This is to certify that the

dissertation entitled

An Assessment of the Impacts
of Alternative Factor Analyses
on the Stability of Cluster Membership

presented by

Sheng Jung Ou

has been accepted towards fulfillment
of the requirements for

PhD degree in the

Department of Park and Recreation Resources

Edward M Mahoney
Major professor

Date 1/3/91

AN ASSESSMENT OF THE IMPACTS OF ALTERNATIVE FACTOR ANALYSES ON THE STABILITY OF CLUSTER MEMBERSHIP

BY

SHENG JUNG OU

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirement
for the degree of

DOCTOR OF PHILOSOPHY

Department of Park and Recreation Resources

1990

ABSTRACT

AN ASSESSMENT OF THE IMPACTS OF ALTERNATIVE FACTOR ANALYSES ON THE STABILITY OF CLUSTER MEMBERSHIP

By

Sheng Jung Ou

Even though the use of factor scores as input data for cluster analysis is a relatively common procedure, there has been very little research on the effect of alternative factor analyses on the results of cluster analysis, especially cluster membership. The primary purpose of the study was to examine the impact of factor analyses on cluster membership when clustering is based on factor scores. Specifically, the study examined the effect of alternative factor solutions (number of factors) and factor rotation on cluster membership.

The study used the importance ratings of 20 different campground attributes/facilities collected in a study of the 1988 National Campers and Hikers Association Campvention. To achieve three study objectives, principal component analysis with and without varimax rotation, cluster analysis (Ward's method using the squared Euclidean as a distance measure), crosstabulation technique, and the entropy (information) measure were employed.

Three major conclusions were drawn from the analyses. First, when factor analysis is used in conjunction with cluster analysis, the factor

solution (number of factors) selected has an effect on the cluster membership. Second, whether or not the initial factors are rotated does not affect cluster membership. However, rotation will effect the interpretation of the clustering results (i.e., the cluster labels). Third, clustering on raw data rather than factor scores results in more stable cluster membership.

The study resulted in two primary recommendations regarding the use of factor analysis and cluster analysis. First, when factor analysis is performed as a preliminary step to cluster analysis, they should not be treated as distinct analyses. Decisions regarding the number of factors should be based on both the factor analysis criteria (eigenvalues greater than one, percentage of variance explained, scree test) and the impact on the cluster solution. Second, researchers may first perform cluster analysis based on raw data for classification (segmentation) purposes, and then use factor analysis as a means of describing clusters.

ACKNOWLEDGEMENTS

I would like to thank my major advisor, Dr. Edward M. Mahoney, for his support and instruction. He was essential in the success of my Ph. D. program at Michigan State University. Without his patience, thoughtful response, and intellectual stimulation, I would not be able to finish this dissertation.

I would also like to thank my committee members Dr. Donald F. Holecek (Department of Park and Recreation Resources), Dr. Rene C. Hinojosa (Department of Urban Planning), and Dr. Paul E. Nickel (Department of Resource Development) for their patience and constructive criticisms of this dissertation. Special thanks are extended to Douglas B. Jester, for his input and assistance, especially his advice regarding research and statistical methods.

I would also like to express my appreciation to Dr. James L. Bristor for his friendship and spiritual support, to Ms. Tsao Fang Yuan for her assistance, and to Mr. Jong Pyng Li for his help with the computer programming.

My deepest and warmest gratitude must go to my parents, fiancee, sisters, and brothers for their support throughout my graduate program. Without their contributions, I would not have achieved my personal goals. Finally, to everyone else who helped me and made my stay at Michigan State University truly enjoyable and unforgettable, thanks.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

Table	Page

Table | Page

Table | Page

LIST OF FIGURES

Figure | Page

Figure Page

CHAPTER I

INTRODUCTION

Cluster analysis is a statistical method commonly used to classify individuals or objects into groups (clusters) based on their similarity with respect to specific characteristics/variables so that the resulting clusters possess high internal (within-cluster) homogeneity and high external (between-cluster) heterogeneity. In addition to the grouping function, cluster analysis can also be used to perform data reduction and to test hypotheses (Anderberg, 1973; Everitt, 1974). Cluster analysis has been applied in many fields such as business, social science, psychology, biology, political science, remote sensing research, and leisure research.

Clustering methods have been recognized throughout this century, but most of the literature on cluster analysis and its application has been written during the past two decades. Cluster analysis was first discussed by social scientists during the 1930s (Driver & Kroeber, 1932; Tryon, 1939; Zubin, 1938). However, it was not until the late 1950s that cluster analysis attracted significant attention. The main stimuli for this increased interest were the publication of Principles of Numerical Taxonomy by Sokal and Sneath (1963), and the development of high-speed computers and cluster analysis software. At least 14 different computer software programs are now available for cluster analysis (Punj & Stewart, 1983), including SPSS (Statistical Package for the Social Sciences), SAS (Statistical Analysis System), BMDP, and CLUSTAN.

Cluster analysis has been utilized extensively to segment various product and service markets including different recreation and tourism markets (Boggis & Held, 1971; Calantone & Johar, 1984; Calantone, Schewe, & Allen, 1980; Crask, 1980; Davis, Allen, & Cosanza, 1988; Ditton, Goodale, & Jonsen, 1975; Funk & Hudon, 1988; Goodrich, 1980; Green, Frank, & Robinson, 1967; Green, Sommers, & Kernan, 1973; Harrigan, 1985; Huszagh, Fox, & Day, 1985; Lessig & Tollefson, 1971; Mazanec, 1984; Perreault, Darden, & Darden, 1977; Saunders, 1985; Sethi, 1971; Shoemaker, 1989; Stynes & Mahoney, 1980; Tatham & Dornoff, 1971; Woodside & Motes, 1981). Besides market segmentation, cluster analysis also has been used in the field of recreation and tourism to classify leisure activities (Devall & Harry, 1981; Ellis & Rademacher, 1987; Tinsley & Johnson, 1984) and to identify different types of experiences, preferences and attributes (Hautaluoma & Brown, 1979; Heywood, 1987; Knopp, Ballman, & Merriam, 1979; Manfredo, Driver, & Brown, 1983).

The increased use of cluster analysis has resulted in greater attention to various clustering/methodological decisions including (a) the clustering algorithm, (b) the similarity measure, and (c) the number of clusters. These decisions are all critical elements in the clustering process. Another primary concern in cluster analysis is the degree of correlation between the clustering variables. Correlation among clustering variables results in an implicit weighting (double counting) problem; correlated variables have more weight in determining the cluster solution. To address the implicit weighting problem, researchers have proposed/used factor analysis (principal component analysis) as a prelude to cluster analysis (Aldenderfer & Blashfield, 1984; Anderberg, 1973; Everitt, 1979; Gorsuch, 1983; Green et al., 1967; Rohlf, 1970; Skinner, 1979; Smith, 1989). Factor analysis is also used as a preparatory step to reduce potential clustering variables to a core set of dimensions in order to make the results more interpretable (Kikuchi, 1986).

Factor (principal component) analysis is a process for grouping variables. It is a multivariate statistical technique in which a large number of interrelated variables is summarized/reduced to a smaller number of factors (dimensions) without appreciable loss of information. By performing factor (principal component) analysis, the original data are reduced to some independent (noncorrelated) dimensions or factors. Factor scores (calculated by multiplying the original raw data measurements by the corresponding factor score coefficients) are often used as input variables in cluster analysis.

In addition to data reduction, there are two additional benefits to clustering based on the principal component analysis rather than raw data (e.g., ratings of attributes). First, the dimensions (factors) are independent, thereby avoiding the collinearity or multicollinearity problem associated with correlated data. Second, the resultant factors are given equal weight which avoids the implicit weighting problem. Although factor scores (derived from principal component analysis) are commonly used as input to clustering algorithms, researchers have raised questions or concerns about this practice. Anderberg (1973) questioned whether the factors reflect the relationship among variables that are actually observed in the clusters. Rohlf (1970) voiced the concern that principal component analysis tends to maintain the representation of widely separated clusters in a reduced space but minimizes the distances between clusters or groups that are not widely separated.

Factor analysis can affect/determine cluster solutions in three potential ways: (a) the number of factors that determine factor scores (Coovert & McNelis, 1988; Zwick & Velicer, 1986), (b) factor rotation (Dielman, Cattell, & Wagner, 1972; Gorsuch, 1983), and (c) factor weighting (DeSarbo, Carroll, & Clark, 1984; Sneath & Sokal, 1973). Relatively little attention has been directed at the potential effects of alternative factor solutions on clustering results. A review of 32 studies in which factor scores were used as the basis for clustering identified only one which analytically compared clustering results based on two different factor solutions (as the bases for clustering) (Day, Fox, & Huszagh, 1988). In another study, Bartko, Strauss, and Carpenter (1971) compared clustering results based on raw data and factor scores. Shutty and DeGood (1987) compared clustering results based on standardized scores and factor scores.

Problem Statement

Although the use of factor scores as input data for cluster analysis is a relatively common procedure, very little research has been done on the effect of factor analysis--number of factors and rotation--on the results of cluster analysis, especially cluster membership. Numerous researchers have raised various methodological questions regarding factor analysis as an independent procedure

(Armstrong & Soelberg, 1968; Bobko & Schemmer, 1984; Browne, 1968b; Hakstian & Muller, 1973; Heeler, Whipple, & Hustad, 1977; Horn, 1965a; Moojjaart, 1985; Tucker, 1971) and cluster analysis (Bayne, Beauchamp, Begovich, & Kane, 1980; Dreger, Fuller, & Lemoine, 1988; Funkhouser, 1983; Krzanowski & Lai, 1988; Lathrop, 1987; Marriott, 1971; McIntyre & Blashfield, 1980; Milligan & Cooper, 1985; Mojena, 1977; Rand, 1971; Skinner, 1978). However, as previously mentioned, only one study was found that examined the effect of alternative factor analyses on clustering when factor scores were the basis for clustering.

Factor analysis and cluster analysis are usually treated as distinct analyses even when used in conjunction with each other (Collins, Cliff, & Cudeck, 1983; Hooper, 1985; Shutty & DeGood, 1987). The factor analysis is performed first; then the factor solution--the number of factors extracted--is decided based on different factoring criteria (e.g., eigenvalues greater than one, percentage of variance explained, scree test, interpretability of factors), and not (also) on the potential effect on the clustering solution--number of clusters, cluster membership, homogeneity of clusters, and identification (description) of clusters (Calantone & Johar, 1884, Crask, 1981; Kikuchi, 1986; Meade, 1987). Although eigenvalues greater than one, percentage of variance explained, and scree test are useful in evaluating and selecting a factor solution, a great deal of subjectivity is still associated with arriving at a factor solution and interpreting the resultant factors.

An important decision in factor analysis is the method to be used in rotating the initial factors that are extracted from the correlation matrix. Rotating the factor matrix redistributes the variance from earlier factors to later ones to achieve a simpler, theoretically more meaningful, factor pattern (Hair, Anderson, & Tatham, 1987; Kim & Mueller, 1989). Rotating factors generally improves the interpretation by reducing some of the ambiguities that often accompany initial unrotated factor solutions. Although rotating the factor matrix may create more interpretable factors, Frank & Green (1968) pointed out that rotation of factor axes also lends a certain arbitrariness to the procedure. Most studies on rotation have focused on alternative methods, either orthogonal or oblique (Arbuckle & Friendly, 1977; Carroll, 1953; Hakstian, 1976; Saunders, 1961); no studies of the effect of rotation on cluster membership were found.

Although the combined use of factor analysis and cluster analysis has been commonly employed in segmentation and classification studies, it has also been used for other purposes, such as differentiating small geographic areas on the basis of well-established sociological constructs, understanding social differentiation in modern industrial society, revealing consumer search patterns, and measuring the concept of social identity.

The use of factor analysis in conjunction with cluster analysis is also wildly used in recreation and tourism, such as segmenting vacationer market based on lifestyle variables, segmenting the tourism market on benefit-seeking choices, exploring aspects of lifestyles with respect to vacation activities, establishing lifestyle profiles of elderly female travelers, and ascertaining the barriers to recreation.

The primary purpose of this study was to assess the effect of different approaches to factor analysis on cluster membership when clustering is based on factor scores. Specifically, the study examined the effect alternative factor solutions (number of factors) and factor rotation on cluster membership. Another purpose was to compare the stability of clusters based on factor scores with the stability of clusters based on raw data.

Study Objectives

To address the aforementioned purposes, three objectives were defined to guide and evaluate this study.

Objective 1. To assess the effect of different factor solutions (number of factors) on cluster membership.

Objective 2. To ascertain the effect of factor rotation on cluster membership.

Objective 3. To compare clustering on factor scores with clustering on raw data.

Organization of the Study

Chapter II is a review of relevant literature, focusing on previous studies, especially in the fields of marketing, recreation, and tourism, that have employed both factor analysis (principal component analysis) and cluster analysis. Chapter III contains a description of the data--ratings of 20 campground attributes--used in the study, including how they were collected, and a discussion of the statistical procedures used for the different objectives. Chapter IV includes descriptive statistics on the ratings of the twenty campground attributes, the appropriateness of data for factor analysis, an assessment of the effect of different factor solutions on the clustering results, an assessment of the effect of rotation on cluster membership, and comparison of clustering on factor scores with clustering on raw data. Chapter V includes a summary of the study, major conclusions, study limitations, and recommendations regarding the combined use of factor analysis and cluster analysis.

CHAPTER II

LITERATURE REVIEW

The primary objective of this chapter is to acquaint the reader with the literature on the combined use of factor analysis and cluster analysis and its application in the fields of marketing (especially market segmentation), recreation, and tourism.

Factor Analysis and Cluster Analysis

Factor Analysis

As mentioned previously, factor analysis is a multivariate statistical tool for exploring the similarity of relationships among variables. The primary purpose of factor analysis is to reconstruct original variables into an underlying multivariate space that specifies the positions of original variables rather than establishing which variables go together (Gorman, 1983; Gorsuch, 1983). Factor analysis starts out with a correlation matrix, which is a table showing the intercorrelations among all variables. The interrelationships between variables are typically determined by Pearson product-moment correlation.

The underlying factors are extracted using either a component model or a common factor model. There are a number of differences between the two models. The major difference is the elements comprising the diagonal of the correlation matrix. The component model uses total variance (unity) in the diagonal of the correlation matrix, whereas the common model uses communalities (common variance). The component model is used to summarize most of the original information (variance) in the minimum number of factors. The common factor model is used to identify underlying factors or dimensions not easily recognized (Hair et al., 1987; Kim & Mueller, 1988).

Although both factoring models are capable of extracting common factors, the initial result seldom represents the final solution because the initial factors are difficult to interpret and may not adequately represent the simple structure. Frequently, the initial factors are rotated. Two rotation procedures are commonly used, orthogonal and oblique. In orthogonal rotation the factors are mutually independent. Three major types of orthogonal rotation--varimax, equimax, and quartimax--are most commonly used in practice. Of the three, varimax rotation is used most frequently (Bieber & Smith, 1986; Norusis, 1988). In oblique rotation the factors are correlated (Bieber & Smith, 1986; Gorsuch, 1983; Hair et al., 1987; Kim & Mueller, 1988). When the result (e.g., factor score) of factor analysis is to be used in subsequent statistical analyses (e.g., cluster analysis), an orthogonal rotation is appropriate because collinearity is eliminated. In contrast, oblique rotation is appropriate if the objective is to obtain theoretically meaningful constructs or dimensions.

There is no agreement in the literature regarding the best rotation method. Bartholomew (1985) indicated that there is no significant difference between orthogonal and oblique rotation procedures in terms of factoring results. Stewart (1981) contends that the basic solutions provided by most rotational programs result in the same factors, thus, the rotation method should have relatively little impact on the interpretation of factor analysis results.

A primary step/decision in factor analysis concerns how many factors should be extracted. Several criteria are typically used to decide on the number of factors. The most common one is the Kaiser criterion (Kaiser, 1960), whereby all factors having eigenvalues greater than one are accepted. This criterion often is used in conjunction with percentage of variance explained and the scree test (Cattell, 1966). Other methods, including significance tests associated with the maximum likelihood and least squares solutions, Horn's (1965b) parallel analysis, Bartlett's (1950, 1951) chi-square test, Velicer's (1976a) minimum average partial method, and interpretability of the factors are also used to determine the number of factors.

Although each criterion has its supporters, Zwick and Velicer (1986) contend that which criterion is most appropriate depends on a number of different factors--sample size, number of variables, component saturation (scale of factor loading), component identification, and special variables (variables having a nonzero loading on more than one component) (Zwick & Velicer, 1986). Based on their research, they concluded that parallel analysis and the minimum average partial method are generally the best across situations. However, a review of factor analysis studies showed that the majority used combined criteria, such as eigenvalue greater than one, percentage of variance explained, and the scree test (Allen, 1982; Beard & Ragheb, 1983; Connelly, 1987; Hollender, 1977; Lounsbury & Hoopes, 1988; Tinsley & Kass, 1979; Wahlers & Etzel, 1985).

Once the number of factors is decided, the next step in factor analysis is to interpret the factor solution. The most common interpretation approach involves analyzing the size and pattern of factor loadings. Factor loadings are key in understanding the nature of factors. A factor loading indicates the relationship between a variable and a factor. The higher the factor loading, the stronger the relationship. Hair et al. (1987) suggested that factor loadings greater than ± 0.30 are significant, those greater than ± 0.40 are more important, and loadings greater than or equal to ± 0.50 are very significant. Their suggestions can be viewed as a rule of thumb. In addition, Gorsuch (1980) indicated there are more exacting but computationally more difficult ways of determining the significant loadings including: Archer and Jennrich's (1973) formulas, Jöreskog's (1978) confirmatory maximum likelihood factor analysis, and Lindell and St. Clair's (1980) jackknife approach.

A review of factor analysis studies showed that the factor loading rule of thumb is used most often. However, researchers contend that valid interpretation of a factor solution should depend on examination of high, medium, and low loadings. High loadings indicate variables which are highly related to a particular factor, whereas low loadings indicate variables which are not related to a particular factor (Bieber & Smith, 1986).

The final stage in factor analysis is to calculate factor scores, which are commonly used as input variables in other statistical analyses such as cluster analysis, discriminant analysis, and regression analysis. There are several different methods for estimating factor scores. According to Tucker (1971), the least squares solution characterized by Horst (1965) and Bartlett (1937) would yield appropriate factor score estimates for evaluating group differences on factors. Thurstone (1935) also suggested that if group membership is to be predicted from factor scores, the regression estimates method would be appropriate. Although Velicer (1976b) found that there is little practical difference among factor score estimates, image scores, and principal component scores, he suggested using principal component or rescaled image scores. However, unless the principal components model is used, factor scores can only be estimated (Kass & Tinsely, 1979; McDonald & Burr, 1967).

Cluster Analysis

The purpose of cluster analysis is to formulate relatively homogeneous groupings of individuals/objects based on one or more similarity criteria. Cluster analysis starts with a similarity measure of the proximity or closeness between all possible pairs of individuals/objects. There are four types of similarity measures:

correlation coefficients, distance measures (e.g., Euclidean distance measure), association coefficients, and probabilistic similarity coefficients (Aldenderfer & Blashfield, 1984). The last two are infrequently used. Although it has been demonstrated that using correlation coefficients as the similarity measure reduces the ratio of misclassification (Hamer & Cunningham, 1981), correlation coefficients are relatively insensitive to differences in the magnitude of the variables and fail to satisfy the triangle inequality (i.e., $d(x,y) \leq d(x,z) + d(y,z)$, given that x, y, and z are different entities). In contrast, distance measures provide the actual distance between cases and satisfy the triangle inequality.

The literature indicated that distance measures are the most commonly used measures of similarity (Aldenderfer & Blashfield, 1984; Bieber & Smith, 1986; Everitt, 1974; Hair et al., 1987). Three types of distance measures are commonly used: Euclidean distance, Manhattan distance, and Mahalanobis D^2. Euclidean distance (assuming that variables are independent) is most commonly used, even though some researchers argue that Mahalanobis D^2 is more versatile in that it can be used even if the clustering variables are correlated. Euclidean distance is often criticized as not having ability to preserve distance ranking (Everitt, 1974). However, this problem can be solved by standardizing the data (Aldenderfer & Blashfield, 1984).

What clustering algorithm to use is obviously an important clustering decision. Most researchers prefer to use hierarchical rather than nonhierarchical clustering algorithms because nonhierarchical clustering algorithms start with the selection of an appropriate starting partition/seed point which is relatively subjective (Blashfield, 1978).

The five popular hierarchical methods--single linkage (minimum distance), complete linkage (maximum distance), average linkage (average distance), Ward's method (minimum variance), and the centroid method (distance between means)--differ in terms of how the distance between clusters is calculated. However, results of a number of studies indicated that Ward's method consistently outperforms the other methods in terms of the accuracy of the cluster solution (Bayne et al., 1980; Blashfield, 1976; Edelbrock, 1979; Edelbrock & McLaughin, 1980; Mojena, 1977).

Ward's (1963) method is used to optimize the minimum variance within clusters. In Ward's procedure, the distance between two clusters is the sum of squares between the two clusters summed over all variables. At each step in the clustering process, the union of every possible pair of clusters is considered. The two clusters whose fusion results in the minimum increase in the error sum of squares become a new cluster (Aldenderfer & Blashfield, 1984; Everitt, 1974; Hair et al., 1987; Norusis, 1988).

Although many researchers recommend Ward's method, it has two problems/limitations. First, it is sensitive to outliers. Also, there is no function for reallocating entities that might have been poorly classified at early clustering stages (Everitt, 1974). Some researchers have suggested that the outlier problem can be eliminated by using both the hierarchical clustering method and the iterative partitioning method (Milligan, 1980; Punj & Stewart, 1983).

A critical step in cluster analysis is deciding on a clustering solution--the number of clusters to form. There are a number of procedures for determining the number of clusters (Aldenderfer & Blashfield, 1988; Dubes & Jain, 1979; Everitt, 1974; Milligan & Cooper, 1985). In many studies, the decision has been based on an examination of different levels of the fusion dendrogram or a similar scree test. A similar scree test involves plotting the fusion coefficients against the number of clusters, which is the numerical value at which various cases merge to form a cluster. Sudden jumps or breaks in the scree plot indicate that two relatively dissimilar clusters have been merged. The solutions (number of clusters) prior to these mergers are likely candidate solutions (Thorndike, 1953). Both the fusion dendrogram and the similar scree test approaches are subjective.

Other less subjective approaches for deciding on cluster solutions have also been discussed (Everitt, 1979; Milligan & Cooper, 1985). For example, Marriot (1971) suggested that a possible criterion for selecting the number of groups/clusters is to take that value of k for which $k^2|W|$ is a minimum, where k is the number of clusters and $|W|$ is the determinant of the pooled within-group variance-covariance matrix. Beale (1969) proposed using a F-ratio to test the hypothesis of the existence of K2 versus K1 cluster in the data (K2 > K1). Wolfe (1970) proposed a likelihood ratio criterion to test the hypothesis of k clusters against k-1 clusters.

Despite the numerous criteria that have been proposed, Everitt (1979) believes that no <u>one</u> completely satisfactory solution is available. The best way to decide on the number of clusters seems to be to utilize a combination of the decision criteria along with the interpretability of results (Bieber & Smith, 1986; Everitt, 1979; Gnanadesikan & Wilk, 1969). Other criteria, such as identifiability, substantiality, variation in responses, and exploitability, are also important in deciding a final cluster solution, especially if the purpose is market segmentation (Kikuchi, 1986; Kotler, 1984; Stynes, 1983).

Comparisons of Factor Analysis and Cluster Analysis

There still is some confusion regarding the differences between factor analysis and cluster analysis. This frequently results in inappropriate applications of both methods.

The major distinction between factor analysis and cluster analysis is that the former detects relationships between variables and thereby reconstructs original variables into fewer dimensions, whereas the latter is concerned with the classification of individuals/objects. Neither method alone may be sufficient if researchers are trying to reduce a large set of data and to classify individuals into groups (on the basis of the reduced data). In this situation, the use of factor analysis in conjunction with cluster analysis is often suggested (Anderberg, 1973; Everitt, 1979; Gorsuch, 1983; Green et al., 1967; Mark, 1980; Punj & Stewart, 1983; Rohlf, 1970; Skinner, 1979; Smith, 1989).

Literature Supporting the Combined Use of Factor Analysis and Cluster Analysis

A number of researchers have determined that factor analysis is helpful in identifying meaningful dimensions/factors on which to cluster individuals/objects. Mark (1980) suggested using principal component analysis as a preparatory step to cluster analysis to identify neighborhoods for preservation and renewal. Swinyard and Struman (1986) found that clustering consumers after a factor analysis, thereby reducing various measures to a fewer factors, resulted in (restaurant/dining) clusters/segments that were easier to describe and act on. Smith (1989) preferred the combined factor-cluster analysis approach over the "a priori" method because it results in more homogeneous clusters. Gorsuch (1983) indicated that factoring before cluster analysis helps clarify the basis on which individuals are grouped, and provides empirical methods of producing typologies. Wind (1978) suggested performing a principal component analysis as a way to obtain a more reliable and meaningful factor structure before clustering.

Combined factor and cluster analysis can be used to solve the problem of independency of variables and to deal with implicit weighting problem in clustering procedures (Green et al., 1967; Punj & Stewart, 1983). In addition, the combined approach can be used to identify a "best" set of dimensions for depicting the relationships among individuals (Skinner, 1979).

Punj and Stewart (1983) contend that when a researcher desires that all dimensions or attributes be given equal weight in the clustering process, it is necessary to correct for interdependencies. They suggested two approaches to correct for interdependencies: (a) using Mahalanobis D^2 or (b) completing a preliminary principal component analysis with orthogonal rotation. Component (factor) scores can then be used as input variables for computing similarity measure in the clustering process.

Studies on the Combined Use of Factor Analysis and Cluster Analysis

As previously stated, combined factor-clustering analysis has been utilized by researchers in many fields, such as marketing, recreation, tourism, psychology, medical science, and sociology. This section contains a review of a number of studies that used factor scores as a basis for clustering, with special attention to the factoring method, criteria for selecting the number of factors, the clustering method, and the criteria for selecting the number of clusters. Table 1 summarizes 22 of the 32 studies which were reviewed.

Day and Heeler (1971) used a randomized block experiment with five strata composed of three stores to test the sales effect of three price-level changes in a new food product. Principal component analysis was first performed on 12 store attributes (e.g., selling area of store, average household income). Five mutually independent factors were identified, which accounted for 77% of the total variance. Factor scores were then calculated to obtain two different similarity measures: modified matching coefficient and Euclidean distance. Both similarity measures were used as the basis for hierarchical and nonhierarchical

Table 1. A summary of studies in which combined factor analysis and cluster analysis was employed.

Author(s)	Nature of Data	Factoring Method And Rotation	Criteria for Selecting The Factor Solution	Clustering Method	Criteria For Selecting The Number Of Clusters	Discussion of The Interaction Of Factor Analysis And Cluster Analysis	Results
Bishara, 1984	63 Dividend Decisions Variables	Principal Component, Varimax Rotation	Percentage of Variance Explained	Ward's Method	Not Specified	No	Two Distinct Company Clusters Were Formed For Each Year (1965, 1970, 1975, and 1979)
Calantone & Johar, 1984	20 Travel Destination Attributes	Not Specified	Eigenvalue $>$ 1, Percentage of Variance Explained	K-means	F-ratio	No	5-6 Distinct Seasonal Benefits Sought Segments
Crask, 1981	15 Vacation Attributes	Principal Component, Varimax Rotation	Eigenvalue $>$ 1, Percentage Variance Explained	Ward's Method	Error Sum Of Square	No	5 Distinct Vacationer Segments
Day & Heeler, 1971	12 Store Attributes	Principal Component, Rotation Not Specified	Percentage of Variance Explained	Not Specified	Not Specified	No	5 Distinct Store Segments

Table 1 (Cont'd.).

Author(s)	Nature of Data	Factoring Method And Rotation	Criteria For Selecting The Factor Solution	Clustering Method	Criteria For Selecting The Number Of Clusters	Discussion Of The Interaction Of Factor Analysis And Cluster Analysis	Results
Day et al., 1988	18 Economic Indicators	Not Specified	Not Specified	K-means	Not Specified	Yes	6 Distinct Country Segments For Different Factor Solutions
Furse et al., 1984	24 new Car Searching Activities	Principal Component, Rotation Not Specified	Eigenvalue > 1, Scree Test	Ward's Method, K-means	Kappa Coefficient	No	6 Distinct Car Searching Segments
Gartner, 1990	90 Entrepreneurship Related Variables	Not Specified	Percentage of Variance Explained	Hierarchical Clustering, K-means	Not Specified	No	Two Entrepreneurship Segments Were Derived From Both Hierarchical Clustering And K-means
Gau, 1978	64 Residential Mortgage Variables	Principal Component, Varimax Rotation	Percentage of Variance Explained	Iterative Partitioning Method	Friedman & Rubin Criterion	No	6 Distinct Residential Mortgage Segments

Table 1 (Cont'd.).

Author(s)	Nature of Data	Factoring Method And Rotation	Criteria For Selecting The Factor Solution	Clustering Method	Criteria For Selecting The Number Of Clusters	Discussion Of The Interaction of Factor Analysis And Cluster Analysis	Results
Hawes, 1988	33 AIO Statements	Principal Component, Varimax Rotation	Eigenvalue > 1, Percentage of Variance Explained, Scree Test	A Priori	In A Priori Cluster Solutions Were Predetermined	No	5 Predefined Age Segments
Henderson & Stalnaker, 1988	55 Recreation Barrier Related Variables	Principal Component, Varimax Rotation	Eigenvalue > 1, Percentage of Variance Explained	A Priori	In A Priori Cluster Solutions Were Predetermined	No	4 Distinct Personality Segments
Hooper, 1985	59 Social Related Variables	Principal Component, Oblique Rotation	Eigenvalue > 1, Scree Test, Interpretation	Not Specified	Ratio Of Between Cluster Variance To Within Cluster Variance	No	13 Distinct Social Identity Segments
Humphrey et al., 1987	60 Sociological Constructs	Principal Component, Varimax Rotation	Minimum Average Partial, Scree Test	Ward's Method	Cubic Clustering Criteria	No	15 Distinct Socioeconomic Status Segments
Jones, 1968	70 Social Composition Variables	Principal Component, Rotation Not Specified	Eigenvalue > 1, Percentage Variance Explained	Centroid Clustering Method	Not Specified	No	20 Distinct Social Segments

Table 1 (Cont'd.).

Author(s)	Nature of Data	Factoring Method And Rotation	Criteria For Selecting The Factor Solution	Clustering Method	Criteria For Selecting The Number Of Clusters	Discussion Of The Interaction of Factor Analysis And Cluster Analysis	Results
Kiel & Layton, 1981	12 Information Searching Variables	Principal Component, Oblique Rotation	Eigenvalue > 1, Percentage of Variance Explained	K-means	Not Specified	No	3 Distinct Car Search Behavior Segments
Kikuchi, 1986	22 Recreational Fishing Attributes & 45 Species-Location Variables	Principal Component, Varimax Rotation	Eigenvalue > 1, Scree Test, Percentage of Variance Explained	Ward's Method, Iterative Partitioning Method	Error Sum Of Square, Managerial Interpretation	No	8 Attributes Sought and 8 Species-location Segments
Kim & Lim, 1988	13 Environmental Variables & 15 Strategic Variables	Principal Component, Rotation Not Specified	Eigenvalue > 1, Percentage of Variance Explained	Ward's Method	Mean-Square Error	No	4 Environmental Segments & 4 Strategic Segments
Krzystofiak et al., 1979	594 Job Related Variables	Common Factor Analysis, Varimax Rotation	Eigenvalue > 1, Percentage of Variance Explained	Ward's Method	Not Specified	No	Cluster Number or Description Was Not Reported
Meade, 1987	11 Physical Variables of Car & Price Variable	Not Specified	Not Specified	Not Specified	Not Specified	No	10 Distinct Car-Purchaser Segments

Table 1 (Cont'd.).

Author(s)	Nature of Data	Factoring Method And Rotation	Criteria For Selecting The Factor Solution	Clustering Method	Criteria For Selecting The Number Of Clusters	Discussion Of The Interaction of Factor Analysis And Cluster Analysis	Results
Perreault et al., 1977	70 Vacation's AIO Statements	Not Specified	Not Specified	Ward's Method	Not Specified	No	5 Distinct Vacation Segments
Rescorla, 1988	73 Clinic Symptoms	Principal Component, Varimax Rotation	Eigenvalue > 1, Factor Loadings	K-means	Not Specified	No	2 to 6 Autistic Cluster Solutions Were Examined
Sorce et al., 1989	8 Lifestyle Dimensions	Principal Component, Varimax Rotation	Eigenvalue > 1	Complete Linkage Clustering Method	Not Specified	No	8 Distinct Lifestyle Segments
Stanley et al., 1987	22 Upscale Financial Service Offerings	Not Specified	Not Specified	Ward's Method	Error Sum Of Squares	No	4 Distinct Financial Services Segments

clustering processes to test the homogeneity and representativeness of strata. Although the factor-cluster approach was used in this study, only one criterion, percentage of variance explained, was used to decide on the number of factors. The authors did not indicate any concern regarding the impact of the factor analysis on the clustering results.

