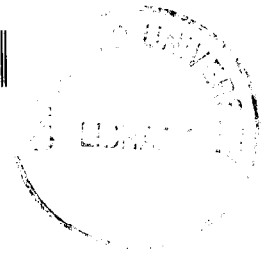




1402801494



**SWP 25/88 ISSUES TO BE ADDRESSED PRIOR TO
UNDERTAKING HIERARCHICAL CLUSTER
ANALYSIS**

DR LESLIE DE CHERNATONY
Lecturer in Marketing
Cranfield School of Management
Cranfield University
Cranfield
Bedford MK43 0AL
United Kingdom

Tel: +44 (0)234 751122

Fax: +44 (0)234 781806

276

ISSUES TO BE ADDRESSED PRIOR TO UNDERTAKING HIERARCHICAL CLUSTER ANALYSIS

DR LESLIE DE CHERNATONY
Lecturer in Marketing
Cranfield School of Management
Cranfield
Bedford MK43 0AL
United Kingdom

(Tel: 0234 - 751122)

School Working Paper 25/88

Copyright Chernatony 1988

Cluster analysis shows how items are allocated to previously undefined groups, such that items in the same group are in some sense similar to each other (Everitt 1979). The (jxk) matrix might first be standardised and is then transformed to a (jxj) matrix of similarities or distances, to which a clustering algorithm is applied to identify any clusters. Some of the advantages of cluster analysis are that no a-priori statement about groupings have to be made, hierarchical algorithms allow the evolutionary clustering process to be seen (cf principal component analysis) and following the pioneering work of Sokal and Sneath (1963) there is a considerable literature on this subject. These advantages make this a most applicable technique for analysing categorisation.

The problem with cluster analysis is that it identifies clusters when there are no natural clusters (Arnold 1979) and there is no universally accepted definition of a cluster (Everitt 1986). As a consequence of this latter weakness, a multitude of clustering algorithms exist (eg Punj and Stewart 1983), each satisfying a different criteria about what constitutes a cluster and, not surprisingly, producing different clustering schema for the same data (eg Mojena 1977). In view of the widespread use of cluster analysis in marketing, it is imperative that market researchers spend some time justifying their cluster analysis strategy (eg Bailey 1974), as is next considered.

WHICH ATTRIBUTES TO USE?

When trying to identify groups of items that display internal cohesion and are externally isolated from other items (ie Cormack's (1971) definition of a cluster), it should be recognised that the attribute statements selected will affect the resulting groupings (Everitt 1986). One way of generating consumer relevant attributes is to use elicitation techniques (eg Kelly repertory grids) followed by principal component analysis to reduce the number of statements. (eg Doyle and Saunders 1985).

A-PRIORI WEIGHTING OF ATTRIBUTES?

Some (eg Morrison 1967) argue that it is right to weight the importance of attributes, in order to satisfy a management orientation. An alternative view is expounded by Sokal and Sneath (1963) who argue that:-

- there are no standard rules as to how attributes should be weighted

the number of items, this problem disappears. When competing items load "heavily" on more than one component the problem then arises as to which cluster the item belongs. Stewart (1981) presents pictorial evidence of a further problem of using the components of Q-type principal component analysis as clusters, shows a two dimensional space defined by the first two components of a principal component analysis. Visually the clustering of items represents two distinct clusters, characterised by their direction and level of association with component 2. A Q-type principal component analysis would define clusters of items in terms of their association with specific components. As there is no notable contribution from component 1, but an overwhelming contribution from component 2, this would suggest that there is only one cluster present. Figure 1B indicates that the items group into one cluster, but defining clusters as the principal components would incorrectly suggest two clusters.

FIGURE 1 ABOUT HERE

As Cattell (1965) recommended, a visual inspection of the inter-item correlation matrix, without factoring, should indicate any groupings (eg Ehrenberg and Goodhardt 1976).

