hierarchical Systems," *The Computer Journal*, 9, 373–80.
——— (1967b), "A General Theory of Classificatory Sorting Strategies II. Clustering Systems," *The Computer Journal*, 10, 271–7.
Landon, E. L. (1974), "Self Concept, Ideal Self Concept, and Consumer Purchase Intentions," *Journal of Consumer Research*, 1, 44–51.
Lessig, V. P. and J. D. Tollefson (1971), "Market Segmentation Through Numerical Taxonomy," *Journal of Marketing Research*, 8 (November), 480–7.
McIntyre, R. M and R K. Blashfield (1980), "A Nearest-Centroid Technique for Evaluating the Minimum-Variance

Clustering Procedure," *Multivariate Behavior Research*, 15, 225–38.
McQuitty, L. L. (1967), "A Mutual Development of Some Typological Theories and Pattern-Analytic Methods," *Educational and Psychological Measurement*, 17, 21–46.
McRae, D. J. (1973), "Clustering Multivariate Observations," unpublished doctoral dissertation, University of North Carolina.
Mezzich, J E. (1978), "Evaluating Clustering Methods for Psychiatric Diagnosis," *Biological Psychiatry*, 13, 265–81.
Milligan, G. W. (1980), "An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms," *Psychometrika*, 45, 325–42.
——— (1981), "A Monte Carlo Study of Thirty Internal Criterion Measures for Cluster Analysis," *Psychometrika*, 46, (2), 187–99.
——— and P. D Isaac (1980), "The Validation of Four Ultrametric Clustering Algorithms," *Pattern Recognition*, 12, 41–50.
——— and V. Mahajan (1980), "A Note on Procedures for Testing the Quality of a Clustering of a Set of Objects," *Decision Sciences*, 11, 669–77.
Mojena, R. (1977), "Hierarchical Grouping Methods and Stopping Rules: An Evaluation," *Computer Journal*, 20, 359–63.
Montgomery, D. B and A J. Silk (1971), "Clusters of Consumer Interests and Opinion Leaders' Sphere of Influence," *Journal of Marketing Research*, 8 (August), 317–21
Morgan, J. N. and J. A Sonquist (1963), "Problems in the Analysis of Survey Data, and a Proposal," *Journal of the American Statistical Association*, 58, 87–93
Moriarty, M. and M. Venkatesan (1978), "Concept Evaluation and Market Segmentation," *Journal of Marketing*, 42, 82–6.
Morrison, B. J. and R C. Sherman (1972), "Who Responds to Sex in Advertising?" *Journal of Advertising Research*, 12, 15–19.
Morrison, D. G. (1967), "Measurement Problems in Cluster Analysis," *Management Science*, 13, B-775–80.
Myers, J. G. and F. M Nicosia (1968), "On the Study of Consumer Typologies," *Journal of Marketing Research*, 5 (May), 182–83.
Neidell, L A. (1970), "Procedures and Pitfalls in Cluster Analysis," *Proceedings*, Fall Conference, American Marketing Association, 107
Nerviano, V J and W F. Gross (1973), "A Multivariate Delineation of Two Alcoholic Profile Types on the 16PF," *Journal of Clinical Psychology*, 29, 370–4.
Peay, E. R (1975), "Nonmetric Grouping Clusters and Cliques," *Psychometrika*, 40, 297–313.
Rand, W M (1971), "Objective Criteria for the Evaluation of Clustering Methods," *Journal of the American Statistical Association*, 66, 846–50
Rogers, G. and J D. Linden (1973), "Use of Multiple Discriminant Function Analysis in the Evaluation of Three Multivariate Grouping Techniques," *Educational and Psychological Measurement*, 33, 787–802.
Rubin, J (1965), "Optimal Taxonomy Program (7090-IBM-0026)," IBM Corporation
Schaninger, C M , V P Lessig, and D B. Panton (1980), "The Complementary Use of Multivariate Procedures to Investigate Nonlinear and Interactive Relationships Between Personality and Product Usage," *Journal of Marketing Research*, 17 (February), 119–24
Sethi, S. P (1971), "Comparative Cluster Analysis for World Markets," *Journal of Marketing Research*, 8 (August), 348–54
Sexton, D. E., Jr (1974), "A Cluster Analytic Approach to Market Response Functions," *Journal of Marketing Research*, 11 (February), 109–14.
Shepard, R. N. (1974), "Representation of Structure in Similarity Data. Problems and Prospects," *Psychometrika*, 39, 373–421.
——— and P Arabie (1979), "Additive Clustering: Representation of Similarities as Combinations of Discrete Overlapping Properties," *Psychological Review*, 86, 87–123
Sherman, L. and J N. Sheth (1977), "Cluster Analysis and Its Applications in Marketing Research," in *Multivariate Methods for Market and Survey Research*, J. N. Sheth, ed Chicago: American Marketing Association.
Shuchman, A. (1967), "Letter to the Editor," *Management Science*, 13, B688–96
Skinner, H. A. (1979), "Dimensions and Clusters. A Hybrid Approach to Classification," *Applied Psychological Measurement*, 3, 327–41
Smith, W. (1956), "Product Differentation and Market Segmentation as Alternative Marketing Strategies," *Journal of Marketing*, 21, 3–8.
Sneath, P H. A. (1957), "The Application of Computer to Taxonomy," *Journal of General Microbiology*, 17, 201–26.
——— and R R Sokal (1973), *Numerical Taxonomy*. San Francisco: W. H Freeman.
Stewart, D. W (1981), "The Application and Misapplication of Factor Analysis in Marketing Research," *Journal of Marketing Research*, 18 (February), 51–62.
Srivastava, R. K , R P. Leone, and A. D Shocker (1981), "Market Structure Analysis: Hierarchical Clustering of Products Based on Substitution-in-Use," *Journal of Marketing*, 45 (Summer), 38–48
———, A D Shocker, and G S. Day (1978), "An Exploratory Study of Usage-Situational Influences on the Composition of Product-Markets," in *Advances in Consumer Research*, H. K Hunt, ed Ann Arbor: Association for Consumer Research, 32–38
Tyron, R. C and D. E Bailey (1970), *Cluster Analysis* New York McGraw-Hill Book Company.
Veldman, D J (1967), *FORTRAN Programming for the Behavioral Sciences*. New York: Holt, Rinehart, and Winston.
Ward, J. (1963), "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, 58, 236–44
Wells, W D (1975), "Psychographics A Critical Review," *Journal of Marketing Research*, 12 (May), 196–213
Wind, Y (1978), "Issues and Advances in Segmentation Research," *Journal of Marketing Research*, 15 (August), 317–37
Wishart, D (1969), "CLUSTAN IA. A FORTRAN Program for Numerical Classification," Computing Laboratory, St Andrew's University, Scotland
Wolf, A. (1926), *Essentials of Scientific Method* New York. Macmillan Company