Wolfe (1978) analyzed data on profiles of 113 occupation groups, using three different clustering procedures: (a) hierarchical grouping of standard scores, (b) hierarchical grouping of orthogonal factor scores, and (c) NORMIX analysis assuming equal covariance matrices for each group. Ward's method and Euclidean distance were used in all three cluster analyses. The hierarchical grouping of standard scores resulted in 13 groups, which were used as the basis of comparison with the results of the other two methods. The results showed that the NORMIX, in which the distance measures were calculated based on component (factor) scores, produced a solution having the most intuitive psychological sense. The results also showed that the hierarchical grouping of orthogonal factor scores provided clustering results nearly as good as NORMIX, whereas the hierarchical grouping of standard scores was the worst of the three approaches in terms of cluster homogeneity. The author did not discuss the impact of alternative factor solutions on the clustering results.

Green et al. (1967) proposed a factor-cluster approach that not only included a data-condensation function but also changed the implicit weighting of characteristics. Principal component analysis was performed on the data matrix first; then objects were clustered, based on principal component scores. They employed this technique to classify 88 cities for the purpose of selecting test markets. Two factors were derived from 14 variables (e.g., population, retail sales, and television coverage), and three clusters were formed. The authors did not provide information on the criteria to decide on the number of clusters, nor did they discuss the potential effect of the factor analyses on the clustering results.

Skinner (1979) presented a hybrid approach to integrate the dimensional and discrete clusters approaches to classification research. Two major steps are involved in this approach. First, a parsimonious set of dimensions is identified by performing a preliminary principal component analysis with orthogonal rotation, and evaluated by replication across samples. Second, relatively homogeneous subgroups are identified (using a clustering or density search algorithm), based on factor scores derived from the first step. This hybrid approach helped Skinner successfully cluster male delinquent adolescents, who had completed the Basic Personality Inventory (i.e., an 11-scale structured inventory of psychology), into three modal profiles (groups). These three groups are similar to what most clinical psychologists would describe. The criteria used to decide on the number of clusters and the potential impact of alternative factor solutions on the clustering results were not discussed.

To develop taxonomies of search behavior by new car buyers, Kiel and Layton (1981) used factor analysis to reduce 12 different search variables (e.g., search time, trips made) to four initial factors. The factors were then rotated by oblique rotation, and the four factors were retained. Factor scores were calculated and used to derive an aggregate search index. A K-means clustering algorithm was used to group buyers, based on the index number. The authors provided no information on the criteria they used to decide on the number of clusters, nor did they discuss the rotational effect of the factor solution on the clustering results.

Stanley, Powell, and Danko (1987) factor analyzed ratings of the desirability of 22 "upscale" financial service offerings (e.g., investment management and advice, immediate access to credit), and developed seven "upscale" financial service factors. Scores for those seven factors were used to categorize financial service customers (using Ward's clustering method) into four clusters/segments. The authors did not report on the factoring method or the criteria for selecting a factor solution. Nor did they discuss the potential impacts of the factor analyses on cluster/segment membership.

To differentiate small geographic areas in Rhode Island on the basis of well-established sociological constructs, Humphrey, Buechner, and Velicer (1987) proposed using combined factor-cluster analysis. Principal component analysis with varimax rotation was performed to reduce 60 original variables (e.g., families with income below poverty level in 1979, females in labor force) to four factors. To demonstrate the clustering procedure, the authors used only two factors (wealth and education factor). Ward's method (using square Euclidean distance) was performed on factor scores. Fifteen socioeconomic status clusters emerged. The potential impacts of alternative factor solutions on the clustering results were not discussed.

To understand social differentiation in modern industrial society, Jones (1968) used combined factor-cluster analysis. Principal component analyses were performed on three domains: socioeconomic status (24 variables), household composition (24 variables), and ethnic composition

(22 variables). Three factors emerged for each domain. Factor scores for each domain were computed to test the independence of the three dimensions. Another principal component analysis was performed, based on 24 variables (eight variables were selected from each dimension). Two constructs/factors were identified (socioeconomic status/ethnicity and household composition). Factor scores for these two factors were used as the basis for clustering. Twenty groups were identified using the centroid clustering method (with the squared Euclidean distance measure). Again, the author did not discuss criteria for selecting the number of clusters or the possible effect of the factor analysis/solutions on the clustering results.

To study the strategic positioning of product (car) range by manufacturers, Meade (1987) employed factor analysis to condense the information contained in 10 observable (e.g., engine capacity, maximum speed) variables to fewer factors. Three factor analyses were performed, which resulted in three-factor, two-factor, and single-factor solutions. The three-factor solution was used only to evaluate pricing policy; no cluster analysis was performed. The two-factor solution was used as the basis for clustering; 10 car segments emerged. The one-factor solution was used to provide the measure for cluster analysis; three groups/segments (small, medium, and large) were formulated. Meade indicated that the combined use of factor analysis and cluster analysis allowed the researcher to superimpose some structure on the ranges of products offered in the market. However, the criteria for deciding on the number of factors or clusters, the factoring method, the clustering method, and the possible effect of factor analysis on the clustering results were not discussed.

Day et al. (1988) used combined factor and cluster analysis to segment the global market for industrial goods based on economic indicators. Two different factor analyses were performed. The first factor analysis was conducted on 18 economic indicators; three factors emerged. In the second factor analysis, two of the original 18 economic indicators were dropped because they did not have any strong affiliation with any of the three factors. Three factors emerged from the second factor analysis on the 16 remaining variables. Factor scores were computed for the factors from both factor analyses. K-means clustering algorithm was used to group countries. Cluster analyses on the factor scores from both the first and second factor analyses resulted in two six-cluster solutions. Comparison of the two solutions indicated that countries were grouped similarly in both analyses. The authors failed to provide information on the criteria they used to decide the number of factors and clusters. However, they examined the clustering results between two different factor solutions (as the bases for clustering).

Sorce, Tyler, and Loomis (1989) employed factor analysis and cluster analysis to segment older Americans based on lifestyle variables. Eight lifestyle dimensions, each containing four to six statements, were submitted to a principal component analysis with varimax rotation. Five factors emerged, which accounted for 31% of the variance. A complete linkage clustering method (using the squared Euclidean distance measure) was used to group the older Americans based on factor scores; eight clusters/segments emerged. The authors did not provide information on the criteria they used to decide the cluster solution, nor did they discuss the potential effects of factor analysis on the clustering results.

In Gartner's study (1990), combined factor and cluster analysis was employed to explore the underlying meanings of entrepreneurship. Ninety different attributes were identified from various definitions of entrepreneurship. Factor analysis was employed to reduce the 90 variables to eight dimensions (factors). Two different clustering methods--hierarchical clustering and the K-means clustering--were then used to discover whether participants (academic researchers in entrepreneurship, business leaders, and politicians) in a Delphi study could be grouped together based on their rating (not factor scores) of the eight entrepreneurship factors. Two groups/clusters emerged from both cluster analyses. The membership of clusters derived from the two clustering methods were compared. The criteria used to decide the number of clusters and the potential impact of alternative factor solutions on the clustering results were not discussed.

Bishara (1984) used combined factor and cluster analysis to investigate whether the size of companies, their organizational structure, or the availability and stability of funds, most influenced the dividend decisions of life insurance companies. Factor analysis with varimax rotation was performed on 63 original variables (e.g., policy loans, income before taxes, ratio of policy loans to total assets); seven factors emerged based on the criterion of percentage of variance explained. Factor scores were computed and submitted to a (Ward's method) cluster analysis for each of the four years selected (1965, 1970, 1975, and 1979). Two clusters were identified for four selected years, with slight changes in cluster membership. Bishara did not discuss the criteria for choosing the cluster solution or the possible impacts of factor solutions on the clustering results.

Gau (1978) undertook factor analysis and cluster analysis to assess the relative levels of default risk inherent in residential mortgages. Sixty-four variables describing the financial, property, and borrower characteristics of residential mortgages were reduced to 28 independent factors using principal component analysis and varimax rotation. Factor scores were then utilized as input in a two-group discriminant analysis. A stepwise-determined subset of 17 factors was employed in the formation of discriminant functions that would differentiate between mortgage defaulters and nondefaulters. After weighting the factor scores on the basis of their respective discriminant coefficients, a nonhierarchical clustering algorithm (iterative partitioning method) was employed to identify a six-cluster solution. Gau did not discuss the potential impact of alternative factor solutions on the clustering results.

Krzystofiak, Newman, and Anderson (1979) used factor-cluster analysis to develop a quantified job analysis system for a power utility firm. Common factor analysis with varimax rotation was performed on 594 job-related items, and 60 factors emerged. Factor scores then were used as the basis for job profiling. Jobs were identified at approximately the same organizational level, and six organizational levels were identified. Within each of the organizational levels, jobs were grouped into job clusters based on Ward's clustering (using Mahalanobis distance). The authors did not provide information on the criteria they used to decide on either the factor analysis or clustering solution, nor did they discuss the potential impact of the factor analyses on the clustering results.

Kim and Lim (1988) concluded that factor analysis and cluster analysis are useful ways to examine the relationship between task environment and strategy. Factor analysis with orthogonal rotation was performed separately on two domains--environmental (e.g., scope of distribution channel, price change of materials/parts) and strategic (e.g., new product development, operating efficiency). Based on the criteria of eigenvalues greater than one and percentage of variance explained, 13 environmental variables were reduced to five factors, and the original 15 strategic variables were reduced to four factors. Ward's method (using the Euclidean distance measure) was performed on factor scores for both the environmental and strategic domains, and four clusters were formulated for both domains. Kim and Lim did not discuss the potential impact of alternative factor solutions on the clustering results.

Using factor analysis and cluster analysis, Furse, Punj, and Stewart (1984) replicated and extended previous research on consumer search patterns. In the first case study (new car buyer study), a principal component analysis was carried out on 24 items related to various search activities (e.g., time spent talking to salespersons, number of different dealers visited). Five factors were extracted and then rotated using both varimax and oblique rotation methods. The rotated factors, both varimax and oblique, were similar to the original factors. The five oblique rotation factors were retained because oblique rotation reduced moderate factor loadings. Factor scores were computed and used as the basis for clustering. Ward's hierarchical clustering method with Euclidean distances then was performed to obtain five to seven candidate cluster solutions, which served as seed points in a K-means clustering procedure; six clusters were formulated. In the second case study (new car dealer salesperson study), same factoring and clustering procedures were performed, and three factors and six clusters were identified. The authors did not discuss the potential impact of alternative factor solutions on the clustering results.

Hooper (1985) utilized combined factor-cluster analysis to measure the concept of social identity more comprehensively and precisely than previous researchers had done. Principal component analysis (with oblique rotation) was performed on 59 sociological variables (e.g., marital status, physical attraction, race). Fifteen factors were extracted. Factor scores were computed and then weighted by multiplying a weighted average of the stimuli defining each social identity according to the importance in the composition of the social-identity factor. The weighted scores then were submitted to cluster analysis. Based on the ratio of between-cluster variance to within-cluster variance and interpretability, 13 clusters were identified. Although Hooper used the weighted scores as the input to cluster analysis, neither weighting scheme, clustering algorithm, nor the relationship between factor and cluster solutions was discussed.

Rescorla (1988) employed combined factor-cluster analysis to explore the major issues of classification regarding autistic children. A principal component analysis with varimax rotation was performed on 73 items derived from Achenbach's Child Behavior Checklist (e.g., child's clinic symptoms--strange behavior, disobedient at home, trouble sleeping). Based on three criteria--eigenvalues greater than one, number of variables with loading above .30, and interpretation, eight factors emerged. Unweighted factor scores were computed by summing each child's scores on the symptom items with loading of .30 or above. Each child's unweighted sums were then converted to T scores. The T scores then were submitted to K-means clustering analysis (using the Euclidean distance measure). Cluster runs were made for 2, 3, 4, 5, and 6 clusters. The relation between cluster assignment and diagnostic grouping was examined. However, the author did not discuss the potential impact of alternative factor solutions on the clustering results.

Calantone and Johar (1984) attempted to segment the tourism market on benefit-seeking choices in different seasons. Factor analysis was first performed for each season on 20 variables (e.g., familiarity with the state, scenery, historical attractions). Based on eigenvalues greater than one and percentage of variance explained, five significant benefits-sought factors emerged for the spring season. Six significant factors were identified for the summer, fall, and winter seasons. Factor scores for the seasonal benefits factors were then used as input for clustering. Ward's method was used in the clustering for each season. Based on the ratio of within-group variance to total variance and interpretation, a five-cluster solution was selected for each season. Calantone and Johar did not discuss the potential impact of alternative factor solutions on the clustering results.

Crask (1981) used both factor analysis and cluster analysis to segment the vacationer market based on lifestyle variables. A principal component analysis with a varimax rotation was performed on 15 vacation attribute statements (e.g., scenic beauty of the area, distance from home, opportunity for fishing and hunting). Based on eigenvalues greater than one and percentage of variance explained, five factors emerged, which accounted for 56.9% of the total variance. Factor scores were computed and submitted to a hierarchical clustering algorithm. Based on within-group variance criteria, five vacationer segments, which had distinct vacation interests and socioeconomic profiles, were identified. Crask did not specify the clustering method, nor did he discuss the possible effect of the factor solution on the clustering results.

Perreault et al. (1977) used factor-cluster analysis to explore aspects of lifestyles with respect to vacation activities. Factor analyses was carried out on 285 vacation-specific statements, and 28 vacation-specific dimensions (factors) emerged. Factor scores were computed and used as input data to Ward's method (using the Euclidean distance measure). Five different vacation segments were identified. The authors did not provide information on the criteria they used to decide on either the number of factors or clusters, nor did they discuss the potential impact of factor solutions on their clustering results.

Kikuchi (1986) used factor-cluster analysis to evaluate two different approaches for segmenting Michigan's sport fishing market: attributes sought and preferred species and locations to fish. For each segmentation approach, factor analysis with varimax rotation was performed before clustering. Based on four criteria--eigenvalues greater than one, scree test, variance explained, and interpretability of factors--five attributes sought and nine species-location factors were identified. Factor scores were computed and used as input to the two-stage clustering process. In the first stage, Ward's method (using the Euclidean distance measure) was performed to obtain preliminary cluster solutions based on the criterion of error sum of squares. In the second stage, these candidate cluster solutions were submitted to a reallocation clustering algorithm to determine the final cluster solution. Eight attributes-sought and eight species-location segments were identified. Kikuchi did not address the potential effects of alternative factor solutions on the clustering results.

Hawes (1988) attempted to establish lifestyle profiles of elderly (50+ years old) female travelers by using both factor analysis and "a priori" cluster analysis. The respondents were categorized into five-year "a priori" age clusters/segments (five clusters). Factor analysis with varimax rotation was performed on 38 variables/characteristics (33 AIO statements and 5 demographic variables) for each of the five age segments. Hawes did not discuss the potential impact of alternative factor solutions on the clustering results.

Henderson and Stalnaker (1988) also used factor analysis and "a priori" cluster analysis to ascertain the barriers to recreation confronting women and to determine the relationship between perceived barriers and gender-role traits. Factor analysis with varimax rotation was performed on 55 barrier-related variables (e.g., work schedule, lack of equipment). Based on eigenvalues greater than one and percentage of variance explained, ten factors emerged. The authors did not discuss the potential effect of factor solutions on the clustering results.

Potential Impact of Factor Solutions on Clustering Results

Very few studies have analytically examined (or mentioned) (1) the differences between clustering solutions based on raw data and factor scores, or (2) the impact of alternative factoring methods or solutions on clustering results. The most critical impact of factor analysis on the clustering results is the change in cluster membership that results from the different input variables (factor scores rather than raw data) to the clustering procedures.

Bartko et al. (1971) compared raw data and factor scores as the basis for clustering and obtained different clustering solutions. Shutty and DeGood (1987) compared clustering on standardized scores and clustering on factor scores and concluded that the results derived from clustering on factor scores might provide a more accurate description of clusters/segments. Schaninger (1986) compared clustering on raw data and clustering on standardized data, and concluded that the standardized data-cluster solution is better than the raw data-cluster solution because the standardized data solution resulted in clearer and more meaningful clusters.

Summary

A review of 32 studies shows that most researchers express little concern about the impact of alternative factor solutions on cluster membership. Some researchers even failed to specify the factoring method, the criteria for selecting a factor solution, the clustering method, or the criteria for deciding a cluster solution.

CHAPTER III

RESEARCH METHODS

This chapter details the methods employed to achieve the study objectives. It begins with a description of the data on which the different factor and cluster analyses were performed. This is followed by a discussion of the different statistical methods employed to achieve the three objectives.

Source and Description of Data

The 1988 Michigan Campvention Study

Several different data sets were evaluated to determine whether they were appropriate with respect to the study objectives. The data obtained from a study of the 1988 National Campers and Hikers Association (NCHA) Campvention were used in this study. The NCHA is one of the largest and most active camping organizations in the country, with more than 25,000 members. Each year the NCHA holds a Campvention. The 1988 Campvention was held from July 8 to July 14 at Highland State Recreation Area, located in southeast Michigan. Approximately 4,000 parties from all over the country attended the Campvention.

The Michigan Association of Private Campground Owners (MAPCO) and State Parks requested that Michigan State University assist them in conducting a marketing and economic study of the Campvention. There were three major purposes for the study: (a) developing a profile of Campvention attendees which could be used to develop and target camping related marketing efforts (see Mahoney, Oh, & Ou, 1989); (b) assessing the economic impact of the Campvention in Michigan; and (c) evaluating a $1.00 off per night of camping sales promotion designed to increase the amount of before and after Campvention camping in Michigan (see Oh, 1990).

Data Collection Methods and Response Rate

Two data-collection methods were employed in the Michigan Campvention study (for a more detailed discussion of the data collection methods, refer to Mahoney et al. (1989) and Oh (1990)). A self administered questionnaire and postage paid return envelope (pretrip) was mailed eight weeks before the 1988 Michigan Campvention to a systematic random sample of 1,575 (33%) of the 4,729 members who were preregistered for the Campvention. One week after the Campvention, the 1,575 persons who had received a pretrip questionnaire were sent a four-page posttrip questionnaire and a postage-paid return envelope. Even if no one in a sampled household had completed the pretrip questionnaire, they were urged to complete the posttrip questionnaire.

The four page pretrip questionnaire was used to collect a variety of information, including: (a) campvention trip plans (i.e., trip length); (b) likelihood that they would take advantage of the $1.00 off sales promotion offer; (c) pretrip perceptions of Michigan campgrounds; (d) their annual volume of camping activity and participation in off-season (before Memorial Day and after Labor Day) camping; (e) the importance they assigned to different attributes when selecting campgrounds; and (f) socioeconomic characteristics--state of residence, gender, work status, marital status, and whether they had children living at home.

Information collected on the posttrip questionnaire included: (a) respondents' evaluation of the Campvention; (b) the number of nights they camped in Michigan before, during, and after the Campvention; (c) posttrip perceptions of Michigan campgrounds; (d) likelihood that they would camp again in Michigan; (e) whether they planned to take advantage of the sales promotion offer; (f) spending on their Campvention trip; (g) membership in camping clubs/organizations and subscription to camping magazines; and (h) additional socioeconomic characteristics, such as family income and education (for detailed information on the development, form, and content of the questionnaires see Oh (1990)).

About fifty percent (794) of the 1,575 pretrip questionnaires were returned; 778 of them were usable. The response rate was somewhat higher for the posttrip questionnaire. A total of 860 (54.6%) posttrip questionnaires were returned; 847 were complete enough to be used in the analysis. A relatively high percentage of the sample (38%) completed and returned both a pretrip and a posttrip questionnaire. Thirty-two percent did not complete either of the questionnaires.

A random sample of 100 (19.6%) of the 510 persons/parties who failed to return either a pretrip or a posttrip questionnaire were mailed an abbreviated questionnaire in an effort to assess possible nonresponse bias. Fifty percent of the nonrespondents returned the "nonresponse bias" questionnaire. The results showed that there was little difference between respondents and nonrespondents in their ratings of the Campvention, the Campvention party size, number of nights on the Campvention trip, likelihood of camping again in Michigan, work status, martial or family status, and presence of children living at home. However, as would be expected, nonrespondents were less likely to have attended the Campvention and less likely to have been aware of or taken advantage of the sales promotion offer.

Profile of Persons Who Completed Questionnaires

The findings from the Michigan Campvention study are detailed in Mahoney et al. (1989) and Oh (1990). The majority of persons who attended the Campvention were retired. Almost all of them (94.6%) were married. Approximately 29% had children living with them at home. Over three quarters (77.2%) percent had family incomes of $20,000 or more. Twenty-seven percent (27%) had incomes of $40,000 or more. This is relatively high given that the majority were retired persons. Almost 80% of the parties were from other states and Canada. About a quarter (22.6%) of the nonresidents traveled from the bordering states of Ohio (12.4%), Indiana (6.4%) and Illinois (3.8%). Thirteen (13.2) percent were from Canada.

They were very active high, volume campers. About 98% camp every year, and they averaged 51 nights of camping annually. About 29% camped 60 or more nights a year. A high proportion of their camping nights (53%, 27 nights) were outside their home state where they resided. On average, they camped in five states in addition to the one where they lived. Most said that selecting where to camp was a family decision. Approximately three quarters (74.8%) subscribed to some camping related magazine/publication/club other than the NCHA. The majority of these were members of Good Sam. Sixty-nine percent (69%) attended camping or outdoor shows.

They were also very active off-season campers. A high percentage camped before Memorial Day (85.8%) or after Labor Day (93.3%). About 83% camped both before Memorial Day and after Labor Day.

More than half (55.8%) had no preference for either public or private campgrounds. About a quarter (25.3%) preferred to stay in private/commercial campgrounds while 18.8% preferred public campgrounds.

Data Used in the Present Study

The factor and cluster analyses were performed on the importance ratings of different campground attributes/facilities (see pretrip questionnaire, Appendix A). Respondents were asked to rank the importance (on a five-point scale, "1" being crucial, and "5" being not important) of 20 campground attributes/facilities: large sites, shaded sites, cleanliness, quietness, site privacy, security, hospitality of campground staff, low price, flush toilets, electricity, showers, laundromat, campground store, water hookups, sewer hookups, natural surroundings, situated on a lake/stream, hiking trails, pool, and playgrounds.

Even though the ratings of the campground attributes are ordinal, it is still appropriate for factor analysis. Usually, an interval or ratio scale is expected for calculating correlation coefficients (e.g., Pearson product-moment correlation coefficient) in factor analysis, because factor analysis is based on linear relationships of variables. However, Gorsuch (1983) indicated that it is not necessary. He pointed out that when rank (ordinal) data are submitted to a standard computer program for Pearson product-moment correlations, the results will be Spearman rank correlation coefficients which is a special case of the Pearson product-moment correlation coefficient and is appropriate for factor analysis.

Only the 424 respondents who rated all 20 attributes were included in this study because missing values on any attribute would have affected the calculation of the correlation matrix and thus have directly affected the parameter estimation (factor loading). However, because of the sample-size limitations of the cluster program and for cross-validation purposes, the total sample was divided into two subsamples, each containing 212 randomly selected cases. T-tests (see Appendix B) showed that there was no statistically significant difference in the importance ratings of different campground attributes/facilities between the two subsamples. Factor analysis was also performed for each subsample. The results of the factor analyses for both subsamples were similar (see Appendix C).

Statistical Methods Used to Achieve the Study Objectives

This section describes the statistical methods which were employed to achieve the study objectives.

The Effects of Different Factor Solutions on Cluster Membership

Objective 1. To assess the effect of different factor solutions (number of factors) on cluster membership.

Procedures

A seven-step procedure was employed to achieve Objective 1.

Step 1: Principal component analyses with varimax rotation were performed on the ratings of the 20 campground attributes/facilities. Nineteen different factor analyses were performed. Each analysis extracted a different number of factors from 20 factors to 2 factors. In the "20 factor" factor solution, each variable represents a factor.

Principal component analysis is a method for extracting principal factors under the component model, which summarizes the data by means of a linear combination of the observed data. The first extracted factor maximizes the variance accounted for in the correlation matrix. Each succeeding factor is extracted to maximize the residual variance explained (Gorsuch, 1983).

A frequent criticism of factor analysis is that the choice of technique is crucial to the final result. However, this criticism has not been supported by empirical evidence comparing the several types of factor analysis (Browne, 1968a; Gorsuch, 1983; Harris & Harris, 1971; Tucker, Koopman, & Linn, 1969). Stewart (1981) also indicated that when communalities are high there are virtually no differences among different factor extracting methods.

There are three primary types of orthogonal factor rotation--varimax, quartimax, and equimax. Varimax rotation is used to simplify the column of the factor matrix. It maximizes the variance of the squared loadings for each factor. Quartimax rotation is used to simplify the row of the factor matrix. Instead of maximizing variance of squared loadings for each factor, it maximizes the variance of the squared loadings for each variable so that a variable loads high on one factor and as low as possible on all other factors. Equimax rotation is a compromise between the varimax and quartimax criteria (Hair et al., 1987).

With the varimax rotational approach, there tend to be some high loadings close to -1 or +1 (indicating a clear association between the variable and the factor) and some loadings near 0 (indicating a clear lack of association) in each column of the matrix. Thus, the results of varimax rotation are easier to interpret than are those of quartimax rotation, which often produces a general factor with high-to-moderate loadings on most variables.

Step 2: Factor scores from the "20 factor" factor analysis were used as input variables for cluster analyses. Factor scores were obtained by multiplying the raw variables (ratings of attributes) by the factor score coefficients. They were treated as independent variables and received equal weight in the clustering procedures.

Step 3: The squared Euclidean distance measure and Ward's method were used to cluster respondents based on factor scores.

Squared Euclidean distance is defined as the square of the distance between two cases. It is generally used along with Ward's method (Norusis, 1988; Saunders, 1985). Ward's method involves a series of clustering steps that begins with N clusters, each containing one case, and ends with one cluster containing all cases. At the first stage, each case is in its own cluster and the error sum of squares (within-groups sum of squares) is 0. In the following stages, the two clusters which increase the least amount value of the sum of squares are merged. This clustering procedure results in a series of fusion coefficients (coefficient of hierarchy). Small increases in the coefficients indicate that fairly homogeneous clusters are being merged. Larger increases of coefficients indicate that clusters containing quite dissimilar members are being combined.

Step 4: The next step was to select a final cluster solution (number of clusters) for the clustering based on the "20 factor" factor solution. The selection criteria were: (a) error sum of squares (coefficient of hierarchy), (b) significance of the inter-cluster differences, and (c) size of clusters.

The coefficient of hierarchy for each clustering stage was plotted, beginning at the 25 cluster solution (see Figure 1 for illustration). The plot was examined to identify break points. A break point indicates a relatively large loss of information resulting from the fusion (of the clusters) at that point/level. Cluster solution(s) immediately preceding a break point(s) are candidates for a final cluster solution.

The three candidate solutions were then examined for significance of the inter-cluster differences. The factor scores centroids for each cluster (for each of the three candidate solutions) were compared using analysis of variance to determine differences between the clusters. The assumptions of ANOVA such as independence, normality, and homogeneity of variances were tested by using Bartlett-Box F test. The tests indicated that the ANOVA assumptions were not violated. The six-cluster solution

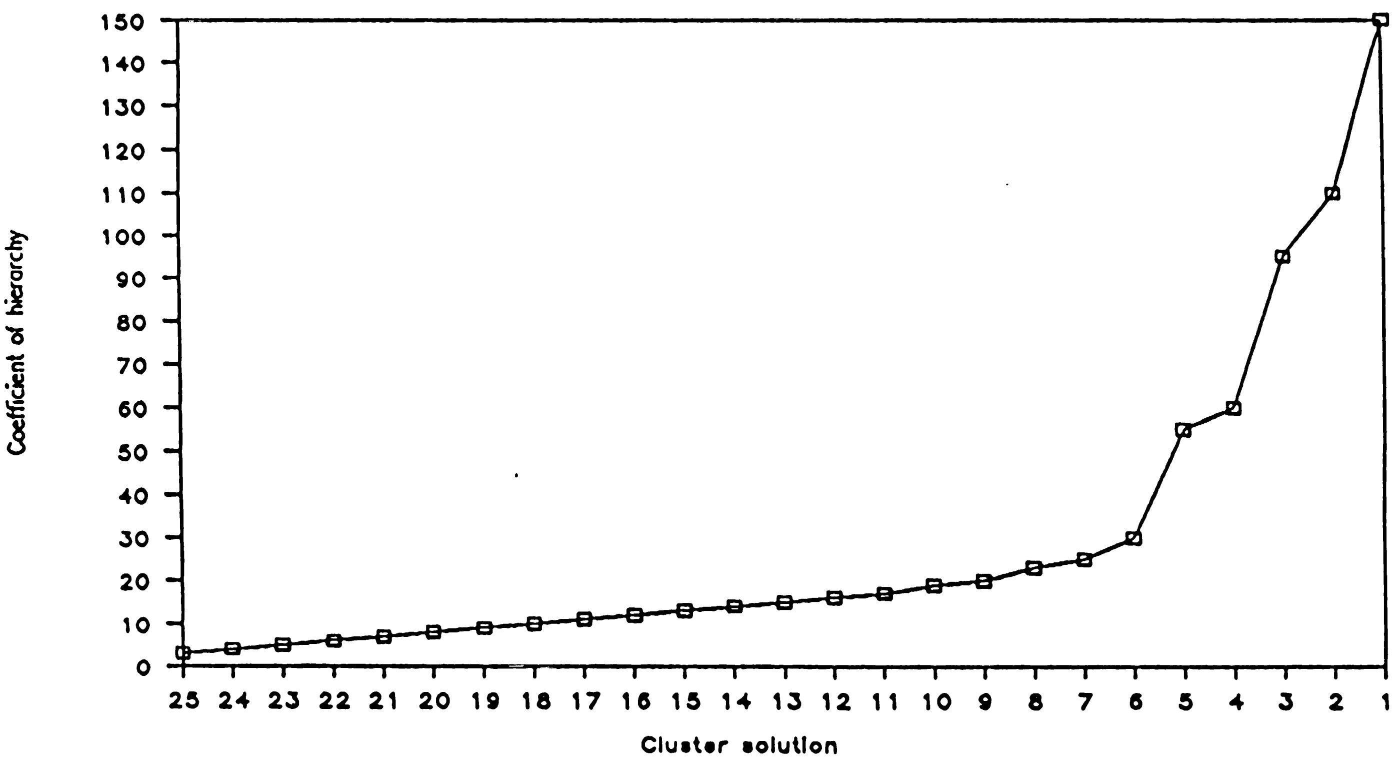


Figure 1. Illustration of a plot of the coefficient of hierarchy by number of clusters.

had the greater significance of the inter-cluster differences and was selected as the final cluster solution.

Step 5: In order to compare the effects of alternative factor solutions on cluster membership, Ward's method (using the squared Euclidean distance) was used to formulate six clusters for each of the other 18 factor analyses (19, 18, ..., 2).

Step 6: Changes in cluster membership across the different factor solutions (20, 19, ..., 2) were assessed by calculating and plotting information/entropy measures derived from crosstabulations of clusters.

Table 2 illustrates how cluster memberships were crosstabulated. It compares membership of clustering based on the "20 factor" factor solution with clustering based on the "19 factor" factor solution and clustering based on the "20 factor" factor solution with clustering based on the "18 factor" factor solution.

Information theory is derived from probability theory. It is concerned with how events/symbols are affected by various processes (Jones, 1979). Jones defined the self-information (I) of the event E_k as the logarithm of the event k's probability (p_k). The mathematical expression is: $I(E_k) = -\log p_k$. The smaller p_k is, the larger $I(E_k)$ is. This means that the rarer an event is, the more information is conveyed by its occurrence. For example, in Table 2 (page 49), the probability of cases being assigned to cluster 1 in the 20-factor solution is 44 (number of cases in cluster 1) divided by 212 (the total sample size); p_1 is 0.208. Therefore, $I(E_1) = -\log 0.208 = 0.682$ is the measure of information in assigning cases to cluster 1.

Table 2. Illustration of the crosstabulations of clusters across different factor solutions.

20-Factor Solution	19-Factor Solution					
	Cluster (percent)[b]					
Cluster/Size[a]	1	2	3	4	5	6
1 (44)	68.2	11.4	4.5	4.5	11.4	0.0
2 (46)	6.5	45.7	6.5	28.3	13.0	0.0
3 (29)	31.0	0.0	17.2	41.4	10.3	0.0
4 (32)	0.0	9.4	46.9	25.0	12.5	6.3
5 (45)	0.0	2.2	6.7	48.9	37.8	4.4
6 (16)	18.8	6.3	0.0	12.5	0.0	62.5
	18-Factor Solution					
	Cluster (percent)[b]					
	1	2	3	4	5	6
1 (44)	40.9	13.6	0.0	36.4	4.5	4.5
2 (46)	30.4	26.1	10.9	6.5	2.2	23.9
3 (29)	34.5	10.3	0.0	31.0	20.7	3.4
4 (32)	50.0	3.1	6.3	3.1	12.5	25.0
5 (45)	8.9	48.9	4.4	24.4	11.1	2.2
6 (16)	6.3	0.0	81.3	0.0	12.5	0.0

[a]Cases in cluster 1 derived from the 20-factor solution.

[b]Percent of cases assigned to the same cluster number in both factor solutions (e.g., 20-19, 20-18).

Information can be seen as the measure of uncertainty. As Donderi (1988) pointed out, information quantifies the effect of choice on uncertainty measured over a finite set of objects. In other words, information is a measure of what you have gained by your choice. Therefore, information gained is uncertainty reduced. For example, assume that a person planning a vacation originally has 8 possible destinations to choose among. After some initial consideration the list of possible destinations is reduced to four. Choosing four destinations reduces the set size from the original eight possible destinations, which required three binary choices (bits) to select a single destination, to a subset of four destinations, which requires only two bits to select a single destination. Narrowing the original eight possible destinations to four results in a gain of one bit of information, which means that the uncertainty has been reduced.