Multidimensional scaling (MDS) describes procedures that investigate a matrix of distances between items to find a configuration in a smaller number of dimensions, with the inter item distances in the reduced dimension space closely reflecting the original inter item distances (Chatfield and Collins 1980). According to whether respondents stated how much they agreed or disagreed with statements describing items (metric MDS) or whether they directly ranked their perceptions of similarity between all pairs of the items (non-metric MDS), a low dimensional map showing the spatial positions of the competing items will be obtained. Limitations have been noted with this technique, eg non-metric methods may produce meaningless results (Green and Tull 1978) and MDS may not give a good representation of the basic structure (Everitt 1986). While MDS is widely used in marketing (Wind 1978) its prime purpose is to produce maps showing the spatial configuration of items rather than acting as a classification method (Cormack 1971).

Recently Jones and Sibson (1987) described the technique of projection pursuit, which while not widely used may be of increasing interest to market researchers in future years as more experience is gained. The technique is concerned with reducing the dimensionality of the data and examining the projections of the data set. By seeking local minima of an index of uninterestingness amongst projection onto one or two dimensions the analyst is able to observe the clustering of items.

INTRODUCTION

At some stage in most market researchers' careers they will come across the problem of having to analyse their data to identify any groupings of products, brands, people or occasions. With the increasing availability of powerful personal computers and statistical software, it is becoming easier for the market researcher to undertake this type of analysis. There is however an increasing danger of some researchers not spending sufficient time appreciating the different options within a particular software package and hence there is a greater tendency to accept the default options without fully appreciating their implications. This paper focuses upon the use of cluster analysis in market research and examines the key issues that market researchers should be aware of before using such computer packages. In view of the popularity of hierarchical clustering methods in marketing (Punj and Stewart 1983), this paper is mainly concerned with hierarchical algorithms.

HOW TO MEASURE GROUPINGS WITHIN DATA?

The raw data from which analysis of any categories starts is a (jxk) matrix showing the way that j entities were evaluated on k attributes. Some of the techniques employed to identify categories are Q-type principal component analysis (Schlinger 1969), multidimensional scaling (Doyle 1975), discriminant analysis (Dillon and Goldstein 1984), and cluster analysis (Saunders 1980). Each of these techniques are based on different theoretical foundations and consequently different categorisation can be recorded from the same (jxk) matrix according to which technique was employed (Mezzich 1978). Discriminant analysis will not be discussed in this paper, since unlike the other categorisation techniques, this starts with a-priori specification of groups.

A Q-type principal component analysis is based around the (jxj) correlation matrix showing the similarity between items. The items of interest are placed in groups characterised by the components on which they have the highest loadings. The prime role of principal component analysis is as a data reduction technique and Dillon and Goldstein (1984) echo the concern of many researchers when stating that this method of identifying groupings is "plagued with a number of problems and ambiguities" (p 43). By analysing the correlations between entities, only the shape of the items' profiles are taken into account, regardless of profile levels. The assumption of an underlying linear model imposes a constraint which is frequently glossed over. The number of components that can be extracted from the (jxj) competing items' correlation matrix will be a maximum of the k attributes less one and thus the number of categories is determined by the number of attributes used. Where though the number of attributes exceeds

AUTOBIOGRAPHICAL NOTE

Dr. Leslie de Chernatony has had over 7 years market research experience. His doctoral research at the City University Business School was concerned with assessing consumers' perceptions of brands and retailer labels. He is currently Lecturer in Marketing at the Cranfield School of Management.

- it is wrong to use weighting criteria which presuppose the existence of clusters
- any weight must be based on intuition which varies between researchers. Subjective decisions would reduce the generality of the research findings.

WHAT MEASURE OF SIMILARITY/DISSIMILARITY?

Three broad options are available ie association measures, correlation measures and distance measures. There is no universal agreement as to which measure should be used (Dillon and Goldstein 1984) and the choice of measure can influence the resulting groupings (eg Edelbrock 1979).