The concept of entropy introduced by Shannon (1948a,1948b) is fundamental in information theory. Entropy can be interpreted either as a measure of how unexpected the event was, or as measure of the information (uncertainty) yielded by the event (Aczél & Daróczy, 1975). Shannon (1948a,1948b) defined entropy (H) as the summation of each event's probability (p_k) individually multiplied by the logarithm of the probability of individual event ($\log p_k$). Jones (1979) integrated the information theory and the concept of entropy. He defined the entropy of system ($H(S)$) as the average of the self-information.

$$H(S) = E(I) = -\sum_{k=1}^{n} p_k * \log p_k \qquad (1)$$

Entropy is either positive or zero because p_k ranges from 0 to 1. When p_k is 0, the value 0 is assigned to $p_k * \log p_k$. When $H(S) = 0$, there is complete certainty the event must occur. In addition, entropy has a limit that entropy (H(S)) should be less than or equal to maximum entropy ($H(S)_{max}$) (Jones, 1979; Krippendorff, 1986). The maximum value of H(S) is attained when the probabilities of events in system S are all equal.

$$0 \leq H(S) \leq H(S)_{max} = \log(\min N_e, n)$$

where:

N_e : the number of events in system S.

n : the sample size.

Entropy as the measure of uncertainty has been applied to different fields, such as biological science, behavioral science, economics, geography, marketing, management, finance, and accounting. For example, Attaran and Guseman (1988) used entropy as a measure of the level of economic activity within the service sector of the United States to assess the changes in employment concentration between or within the manufacturing and service sectors over a 20-year period. Attaran and Zwick (1987) demonstrated that entropy is a useful measure for comparing industrial diversity either among regions or for a particular region over time. Lesser (1988) used entropy to predict the relationship between belief-behavior prediction and shopping style. Starr (1980) proposed a unique modification of the entropy level measure to explain switching patterns of loyalty. Beecher (1989) used entropy to measure the information capacity of an animal's "signature system" (the set of cues by which individuals are identified). Love (1986) used entropy to detect the relationship between concentration and export instability. Garrison (1974) applied an entropy measure of geographical concentration to examine the extent to which rural and small-town counties competed with urban areas for manufacturing employment in the Tennessee Valley region.

Conditional self-information (entropy) was used to measure the stability of cluster membership across different factor solutions (20 vs. 20, 20 vs. 19, 20 vs. 18, ..., 20 vs. 2). Similar to self-information, conditional self-information is based on conditional probability (the probability of event E, given that event F has occurred). Conditional entropy is likewise an analogue of entropy, obtained by taking the average of conditional self-information over all pairs of events, one from each system. Jones (1979) defined the conditional self-information $I(E_j \mid F_k)$ of E_j given that F_k has occurred (see Formula 2) and the conditional entropy $H(S1 \mid S2)$ (see Formula 3).

$$I(E_j \mid F_k) = - \log P(E_j \mid F_k) = -\log(p_{jk}/q_k) \qquad (2)$$

$$H(S_1 \mid S_2) = \sum_{j=1}^{n} \sum_{k=1}^{m} p_{jk} * I(E_j \mid F_k) = - \sum_{j=1}^{n} \sum_{k=1}^{m} p_{jk} * \log (p_{jk}/q_k) \qquad (3)$$

$$P(E_j \mid F_k) = P(E_j \cap F_k) \,/\, P(F_k) = p_{jk} \,/\, q_k \qquad (4)$$

Where:

$I(E_j \mid F_k)$: conditional self-information of E_j given that F_k has occurred

$H(S_1 \mid S_2)$: conditional entropy between system 1 and system 2

E_j : events within system one $\quad j = 1, 2 \ldots\ldots, n$

F_k : events within system two $\quad k = 1, 2 \ldots\ldots, m$

p_j : probabilities associated with E_j

q_k : probabilities associated with F_k

p_{jk}: probabilities of the connection between two systems, $P(E_j \cap F_k)$

To assess the changes in cluster membership across different factor solutions, Formula 3 was employed. Table 3 presents an illustration of major elements in calculating conditional entropy.

Based on Formula 3 and Table 3, the information measure for the changes in cluster membership between the 20-factor solution and the 19-factor solution is 0.5181. The calculation process is presented in Table 4.

In all, there were 19 information measures/entropy measures in this study (20-factor solution vs. 20-factor solution, 20-factor solution vs. 19-factor solution, ..., 20-factor solution vs. 2-factor solution). Nineteen information measures/entropy measures (see Table 5 for illustration) were plotted (see Figure 2) to assess the changes in cluster membership. The larger the entropy is between units (i.e., 20 vs. 20 and 20 vs. 19), the more uncertainty of change in cluster membership there is.

The information measure (entropy) as a measure of uncertainty was employed in this study for three reasons. First, the researchers were uncertain that the cluster 1 derived from the 20-factor solution was most similar to the cluster 1 derived from the 19-factor solution. The same uncertainty also extended to the other clusters (cluster 2, 3, 4, 5, 6). The information measure serves as an indicator showing how many bits of information are needed to clarify the uncertainty situation of the cluster structure. Second, it serves as an indicator assessing the changes of cluster membership in different situations. For example, based on Table 2, bits of information needed to clarify the uncertainty of the cluster structure in the crosstabulation of the 20-factor solution and the 19-factor solution are different from the

Table 3. Illustration of major elements in calculating conditional entropy.

E		System 1 E_1	E_2	E3	E_4	E_5	E_6
F		p_1	p_2	p_3	p_4	p_5	p_6
F1	q_1	$P(E_1 \mid F_1) = 0.682$	$P(E_2 \mid F_1) = 0.114$	$P(E_3 \mid F_1) = 0.045$	$P(E_4 \mid F_1) = 0.045$	$P(E_5 \mid F_1) = 0.114$	$P(E_6 \mid F_1) = 0.000$
F2	q_2	$P(E_1 \mid F_2) = 0.065$	$P(E_2 \mid F_2) = 0.457$	$P(E_3 \mid F_2) = 0.065$	$P(E_4 \mid F_2) = 0.283$	$P(E_5 \mid F_2) = 0.130$	$P(E_6 \mid F_2) = 0.000$
F_3	q_3	$P(E_1 \mid F_3) = 0.310$	$P(E_2 \mid F_3) = 0.000$	$P(E_3 \mid F_3) = 0.172$	$P(E_4 \mid F_3) = 0.414$	$P(E_5 \mid F_3) = 0.103$	$P(E_6 \mid F_3) = 0.000$
F_4	q_4	$P(E_1 \mid F_4) = 0.000$	$P(E_2 \mid F_4) = 0.094$	$P(E_3 \mid F_4) = 0.469$	$P(E_4 \mid F_4) = 0.250$	$P(E_5 \mid F_4) = 0.125$	$P(E_6 \mid F_4) = 0.063$
F_5	q_5	$P(E_1 \mid F_5) = 0.000$	$P(E_2 \mid F_5) = 0.022$	$P(E_3 \mid F_5) = 0.067$	$P(E_4 \mid F_5) = 0.489$	$P(E_5 \mid F_5) = 0.378$	$P(E_6 \mid F_5) = 0.044$
F_6	q_6	$P(E_1 \mid F_6) = 0.188$	$P(E_2 \mid F_6) = 0.063$	$P(E_3 \mid F_6) = 0.000$	$P(E_4 \mid F_6) = 0.125$	$P(E_5 \mid F_6) = 0.000$	$P(E_6 \mid F_6) = 0.625$

$p_{jk} = q_k * P(E_j \mid F_k)$ System 1: 19-factor solution System 2: 20-factor solution

E_1: assigning cases to cluster 1 under the condition of the 19-factor solution
E_2: assigning cases to cluster 2 under the condition of the 19-factor solution
E_3: assigning cases to cluster 3 under the condition of the 19-factor solution
E_4: assigning cases to cluster 4 under the condition of the 19-factor solution
E_5: assigning cases to cluster 5 under the condition of the 19-factor solution
E_6: assigning cases to cluster 6 under the condition of the 19-factor solution
F_1: assigning cases to cluster 1 under the condition of the 20-factor solution
F_2: assigning cases to cluster 2 under the condition of the 20-factor solution
F_3: assigning cases to cluster 3 under the condition of the 20-factor solution
F_4: assigning cases to cluster 4 under the condition of the 20-factor solution
F_5: assigning cases to cluster 5 under the condition of the 20-factor solution
F_6: assigning cases to cluster 6 under the condition of the 20-factor solution
p_1: the probability of assigning cases to cluster 1 associated with E_1
p_2: the probability of assigning cases to cluster 2 associated with E_2
p_3: the probability of assigning cases to cluster 3 associated with E_3
p_4: the probability of assigning cases to cluster 4 associated with E_4
p_5: the probability of assigning cases to cluster 5 associated with E_5
p_6: the probability of assigning cases to cluster 6 associated with E_6
q_1: the probability of assigning cases to cluster 1 associated with F_1
q_2: the probability of assigning cases to cluster 2 associated with F_2
q_3: the probability of assigning cases to cluster 3 associated with F_3
q_4: the probability of assigning cases to cluster 4 associated with F_4
q_5: the probability of assigning cases to cluster 5 associated with F_5
q_6: the probability of assigning cases to cluster 6 associated with F_6

Table 4. The calculation process for the information measure (changes in cluster membership) between the 20-factor solution and the 19-factor solution.

	$P(E_j \mid F_k)$	$P(E_j \cap F_k)$	$I(E_j \mid F_k)$	$H(S_1 \mid S_2)$
q_k	p_{jk}/q_k	p_{jk}	$-\log(p_{jk}/q_k)$	$-p_{jk} * \log(p_{jk}/q_k)$
44/212	0.682	0.142	0.166	0.0235
46/212	0.114	0.025	0.943	0.0233
29/212	0.045	0.006	1.347	0.0082
32/212	0.045	0.007	1.347	0.0091
45/212	0.114	0.024	0.943	0.0228
16/212	0.000	0.000	0.000	0.0000
Subtotal				0.0871
44/212	0.065	0.013	1.187	0.0160
46/212	0.457	0.099	0.340	0.0337
29/212	0.065	0.009	1.187	0.0106
32/212	0.283	0.043	0.548	0.0234
45/212	0.013	0.003	1.886	0.0052
16/212	0.000	0.000	0.000	0.0000
Subtotal				0.0889
44/212	0.310	0.064	0.509	0.0327
46/212	0.000	0.000	0.000	0.0000
29/212	0.172	0.024	0.764	0.0180
32/212	0.414	0.062	0.383	0.0239
45/212	0.103	0.022	0.987	0.0216
16/212	0.000	0.000	0.000	0.0000
Subtotal				0.0962
44/212	0.000	0.000	0.000	0.0000
46/212	0.094	0.020	1.027	0.0209
29/212	0.469	0.064	0.329	0.0211
32/212	0.250	0.038	0.602	0.0227
45/212	0.125	0.027	0.903	0.0240
16/212	0.063	0.005	1.201	0.0057
Subtotal				0.0944
44/212	0.000	0.000	0.000	0.0000
46/212	0.022	0.005	1.658	0.0079
29/212	0.067	0.009	1.174	0.0108
32/212	0.489	0.074	0.311	0.0229
45/212	0.378	0.080	0.422	0.0339
16/212	0.044	0.003	1.357	0.0045
Subtotal				0.0800
44/212	0.188	0.039	0.726	0.0283
46/212	0.063	0.014	1.201	0.0164
29/212	0.000	0.000	0.000	0.0000
32/212	0.125	0.019	0.903	0.0170
45/212	0.000	0.000	0.000	0.0000
16/212	0.625	0.047	0.204	0.0096
Subtotal				0.0096
Total				0.5181

Table 5. Artificial data for information (entropy) measure.

Comparison of Factor Solutions	Entropy
20 - 20	0.00
20 - 19	0.52
20 - 18	0.33
20 - 17	0.67
20 - 16	0.70
20 - 15	0.75
20 - 14	0.83
20 - 13	0.80
20 - 12	0.88
20 - 11	0.81
20 - 10	0.70
20 - 09	0.50
20 - 08	0.15
20 - 07	0.38
20 - 06	0.47
20 - 05	0.53
20 - 04	0.37
20 - 03	0.20
20 - 02	0.40

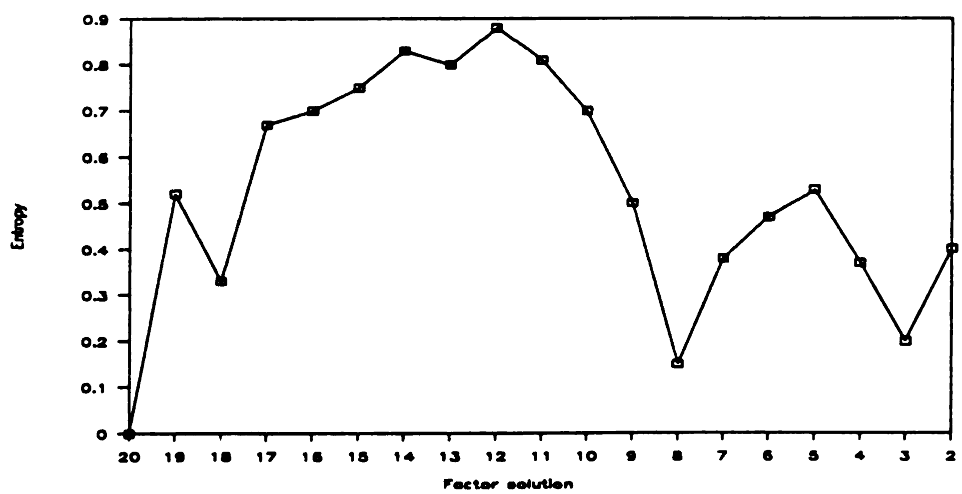


Figure 2. Illustration of the plot of 19 entropy measures.

crosstabulation of the 20-factor solution and the 18-factor solution. The difference in bits of information indicates how cluster membership has been changed during the process of reducing the factor solution (i.e., reducing factor solution from 20 to 19 and from 19 to 18). Third, the information (entropy) serves as an indicator assessing the stability of cluster membership. Because the level of changes in cluster membership is uncertain during the process of reducing the factor solution, plotting all the information measures (derived from the crosstabulation of the 20- and 19-factor solution, 20- and 18-, ..., 20-and 2-factor solution) will provide the stability/change pattern of cluster membership.

Step 7: In order to assess the stability of the (factor) centroids for each cluster, the (factor score) centroids of each of the six clusters was calculated for each of the 19 factor analyses (see Table 6 for illustration). The (factor score) centroids of the six clusters were then plotted for the 19 different factor solutions (see Figure 3 for illustration).

The Effects of Factor Rotation on Cluster Membership

Objective 2. To ascertain the effect of factor rotation on cluster membership.

Procedures

A four-step procedure was used to achieve Objective 2. The first two steps, factor analysis and clustering on the factor scores, were the

Tab

Fa
Sc

1
1
1
1
1
1
1
1
1
1
1
2

1
1
1
1
1
1
1
1
1
1
2

Table 6. Illustration of (factor score) centroids for each of the six clusters across different factor solutions.

Factor Solution	Cluster 1	2	3	4	5	6
	(Factor 1 Factor Score Centroid)					
2	.655	.048	.567	-.698	-.953	-1.800
3	.866	-.129	.573	-.860	-.866	-.201
4	-.777	.338	.811	-1.213	-.369	1.956
5	-.686	.343	.808	-1.333	-.372	1.736
6	-.716	.354	.775	-1.199	-.379	1.767
7	-.662	.173	.782	-1.079	-.342	1.970
8	-.700	.090	.775	-.945	-.302	2.192
9	-.683	.160	.785	-1.055	-.325	1.992
10	-.665	.138	.799	-1.119	-.288	1.878
11	-.693	.137	.768	-.923	-.268	1.547
12	-.697	.134	.769	-.945	-.245	1.513
13	-.591	.316	.676	-.946	-.304	1.449
14	-.583	.217	.667	-.941	-.302	1.437
15	-.535	.236	.631	-.909	-.327	1.363
16	-.533	.247	.620	-.913	-.325	1.369
17	-.523	.265	.610	-.908	-.342	1.373
18	.733	-.754	.500	-.923	-.170	-.195
19	.721	-.752	.490	-.898	-.156	-.209
20	-.436	-.176	.540	-.452	-.053	1.125
	Cluster 1	2	3	4	5	6
	(Factor 2 Factor Score Centroid)					
2	-.484	.133	.916	-1.619	-.349	1.635
3	-.500	.106	.868	-1.574	-.269	1.715
4	.742	-.421	.554	-1.353	-.289	-.928
5	-.006	.099	.291	-.363	-.328	-.035
6	.411	.343	.377	-.230	-.978	-1.780
7	.436	.081	.363	-.032	-.867	-1.338
8	.449	.210	.355	-.143	-.918	-1.571
9	.441	.167	.329	-.067	-.888	-1.413
10	.435	.184	.332	-.048	-.926	-1.332
11	.446	.174	.326	-.066	-.926	-1.217
12	.433	.201	.308	-.008	-.959	-1.135
13	.458	.198	.293	.004	-.969	-1.122
14	.472	.261	.255	.024	-1.017	-1.058
15	.464	.245	.262	-.014	-1.001	-.931
16	.731	-.752	.503	-.907	-.197	-.099
17	.736	-.746	.503	-.906	-.208	-.104
18	-.470	.253	.520	-.910	-.196	.975
19	-.262	.194	.309	-.848	-.034	.659
20	.435	-.456	.390	-.652	-.145	-.298

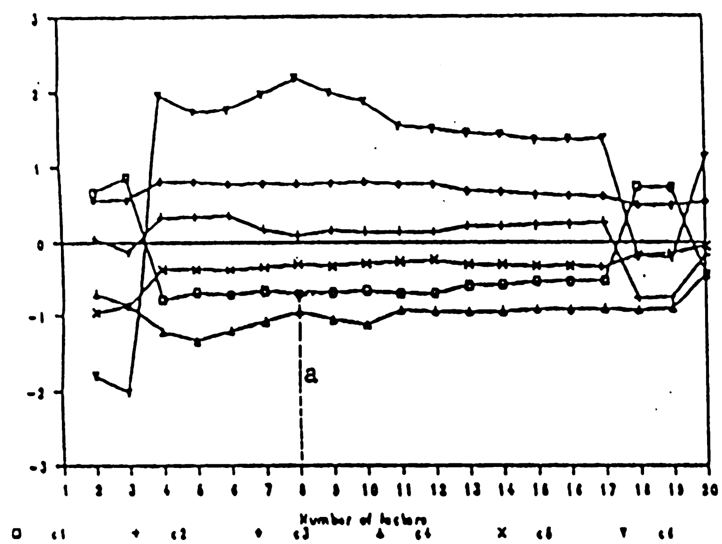


Figure 3. Illustration of the plot of factor centroids (a = factor 1 factor score centroid of cluster 1 for the "8 factor" factor solution, b = factor 2 factor score centroid of cluster 3 for the "15 factor" factor solution; c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

same as steps 1 and 2 used to achieve Objective 1 except that the initial factors were not rotated.

Step 3: The clusters (memberships) formulated on the basis of unrotated factor scores were compared (crosstabulated) with cluster (memberships) formulated on the basis of rotated factor scores. Table 7 illustrates how the comparison was performed.

Step 4: The cell percentages were analyzed to determine the degree of similarity in cluster memberships. If the diagonal percentages equaled 100%, the cluster memberships were the same. The greater the deviation from 100%, the greater the difference in cluster memberships.

Comparison of Different Clustering Approaches

Objective 3. To compare clustering on factor scores with clustering on raw data.

Procedures

A seven-step procedure was employed to achieve Objective 3.

Step 1: Respondents were first clustered on the raw data (importance ratings of the 20 attributes). Ward's method (using the squared Euclidean distance measure) was employed. The error sum of squares, significance of the inter-cluster difference, and size of clusters were again used as the criteria to decide a cluster solution. A six cluster solution was selected.

Step 2: Nineteen principal component analyses with varimax rotation were performed on the rating of the 20 campground

Table 7. Illustration of crosstabulation comparison of the memberships of clusters derived from rotated factor scores with clusters derived from unrotated factor scores.

Rotated Factor Analysis (20, 19, 18, ..., 2)	Unrotated Factor Analysis (20, 19, 18, ..., 2)					
	Clusters (percent)[a]					
Clusters	1	2	3	4	5	6
1	%	%	%	%	%	%
2	%	%	%	%	%	%
3	%	%	%	%	%	%
4	%	%	%	%	%	%
5	%	%	%	%	%	%
6	%	%	%	%	%	%

[a]percentage of cases assigned to cluster 1 in both the rotated and unrotated factor analysis.

attributes/facilities, as was done in step 1 for Objective 1 (see page 44). Nineteen different factor analyses were performed. Each factor analysis extracted a different number of factors from 20 factors to 2 factors.

Step 3: The (factor score) centroids for each of the six clusters were calculated for each of the 19 factor analyses (see Table 6 for illustration). The (factor score) centroids of each of the six clusters were then plotted for each factor solution (see Figure 3 for illustration).

Step 4: The sum of squared distance for each cluster on each factor (factor score) centroid was computed when clustering on raw data. For example, in Table 8, the sum of squared distance for cluster 1 on "factor 1" factor score centroid is calculated by adding the squared

Table 8. Illustration of the calculation of the sum of squared distance.

						Cluster						
	1		2		3		4		5		6	
Factor Solution		D_1		D_2		D_3		D_4		D_5		D_6
				(Factor 1 Factor Score Centroid)								
2	1		3		2		2		1		4	
3	1	0	2	1	0	4	1	1	3	4	1	9
4	3	4	0	4	1	1	3	4	0	9	1	0
5	2	1	0	0	2	1	1	4	1	1	2	1
6	0	4	1	1	2	0	3	4	2	1	1	1
7	3	9	1	0	1	1	3	0	3	1	2	1
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
20	.	.	.	.	.	.	.	.	.	.	.	.
Sum of Squared Distance		18		6		7		13		16		12

Note: For illustration purpose, this table only shows five squared distances.

D_1 means the squared difference of factor 1 factor score centroid between different factor solutions for cluster 1.

D_2 means the squared difference of factor 1 factor score centroid between different factor solutions for cluster 2.

D_3 means the squared difference of factor 1 factor score centroid between different factor solutions for cluster 3.

D_4 means the squared difference of factor 1 factor score centroid between different factor solutions for cluster 4.

D_5 means the squared difference of factor 1 factor score centroid between different factor solutions for cluster 5.

D_6 means the squared difference of factor 1 factor score centroid between different factor solutions for cluster 6.

distance of centroid points between a 2-factor solution and a 3-factor solution, the squared distance of centroid points between a 3-factor solution and a 4-factor solution, ..., and the squared distance of centroid points between a 19-factor solution and a 20-factor solution.

Step 5: The sum of squared distance for each cluster on each factor (factor score) centroid was also computed when clustering on factor scores.

Step 6: The similarity of each of the clusters formulated on raw data and factor scores was assessed using a specially designed computer program (see Appendix D). The program identified the best set of matched clusters for each factor (factor score) centroid. For example, in factor 1 factor score centroid, the cluster 6 derived from clustering on factor scores is most similar to the cluster 1 derived from clustering on raw data (see Table 9).

The program was specially written to determine the best set of matched clusters between the two clustering approaches--raw data and factor scores. The sum of squared distances calculated in step 4 and step 5 were used as input to this computer program. In each iteration, the program generates a set of matched clusters. For example, cluster 1 (based on raw data) matches with cluster 6 (derived from factor scores), which marked as C_{16}; cluster 2 (based on raw data) with cluster 5 (derived from factor scores), marked as C_{25}; the other matched clusters were marked as C_{34}, C_{42}, C_{55}, and C_{61}.

The difference of the sum of squared distance is then calculated for each of the six matches (e.g., C_{16}, C_{25}, ..., C_{61}) and summed. The computer program then generates other sets of matched clusters. For each set of cluster match, the total difference of the sum of squared

Table 9. Illustration for the measure of cluster similarity.

	Clustering On Factor Scores			Clustering On Raw Data	
Cluster	Sum of Distance	Standard Deviation	Cluster	Sum of Distance	Standard Deviation
6	12.783	1.3	1	5.686	0.8
4	9.453	2.1	2	1.672	0.6
3	6.656	1.5	3	0.084	1.1
2	6.909	1.2	4	0.472	0.5
1	4.612	0.5	5	0.305	1.4
5	15.527	1.7	6	20.342	0.7

distance is calculated. Based on the criterion of minimum total difference of the sum of squared distance, the computer program identifies the best set of matched clusters.

Step 7: The standard deviations of factor score centroids for each cluster across different factor solutions were calculated. The values of the standard deviation for each of the six matched clusters were used as the basis for comparing the stability of each factor score centroid between clustering on raw data and clustering on factor scores. Six sets of stability comparisons were made. The higher the standard deviation, the more unstable the cluster membership (factor score centroid). The "best" approach results in more stable clusters.

To demonstrate how the stability comparisons were made, the following example is presented. The computer program identified a set of matched clusters: C_{16}, C_{25}, C_{34}, C_{42}, C_{55}, C_{61}. As stated above, standard deviations were calculated for each of the six matched clusters. Suppose that the standard deviation of the cluster 1 (based on raw data) is 0.8 and the standard deviation of the cluster 6 (based on factor scores) is 1.3, the cluster membership of the cluster 1 (based on raw data) is more stable than the cluster 6 (based on factor scores). The other five matched clusters were also compared based on the value of standard deviations. If clustering on raw data has more stable clusters than that of clustering on factor scores, clustering on raw data is identified as a better approach.

CHAPTER IV

RESULTS

The chapter is divided into five sections dealing with (1) the importance ratings of the twenty different campground attributes, (2) the appropriateness of data for factor analysis, (3) an assessment of the effect of different factor solutions on the clustering results, (4) an assessment of the effect of rotation on cluster membership, and (5) a comparison of clustering on factor scores with clustering on raw data.

Importance Ratings of 20 Campground Attributes

The importance ratings assigned to the 20 campground attributes/facilities by respondents are shown in Table 10. The ratings ranged from crucial (1) to not important (5). The distribution of ratings, mean and median scores, and standard deviation for each attribute are also reported in Table 10.

Cleanliness of a campground (mean = 1.877) was the most important attribute. This was followed by security (mean = 2.160), hospitality of campground staff (mean=2.500), quietness (mean = 2.759), electricity (mean = 2.750), and low price (mean = 2.896). Campers as a whole were

Table 10. Importance ratings (assigned the campground attributes) which were used in the factor analyses and cluster analyses.

Campground Attributes	Importance Rating[a] 1	2	3 (percent)	4	5	Mean	Median	Standard Deviation
Large sites	6.6	17.9	41.5	25.9	8.0	3.108	3.0	1.008
Shaded sites	1.9	20.8	40.6	29.2	7.5	3.198	3.0	0.918
Cleanliness	30.2	55.2	11.8	2.4	0.5	1.877	2.0	0.738
Quietness	6.1	32.1	45.3	12.7	3.8	2.759	3.0	0.889
Site privacy	2.4	17.5	37.3	30.2	12.7	3.335	3.0	0.986
Security	23.1	46.2	22.6	7.5	0.5	2.160	2.0	0.883
Hospitality of campground staff	12.3	41.5	33.0	10.4	2.8	2.500	2.0	0.936
Low price	8.5	26.4	35.8	25.5	3.8	2.896	3.0	1.002
Flush toilets	6.1	18.9	29.7	25.9	19.3	3.335	3.0	1.167
Electricity	13.2	29.2	32.5	19.3	5.7	2.750	3.0	1.088
Showers	9.0	25.9	31.1	23.6	10.4	3.005	3.0	1.129
Laundromat	1.9	5.7	24.5	34.0	34.0	3.925	4.0	0.990
Campground store	1.4	9.4	20.8	43.4	25.0	3.811	4.0	0.965
Water hookups	9.4	26.4	25.5	22.2	16.5	3.099	3.0	1.233
Sewer hookups	4.7	11.3	23.6	25.9	34.4	3.741	4.0	1.182
Natural surroundings	4.7	20.8	34.9	27.4	12.3	3.217	3.0	1.058
Situated on a lake/stream	1.4	8.0	18.4	30.2	42.0	4.033	4.0	1.028
Hiking trails	1.4	9.4	15.1	35.8	38.2	4.000	4.0	1.021
Pool	1.4	10.4	20.3	25.0	42.9	3.976	4.0	1.086
Playgrounds	0.9	6.6	8.5	15.1	68.9	4.443	5.0	0.965

[a]The importance ratings of campground attributes ranged from crucial (1) to not important (5).

less concerned with whether a campground had a laundromat (mean = 3.925), a swimming pool (mean = 3.976), or a hiking trail (mean = 4.000), whether it was situated on lake/stream (mean = 4.033), and whether it had playgrounds (mean = 4.443).

Appropriateness of the Data for Factor Analysis

Prior to performing a factor analysis, the data (importance ratings) were examined with respect to their appropriateness (sample size and correlation between variables) for factor analysis. A number of criteria for determining whether a factor analysis should be applied to a set of data were reviewed. A common criterion is the size of the sample. Comrey (1973) suggested that if the sample size is equal to 100, the appropriateness for factor analysis is poor; 200 it is fair; 300 it is good; 500 it is very good; and 1000 it is excellent. Stewart (1981) suggested six methods of determining whether the data are appropriate for factor analysis. These include the examination of the correlation matrix, the plotting of the eigenvalues obtained from matrix decomposition, the examination of communality estimates, the inspection of the off-diagonal elements of the anti-image covariance or correlation matrix, Bartlett's test of sphericity, and the Kaiser-Meyer-Olkin measure of sampling adequacy (MSA).

The criteria used were (a) the sample size, (b) Bartlett's test of sphericity, and (c) the Kaiser-Meyer-Olkin measure of sampling adequacy (MSA). In the present study, there are two split subsamples each containing 212 cases, which is an adequate size for factor analysis.

Bartlett's test of sphericity was used to test (using a chi-square test) the hypothesis that the correlation matrix is an identity matrix (e.g., variables correlate perfectly with themselves, but are uncorrelated with other variables). That is, all diagonal terms are 1 and all off-diagonal terms are 0. Rejecting the hypothesis indicates that the data are appropriate for factor analysis (Bartlett, 1950, 1951).

Bartlett's test of sphericity was performed. The chi-square value is 1441 (with 190 degrees of freedom) that is highly significant. Thus, based on this test, the data is appropriate for factor analysis.

Kaiser-Meyer-Olkin measure of sampling adequacy (MSA) provides a measure of the extent to which the variables belong together (Kaiser, 1970). Small value for the MSAs (less than .50) indicate that data may not be appropriate for factor analysis because correlation between pairs of variables can not be explained by the other variables (Norusis,1988). In this study, the MSA is 0.81, which indicates that data is appropriate for factor analysis (Kaiser & Rice, 1974).

Assessment of the Effect of Different Factor Solutions on the Clustering Results

Factoring Results

Nineteen (20, 19, 18, ..., 2 factors) different principal component analyses with varimax rotation were performed. The eigenvalues and percentages of variance explained are reported in Table 11 along with the cumulative percentage of variance explained by the

Table 11. Eigenvalue, percent of variance explained, and cumulative percent of variance explained for 20 campground attributes.

Factor	Eigenvalue	Percent of Variance Explained	Cumulative Percent of Variance Explained
1	5.60131	28.0	28.0
2	1.93845	9.7	37.7
3	1.69936	8.5	46.2
4	1.32863	6.6	52.8
5	1.16849	5.8	58.7
6	1.09119	5.5	64.1
7	1.02010	5.1	69.2
8	0.80158	4.0	73.2
9	0.67725	3.4	76.6
10	0.61859	3.1	79.7
11	0.57406	2.9	82.6
12	0.54578	2.7	85.3
13	0.50535	2.5	87.9
14	0.47601	2.4	90.2
15	0.44025	2.2	92.4
16	0.38502	1.9	94.4
17	0.32611	1.6	96.0
18	0.29376	1.5	97.5
19	0.27759	1.4	98.8
20	0.23112	1.2	100.0

different number of factors. For each factor, the eigenvalue is the sum of squared factor loadings. Eliminating factors one at a time starting from the 20 factor reduced the percentage of total variance explained. The eigenvalues and percentages of variance explained in proportion to the eigenvalues of the factors eliminated from the solution remained the same. For example, the first 18 eigenvalues of the "19 factor" principal component analysis are identical to the 18 eigenvalues of the "18 factor" principal component analysis.

The next step was to identify the "best" factor solution based on factor analysis criteria. The scree test/plot which was used to select candidate factor solutions is presented in figure 4. The scree plot identified three candidate factor solutions (2 factors, 4 factors, and 7 factors). A seven-factor solution was selected from among all possible solutions because (a) eigenvalues from factor 1 to factor 7 were greater than 1, and (b) the percentage of total variance explained was about 70%. In many studies, the seven-factor solution would have been used as the basis for clustering. However, the purpose of this study was to assess the effects of alternative factor solutions on the clustering results, so the seven-factor solution was only one of 19 different factor solutions which were considered.