There are numerous association coefficients, each based on a different matching criterion. For the same data set Everitt (1986) shows that the different association coefficients are not monotonic in the sense that if all the values for pairs of items on one association coefficient are ordered so they form a series arranged in increasing or decreasing value, the values of the individual pairs taken by another association coefficient will not be in increasing or decreasing order.

Correlation coefficients have an advantage over association coefficients, since by considering deviations from a mean, the magnitude of mismatches between items is taken into account. This can be viewed as being a weakness in cluster analysis though, since each item is being averaged over disparate attributes, and for this reason some condemn this measure (eg Jardine and Sibson 1971). As Fleiss and Zubin (1969) observed, understanding the meaning of the value of a correlation coefficient between items presents problems - we do not talk about the "independence" of competing items when there is a correlation coefficient of zero. Nor does this mean the items are dissimilar, since this is what a correlation coefficient of minus 1 means.

Distance measures are frequently used in cluster analysis, the most widely used form being Euclidean distance (Sherman and Sheth 1977). Unlike correlation coefficients they are invariant under alterations in the direction of coding attributes (Minkoff 1965). They are also easier to interpret since, for example, a value of zero means that two competing items' profiles on the same attribute are identical. This measure is not without problems, eg the measuring scale influences the magnitude of the scores and no corrections are made if there are any correlations between the attributes over which the items were scored (albeit this would have been corrected for when using principal component analysis to help select the attributes).

A policy of selecting correlation coefficients because of others' comparative results is regarded as being a weak procedure. There must be some rational basis for selection, other than it happens to work. With a firm reason for choosing a particular similarity/dissimilarity measure, there is a logical basis to analyse unexpected results, rather than having to admit to little more than surprise. Also some of the findings are only useful if one has a data set similar to that tested by others. Thus Edelbrock's (1979) finding that correlation measures produced more accurate clusters than Euclidean distances, is best applicable when the market researcher has a data set similar to that generated by Edelbrock. One of the reasons for this result was that the data set had several items that scored at the extremes of the attribute scales (ie were outliers). Since Euclidean distances are sensitive to outliers, these would not have clustered until quite late in the process.

The issues raised in this section should make the market researcher recognise that the properties of the similarity/dissimilarity measure will have an impact on the resulting clusters. The choice of an appropriate measure must be guided not only by an understanding of the properties of that measure, but it must also be appropriate to the scaling of the data and the specific clustering algorithm employment. Some generalisation can be made about the types of similarity/dissimilarity measures for different types of data and their applicability to different clustering algorithms.

When observations are nominally scaled, association coefficients should be employed. Thus when grouping manufacturers on the basis of whether or not they have certain attributes, one of the variety of association coefficients reviewed by Sokal and Sneath (1983) can be used. Where the observations are based upon a mixture of nominal and interval scales, Gower's similarity coefficient is best used. These association measures can be used with most hierarchical techniques except Wards method and some of the average link algorithms (Centroid and Median).

The other similarity measures the coefficient of rank correlation and the product-moment correlation coefficient can be used for ordinal and interval data respectively. The limitations on this applicability in hierarchical clustering techniques are similar to those of association coefficients. For data having metric properties, a distance type measure can be used. Distance measures are appropriate for all forms of hierarchical agglomerative clustering algorithms, albeit the computational form of the distance measure may vary according to the algorithm (eg the error sum of squares used in Wards method).

SHOULD STANDARDISED OR RAW DATA BE USED?

Calculating the distance matrix between items is influenced by the differing scales that may have been used for each of the attributes, by the non-comparability of attributes (eg how meaningful is it to combine scores on pricing with packaging?) and by the extent to which the full width of the scales were used. To overcome these issues some (eg Wishart 1978; Schaninger and Buss 1986) recommend reducing all attributes to a standard form to enable a more logical combination of item scores. It should be realised that as standardisation does not equally transform all of the item scores, this may alter the relationship between scores across all of the attributes for a particular item. While there are cases reported where standardisation did not affect the resulting cluster (eg Rohlf and Sokal 1965; Edelbrock 1979), the analyst should seriously consider the implications of standardisation and "not rigidly adhere to such a practice" (p 103, Anderberg 1973).