Next, one factor at a time was eliminated beginning with the 20-factor solution. The impact of the "one at a time" factor elimination on the factor pattern matrix are shown in Tables 12-30. Only the loadings of variables with a factor loading of 0.40 or greater are shown in the tables. For example, Table 12 shows the factor pattern matrix for the 20 factor principal component analysis (with varimax

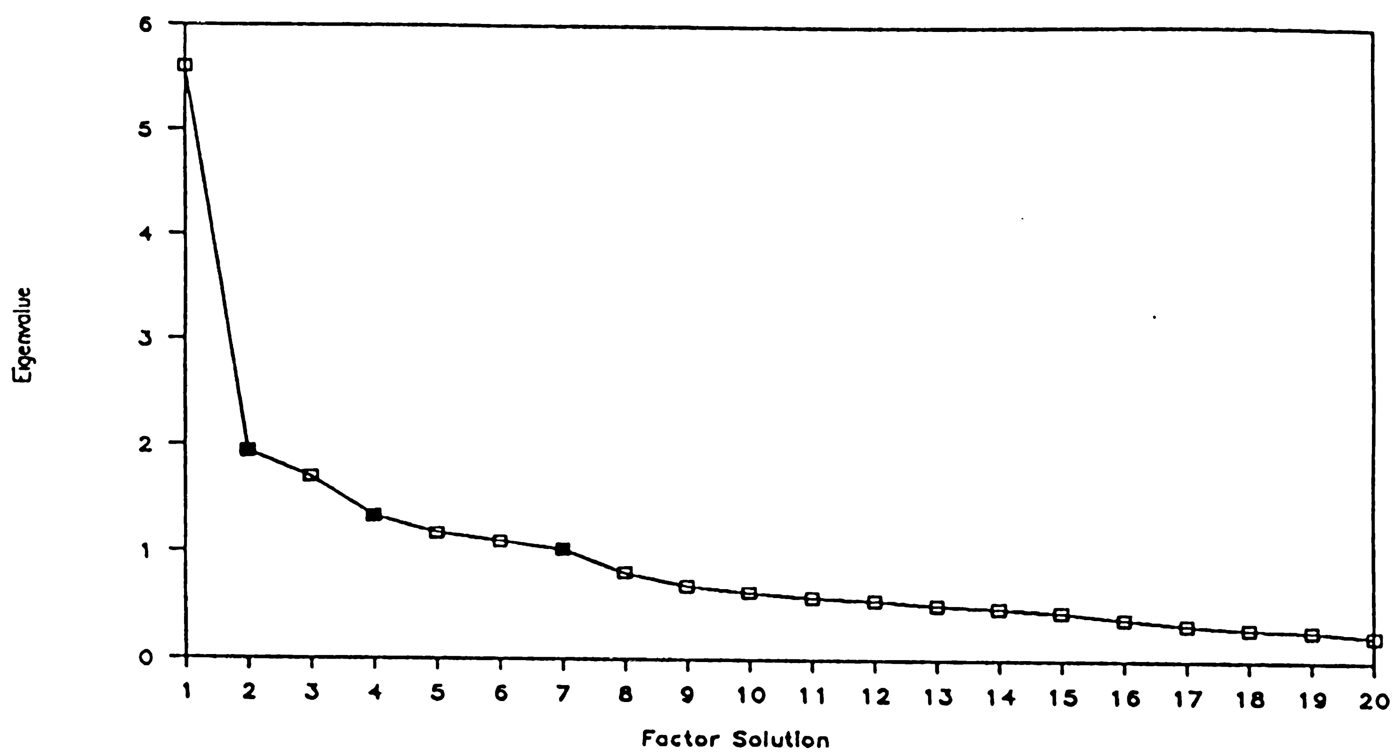


Figure 4. Scree test for selecting candidate factor solutions (■: based on eigenvalue > 1 and percentage of total variance explained, this would be the factor solution).

Table 12. Campground attribute sought factor pattern matrix for "20 factor" principal component analysis with varimax rotation.

Campground Attributes	Factor (Factor Loadings) 01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Large sites					.96															
Shaded sites						.96														
Cleanliness													.91							
Quietness															.89					
Privacy										.91										
Security									.92											
Hospitality								.92												
Low price				.96																
Flush toilets		.89																		
Electricity	.92																			
Shower																				.85
Laundromat														.89						
Store																		.88		
Water hookups																			.86	
Sewer hookups																.87				
Natural surroundings											.90									
Lake/stream												.89								
Hiking trail																	.88			
Swimming pool							.93													
Playgrounds			.95																	

Note: Only variables whose loadings are greater than .04 are shown.

Table 13. Campground attribute sought factor pattern matrix for "19 factor" principal component analysis with varimax rotation.

	Factor																		
Campground Attributes	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
	(Factor Loadings)																		
Large sites				.96															
Shaded sites							.96												
Cleanliness															.90				
Quietness																	.89		
Privacy								.90											
Security									.92										
Hospitality										.92									
Low price					.96														
Flush toilets	.90																		
Electricity			.91																
Shower	.84																		
Laundromat														.89					
Store																		.88	
Water hookups																			.79
Sewer hookups		.88																	
Natural surroundings											.89								
Lake/stream													.89						
Hiking trail																.88			
Swimming pool												.91							
Playgrounds						.94													

Note: Only variables whose loadings are greater than .04 are shown.

Table 14. Campground attribute sought factor pattern matrix for "18 factor" principal component analysis with varimax rotation.

Campground Attributes	Factor 01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18
	(Factor Loadings)																	
Large sites					.96													
Shaded sites							.96											
Cleanliness														.90				
Quietness																	.86	
Privacy								.90										
Security										.91								
Hospitality									.91									
Low price						.96												
Flush toilets	.89																	
Electricity				.87														
Shower	.86																	
Laundromat															.87			
Store																		.86
Water hookups		.72																
Sewer hookups		.91																
Natural surroundings											.88							
Lake/stream													.89					
Hiking trail																.88		
Swimming pool												.89						
Playgrounds			.94															

Note: Only variables whose loadings are greater than .04 are shown.

Table 15. Campground attribute sought factor pattern matrix for "17 factor" principal component analysis with varimax rotation.

Campground Attributes	Factor 01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
	(Factor Loadings)																
Large sites					.96												
Shaded sites									.95								
Cleanliness				.90													
Quietness			.64	.42													
Privacy			.90														
Security													.90				
Hospitality						.91											
Low price								.96									
Flush toilets		.89															
Electricity	.43																.82
Shower		.86															
Laundromat														.86			
Store																.85	
Water hookups	.86																
Sewer hookups	.86																
Natural surroundings											.88						
Lake/stream										.89							
Hiking trail															.87		
Swimming pool												.89					
Playgrounds							.94										

Note: Only variables whose loadings are greater than .04 are shown.

Table 16. Campground attribute sought factor pattern matrix for "16 factor" principal component analysis with varimax rotation.

Campground Attributes	Factor 01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16
	(Factor Loadings)															
Large sites						.96										
Shaded sites										.95						
Cleanliness					.90											
Quietness				.65	.42											
Privacy				.89												
Security													.90			
Hospitality							.90									
Low price								.96								
Flush toilets		.89														
Electricity	.44															.80
Shower		.86														
Laundromat												.86				
Store															.84	
Water hookups	.85															
Sewer hookups	.87															
Natural surroundings														.85		
Lake/stream			.90													
Hiking trail			.76													
Swimming pool											.88					
Playgrounds									.94							

Note: Only variables whose loadings are greater than .04 are shown.

Table 17. Campground attribute sought factor pattern matrix for "15 factor" principal component analysis with varimax rotation.

	Factor														
Campground Attributes	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
	(Factor Loadings)														
Large sites								.96							
Shaded sites											.95				
Cleanliness						.89									
Quietness				.65											
Privacy				.90											
Security													.86		
Hospitality							.90								
Low price									.94						
Flush toilets			.89												
Electricity	.45													.80	
Shower			.86												
Laundromat					.85										
Store					.69										
Water hookups	.86														
Sewer hookups	.86														
Natural surroundings															.81
Lake/stream		.85													
Hiking trail		.82													
Swimming pool												.89			
Playgrounds										.94					

Note: Only variables whose loadings are greater than .04 are shown.

Table 18. Campground attribute sought factor pattern matrix for "14 factor" principal component analysis with varimax rotation.

	Factor													
Campground Attributes	01	02	03	04	05	06	07	08	09	10	11	12	13	14
	(Factor Loadings)													
Large sites							.95							
Shaded sites												.94		
Cleanliness						.90								
Quietness					.62									
Privacy					.92									
Security						.50								
Hospitality								.90						
Low price									.92					
Flush toilets			.89											
Electricity	.56													.68
Shower			.86											
Laundromat				.83										
Store				.75										
Water hookups	.87													
Sewer hookups	.85													
Natural surroundings													.79	
Lake/stream		.85												
Hiking trail		.82												
Swimming pool											.88			
Playgrounds										.93				

Note: Only variables whose loadings are greater than .04 are shown.

Table 19. Campground attribute sought factor pattern matrix for "13 factor" principal component analysis with varimax rotation.

Campground Attributes	01	02	03	04	05	06	07	08	09	10	11	12	13
	(Factor Loadings)												
Large sites								.95					
Shaded sites									.90				
Cleanliness				.77									
Quietness				.41		.57							-.40
Privacy						.91							
Security				.73									
Hospitality							.76						
Low price										.91			
Flush toilets			.88										
Electricity	.57												.67
Shower			.86										
Laundromat					.83								
Store					.75								
Water hookups	.87												
Sewer hookups	.85												
Natural surroundings		.56					.63						
Lake/stream		.84											
Hiking trail		.85											
Swimming pool												.88	
Playgrounds											.93		

Note: Only variables whose loadings are greater than .04 are shown.

Table 20. Campground attribute sought factor pattern matrix for "12 factor" principal component analysis with varimax rotation.

Campground Attributes	Factor 01	02	03	04	05	06	07	08	09	10	11	12
	(Factor Loadings)											
Large sites									.92			
Shaded sites								.90				
Cleanliness						.44	.63					
Quietness				.79								
Privacy				.80								
Security							.75					
Hospitality						.81						
Low price											.91	
Flush toilets			.86									
Electricity	.80											
Shower			.87									
Laundromat					.79							
Store					.78							
Water hookups	.83											
Sewer hookups	.80											
Natural surroundings		.61				.53						
Lake/stream		.84										
Hiking trail		.85										
Swimming pool												.87
Playgrounds										.92		

Note: Only variables whose loadings are greater than .04 are shown.

Table 21. Campground attribute sought factor pattern matrix for "11 factor" principal component analysis with varimax rotation.

Campground Attributes	Factor 01	02	03	04	05	06	07	08	09	10	11
	(Factor Loadings)										
Large sites								.92			
Shaded sites										.90	
Cleanliness				.73							
Quietness				.42	.77						
Privacy					.82						
Security					.43		.42				
Hospitality				.81							
Low price							.87				
Flush toilets			.86								
Electricity	.80										
Shower			.86								
Laundromat						.79					
Store						.75					
Water hookups	.82										
Sewer hookups	.80										
Natural surroundings		.71									
Lake/stream		.83									
Hiking trail		.83									
Swimming pool											.78
Playgrounds									.89		

Note: Only variables whose loadings are greater than .04 are shown.

Table 22. Campground attribute sought factor pattern matrix for "10 factor" principal component analysis with varimax rotation.

Campground Attributes	Factor 01	02	03	04	05	06	07	08	09	10
	(Factor Loadings)									
Large sites								.92		
Shaded sites									.89	
Cleanliness				.74						
Quietness					.77					
Privacy					.82					
Security				.47	.42					-.42
Hospitality				.78						
Low price						.88				
Flush toilets			.86							
Electricity	.77									
Shower			.87							
Laundromat	.48					.43				
Store	.44					.48	.47			
Water hookups	.83									
Sewer hookups	.82									
Natural surroundings		.70								
Lake/stream		.84								
Hiking trail		.83								
Swimming pool										.76
Playgrounds							.86			

Note: Only variables whose loadings are greater than .04 are shown.

Table 23. Campground attribute sought factor pattern matrix for "9 factor" principal component analysis with varimax rotation.

	Factor								
Campground Attributes	01	02	03	04	05	06	07	08	09
	(Factor Loadings)								
Large sites								.92	
Shaded sites									.85
Cleanliness					.72				
Quietness				.79					
Privacy				.79					
Security				.52	.42		.40		
Hospitality					.79				
Low price							.82		
Flush toilets			.86						
Electricity	.75								
Shower			.86						
Laundromat	.44						.51		
Store	.41					.41	.55		
Water hookups	.85								
Sewer hookups	.81								
Natural surroundings		.69							
Lake/stream		.84							
Hiking trail		.82							
Swimming pool						.72			
Playgrounds						.77			

Note: Only variables whose loadings are greater than .04 are shown.

Table 24. Campground attribute sought factor pattern matrix for "8 factor" principal component analysis with varimax rotation.

Campground Attributes	Factor 01	02	03	04	05	06	07	08
	(Factor Loadings)							
Large sites							.77	
Shaded sites								.84
Cleanliness						.68		
Quietness			.75					
Privacy			.72					
Security			.73					
Hospitality						.80		
Low price					.72			
Flush toilets				.86				
Electricity	.79							
Shower				.86				
Laundromat	.41				.58			
Store					.70			
Water hookups	.83							
Sewer hookups	.78							
Natural surroundings		.71						
Lake/stream		.82						
Hiking trail		.83						
Swimming pool							.62	
Playgrounds					.42			

Note: Only variables whose loadings are greater than .04 are shown.

Table 25. Campground attribute sought factor pattern matrix for "7 factor" principal component analysis with varimax rotation.

				Factor			
Campground	0	0	0	0	0	0	0
Attributes	1	2	3	4	5	6	7
				(Factor Loadings)			
Large sites					.73		
Shaded sites							.44
Cleanliness			.75				
Quietness			.56		.60		
Privacy					.69		
Security			.55		.47		
Hospitality			.75				
Low price							.81
Flush toilets				.83			
Electricity	.77						
Shower				.82			
Laundromat	.47						.45
Store	.42					.41	.55
Water hookups	.84						
Sewer hookups	.81						
Natural surroundings		.75					
Lake/stream		.80					
Hiking trail		.80					
Swimming pool						.70	
Playgrounds						.73	

Note: Only variables whose loadings are greater than .04 are shown.

Table 26. Campground attribute sought factor pattern matrix for "6 factor" principal component analysis with varimax rotation.

Campground Attributes	Factor 0 1	0 2	0 3	0 4	0 5	0 6
	(Factor Loadings)					
Large sites					.60	
Shaded sites						.45
Cleanliness				.76		
Quietness				.59	.55	
Privacy					.72	
Security				.57		
Hospitality				.73		
Low price						.81
Flush toilets			.80			
Electricity	.71					
Shower			.83			
Laundromat	.53					.45
Store	.49					.55
Water hookups	.83					
Sewer hookups	.81					
Natural surroundings		.72				
Lake/stream		.78				
Hiking trail		.80				
Swimming pool			.56			
Playgrounds		.48				

Note: Only variables whose loadings are greater than .04 are shown.

Table 27. Campground attribute sought factor pattern matrix for "5 factor" principal component analysis with varimax rotation.

	Factor				
Campground Attributes	01	02	03	04	05
	(Factor Loadings)				
Large sites					-.54
Shaded sites		.42			
Cleanliness		.65			
Quietness		.78			
Privacy		.65			
Security		.72			
Hospitality					
Low price					.64
Flush toilets				.81	
Electricity	.72				
Shower				.83	
Laundromat	.61				
Store	.56				
Water hookups	.83				
Sewer hookups	.82				
Natural surroundings			.74		
Lake/stream			.78		
Hiking trail			.79		
Swimming pool				.53	
Playgrounds			.50		

Note: Only variables whose loadings are greater than .04 are shown.

Table 28. Campground attribute sought factor pattern matrix for "4 factor" principal component analysis with varimax rotation.

Campground Attributes	Factor 0 1	0 2	0 3	0 4
	(Factor Loadings)			
Large sites				
Shaded sites			.42	
Cleanliness			.66	
Quietness			.77	
Privacy			.62	
Security			.73	
Hospitality			.42	
Low price		.51		
Flush toilets		.65		
Electricity	.74			
Shower		.75		
Laundromat	.56	.49		
Store	.46	.54		
Water hookups	.82			
Sewer hookups	.78			
Natural surroundings				.72
Lake/stream				.78
Hiking trail				.77
Swimming pool		.53		
Playgrounds		.48		.44

Note: Only variables whose loadings are greater than .04 are shown.

Table 29. Campground attribute sought factor pattern matrix for "3 factor" principal component analysis with varimax rotation.

Campground Attributes	Factor 0 1	0 2	0 3
	(Factor Loadings)		
Large sites			
Shaded sites			
Cleanliness			.61
Quietness			.78
Privacy			.67
Security			.72
Hospitality			.41
Low price			
Flush toilets	.53		
Electricity		.73	
Shower	.60		
Laundromat		.69	
Store		.60	
Water hookups		.77	
Sewer hookups		.75	
Natural surroundings	.65		
Lake/stream	.61		
Hiking trail	.65		
Swimming pool	.60		
Playgrounds	.65		

Note: Only variables whose loadings are greater than .04 are shown.

Table 30. Campground attribute sought factor pattern matrix for "2 factor" principal component analysis with varimax rotation.

Campground Attributes	Factor 01 (Factor Loadings)	Factor 02
Large sites		
Shaded sites	.52	
Cleanliness	.56	
Quietness	.45	
Privacy	.44	
Security	.50	
Hospitality	.52	
Low price		
Flush toilets	.46	
Electricity		.74
Shower	.48	
Laundromat		.70
Store		.61
Water hookups		.79
Sewer hookups		.77
Natural surroundings	.71	
Lake/stream	.67	
Hiking trail	.70	
Swimming pool		
Playgrounds	.48	

Note: Only variables whose loadings are greater than .04 are shown.

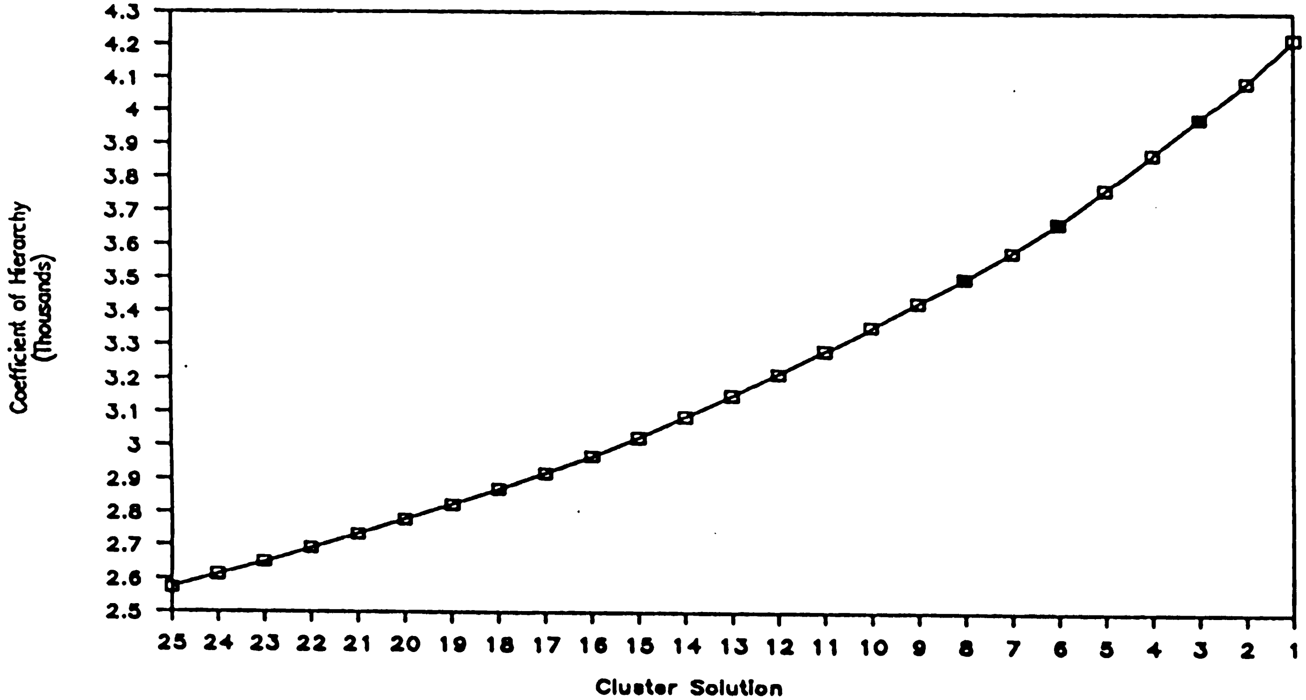
rotation). Only one variable was significantly loaded on each of the 20 factors.

Tables 12 through 30 reveal two major changes as the number of factors are reduced from 20 to 2. First, the size of factor loadings change. Second, certain factors will have two or more variables with significant (>.40) loadings. Changes in factor loadings and the number of variables with significant loadings on different factors result in different factor interpretation and different factor scores. When factor scores are used as the basis for clustering process, the clustering results (cluster membership and cluster description) would be different for different factor solutions (20, 19,..., 2).

Clustering Results

Factor scores were computed for each factor in each of the 19 different principal component analyses. The regression estimates method was used to obtain the factor scores. The original raw data measurements were multiplied by the corresponding factor score (regression) coefficients. The factor scores were used as the basis for clustering.

The factor scores from the "20 factor" principal component analysis were used as input data to Ward's clustering method with the squared Euclidean distance as the distance measure. Figure 5 shows the increase in the coefficient of hierarchy (which resulted from fusion of clusters) plotted against the number of clusters. As stated previously, the break points along the plot mean that a relatively large loss of information resulted from the fusion of two clusters. Based on the



Cluster Solution	Coefficient of Hierarchy
25	2572
24	2611
23	2650
22	2691
21	2734
20	2778
19	2823
18	2869
17	2917
16	2968
15	3025
14	3087
13	3150
12	3214
11	3284
10	3354
9	3426
8	3498
7	3577
6	3665
5	3766
4	3870
3	3979
2	4090
1	4220

Figure 5. Coefficient of hierarchy by number of attribute sought clusters when clustering is based on factor scores (■: candidate solution).

coefficient of hierarchy and the examination of plot slopes, three candidate cluster solutions were identified: eight clusters, six clusters, and three clusters.

The three candidate solutions were evaluated on (a) the significance of inter-cluster differences and (b) the size of clusters. ANOVA was used to test for inter-cluster differences. The results of the ANOVA tests on the three candidate cluster solutions are presented in Table 31-33. In the eight-cluster solution (Table 31), there were significant differences across clusters on all but two (flush toilet and campground store) of the 20 factors/variables. The six clusters differed significantly on 16 of the 20 factors/variables (Table 32). The three-cluster solution showed the least amount of inter-cluster differences (Table 33); clusters differed significantly on only 10 of the 20 factors/variables.

Even though the eight cluster solution exhibited more inter-cluster differences. The six-cluster solution was selected as the final solution because one of the 8 clusters was disproportionally small; it only had 5 (2.4%) cases (see Table 34). In the six cluster solution, the smallest cluster contained 16 (7.5%) cases.

Factor Score Pattern

The (factor score) centroids for each of the six clusters were calculated for each of the 19 principal component analyses (20, 19, 18, ..., 2). The (factor score) centroids are graphically presented in Figures 6-25. Each graph shows the factor centroids for each cluster

Table 31. Mean attribute sought factor scores for the eight-cluster candidate solution when clustering on factor scores.

	Cluster								
Factor	1	2	3	4	5	6	7	8	F-ratio
Electricity	-.48	.06	-.49	.54	.31	.13	-.17	-.01	3.72*
Toilet	.02	.22	.15	-.20	.36	-.21	-.21	-.69	1.67
Playground	.36	.17	.33	-.14	.23	.16	-2.29	.47	24.21*
Price	.12	-.22	.23	.81	-.16	-.31	-.21	-.18	4.37*
Large sites	-.46	-.02	.10	.53	.14	-.24	-.05	.82	2.98*
Shade sites	-.12	.83	-.51	-.11	-.13	-.32	-.43	1.26	10.39*
Pool	-.74	.02	.07	.50	-.31	.43	-.70	-.09	6.57*
Hospitality	.30	.26	.30	-.68	.28	-.28	-.14	-.01	4.10*
Security	-.33	.29	-.10	-.39	.09	.31	-.19	-.76	2.87*
Privacy	.07	.09	-.32	.63	-.00	-.37	-.25	1.42	5.05*
Natural surr.	.45	.35	-.37	-.33	.41	-.27	.09	-.93	4.56*
Lake/stream	.31	.03	-.72	.08	-.33	.48	-.05	-1.03	5.85*
Cleanliness	.00	-.45	-.27	-.29	1.78	.01	-.08	.69	16.69*
Laundromat	.64	-.41	.01	.16	.15	-.24	-.04	1.40	5.13*
Quietness	.14	-.05	-.15	.47	.62	-.26	-.15	-1.53	4.80*
Sewer hookups	.49	.26	-.14	.23	.39	-.45	-.22	-1.92	7.34*
Natural trail	.59	-.22	-1.16	.09	.41	.47	-.21	.28	12.74*
Store	.35	-.21	.22	-.25	-.27	.24	-.11	-.54	2.02
Water hookups	-.92	.09	.18	.19	.54	-.11	.17	.11	4.82*
Shower	-.03	-.48	.36	.18	-.09	.31	-.18	-.41	3.25*

* Significant at .05 level.

Table 32. Mean attribute sought factor scores for the six-cluster candidate solution when clustering on factor scores.

	Cluster						
Factor	1	2	3	4	5	6	F-ratio
Electricity	-.14	.06	-.49	.45	.13	-.17	3.37*
Toilet	.17	.22	.15	-.27	-.21	-.21	1.87
Playground	.03	.17	.33	-.05	.16	-2.29	33.08*
Price	.00	-.22	.23	.65	-.31	-.21	4.92*
Large sites	-.20	-.02	.10	.58	-.24	-.05	3.22*
Shade sites	-.12	.83	-.51	.10	-.32	-.43	11.98*
Swimming pool	-.56	.02	.07	.41	.43	-.70	8.31*
Hospitality	.29	.26	.30	-.57	-.28	-.14	5.32*
Security	-.15	.29	-.10	-.45	.31	-.19	3.48*
Privacy	.04	.09	-.32	.75	-.37	-.25	6.42*
Natural surr.	.43	.35	-.37	-.43	-.27	.09	6.05*
Lake/stream	.04	.03	-.72	-.09	.48	-.05	5.70*
Cleanliness	.77	-.45	-.27	-.14	.01	-.08	9.19*
Laundromat	.43	-.41	.01	.35	-.24	-.04	4.94*
Quietness	.35	-.05	-.15	.16	-.26	-.15	2.11
Sewer hookups	.45	.26	-.14	-.11	-.45	-.22	5.00*
Natural trail	.51	-.22	-1.16	.12	.46	-.21	17.80*
Store	.08	-.21	.22	-.30	.24	-.11	1.88
Water hookups	-.29	.09	.18	.18	-.11	.17	1.40
Shower	-.06	-.48	.36	.09	.31	-.16	4.24*

* Significant at .05 level.

Table 33. Mean attribute sought factor scores for the three-cluster candidate solution when clustering on factor scores.

	Cluster			
Factor	1	2	3	F-ratio
Electricity	-.14	.06	.03	0.58
Toilet	.17	.22	-.14	3.01
Playground	.03	.17	-.17	4.71*
Price	.00	-.22	.08	1.59
Large sites	-.20	-.02	.08	1.29
Shade sites	-.12	.83	-.27	25.12*
Pool	-.56	.02	.19	9.82*
Hospitality	.29	.26	-.20	6.30*
Security	-.15	.29	-.05	2.61
Privacy	.04	.09	-.05	0.35
Natural surr.	.43	.35	-.29	13.30*
Lake/stream	.04	.03	-.02	0.09
Cleanliness	.77	-.45	-.11	22.21*
Laundromat	.43	-.41	.00	8.53*
Quietness	.35	-.05	-.11	3.53*
Sewer hookups	.45	.26	-.26	10.93*
Natural trail	.51	-.22	-.10	8.05*
Store	.08	-.21	.05	1.27
Water hookups	-.29	.09	.07	2.35
Shower	-.06	-.48	.20	8.53

* Significant at .05 level.

Table 34. Number of respondents in each of the cluster candidate solutions when clustering on factor scores.

Cluster	Number of Respondents	Relative Size (percent)
Eight Cluster Solution		
1	25	11.8
2	46	21.7
3	29	13.7
4	27	12.7
5	19	9.0
6	45	21.2
7	16	7.5
8	5	2.4
Total	212	100.0
Six Cluster Solution		
1	44	20.8
2	46	21.7
3	29	13.7
4	32	15.1
5	45	21.2
6	16	7.5
Total	212	100.0
Three Cluster Solution		
1	44	20.8
2	46	21.7
3	122	57.5
Total	212	100.0

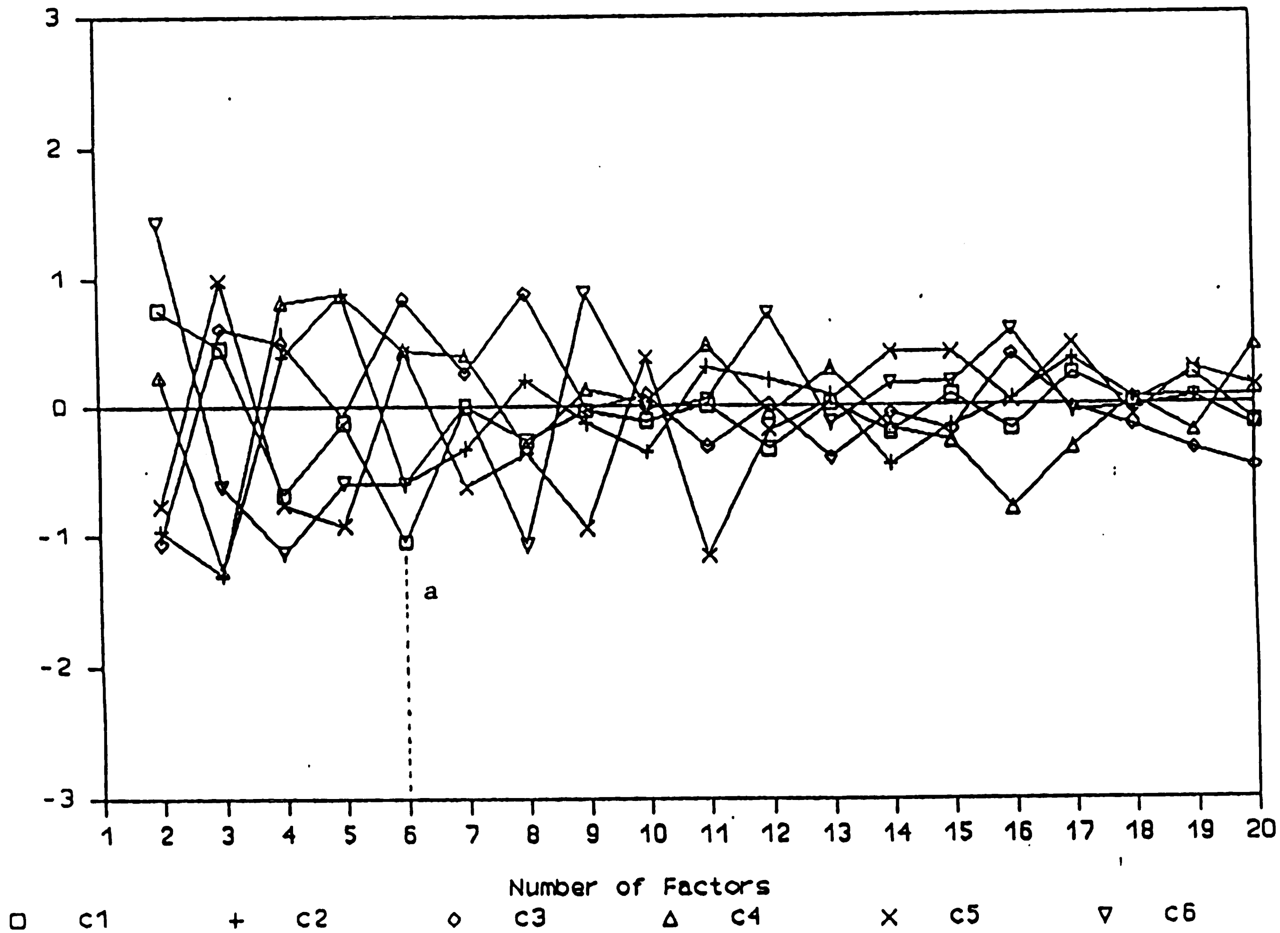


Figure 6. The "factor 1" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6, a: factor 1 factor score centroid of cluster 1 for the "6 factor" factor solution).