WHAT CLUSTERING ALGORITHM?

There are many clustering algorithms (eg Anderberg 1973; Everitt 1986), each satisfying a different definition of a cluster. In the broadest terms, there are 2 main classes of cluster analysis, hierarchical methods where, over a series of separate clustering cycles, the evolution of clusters is shown and non-hierarchical methods where the user only sees the final grouping of items. The hierarchical methods are particularly powerful, since one can see the order in which the clusters emerge and thus the rest of this section focuses on hierarchical methods.

When using a hierarchical technique there is the choice of an agglomerative method, where on each clustering cycle items are amalgamated into clusters on an incremental similarity basis or a divisive method where the total group of items are successively divided until each item is in its own separate group. The large number of ways of dividing the items imposes restrictions on the number of clusters that can be considered when analysing very large numbers of entities (Dillon and Goldstein 1984) and in part related to this, agglomerative techniques tend to be more frequently used (Mezzich 1978). Blashfield and Aldenderfer (1978) reported that hierarchical agglomerative techniques are the more frequently employed clustering techniques, partly because they are so established and because they have been the subject of considerable investigation (eg Cunningham and Ogilvie 1972; Kuiper and Fisher 1975; Edelbrock 1979; Milligan and Isaac 1980).

The four more popular agglomerative hierarchical methods are single link, complete link, average link and minimum variance (Blashfield 1976). As Johnson (1967) argues, the main difference between these algorithms is the way that the similarity-dissimilarity relation between clusters and remaining items is used to form new clusters. Lance and Williams (1967) developed a general equation showing how the different agglomerative methods calculate the different levels of cluster compositions. The resulting hierarchical clustering schema is displayed on a dendrogram, ie a hierarchical tree showing the levels of similarity/dissimilarity at which each of the clusters formed.

Single link cluster analysis (also referred to as nearest neighbour) regards a cluster as a group of items where each member of the cluster is more similar to at least one member of the same cluster than to any member of another cluster. Williams et al (1971a) show that the single link method is "space contracting", ie as cluster membership increases, the space containing items contracts since the clusters appear to approach unclustered items. This leads to the characteristic of "chaining" where, instead of the cluster being compact points in space, they take on a chain like form. All of the usual similarity/dissimilarity measures can be employed within this algorithm.

Complete link (or furthest neighbour) is the opposite to single link. It defines a cluster such that membership is based on items being more similar to all members of the same cluster, than to all members of any other cluster. This algorithm is space dilating in the sense that as clusters grow they move away from unclustered items which are more likely to form new clusters, rather than add to existing items. Again all of the usual similarity/dissimilarity measures are applicable.

In average link, clusters are formed on the basis of each group member having a greater average similarity with all members of the same group than with all members of any other group. Various methods result as a consequence of the way that averages are calculated (eg centroid, median, group average methods). These methods are space conserving. Only distance measures can be used for the Centroid and Median methods, while for the group average method all of the similarity/dissimilarity methods can be employed.

The minimum variance method (also known as Wards method) seeks clusters which satisfy the criterion that the error sum of squares among members of each cluster is a minimum. Only distance measures can be used with this algorithm. Further details about the cluster structures that each of these methods tend to find can be found in Anderberg (1973).

With numerous algorithms to choose from, it would be wise for market researchers to clarify what definition of a cluster they feel to be most appropriate and then select a suitable algorithm. In reality though this procedure is normally reversed. Some (eg Jardin and Sibson 1968; Fisher and van Ness 1971) list mathematical criteria that clustering methods must meet to warrant their selection. A problem with this approach is that there is no agreement about which criteria are important (eg Williams et al 1971b). Numerous empirical studies have been reported comparing the performance of various clustering algorithms; (Everitt 1986; Doyle and Saunders 1985; Edelbrock 1979; Kuiper and Fisher 1975; Cunningham and Ogilvie 1972), but it would be wrong to select an algorithm solely on the basis of these studies. The comparative results are specific to the distributions generated and the researcher would need to make assumptions about distribution similarities.

In conclusion as no one clustering algorithm has emerged as the accepted norm (Punj and Stewart 1983), market researchers should start by clarifying what they understand by a cluster and select an algorithm that best meets their perspective. For example, where the researcher feels there are only small distances between the items under investigation (eg the way some grocery own labels have become similar to competing brands) then the use of techniques such as single link, with the characteristic of showing chaining, may be close to the researchers view of what is meant by a cluster. By contrast if the market researcher intuitively feels that he is investigating compact groups of items, well separated from other groups of items then complete link cluster analysis may be most appropriate.

A problem that all market researchers face when using cluster analysis is that clusters will be found by all algorithms irrespective of the presence or absence of "natural clusters" in the data (eg Rand 1971; Arnold 1979). This problem arises in part from the fact that there is no universal agreement about what constitutes a natural cluster and clustering algorithm is attempting to identify clusters which satisfy the definition of a cluster for that particular algorithm. For this reason it is essential that the analyst be clear about his understanding of the term cluster since he will need to apply subjective and intuitive judgement to interpret the cluster results and their significance (Anderberg 1973).

An alternative approach to the pragmatic method of judging the cluster compositions is to examine the literature for tests on the significance of clusters. Again no universal test has been accepted, in part because of different assumptions about what constitutes a cluster and the programme of the distribution form of the data. McClain and Raro (1975) proposed that statistics based upon the ratio of the average within - cluster proximity to the average between cluster

proximity could be used to determine whether a clustering of the data set was significant. This observed statistic was compared against an approximate sampling distribution based on a specific null hypothesis. As Milligan and Mahajan (1980) showed there is a tendency to indicate significant cluster structure in data that do not possess such structure because the sampling distribution for the null hypothesis is strongly specified. Aronld (1979) proposed that the statistics

$$C = \log (\max ITI / IWI)$$

be used as a test of the statistical significance of a cluster solution. In this formula ITI is the determinant of the total covariance matrix and IWI is the determinant of the pooled within groups covariance matrix. He generated distributions of the C statistic and indicated values of C which allow rejection of the null hypotheses that the data arise from uniform distributors. While this test can be quite widely applied, Sneath (1977) developed a significance test specifically for clusters obtained from the average link algorithm when squared Euclidean distance measures are employed. His test is based upon an assessment of whether the observed overlap between 2 clusters is significantly less than some critical level of overlap.

An alternative approach to assessing the significance of the clusters is by estimating how well the cluster solution matches the "true" structure within the data set. The problem that most market researchers will face with this approach is that cluster analysis is normally being used because the "true" structure of the data set is not known. By comparing the composition of the clusters obtained from a cluster algorithm against the known cluster structure, the Rand statistic (Rand 1971) or Kappa statistic (Blashfield 1986) provides some guidance as to the similarity of structures.

INTERPRETING AND COMPARING CLUSTERING SCHEMA

Having reached the state where the hierarchical clustering of the items is displayed on a dendrogram, the next task is to decide how many clusters to choose (ie which level on the dendrogram to choose). While tests have been developed for assessing the number of clusters, a review of the cluster analysis literature reveals that for the more common agglomerative hierarchical techniques there is no accepted solution as to the most appropriate number of clusters (Rao and Sabavala 1981; Milligan and Mahajan 1980; Everitt 1979). Several reasons exist