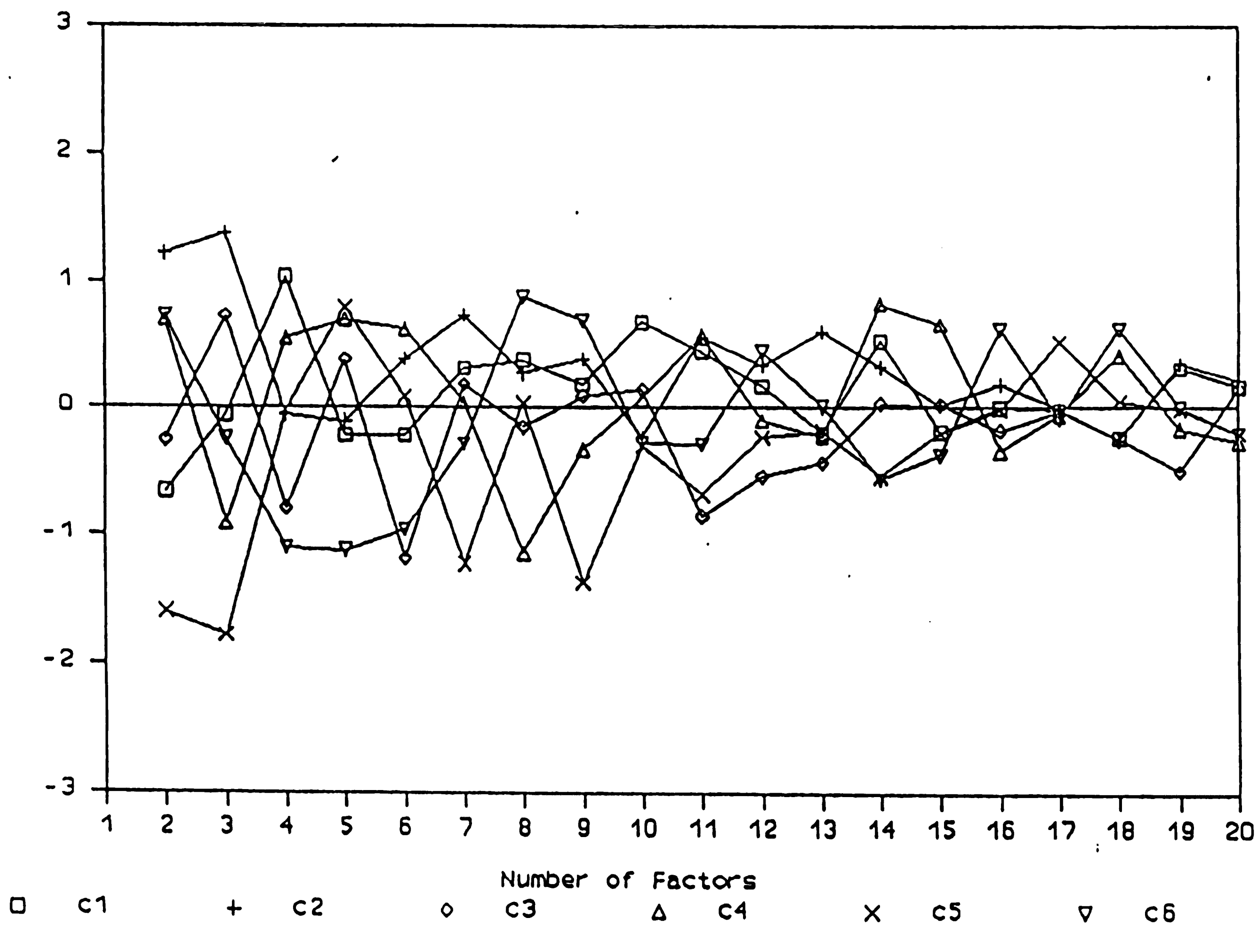


Figure 7. The "factor 2" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

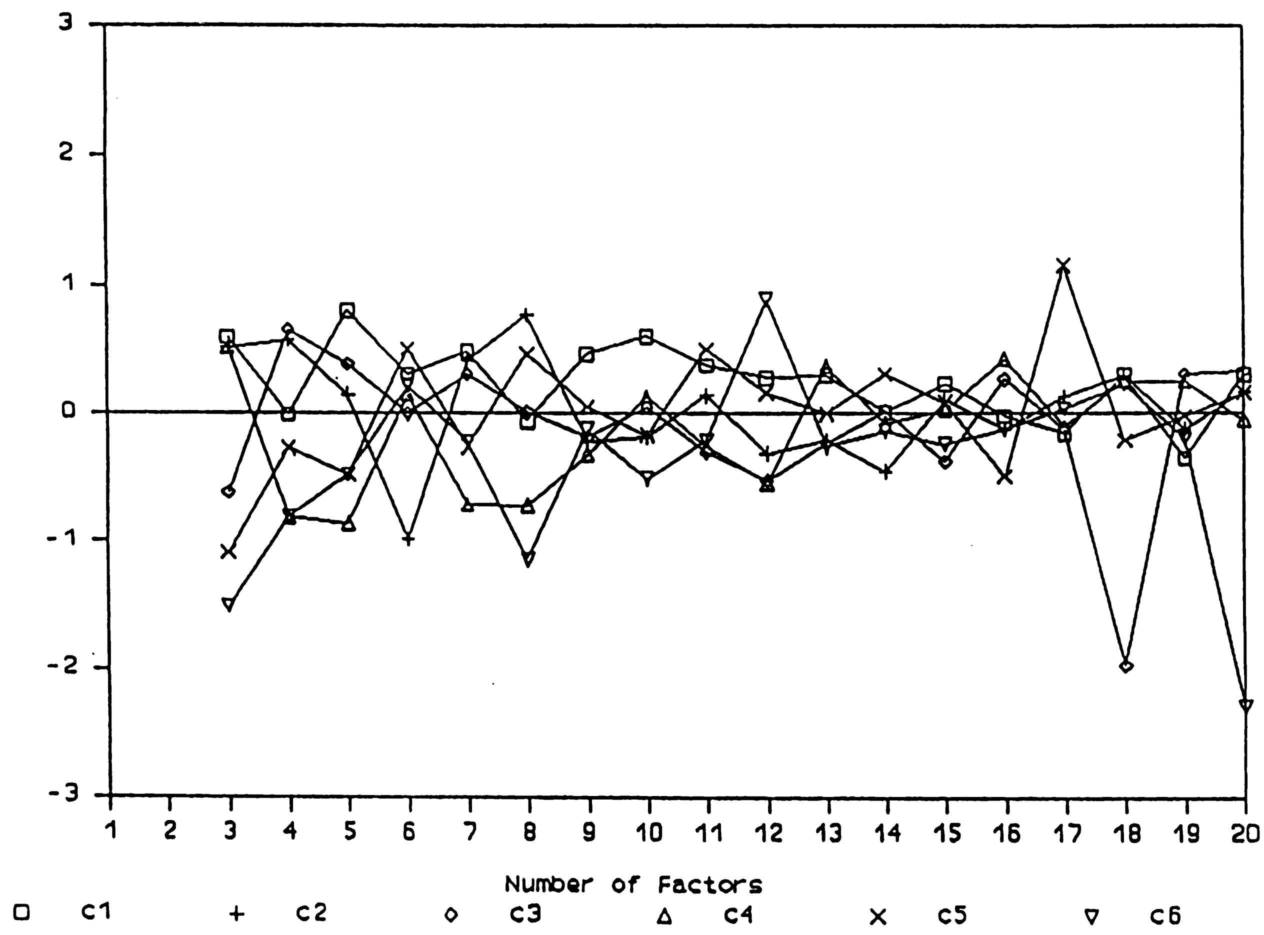


Figure 8. The "factor 3" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

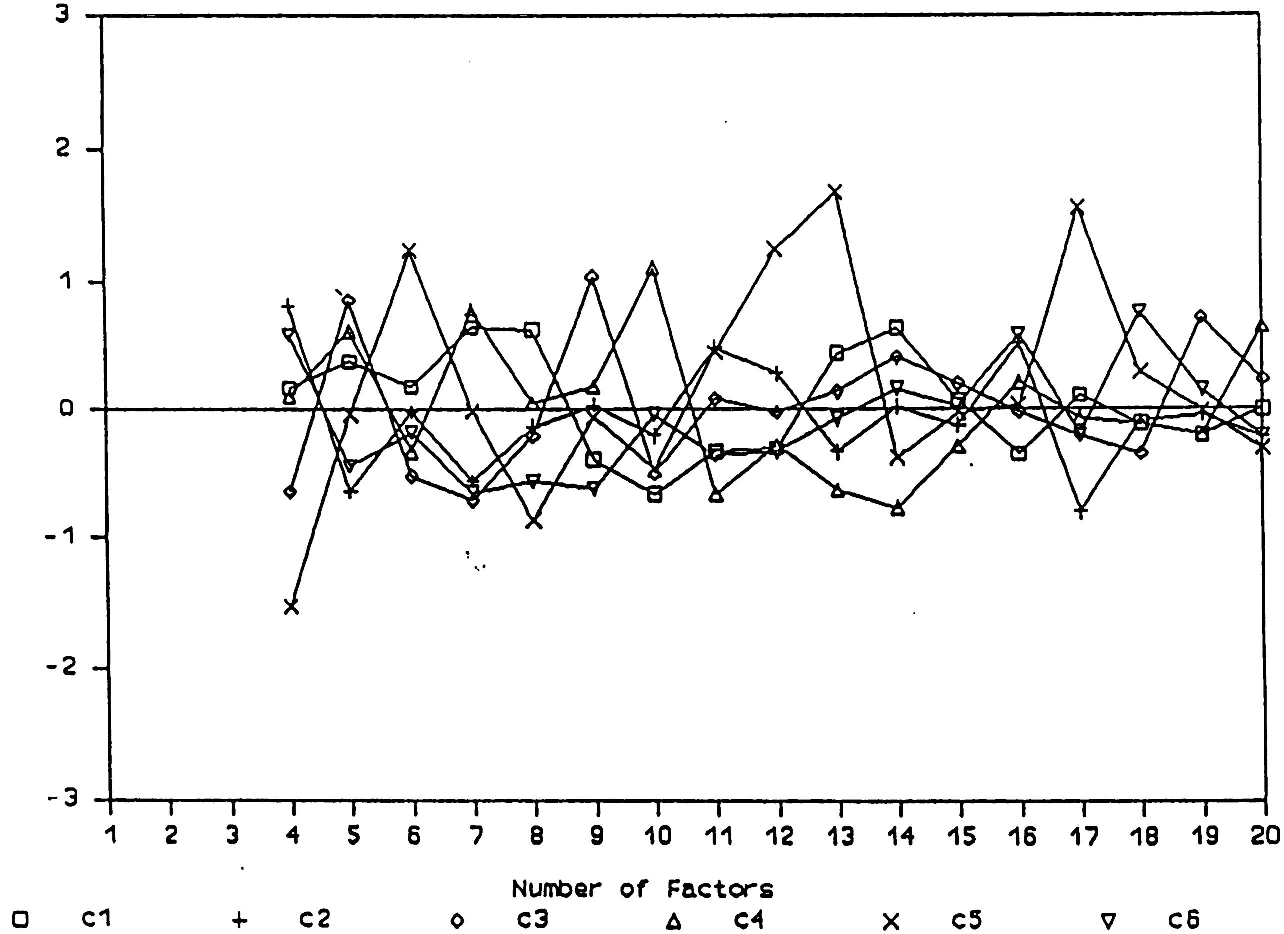


Figure 9. The "factor 4" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

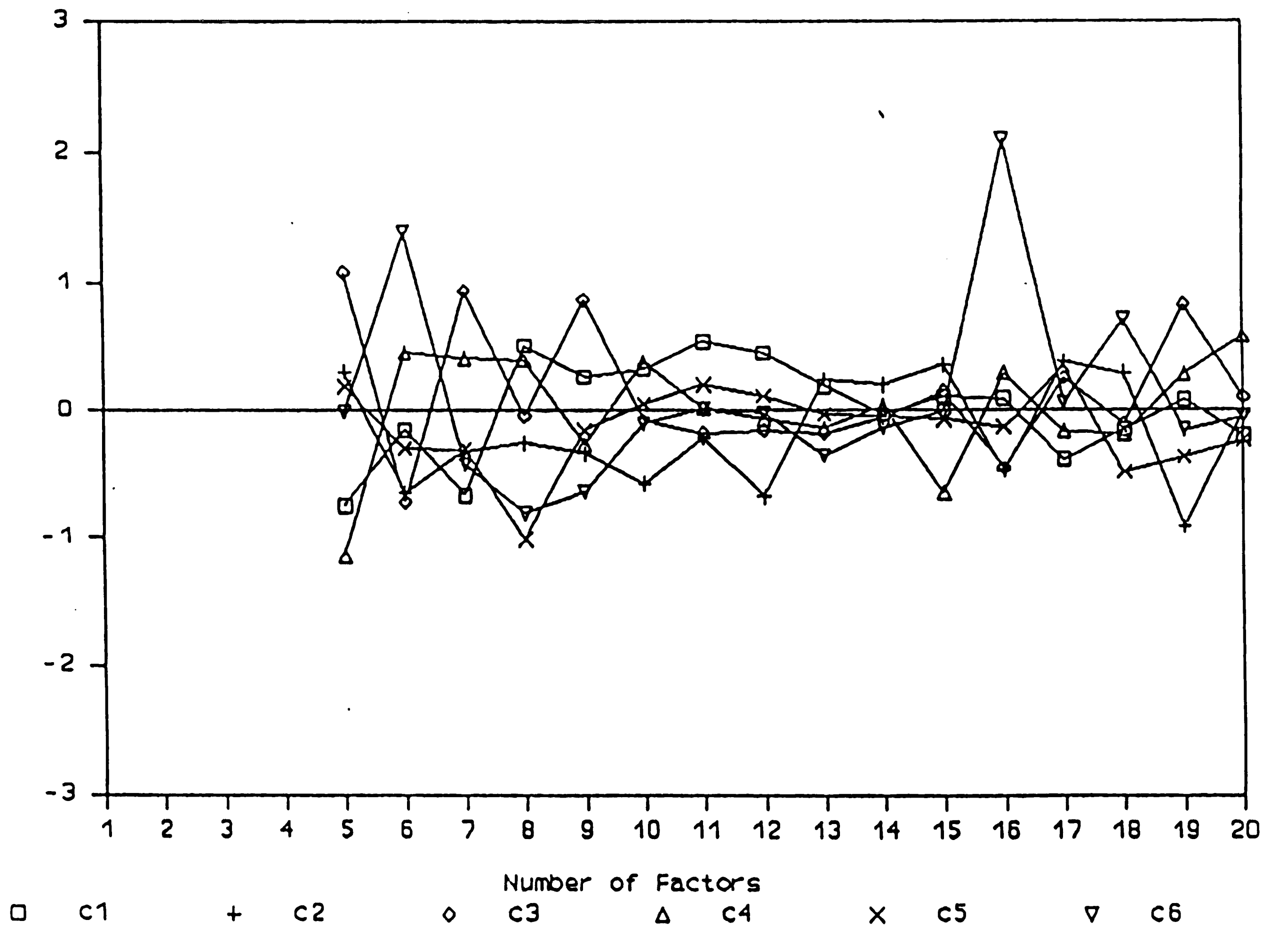


Figure 10. The "factor 5" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

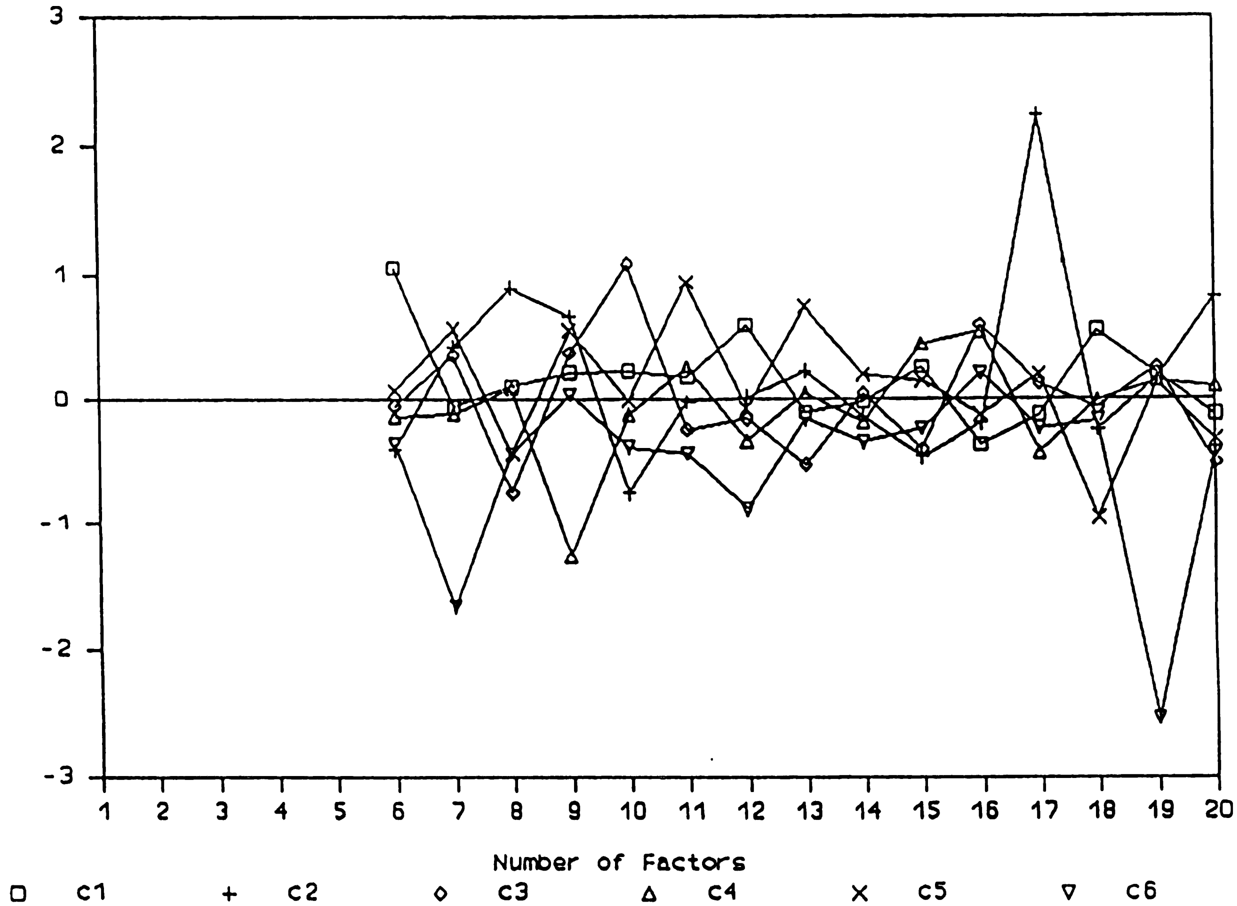


Figure 11. The "factor 6" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

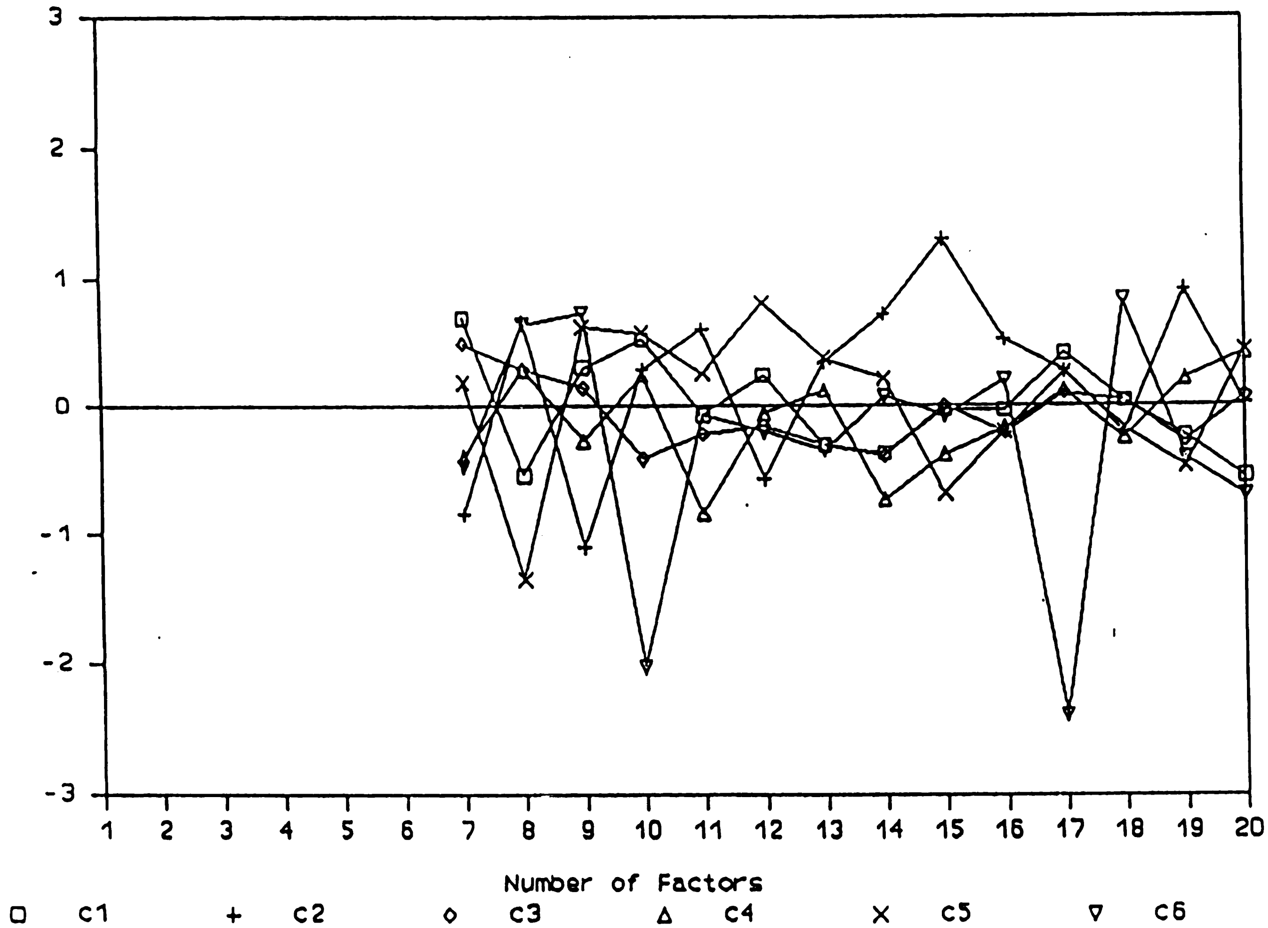


Figure 12. The "factor 7" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

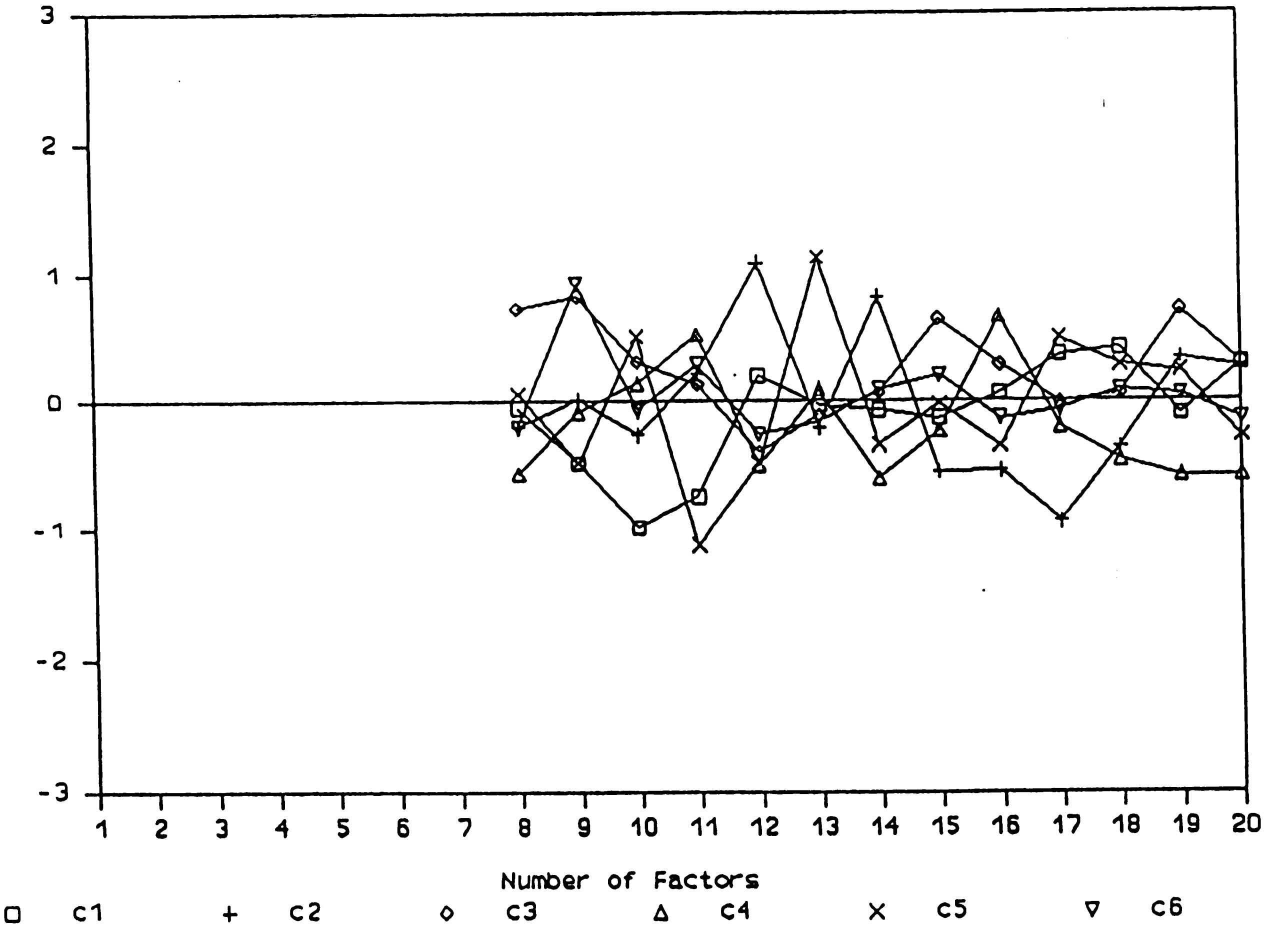


Figure 13. The "factor 8" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

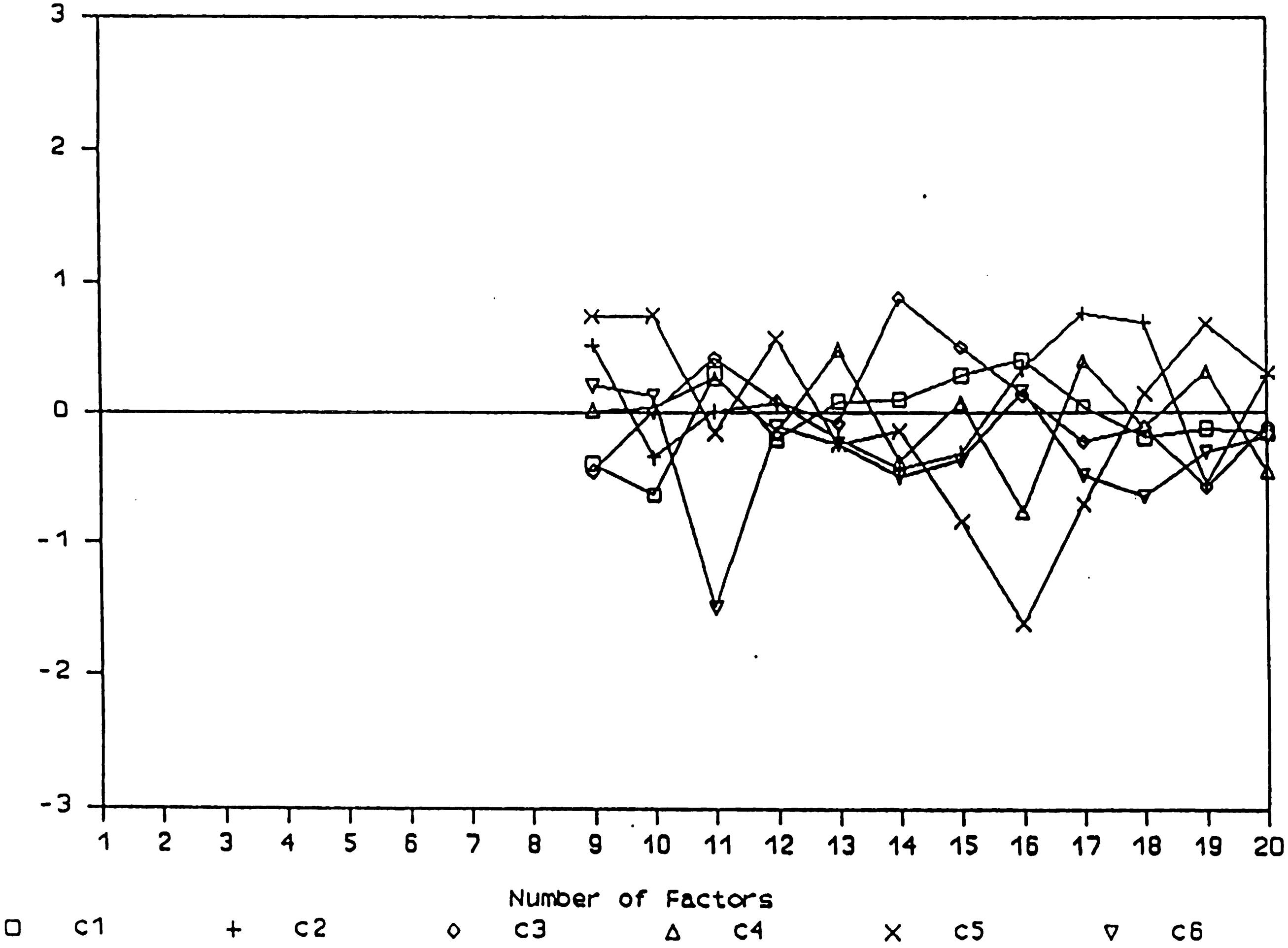


Figure 14. The "factor 9" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

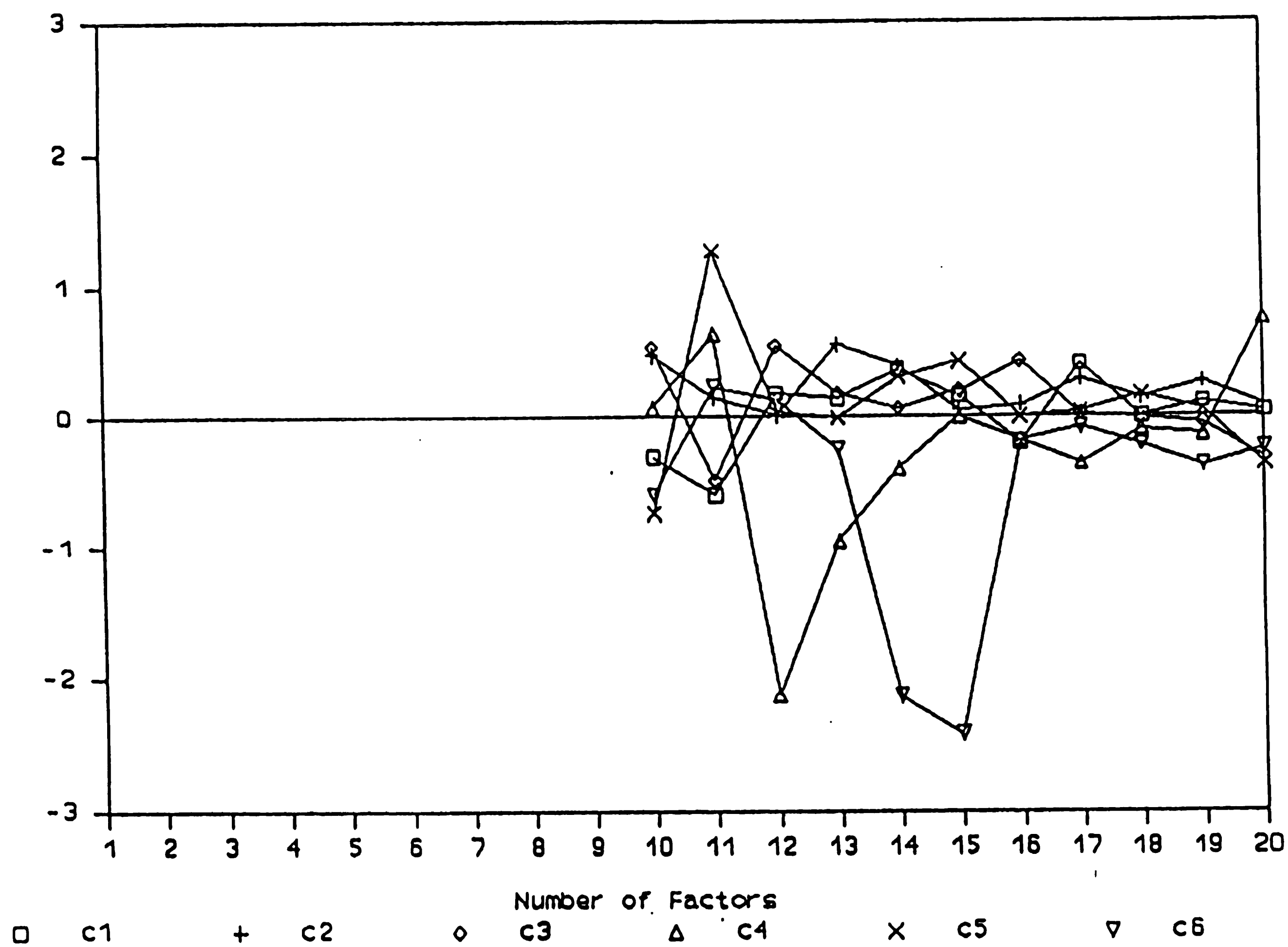


Figure 15. The "factor 10" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

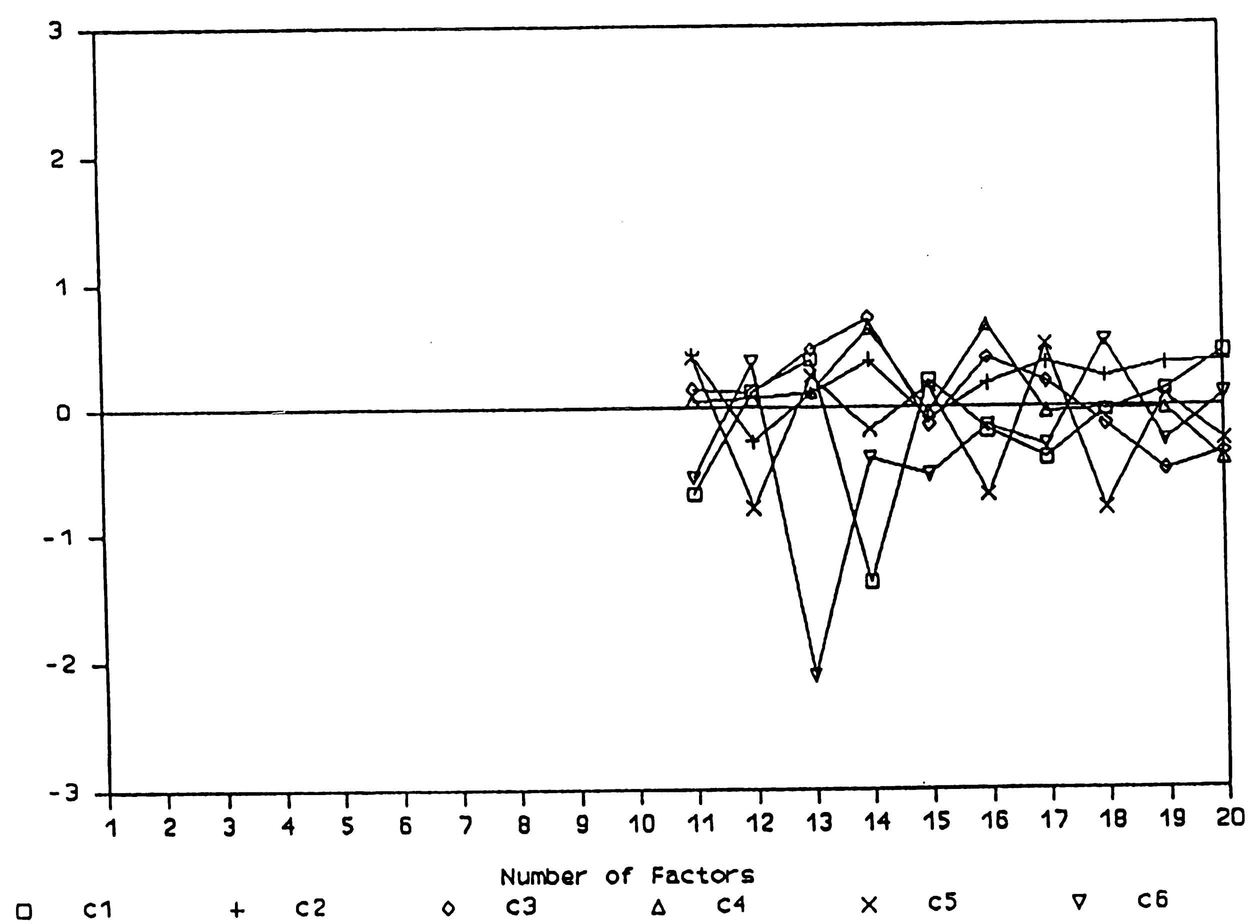


Figure 16. The "factor 11" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

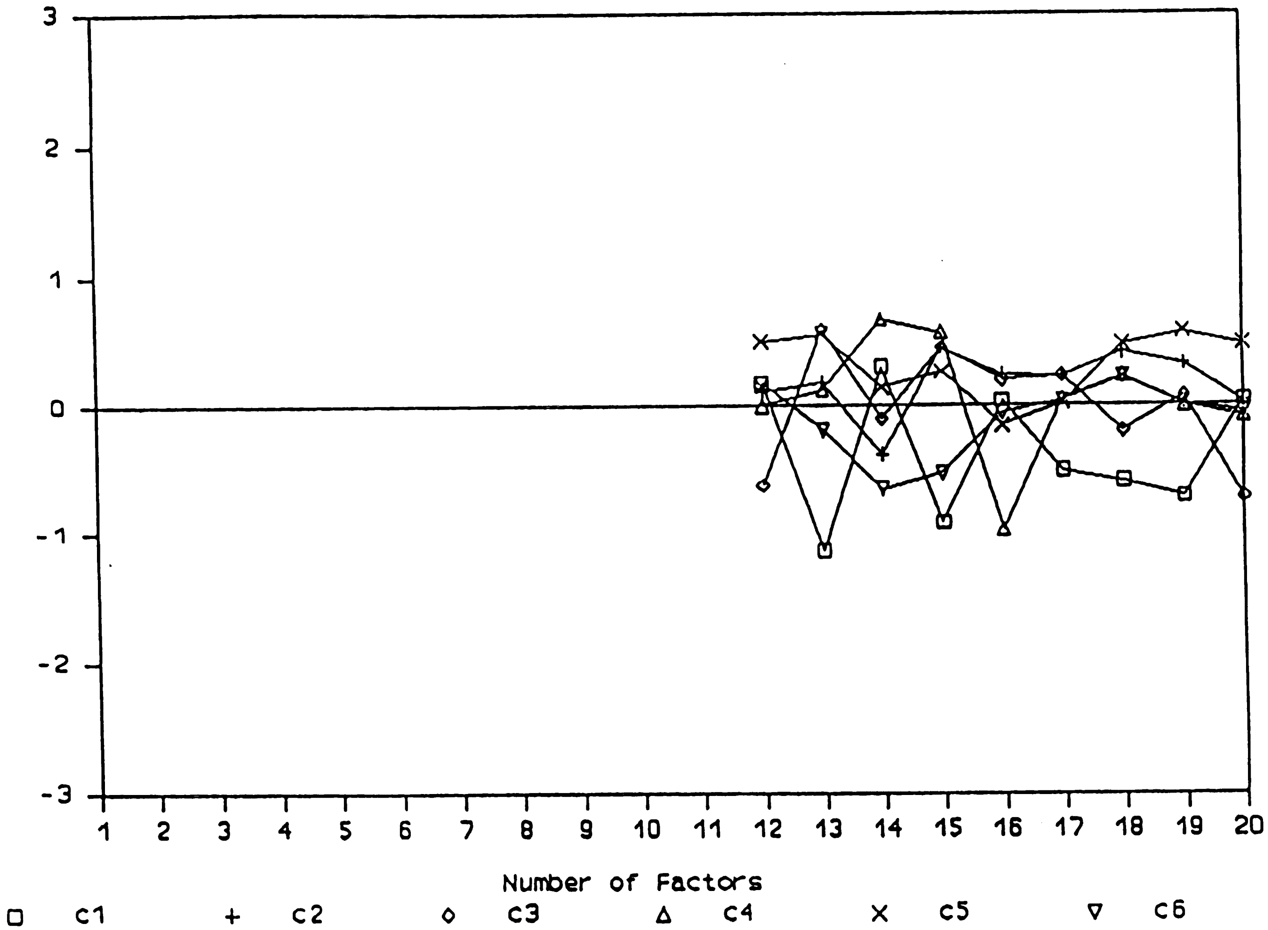


Figure 17. The "factor 12" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