for this. Techniques such as single link are not built on any formal statistical framework, the definition of a cluster is unique to a particular algorithm and there are problems identifying the sampling distribution of the inter-item distances along with developing a flexible test procedure (Lenington and Flake 1975). Gower (1975) suggests a test whereby the levels at which clusters form are plotted against the number of clusters. The point where a sharp change in gradient occurs indicates the number of clusters (cf scree test in factor analysis). Everitt (1979) reported though that there are problems in interpreting these graphs. Mojena (1977) developed a rule based upon the distribution of the fusion levels at which the items form clusters. The mean and standard deviation of the cluster fusion levels are calculated and the number of clusters present is that for which the value of the fusion distance exceeds a level specified by the mean fusion level plus a multiple of the standard deviation. The problem with this procedure is it has no theoretical basis and as Mojena (1971) stated, it was developed on the basis that it "appears to give good results" (p 68).

An approach which has greater managerial orientation, is for the market researcher to examine each level of the dendrogram and to consider those cluster structures that have managerial meaning, presenting several levels of the dendrograms to management. Examples of this procedure being followed are provided by de Chernatony (1987) and Doyle and Saunders (1985). In de Chernatony's (1987) study he was interested in observing how consumers perceived the competitive structure of product fields. Since marketers talk about resource allocation in terms of brands, own labels and generics he examined the cluster compositions at the 3 cluster level.

To compare the hierarchical clustering of items between several groups of respondents, the composition of the clusters at different levels on the dendrogram should be visually inspected. Further guidance is provided by using the cophenetic correlation coefficient (Sokal and Rohlf 1962). The inter-item distance matrices of the dendrograms being compared are each strung out in the form of single columns and the product moment correlation coefficient is calculated - this being referred to as the cophenetic correlation coefficient. This gives an indication of the similarity of the structures in terms of the distance at which the levels of the dendrograms occur, but may conceal small differences in terms of the composition of the clusters. As an example of this, figure 2 shows the dendrograms for the split half test of a group of respondents evaluating competing items A to H, after being subject to single link cluster analysis. The cluster composition at the 3 cluster level for each dendrogram (ie below the dotted lines) are:

(A,B,C,D) (E) (F,G,H) : top dendrogram

(A,B,C,D,E) (F) (G,H) : bottom dendrogram

The cophenetic correlation coefficient between these 2 dendrograms is 0.9956 indicates a mirror image, yet as can be seen the composition of the clusters at the 3 cluster level is different between the 2 dendrograms.

FIGURE 2 ABOUT HERE

VALIDITY AND RELIABILITY ISSUES

To assess the construct validity of the algorithm selected, ie how validly does the chosen algorithm measure groupings, an approach recommended (eg Everitt 1986; De Sarbo 1982) is to apply several other clustering algorithms to the same data. Such an approach can be relatively quickly undertaken, but since each algorithm is based on a different concept of a cluster, employing several clustering algorithms to the same data will not necessarily produce similar results (eg Mojena 1977).

The extent to which the resulting clusters can be replicated, ie show some degree of reliability, can be assessed using the split half method (eg Funkhouser 1983). If respondents' perceptions are being measured, by randomly splitting them into two equally sized groups and then applying the same algorithm to the inter-item similarity/dissimilarity matrices, a comparison of the resulting dendrograms is an indicator of reliability. This should be treated as an indicator, since it could be assessing how well the random divider matched the respective halves.

COMPUTATION

The most comprehensive suite of cluster analysis programs in the United Kingdom is that provided by CLUSTAN (Wishart 1987) which offers the user over 30 different clustering methods and which is frequently being revised. This suite provides over 40 types of similarity/dissimilarity measures, provides analysis statistics (eg cophenetic correlation coefficient) and presents good graphic displays. It also facilitates analysis by reading SPSS data files. Some of the general statistical software packages also offer cluster analysis facilities eg BMDP, SAS and SPSS.

CONCLUSIONS

This paper has identified a series of issues that market researchers need to address before undertaking cluster analysis of their data. It has been shown that according to the cluster analysis strategy adopted, so different clustering results may be recorded. There is no standard strategy which exists and market researchers should develop their own justification for the route they have followed - and not simply accept the software default options!