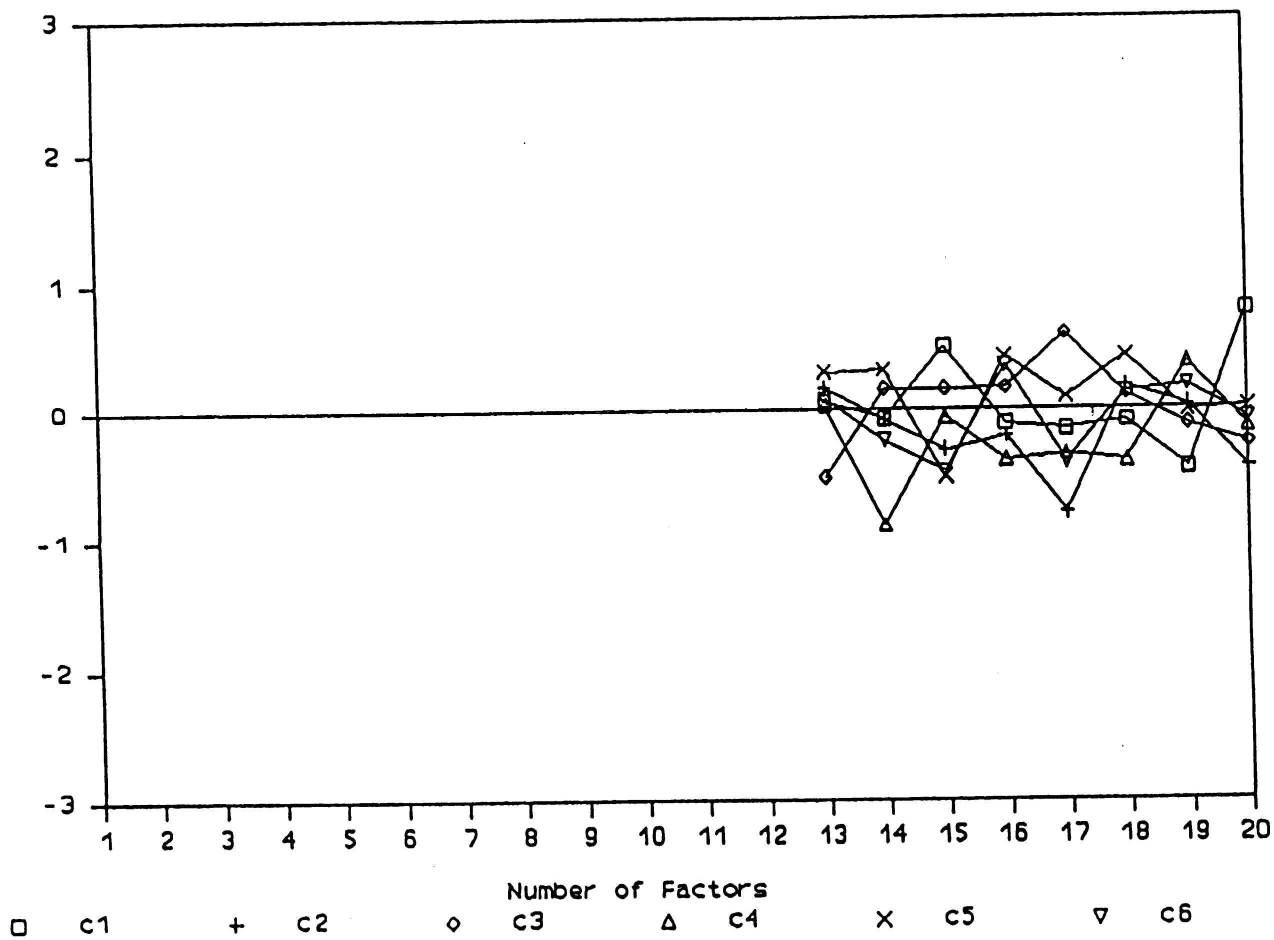


Figure 18. The "factor 13" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

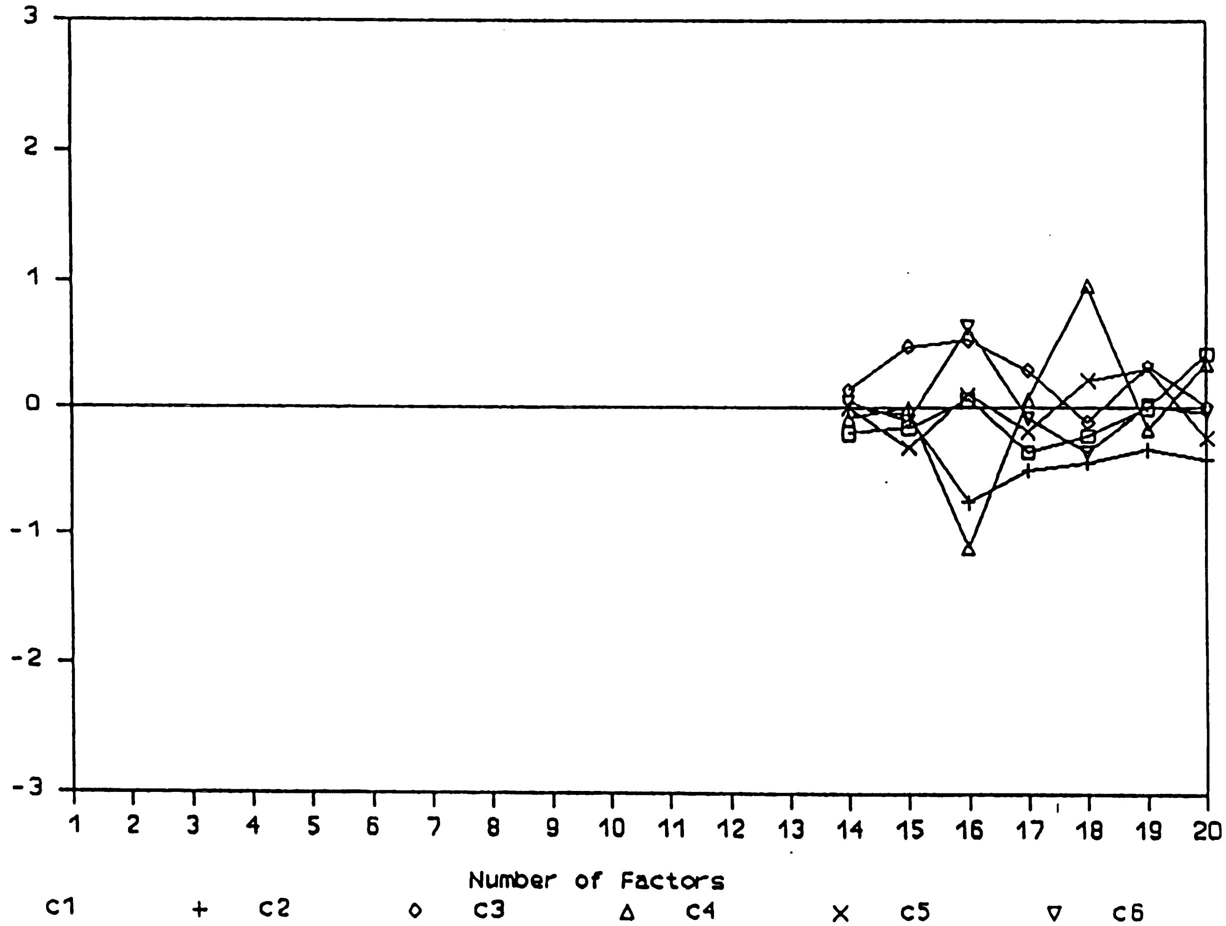


Figure 19. The "factor 14" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster_2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

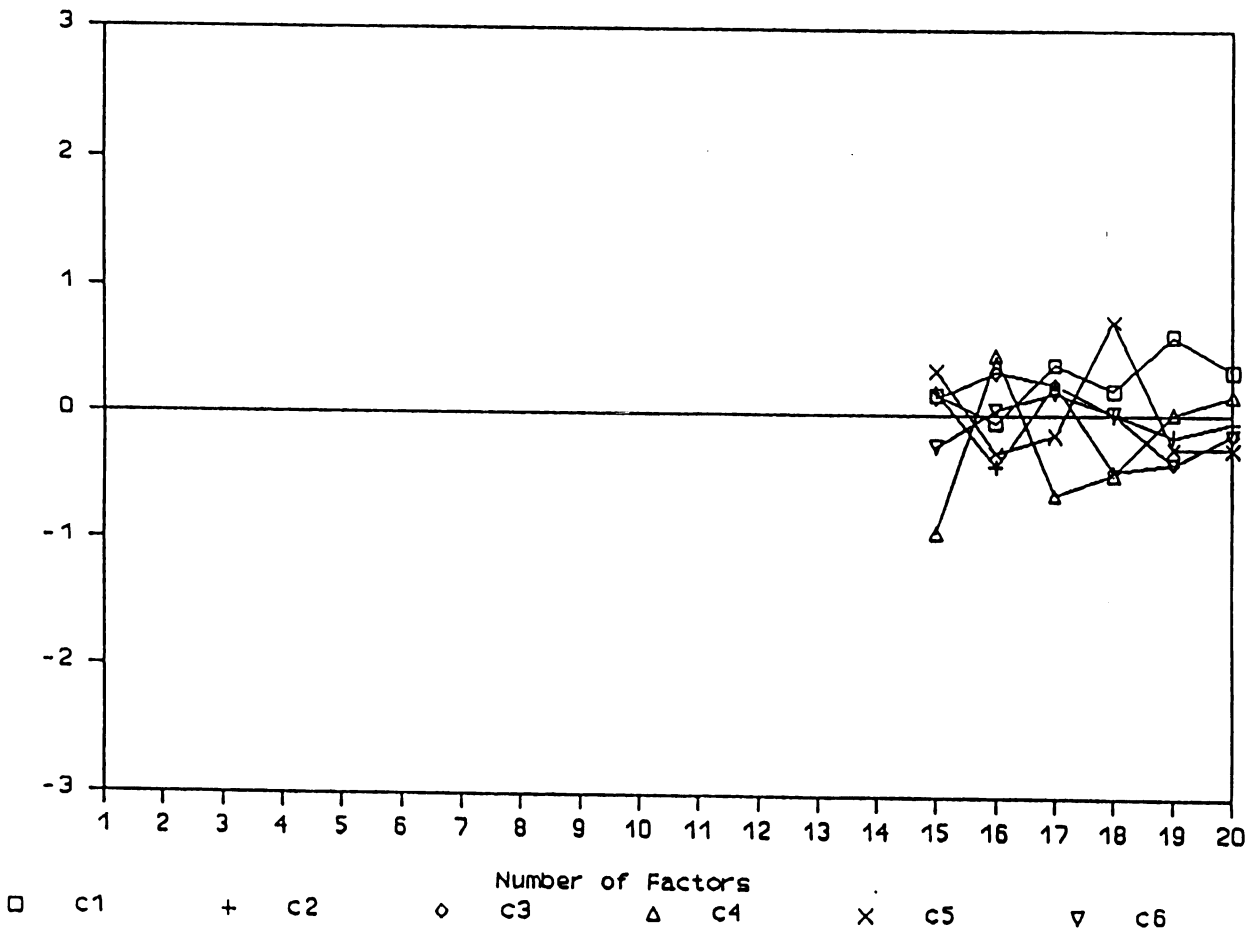


Figure 20. The "factor 15" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

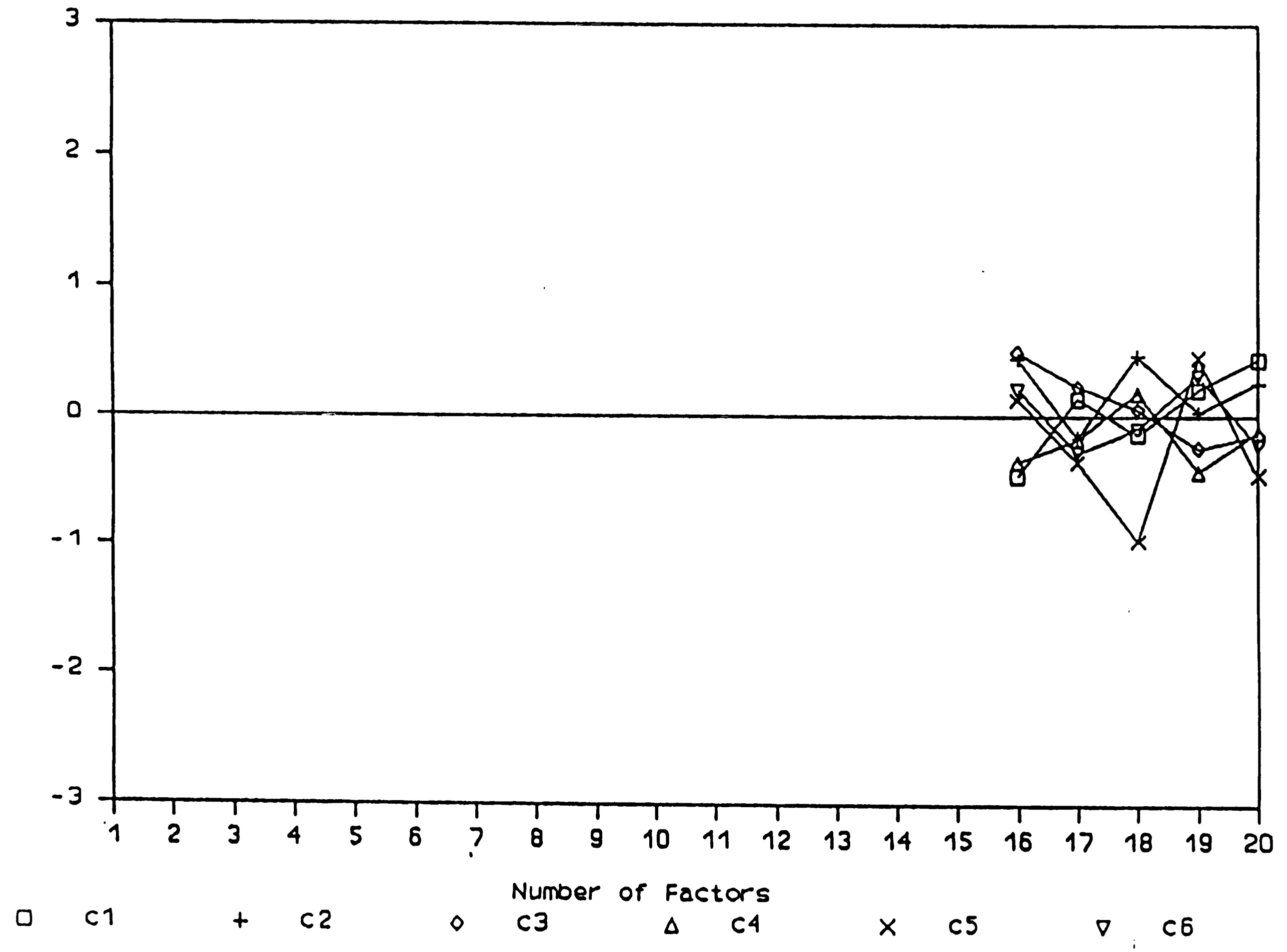


Figure 21. The "factor 16" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

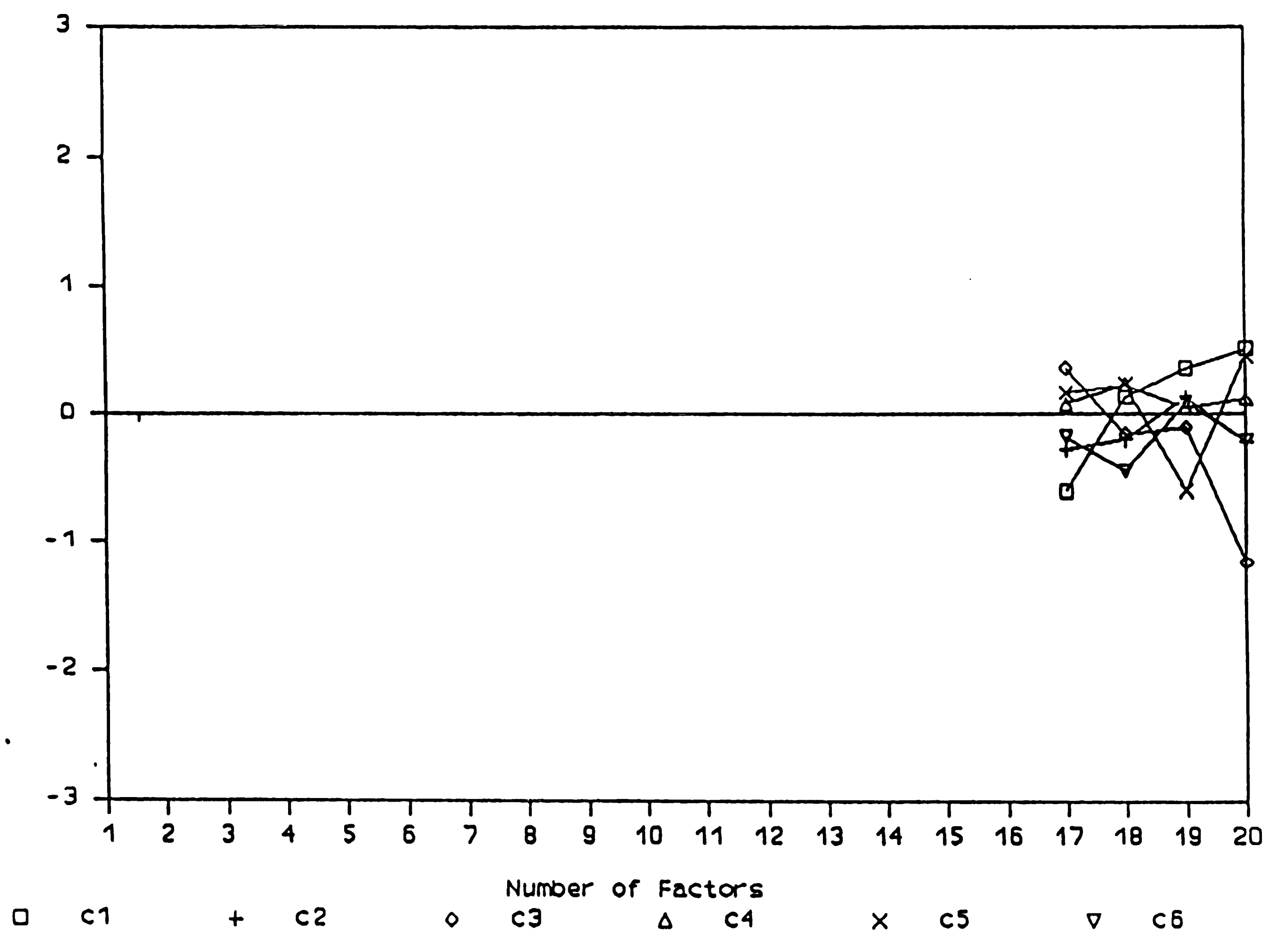


Figure 22. The "factor 17" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

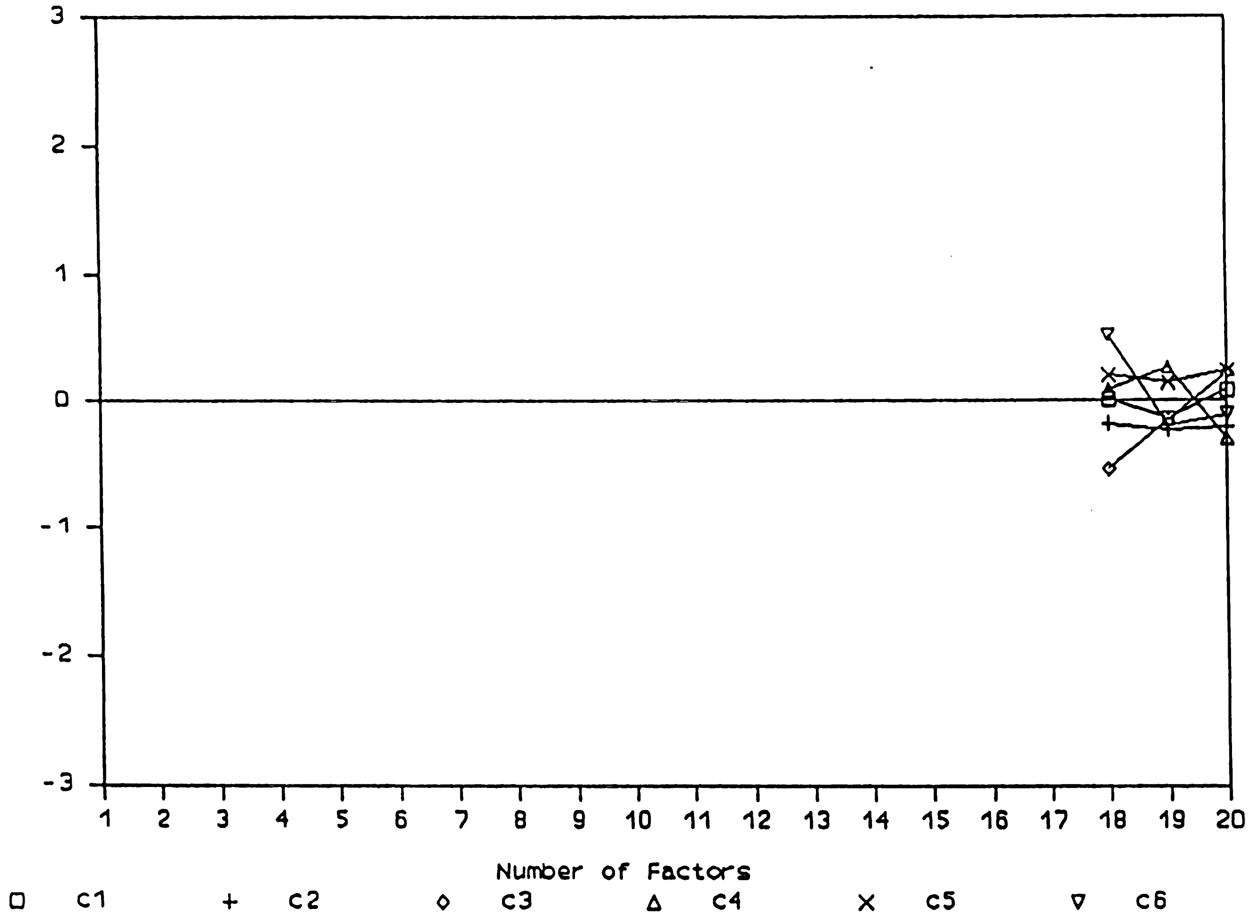


Figure 23. The "factor 18" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

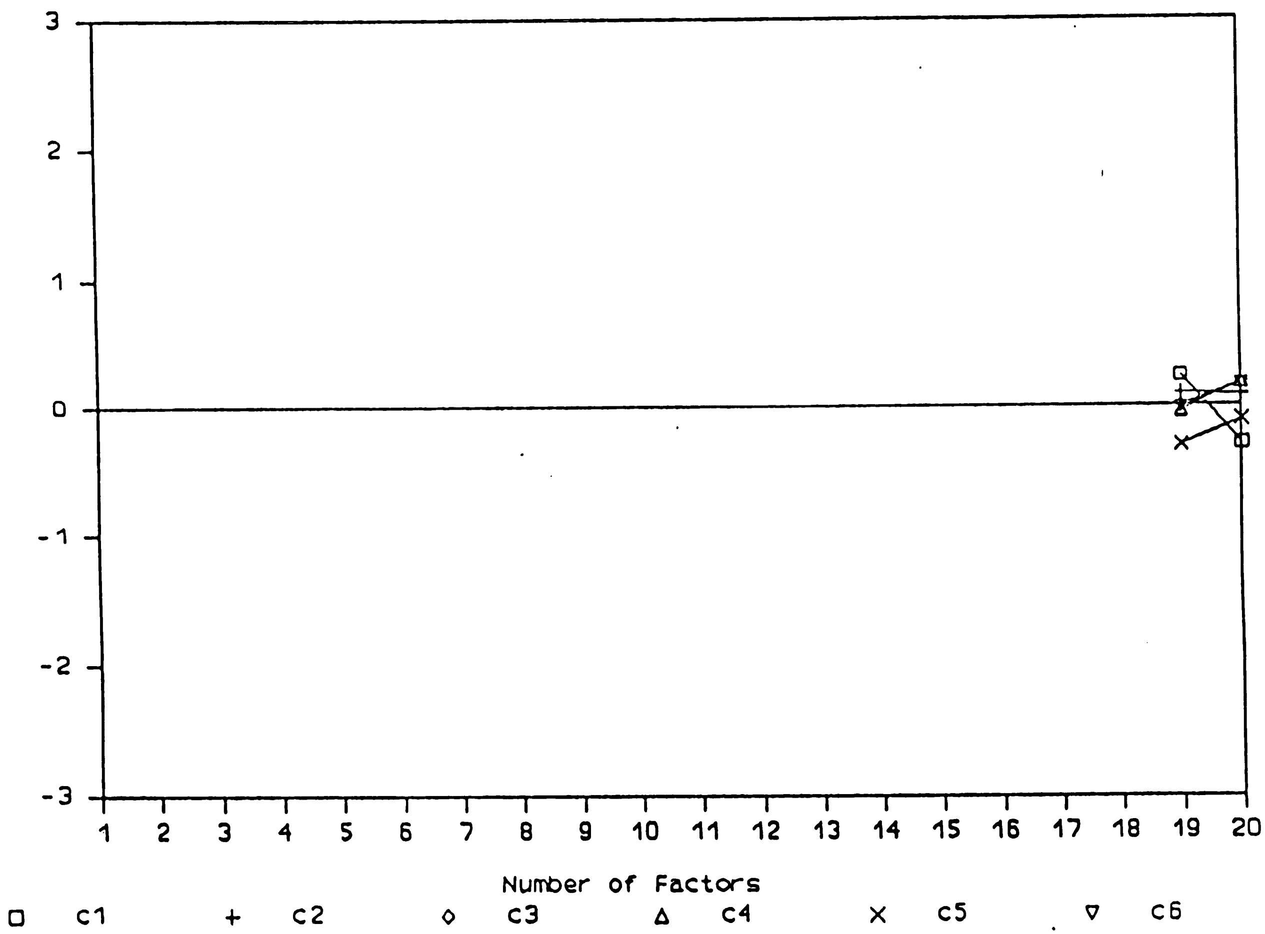


Figure 24. The "factor 19" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

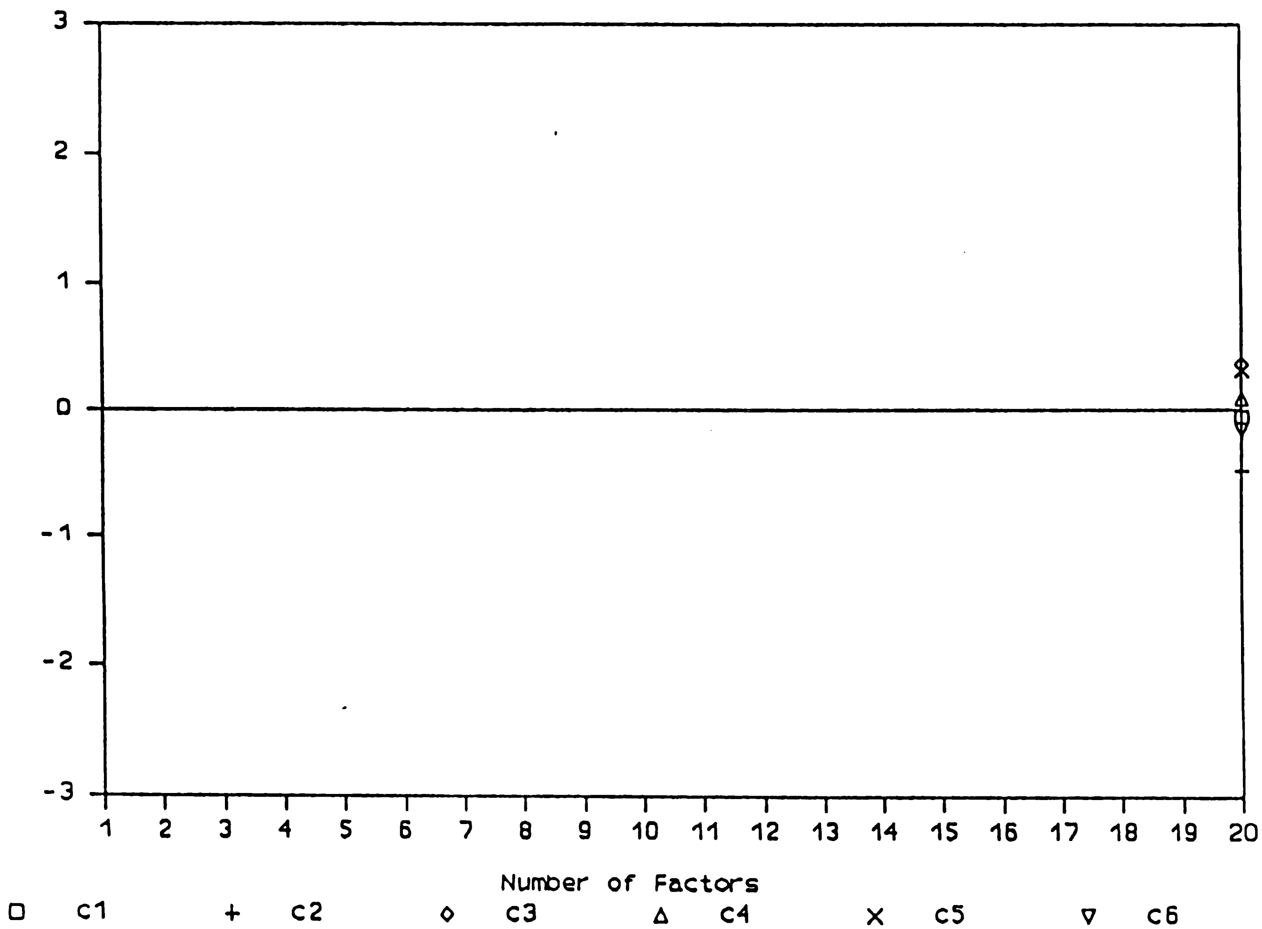


Figure 25. The "factor 20" factor score centroids for six clusters across the different factor solutions when clustering on factor scores (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

for each factor solution. For example, Figure 6 shows "factor 1" factor score centroids for the six clusters across different factor solutions.

The graphs show that factor score centroids differ markedly across the different factor solutions. In Figure 6, "factor 1" factor score centroid for cluster 1 changes significantly across the 19 different factor solutions. The same is true for the centroids of the other five clusters. Figures 7 ("factor 2" factor score centroids) to 24 ("factor 19" factor score centroids) show similar instability of factor score centroids across factor (20, 19, 18, ..., 2) solutions. In Figure 7, the "factor 2" factor score centroid for cluster 1 changes across the different factor solutions. The results indicate that when clustering on factor scores different factor solutions yield very different clustering results in terms of cluster membership and cluster description.

Comparison Of Cluster Membership

As described in Chapter III, a crosstabulation technique and entropy (information) measure was employed to assess the effects of alternative factor solutions on cluster membership. Tables 35 to 53 show the crosstabulation of cluster membership. In each table, the "20 factor" factor solution serves as the basis for (cluster membership) comparison. Crosstabulations serve two primary functions. First, the crosstabulations show the percentage of cases assigned to the same cluster numbering (e.g., cluster 1) in two different clustering analyses each based on factor scores from a different factoring solution (e.g., "20 factor" factor solution vs. "19 factor" factor solution). For

Table 35. Cluster membership crosstabulation of the 20-factor solution and the 20-factor solution.

20-Factor Solution	20-Factor Solution (percent) 1	2	3	4	5	6
1	100.0	0.0	0.0	0.0	0.0	0.0
2	0.0	100.0	0.0	0.0	0.0	0.0
3	0.0	0.0	100.0	0.0	0.0	0.0
4	0.0	0.0	0.0	100.0	0.0	0.0
5	0.0	0.0	0.0	0.0	100.0	0.0
6	0.0	0.0	0.0	0.0	0.0	100.0

Table 36. Cluster membership crosstabulation of the 20-factor solution and the 19-factor solution.

20-Factor Solution	19-Factor Solution (percent) 1	2	3	4	5	6
1	68.2	11.4	4.5	4.5	11.4	0.0
2	6.5	45.7	6.5	28.3	13.0	0.0
3	31.0	0.0	17.2	41.4	10.3	0.0
4	0.0	9.4	46.9	25.0	12.5	6.3
5	0.0	2.2	6.7	48.9	37.8	4.4
6	18.8	6.3	0.0	12.5	0.0	62.5

Table 37. Cluster membership crosstabulation of the 20-factor solution and the 18-factor solution.

20-Factor Solution	18-Factor Solution (percent) 1	2	3	4	5	6
1	40.9	13.6	0.0	36.4	4.5	4.5
2	30.4	26.1	10.9	6.5	2.2	23.9
3	34.5	10.3	0.0	31.0	20.7	3.4
4	50.0	3.1	6.3	3.1	12.5	25.0
5	8.9	48.9	4.4	24.4	11.1	2.2
6	6.3	0.0	81.3	0.0	12.5	0.0

Table 38. Cluster membership crosstabulation of the 20-factor solution and the 17-factor solution.

20-Factor Solution	17-Factor Solution (percent) 1	2	3	4	5	6
1	34.1	4.5	29.5	9.1	22.7	0.0
2	17.4	17.4	23.9	39.1	2.2	0.0
3	24.1	0.0	20.7	51.7	3.4	0.0
4	9.4	0.0	37.5	31.3	12.5	9.4
5	22.2	0.0	55.6	17.8	0.0	4.4
6	0.0	0.0	6.3	25.0	0.0	68.8

Table 39. Cluster membership crosstabulation of the 20-factor solution and the 16-factor solution.

20-Factor Solution	16-Factor Solution 1	2	3 (percent)	4	5	6
1	54.5	6.8	9.1	0.0	0.0	29.5
2	23.9	19.6	34.8	2.2	19.6	0.0
3	62.1	17.2	10.3	6.9	3.4	0.0
4	12.5	31.3	37.5	15.6	3.1	0.0
5	51.1	11.1	22.2	13.3	2.2	0.0
6	6.3	0.0	0.0	12.5	81.3	0.0

Table 40. Cluster membership crosstabulation of the 20-factor solution and the 15-factor solution.

20-Factor Solution	15-Factor Solution 1	2	3 (percent)	4	5	6
1	56.8	18.2	4.5	4.5	15.9	0.0
2	32.6	19.6	13.0	13.0	21.7	0.0
3	24.1	24.1	17.2	6.9	27.6	0.0
4	15.6	3.1	56.3	9.4	9.4	6.3
5	8.9	8.9	31.1	40.0	11.1	0.0
6	12.5	0.0	0.0	6.3	6.3	75.0

Table 41. Cluster membership crosstabulation of the 20-factor solution and the 14-factor solution.

20-Factor Solution	14-Factor Solution (percent) 1	2	3	4	5	6
1	29.5	11.4	18.2	4.5	36.4	0.0
2	17.4	21.7	8.7	13.0	39.1	0.0
3	0.0	6.9	27.6	0.0	55.2	10.3
4	3.1	18.8	31.3	9.4	31.3	6.3
5	13.3	33.3	31.3	17.8	4.4	0.0
6	0.0	0.0	0.0	6.3	12.5	81.3

Table 42. Cluster membership crosstabulation of the 20-factor solution and the 13-factor solution.

20-Factor Solution	13-Factor Solution (percent) 1	2	3	4	5	6
1	29.5	31.8	6.8	15.9	15.9	0.0
2	19.6	30.4	10.9	30.4	4.3	4.3
3	13.8	13.8	41.4	27.6	3.4	0.0
4	12.5	37.5	9.4	18.8	15.6	6.3
5	15.6	33.3	33.3	8.9	0.0	8.9
6	0.0	0.0	0.0	12.5	0.0	87.5

Table 43. Cluster membership crosstabulation of the 20-factor solution and the 12-factor solution.

20-Factor Solution	12-Factor Solution (percent) 1	2	3	4	5	6
1	50.0	11.4	18.2	0.0	18.2	2.3
2	26.1	39.1	26.1	0.0	2.2	6.5
3	27.6	0.0	51.7	3.4	3.4	13.8
4	12.5	15.6	15.6	9.4	18.8	28.1
5	40.0	4.4	17.8	4.4	11.1	22.2
6	0.0	6.3	6.3	87.5	0.0	0.0

Table 44. Cluster membership crosstabulation of the 20-factor solution and the 11-factor solution.

20-Factor Solution	11-Factor Solution (percent) 1	2	3	4	5	6
1	29.5	36.4	6.8	9.1	11.4	6.8
2	13.0	39.1	8.7	23.9	2.2	13.0
3	6.9	6.9	65.5	0.0	13.8	6.9
4	6.3	28.1	28.1	18.8	6.3	12.5
5	22.2	26.7	35.6	11.1	2.2	2.2
6	12.5	6.3	18.8	0.0	0.0	62.5

Table 45. Cluster membership crosstabulation of the 20-factor solution and the 10-factor solution.

20-Factor Solution	10-Factor Solution (percent) 1	2	3	4	5	6
1	20.5	2.3	6.8	54.5	13.6	2.3
2	6.5	17.4	13.0	21.7	30.4	10.9
3	3.4	48.3	17.2	24.1	6.9	0.0
4	9.4	6.3	43.8	15.6	21.9	3.1
5	17.8	37.8	15.6	17.8	8.9	2.2
6	0.0	6.3	18.8	0.0	0.0	75.0

Table 46. Cluster membership crosstabulation of the 20-factor solution and the 9-factor solution.

20-Factor Solution	9-Factor Solution (percent) 1	2	3	4	5	6
1	36.4	11.4	20.5	15.9	11.4	4.5
2	19.6	34.8	10.9	26.1	2.2	6.5
3	41.4	10.3	6.9	10.3	27.6	3.4
4	9.4	6.3	15.6	31.3	12.5	25.0
5	46.7	17.8	15.6	4.4	2.2	13.3
6	12.5	0.0	0.0	81.3	0.0	6.3

Table 47. Cluster membership crosstabulation of the 20-factor solution and the 8-factor solution.

20-Factor Solution	8-Factor Solution (percent) 1	2	3	4	5	6
1	38.6	36.4	6.8	4.5	11.4	2.3
2	21.7	23.9	30.4	6.5	8.7	8.7
3	17.2	6.9	6.9	69.0	0.0	0.0
4	9.4	18.8	59.4	12.5	0.0	0.0
5	46.7	15.6	4.4	11.1	4.4	17.8
6	18.8	0.0	25.0	0.0	56.3	0.0

Table 48. Cluster membership crosstabulation of the 20-factor solution and the 7-factor solution.

20-Factor Solution	7 Factor Solution (percent) 1	2	3	4	5	6
1	36.4	15.9	27.3	13.6	2.3	4.5
2	15.2	21.7	17.4	28.3	8.7	8.7
3	13.8	0.0	6.9	10.3	62.1	6.9
4	9.4	3.1	40.6	31.3	12.5	3.1
5	22.2	28.9	13.3	22.2	8.9	4.4
6	6.3	0.0	0.0	18.8	0.0	75.0

Table 49. Cluster membership crosstabulation of the 20-factor solution and the 6-factor solution.

20-Factor Solution	6-Factor Solution 1 (percent)	2	3	4	5	6
1	22.7	22.7	0.0	22.7	29.5	2.3
2	19.6	19.6	4.3	43.5	4.3	8.7
3	13.8	10.3	37.9	3.4	24.1	10.3
4	6.3	0.0	15.6	53.1	3.1	21.9
5	6.7	26.7	11.1	37.8	13.3	4.4
6	6.3	31.3	31.3	12.5	0.0	18.8

Table 50. Cluster membership crosstabulation of the 20-factor solution and the 5-factor solution.

20-Factor Solution	5-Factor Solution 1 (percent)	2	3	4	5	6
1	11.4	25.0	34.1	11.4	15.9	2.3
2	34.8	21.7	4.3	6.5	23.9	8.7
3	0.0	24.1	13.8	6.9	24.1	31.0
4	18.8	31.3	0.0	25.0	12.5	12.5
5	31.1	20.0	11.1	0.0	11.1	26.7
6	0.0	31.3	6.3	6.3	25.0	31.3

Table 51. Cluster membership crosstabulation of the 20-factor solution and the 4-factor solution.

20-Factor Solution	4-Factor Solution 1	2	3 (percent)	4	5	6
1	29.5	36.4	13.6	4.5	6.8	9.1
2	10.9	37.0	17.4	15.2	10.9	8.7
3	20.7	0.0	17.2	24.1	37.9	0.0
4	3.1	28.1	31.3	21.9	15.6	0.0
5	22.2	22.2	4.4	26.7	2.2	22.2
6	6.3	0.0	43.8	12.5	12.5	25.0

Table 52. Cluster membership crosstabulation of the 20-factor solution and the 3-factor solution.

20-Factor Solution	3-Factor Solution 1	2	3 (percent)	4	5	6
1	61.4	4.5	18.2	9.1	4.5	2.3
2	52.2	6.5	15.2	17.4	6.5	2.2
3	20.7	10.3	27.6	3.4	3.4	34.5
4	28.1	15.6	28.1	18.8	0.0	9.4
5	40.0	0.0	28.9	8.9	6.7	15.6
6	0.0	18.8	12.5	56.3	0,0	12.5

Table 53. Cluster membership crosstabulation of the 20-factor solution and the 2-factor solution.

20-Factor Solution	2-Factor Solution 1	2	3 (percent)	4	5	6
1	38.6	0.0	9.1	29.5	2.3	20.5
2	39.1	10.9	8.7	23.9	8.7	8.7
3	6.9	27.6	34.5	20.7	10.3	0.0
4	9.4	25.0	21.9	28.1	3.1	12.5
5	40.0	8.9	13.3	17.8	15.6	4.4
6	0.0	25.0	31.3	6.3	37.5	0.0

example, in Table 36, about sixty-eight percent (68.2%) of the cases which were grouped into cluster 1 when clustering was based on"20 factor" factor scores and was also assigned to cluster 1 when clustering was based on "19 factor" factor scores. And, as indicated in the methods chapter, the crosstabulations were also used as the basis for calculating entropy measures.

Table 35 shows the comparison of cluster membership between the "20 factor" factor solution and "20 factor" factor solution when clustering on 20 factor scores. The reason for this self-comparison is to serve as a foundation (starting point) for calculating the entropy measure. This self-comparison shows complete certainty (entropy is 0) because all the elements of diagonal in Table 35 are 100% which means that cluster one in "20 factor" factor solution is exactly the same as the cluster one in "20 factor" factor solution.

The membership crosstabulations (Tables 36 to 53) reveal two major things about clustering and the membership of clusters. First, numbering of the different clusters appears to have changed across different cluster analyses. For example, in Table 36, cluster 3 formulated on factor scores from the "20 factor" factor solution is likely not to be the same as cluster 3 formulated on the "19 factor" factor scores. Only 17.2% of the cases assigned to cluster 3 are the same for the "20 factor" and "19 factor" factor solution. Cluster 3 in the "20 factor" factor solution is more likely cluster 4 in the "19 factor" factor solution. About forty-one percent (41.3%) of cluster 3 (20 factor factor solution) members are also in cluster 4 (19 factor factor solution). This created a problem when it came to assessing the impacts of factor-cluster solution on the stability of clusters.

Second, cluster membership is not stable; it changes across different factor solutions (e.g., "19 factor" factor solution vs. "18 factor" factor solution). The percentage of cases assigned to clusters changed significantly. For example, comparing Table 36 with Table 37, the percentage of cases (68.2%) assigned to cluster 1 when clustering was based on the "20 factor" factor scores and "19 factor" factor scores (see Table 36) changed to 40.9% (percentage of cases assigned to cluster 1) when clustering on "20 factor" factor scores and "18 factor" factor scores (see Table 37). About twenty-seven percent (27.3%) of cases were redistributed to other clusters.

Both the uncertainty of cluster numbering and the shift of cluster membership lead to the use of entropy measure to assess the effects of alternative factor solutions on cluster membership.

Based on the crosstabulation results (Table 35 to 53, page 120-129) and Formula 3 (discussed in Chapter III, page 52), an entropy measure was calculated for each crosstabulation/comparison. The entropy measures are presented in Table 54. The lower the entropy value, the

Table 54. Entropy measures (using the 20 factor solution as a basis of comparison) of cluster membership for different factor solutions.

Factor Solution Comparison	Entropy
20 - 20	0.0000
20 - 19	0.5181
20 - 18	0.5756
20 - 17	0.5170
20 - 16	0.5371
20 - 15	0.6174
20 - 14	0.5849
20 - 13	0.5964
20 - 12	0.5487
20 - 11	0.6083
20 - 10	0.5979
20 - 09	0.6112
20 - 08	0.8245
20 - 07	0.5942
20 - 06	0.6377
20 - 05	0.6788
20 - 04	0.6572
20 - 03	0.5727
20 - 02	0.6552

less the uncertainty of cluster membership between two different factor-cluster analytic solutions. That is, when the entropy value is low,changes in cluster membership between two different factor-cluster analytic solutions is small. Cluster membership (having lower entropy value) is relatively stable. Large entropy values indicate instability and that the membership of clusters based on different factor solutions is very different. For example, the uncertainty (membership instability) of cluster membership increases when basis for clustering is the "16 factor" factor solution rather than the "15 factor" factor solution. Uncertainty (membership instability) decreases when the clustering basis changes from the "13 factor" factor solution to the "12 factor" factor solution.

The entropy measures for different factor solution comparisons are plotted in Figure 26. The sudden downward or upward movement/change in the plot indicates that cluster membership is very instable across factor solutions. The result also indicates that the greatest instability occurs between the "9 factor" factor solution and the "7 factor" factor solution. Selecting a "9 factor" factor solution would result in a clustering solution that is very different from a clustering solution based on "8 factor" factor scores.

The entropy (information) measures indicate that cluster membership is very unstable across clustering solutions based on different factor scores (solutions). Thus, when clustering on factor scores, different factor solutions (number of factors) will affect cluster membership. The implication is that alternative factor solutions (number of factors) will result in different clustering results.

Assessment of the Effect of Rotation on Cluster Membership

Objective two was to ascertain the effect of factor rotation on the results of clustering (cluster membership). Nineteen (20, 19, 18, ..., 2) principal component analyses were again performed on the importance ratings of the 20 campground attributes. However, the initial factors were not rotated. The eigenvalues and percent of

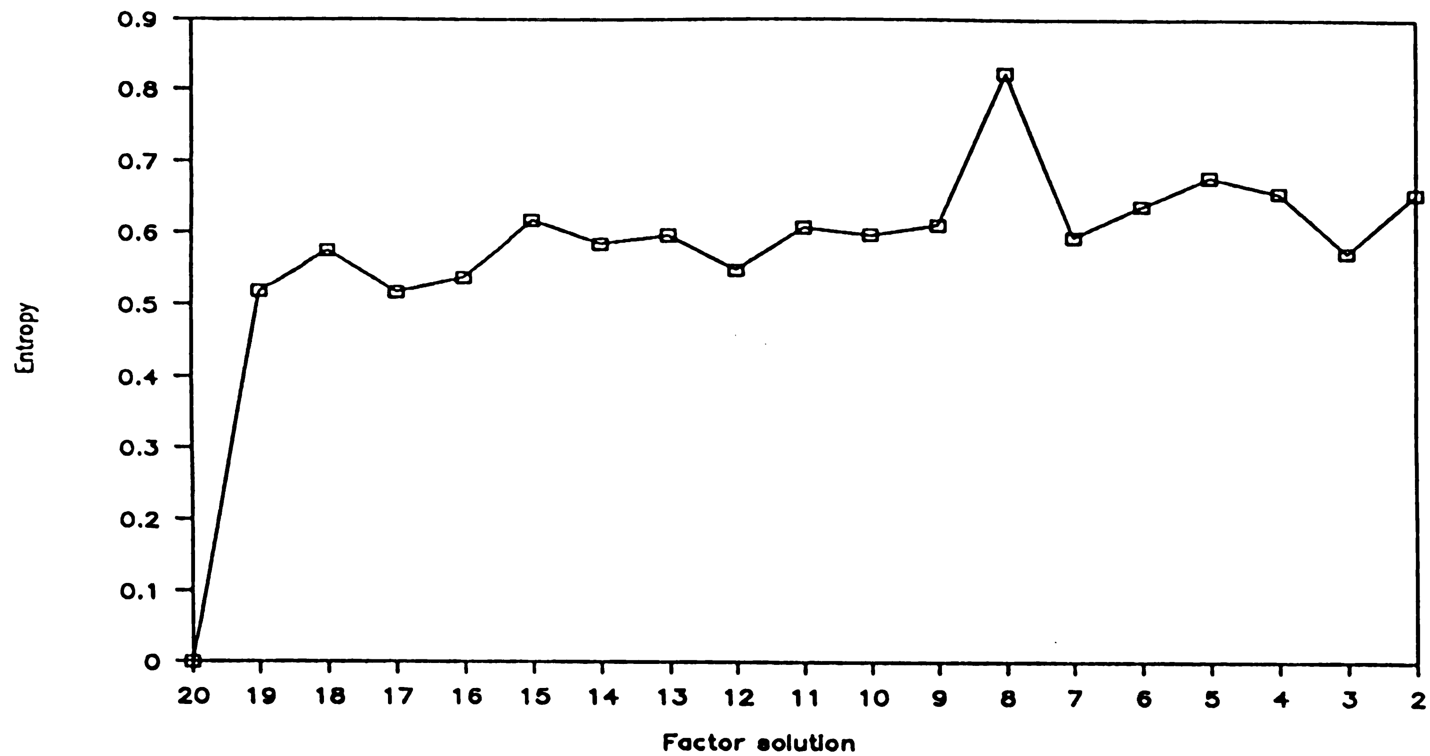


Figure 26. Entropy pattern of cluster membership across the different factor solutions.

variance explained for the factors are the same as the results derived from factor analysis with varimax rotation (see Table 11). Factor scores were again calculated using regression estimate method.

The factor scores were again used as input variables for a Ward's clustering method (using squared Euclidean distance). Nineteen different cluster analysis were performed; one on factor scores for each of the 19 (nonrotated) factor analyses. In each case, a six-cluster solution was selected to permit comparison of cluster membership with the clusters generated on rotated factor scores (see previous section).

Table 55 shows the results of crosstabulation of clusters based on rotated and nonrotated factor scores for the "20 factor" factor solution. It shows that there is <u>no</u> difference in cluster membership. The same is true for the other factor solutions (19, 18, 17, ..., 2). Rotation (or nonrotation) of factors does not affect clustering results when clustering based on factor scores. Clustering results do not change because rotating factors does not affect the goodness of fit of a factor solution. This is because the communalities and the percentage of total variance explained do not change.

Although rotation changes the factor matrix, the cluster (membership) solution does not change because rotation does not change the original relationship between variables. The distance between cases for each variable is not changed by rotation.

However, rotation of the initial factors can lead to a different interpretation of clustering solutions because of the difference in factor scores. Table 56 presents a comparison of factor score centroids for clusters based on rotated and nonrotated factor scores for the "20 factor" solution. It shows that the cluster centroids are different for

Table 55. Crosstabulation of clustering results based on rotated and nonrotated factors.

20 Rotated Factors	20 Nonrotated Factors (percent) 1	2	3	4	5	6
1	100.0	0.0	0.0	0.0	0.0	0.0
2	0.0	100.0	0.0	0.0	0.0	0.0
3	0.0	0.0	100.0	0.0	0.0	0.0
4	0.0	0.0	0.0	100.0	0.0	0.0
5	0.0	0.0	0.0	0.0	100.0	0.0
6	0.0	0.0	0.0	0.0	0.0	100.0

Table 56. Comparison of factor score centroids for clusters based on rotated and nonrotated factor scores for the "20 factor" solution.

	Rotated Approach							Nonrotated Approach					
	Cluster							Cluster					
Factor	1	2	3	4	5	6	Factor	1	2	3	4	5	6
1	-.14	.06	-.49	.45	.13	-.17	1	.57	.09	-.51	.25	-.10	-1.13
2	.17	.22	.15	-.27	-.21	-.21	2	.34	.29	-.61	-.55	.32	-.48
3	.30	.17	.33	-.05	.16	-2.29	3	.20	.18	-.36	.16	-.55	.85
4	.00	-.22	.23	.65	-.31	-.21	4	.07	-.27	.93	-.10	-.23	-.24
5	-.20	-.02	.10	.58	-.24	-.05	5	-.39	-.02	.07	.51	-.03	.07
6	-.12	.83	-.51	.10	-.32	-.43	6	-.24	.04	-.46	.83	-.09	-.01
7	-.56	.02	.07	.41	.43	-.70	7	-.17	-.25	.72	.25	.16	-1.08
8	.29	.26	.30	-.58	-.28	-.14	8	-.30	.44	.03	.40	-.48	.03
9	-.15	.29	-.10	-.45	.31	-.19	9	-.19	-.34	.03	-.10	.30	.80
10	.04	.09	-.32	.75	-.37	-.25	10	.46	.17	-.13	-.50	.12	-.86
11	.43	.35	-.37	-.43	-.27	.09	11	-.59	.44	-.02	.39	.23	-1.05
12	.04	.03	-.72	-.09	.48	-.05	12	.11	.35	.60	.03	-.74	-.36
13	.77	-.45	-.27	-.14	.01	-.08	13	-.12	.02	-.18	.34	.05	-.22
14	.43	-.41	.01	.35	-.24	-.04	14	-.37	.64	.16	-.37	-.07	-.15
15	.35	-.05	-.15	.16	-.26	-.15	15	.16	-.31	-.02	.27	.08	-.25
16	.45	.26	-.14	-.11	-.45	-.22	16	.46	-.41	-.27	.55	-.15	-.25
17	.51	-.22	-1.16	.12	.46	-.21	17	.34	-.16	-.49	.27	.04	-.25
18	.08	-.21	.22	-.30	.24	-.11	18	-.47	-.03	-.12	.16	.47	-.08
19	-.29	.09	.18	.18	-.11	.17	19	-.27	-.11	-.05	.06	.34	.12
20	-.06	-.48	.36	.09	.31	-.16	20	.19	.49	-.30	-.36	-.23	-.05

clusters on rotated and nonrotated factor scores (because the factor matrix changes), even though the cluster membership is the same. Since cluster centroids are used to label/describe clusters, rotating factors will affect the interpretation of the clustering results. For example, cluster 1 based on rotated factor scores would be labeled based on factor 13 (.77), factor 17 (.51), and factor 7 (-.56). Cluster 1 formulated on nonrotated factor scores would be labeled based on factor 1 (.57), factor 16 (.46), and factor 11 (-.59). So, clusters comprised of the same members would be described differently depending on whether the clusters are based on rotated or nonrotated factor scores.

Comparison of Clustering on Factor Scores with Clustering on Raw Data

As mentioned previously, factor analysis is often performed as a preliminary step to clustering in order to reduce a large number of variables and make it easier to describe/label the resultant clusters. Shutty and DeGood (1987) contended that clustering on factor scores results in clusters which can be described more accurately. However, reducing variables to a smaller number of dimensions also results in a loss of information (e.g., percentage of total variance explained) which affects the clustering results (e.g., membership). This section compares clustering based on factor scores with clustering on raw data (the importance ratings assigned different campground attributes).

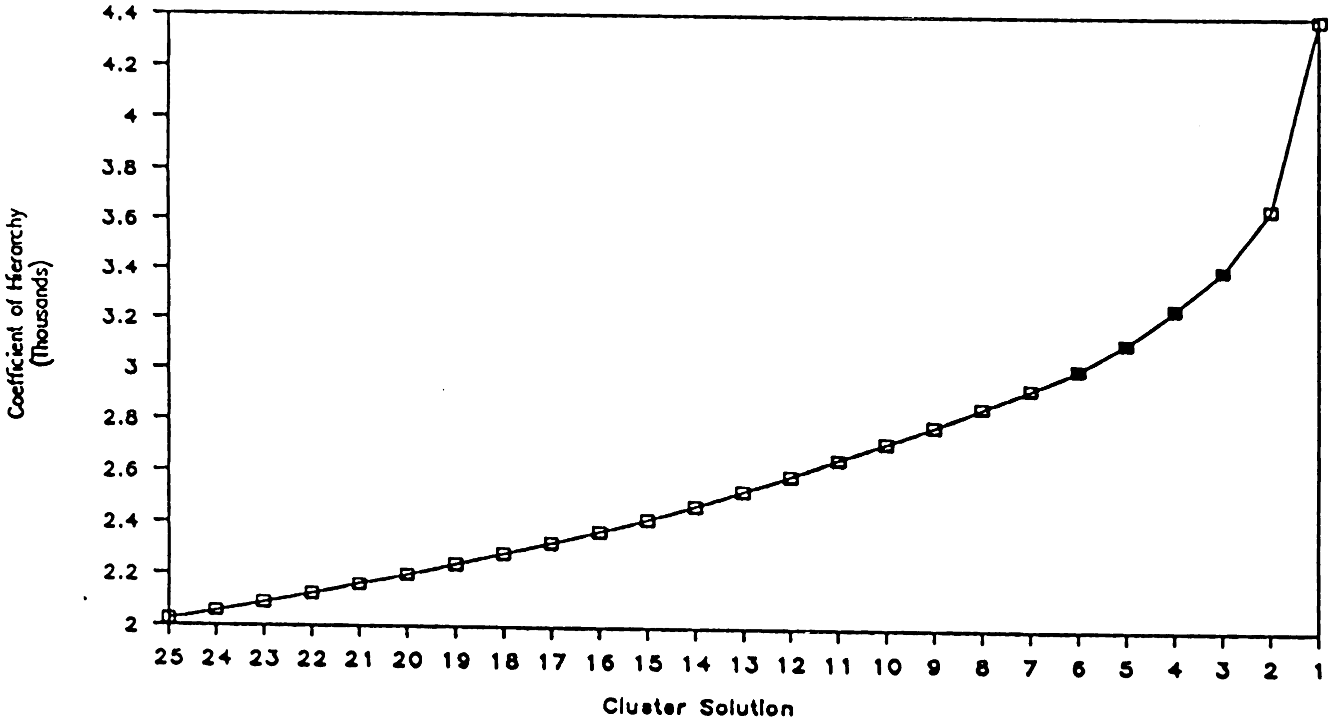
Clustering Results

Ward's clustering method (with squared Euclidean distance as the distance measure) was used to group respondents based on the importance they assigned to the 20 different campground attributes. Figure 27 shows the increase in coefficient of hierarchy (which resulted from fusion of clusters) plotted against the number of the clusters. Four candidate cluster solutions were identified: six clusters, five clusters, four clusters, and three clusters.

ANOVAs were conducted to determine the extent of inter-cluster differences among the four potential cluster solutions. For each of the four potential solutions, there were statistically significant differences among clusters on all 20 attributes (see Tables 57-60). The primary weakness of the six-cluster solution is that one of the clusters has less than 10 cases (see Table 61). However, the six-cluster solution was still selected to enable comparisons with the six cluster formulated on factor scores.

Comparisons Between Clustering Approaches

Nineteen principal component analyses with varimax rotation were performed on the importance ratings of the 20 campground attributes. Again, the regression estimates method was used to calculate factor scores. The (factor score) centroids for each of the six clusters (based on raw data) were then calculated for each of 19 factor analyses (20, 19, 18, ..., 2). They are graphically presented in Figures 28-47.



Cluster Solution	Coefficient of Hierarchy
25	2024
24	2057
23	2090
22	2125
21	2162
20	2200
19	2240
18	2280
17	2324
16	2369
15	2416
14	2471
13	2528
12	2588
11	2652
10	2717
9	2784
8	2857
7	2931
6	3007
5	3113
4	3252
3	3408
2	3654
1	4389

Figure 27. Coefficient of hierarchy by number of clusters when clustering is based on raw data (■: candidate solution)

Table 57. Mean attribute sought factor scores for the six-cluster candidate solution when clustering on raw data.