REFERENCE

- Anderberg, M. R. (1973). Cluster analysis for applications (New York: Academic Press).
- Arnold, S. J. (1979). A test for clusters. Journal of Marketing Research 16 (4): 545-551.
- Bailey, K. D. (1974). Cluster analysis. In: Sociological methodology 1975, ed by D. R. Heise (San Francisco: Jossey-Bass).
- Blashfield, R. K. (1976). Mixture model tests of cluster analysis: accuracy of four agglomerative hierarchical methods. Psychological Bulletin 83 (3): 377-388.
- Blashfield, R. K. and Aldenderfer, M. S. (1978). The literature on cluster analysis. Multivariate Behavioural Research 13 (July): 271-295.
- Cattell, R. B. (1965). Factor analysis: an introduction to essentials. 2. The role of factor analysis in research Biometrics 21 (2): 405-435.
- Chatfield, C and Collins, A. J. (1980). Introduction to multivariate analysis (London: Chapman and Hall).
- Cormack, R. M. (1971). A review of classification. Journal of the Royal Statistical Society Series A 134 (3): 321-367.
- Cunningham, K. M. and Ogilvie, J. C. (1972). Evaluation of hierarchical grouping techniques: a preliminary study. The Computer Journal 15 (3): 209-213.
- de Chernatony, L. (1978). The impact of the changed balance of power from manufacturer to retailer in the United Kingdom packaged groceries market. Fourth International Conference on Distribution, University Bocconi, Milan.
- de Sarbo, W. S. (1982). Clustering consistency analysis Journal of the Academy of Marketing Science 10 (3): 217-223.
- Dillon, W. R. and Goldstein, M. (1984). Multivariate analysis. (New York: John Wiley & Sons).
- Doyle, P. (1975). Brand positioning using multidimensional scaling. European Journal of Marketing 9 (1): 20-34.
- Doyle, P. and Saunders, J. (1985). Market segmentation and positioning in specialized industrial markets. Journal of Marketing 49 (2): 24-32.
- Edelbrock, C. (1979). Mixture model tests of hierarchical clustering algorithms: the problem of classifying everybody. Multivariate Behavioral Research 14 (July): 367-384.
- Ehrenberg, A. S. C. and Goodhardt, G. J. (1976). Factor analysis: limitations and alternatives. Marketing Science Institute Report No. 76-116.
- Everitt, B. S. (1979). Unresolved problems in cluster analysis. Biometrics 35 (March): 169-181.
- Everitt, B. S. (1986). Cluster analysis (Aldershot: Gower Publishing Company).
- Fisher, L. and van Ness, J. W. (1971). Admissible clustering procedures. Biometrika 58 (1): 91-104.

- Fleiss, J. L. and Zubin, J. (1969). On the methods and theory of clustering. Multivariate Behavioral Research 4 (April): 235-250.
- Funkhouser, G. R. (1983). A note on the reliability of certain clustering algorithms. Journal of Marketing Research 20 (1): 99-102.
- Gower, J. C. (1975). Goodness of fit criteria for classification models and other patterned structures. In: Proceedings of the eighth international conference on numerical taxonomy, ed. by G. F. Estabrook. San Francisco: W. H. Freeman and Company pp 38-62.
- Green, P. E. and Tull, D. S. (1978). Research for marketing decisions. (Englewood Cliffs: Prentice Hall International).
- Jardine, N. and Sibson, R. (1968). The construction of hierarchic and non-hierarchic classifications. The Computer Journal 11 (2): 177-184.
- Jardine, N. and Sibson, R. (1971). Mathematical taxonomy (London: John Wiley and Son).
- Johnson, S. C. (1967). Hierarchical clustering schemes. Psychometrika 32 (3): 241-254.
- Jones, M. C. and Sibson, R. (1987). What a projection pursuit? Journal of the Royal Statistical Society. Series A 150 (1): 1-36.
- Kuiper, F. K. and Fisher, L. (1975). A Monte Carlo comparison of six clustering procedures. Biometrics 31 (September): 777-783.
- Lance, G. N. and Williams, W. T. (1967). A general theory of classificatory sorting strategies. 1. Hierarchical systems. The Computer Journal 9 (4): 373-380.
- Lenington, R. K. and Flake, R. H. (1975). Statistical evaluation of a family of clustering methods. In: Proceedings of the eighth international conference on numerical taxonomy, ed by G. F. Estabrook San Francisco: W. H. Freeman and Company pp 1-37.
- McClain, J. O. and Rao, V. R. (1975). CLUSTISZ: A programme to test for the quality of clustering of a set of objects. Journal of Marketing Research 12 (4): 456-460.
- Mezzich, J. E. (1978). Evaluating clustering methods for psychiatric diagnosis. Biological Psychiatry 13 (2): 265-281.
- Milligan, G. W. and Isaac, P. D. (1980). The validation of four ultrametric clustering algorithms. Pattern Recognition 12: 41-50.
- Milligan, G. W. and Mahajan, V. (1980). A note on procedures for testing the quality of a clustering of a set of objects. Decision Science 11: 669-677.
- Minkoff, E. C. (1965). The effects on classification of slight alteration in numerical technique. Systematic Zoology 14: 196-213.
- Mojena, R. (1971). An approach to the problem of evaluation in multivariate classification. PhD thesis: University of Cincinnati.
- Mojena, R. (1977). Hierarchical grouping methods and stopping rules: an evaluation. The Computer Journal 20 (4): 359-363.

- Morrison, D. G. (1967). Measurement problems in cluster analysis. Management Science 13: 775-780.
- Punj, G. and Stewart, D. W. (1983). Cluster analysis in marketing research: review and suggestions for applications. Journal of Marketing Research 20 (2): 134-148.
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66 (336) : 846-850.
- Rao, V. R. and Sabavala, D. J. (1981). Inference of hierachical choice processes from panel data. Journal of Consumer Research 8 (1): 85-96.
- Rohlf, F. J. and Sokal, R. R. (1965). Coefficients of correlation and distance in numerical taxonomy. The University of Kansas Science Bulletin 45 (1): 3-27.
- Saunders, J. A. (1980). Cluster analysis for market segmentation. European Journal of Marketing 14 (7): 422-435.
- Schaninger, C. M. and Buss, W. C. (1986). Removing response style effects in attribute determinance ratings to identify market segments. Journal of Business Research 14 (3): 237-252.
- Schlinger, M. J. (1969). Cues on Q-technique. Journal of Advertising Research 9 (3): 53-60.
- Sherman, L. and Sheth, J. N. (1977). Cluster analysis and its applications in marketing research. In: Multivariate methods for market and survey research, ed. by J. N. Sheth. (Chicago: American Marketing Association).
- Sneath, P. (1977). A significance test for cluster in UPGMA phenograms obtained from squared Euclidean distances. Classification Society Bulletin 4 (1) : 2-14.
- Sokal, R. R. and Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. Taxon 11 (2): 33-40.
- Sokal, R. and Sneath, P. (1963). Principles of 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 (1): 51-62.
- Williams, W. T., Clifford, H. T. and Lance, G. N. (1971a). Group size dependence: a rationale for choice between numerical classification. The Computer Journal 14 (2): 157-162.
- Williams, W. T., Lance, G. N., Dale, M. B. and Clifford, H. T. (1971b). Controversy concerning the criteria for taxonomic strategies. The Computer Journal 14 (2): 162-165.
- Wind, Y. J. (1978). Issues and advances in segmentation research. Journal of Marketing Research 15 (3): 317-337.
- Wishart, D. (1978). CLUSTAN user manual. (Edinburgh: CLUSTAN Ltd).