Factor	Cluster 1	2	3	4	5	6	F-ratio
Large sites	3.27	3.39	3.60	2.48	2.75	3.50	8.80*
Shaded sites	3.53	2.91	3.55	2.81	2.90	2.83	6.23*
Cleanliness	2.02	1.77	2.24	1.52	1.55	1.50	7.09*
Quietness	2.70	3.02	3.05	2.20	2.32	3.33	7.09*
Site privacy	3.00	3.61	3.81	2.67	3.05	3.33	8.17*
Security	2.23	2.43	2.43	1.76	1.70	1.50	6.41*
Hospitality	2.67	2.68	2.88	1.95	1.92	2.00	8.66*
Low price	3.11	2.98	3.10	1.95	2.85	2.33	5.66*
Flush toilets	3.98	2.77	4.19	1.95	2.80	3.00	32.31*
Electricity	2.21	2.61	3.67	1.67	2.48	4.33	30.11*
Showers	3.81	2.30	3.76	1.67	2.58	2.67	37.93*
Laundromat	3.79	4.07	4.60	2.48	3.58	4.67	26.34*
Campground store	3.98	3.95	4.41	2.48	3.28	4.00	23.98*
Water hookups	2.44	3.34	4.16	1.86	2.42	4.67	36.68*
Sewer hookups	3.30	4.09	4.87	2.14	3.28	4.83	30.16*
Natural surr.	3.58	3.11	3.76	2.76	2.58	2.00	11.91*
Lake/stream	4.47	4.41	4.48	3.57	2.85	3.33	27.60*
Hiking trails	4.42	4.09	4.53	3.57	3.10	2.67	19.79*
Pool	4.21	4.23	4.60	3.14	3.02	3.67	19.26*
Playground	4.72	4.75	4.88	3.38	4.10	2.00	30.00*

* Significant at .05 level.

Table 58. Mean attribute sought factor scores for the five-cluster candidate solution when clustering on raw data.

Factor	Cluster 1	2	3	4	5	F-ratio
Large sites	2.74	3.39	3.60	2.48	2.85	10.03*
Shaded sites	3.53	2.91	3.55	2.81	2.89	7.81*
Cleanliness	2.02	1.77	2.24	1.52	1.54	8.89*
Quietness	2.70	3.02	3.05	2.19	2.46	6.73*
Site privacy	3.00	3.61	3.81	2.67	3.09	10.12*
Security	2.23	2.43	2.43	1.76	1.67	7.96*
Hospitality	2.67	2.68	2.88	1.95	1.93	10.86*
Low price	3.11	2.98	3.10	1.95	2.78	6.67*
Flush toilets	3.98	2.77	4.19	1.95	2.83	40.47*
Electricity	2.21	2.61	3.67	1.67	2.71	27.88*
Showers	3.81	2.30	3.76	1.67	2.59	47.60*
Laundromat	3.79	4.07	4.60	2.48	3.71	29.10*
Campground store	3.98	3.95	4.41	2.48	3.37	28.36*
Water hookups	2.44	3.34	4.16	1.86	2.71	32.99*
Sewer hookups	3.30	4.09	4.59	2.14	3.48	31.68*
Natural surr.	3.58	3.11	3.76	2.76	2.50	14.33*
Lake/stream	4.47	4.41	4.48	3.57	2.91	33.89*
Hiking trails	4.42	4.09	4.53	3.57	3.04	24.36*
Pool	4.21	4.23	4.60	3.14	3.11	23.24*
Playground	4.72	4.75	4.88	3.38	3.83	22.62*

* Significant at .05 level.

Table 59. Mean attribute sought factor scores for the four-cluster candidate solution when clustering on raw data.

Factor	Cluster 1	2	3	4	F-ratio
Large sites	2.74	3.39	3.60	2.73	12.54*
Shaded sites	3.53	2.91	3.55	2.87	10.42*
Cleanliness	2.02	1.77	2.24	1.54	11.91*
Quietness	2.70	3.02	3.05	2.37	8.48*
Site privacy	3.00	3.61	3.81	2.96	12.34*
Security	2.23	2.43	2.43	1.70	10.60*
Hospitality	2.67	2.68	2.88	1.94	14.55*
Low price	3.12	2.98	3.10	2.52	4.99*
Flush toilets	3.98	2.77	4.19	2.55	46.32*
Electricity	2.21	2.61	3.67	2.39	27.82*
Showers	3.81	2.30	3.76	2.30	53.11*
Laundromat	3.79	4.07	4.60	3.33	23.43*
Campground store	3.98	3.95	4.41	3.09	29.07*
Water hookups	2.44	3.34	4.16	2.45	38.33*
Sewer hookups	3.30	4.09	4.59	3.06	28.66*
Natural surr.	3.58	3.11	3.76	2.58	18.73*
Lake/stream	4.47	4.41	4.48	3.12	40.32*
Hiking trails	4.42	4.09	4.53	3.21	29.97*
Pool	4.21	4.23	4.60	3.12	31.13*
Playground	4.72	4.75	4.88	3.69	28.26*

* Significant at .05 level.

Table 60. Mean attribute sought factor scores for the three-cluster candidate solutions when clustering on raw data.

Factor	Cluster 1	2	3	F-ratio
Large sites	3.07	3.60	2.73	13.09*
Shaded sites	3.22	3.55	2.87	9.42*
Cleanliness	1.90	2.24	1.53	16.27*
Quietness	2.86	3.05	2.37	10.99*
Site privacy	3.31	3.81	2.96	13.07*
Security	2.33	2.43	1.70	15.25*
Hospitality	2.68	2.88	1.94	21.93*
Low price	3.05	3.10	2.52	7.29*
Flush toilets	3.37	4.19	2.55	42.84*
Electricity	2.41	3.67	2.39	39.07*
Showers	3.05	3.76	2.30	34.34*
Laundromat	3.93	4.60	3.33	33.81*
Campground store	3.97	4.41	3.09	43.80*
Water hookups	2.90	4.16	2.45	45.05*
Sewer hookups	3.70	4.59	3.06	34.20*
Natural surr.	3.34	3.76	2.58	24.92*
Lake/stream	4.44	4.48	3.12	60.68*
Hiking trails	4.25	4.53	3.21	42.93*
Pool	4.22	4.60	3.12	46.91*
Playground	4.74	4.88	3.69	42.57*

* Significant at .05 level.

Table 61. Number of respondents in each of the cluster candidate solution when clustering on raw data.

Cluster	Number of Respondents	Relative Size (percent)
Six Cluster Solution		
1	43	20.3
2	44	20.8
3	58	27.4
4	21	9.9
5	40	18.9
6	6	2.8
Total	212	100.1
Five Cluster Solution		
1	43	20.3
2	44	20.8
3	58	27.4
4	21	9.9
5	46	21.7
Total	212	100.1
Four Cluster Solution		
1	43	20.3
2	44	20.8
3	58	27.4
4	67	31.6
Total	212	100.1
Three Cluster Solution		
1	87	41.0
2	58	27.4
3	67	31.6
Total	212	100.0

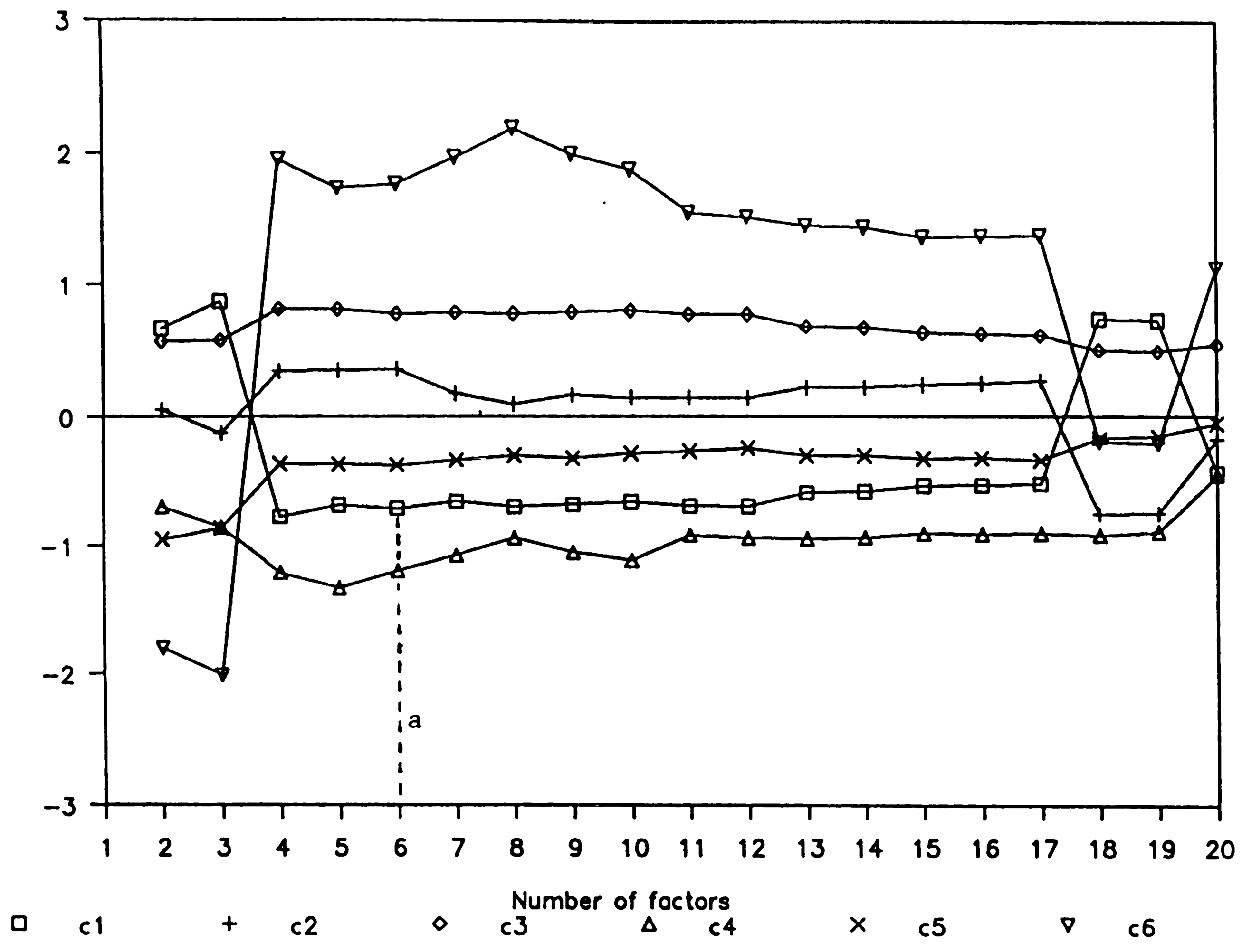


Figure 28. The "factor 1" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6, a: factor 1 factor score centroid of cluster 1 for the "6 factor" factor solution).

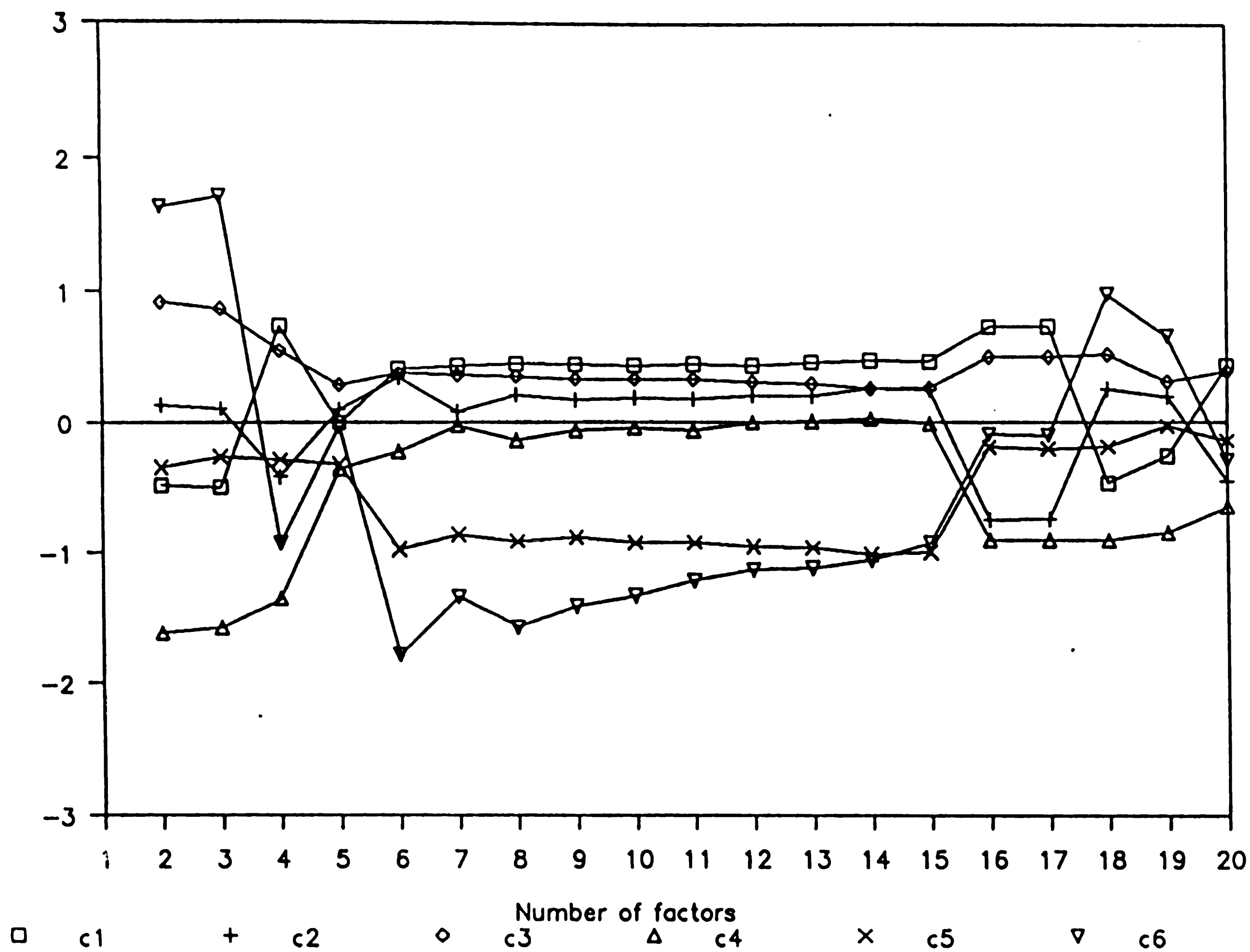


Figure 29. The "factor 2" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

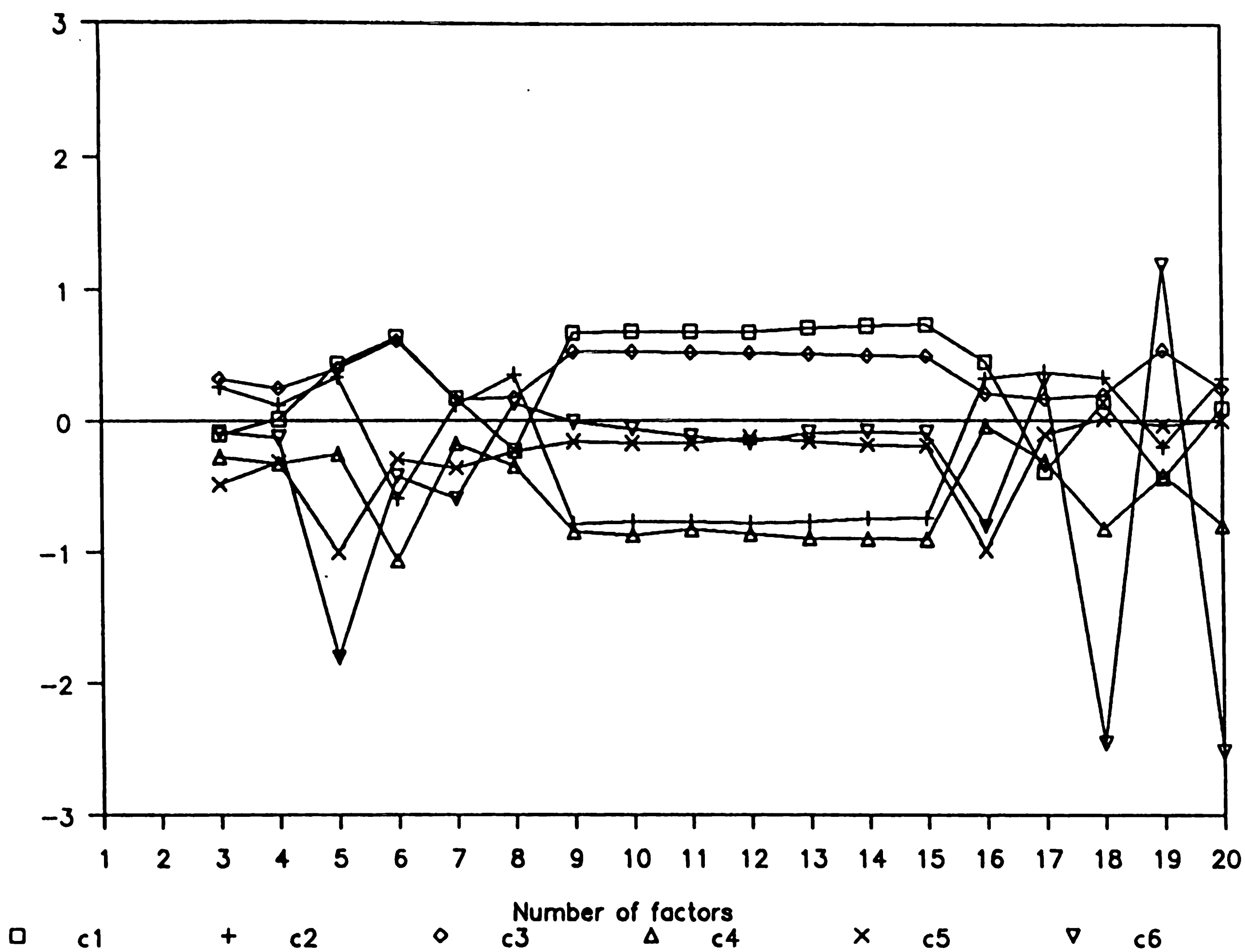


Figure 30. The "factor 3" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

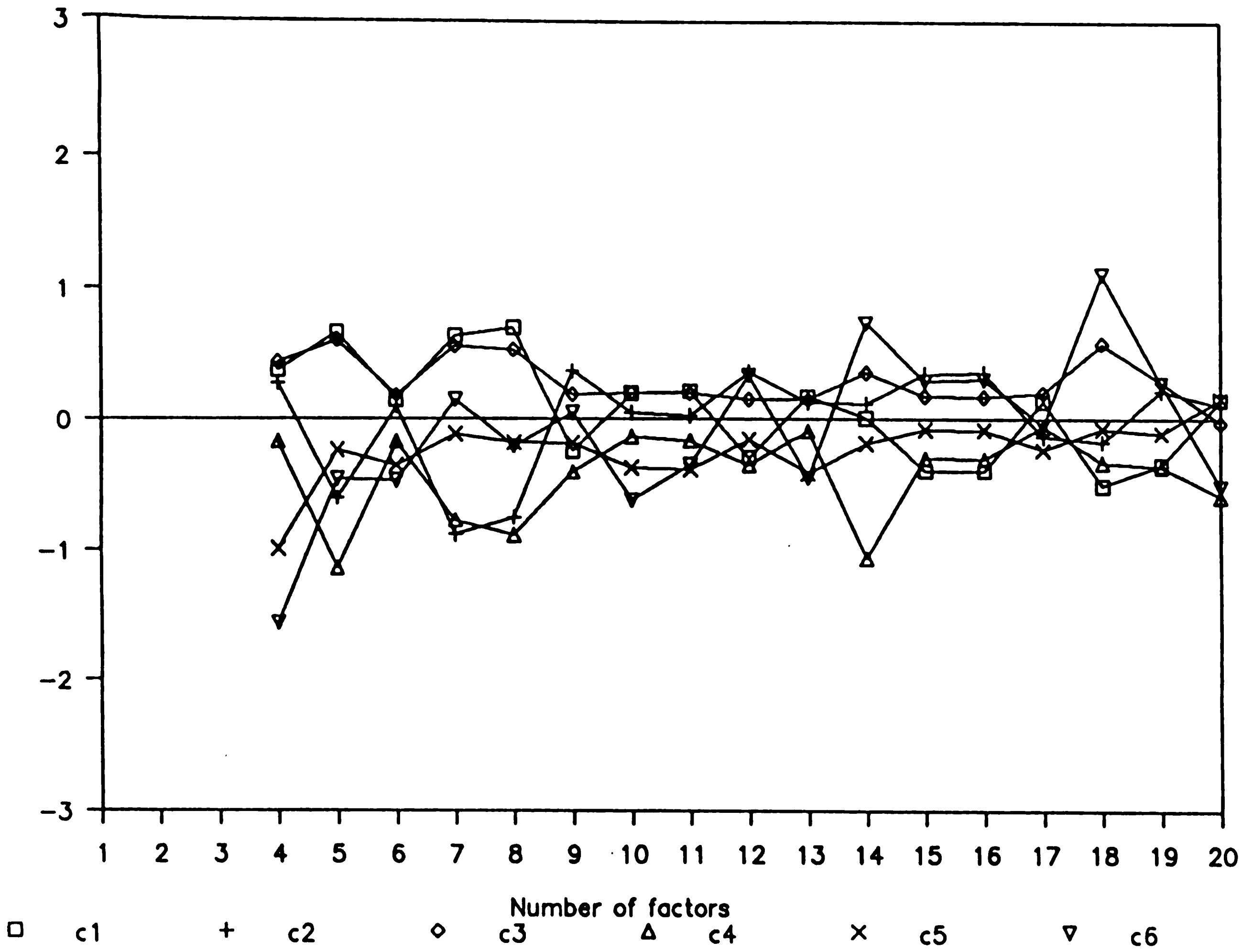


Figure 31. The "factor 4" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

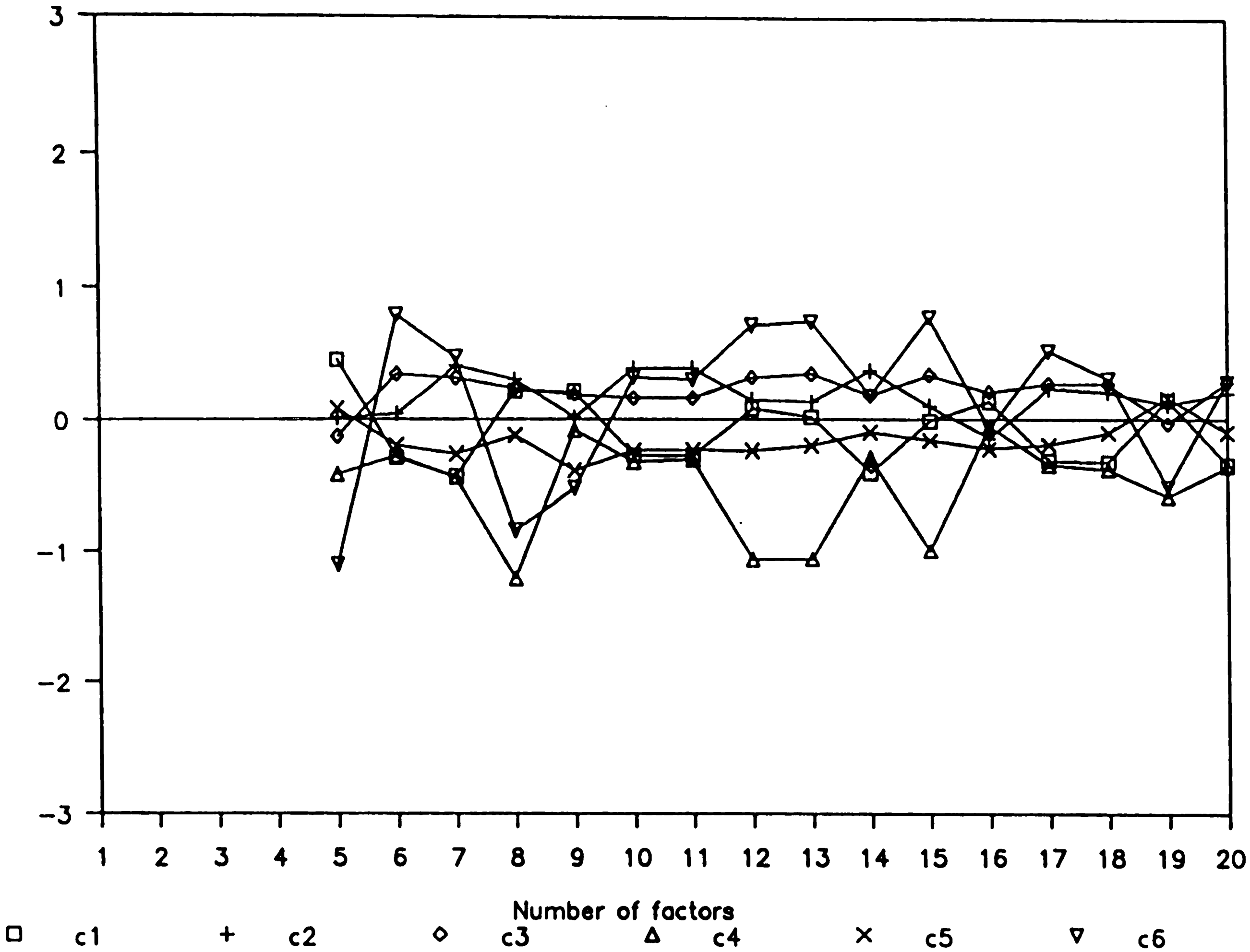


Figure 32. The "factor 5" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

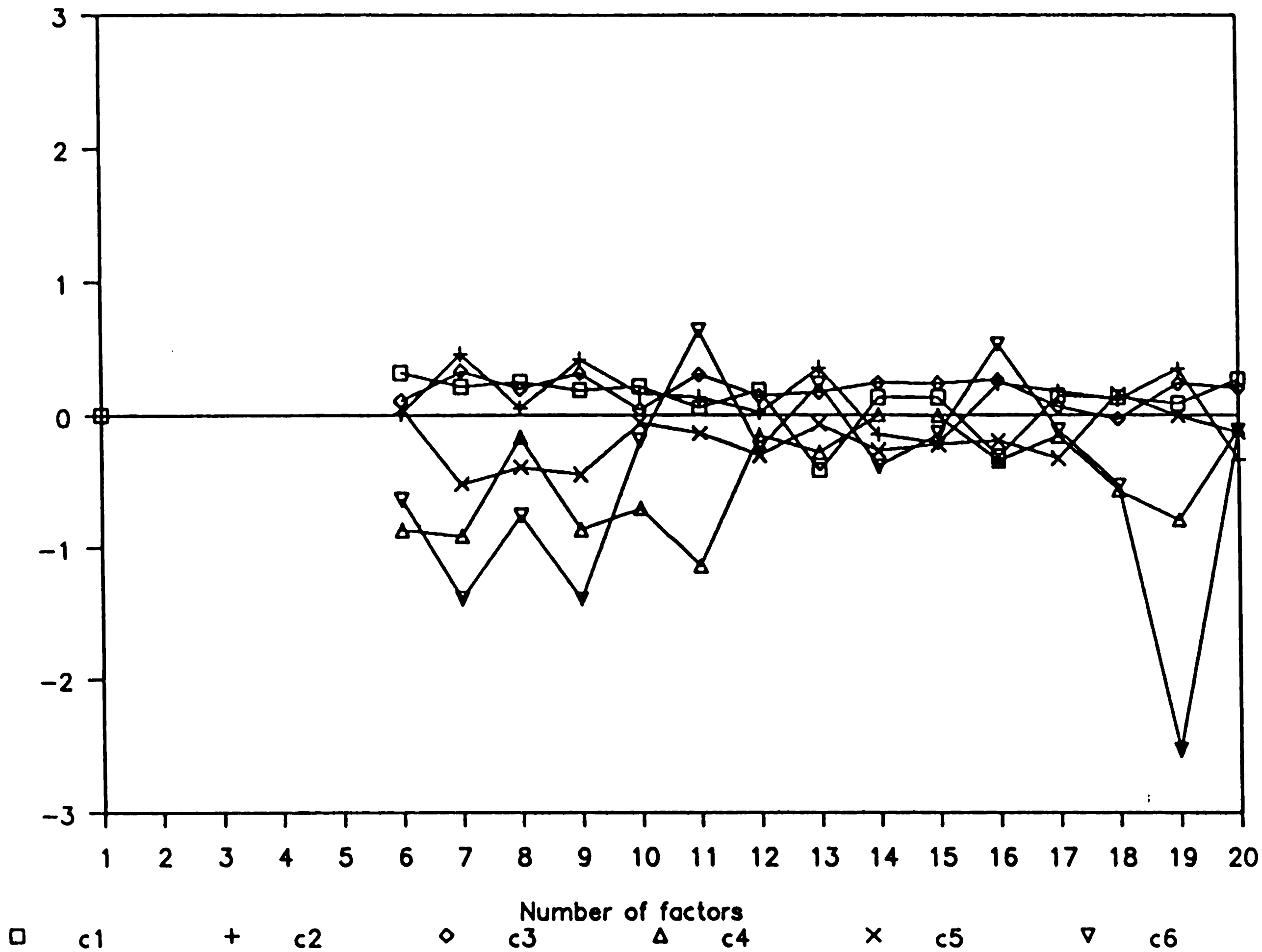


Figure 33. The "factor 6" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

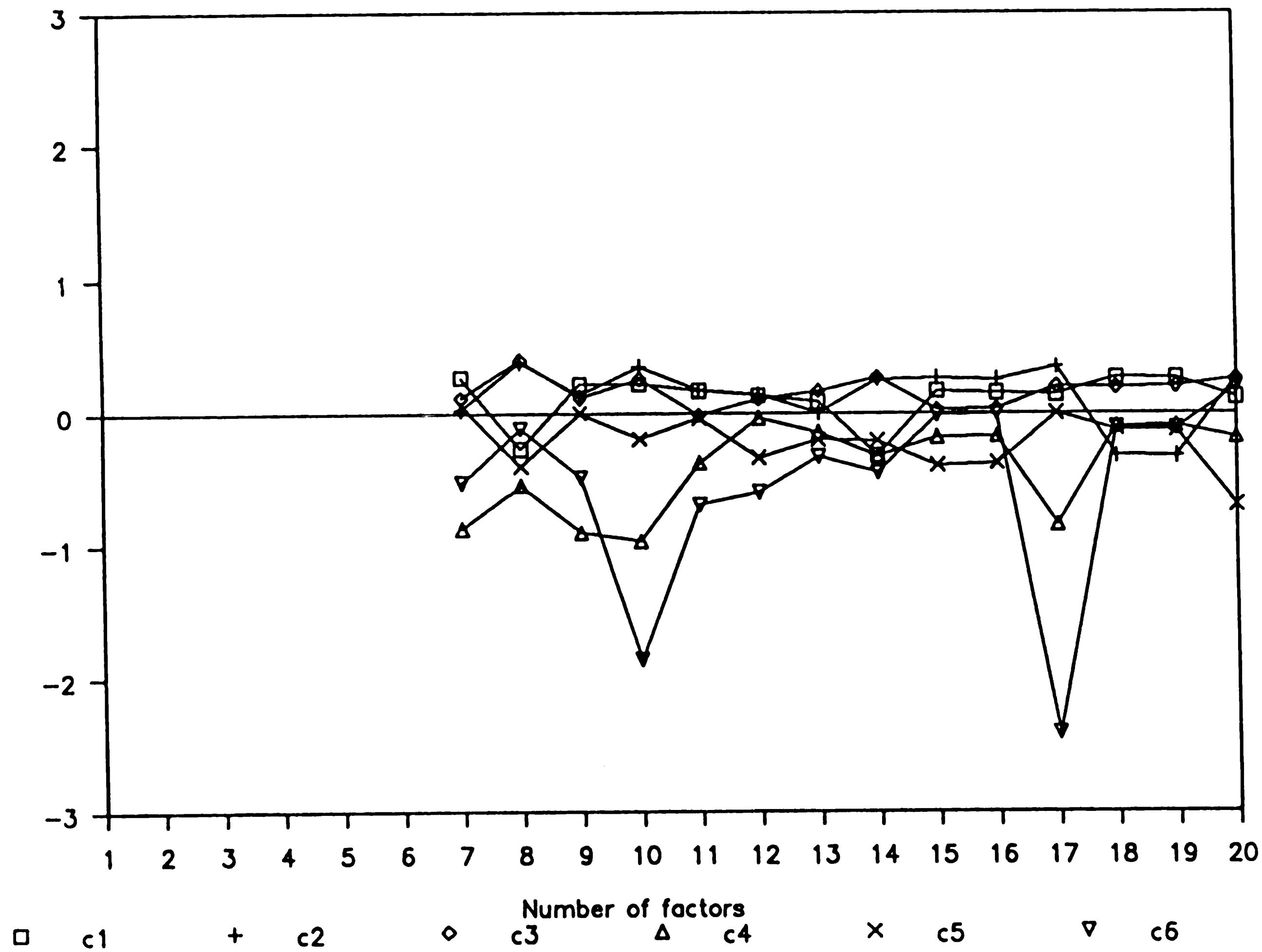


Figure 34. The "factor 7" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

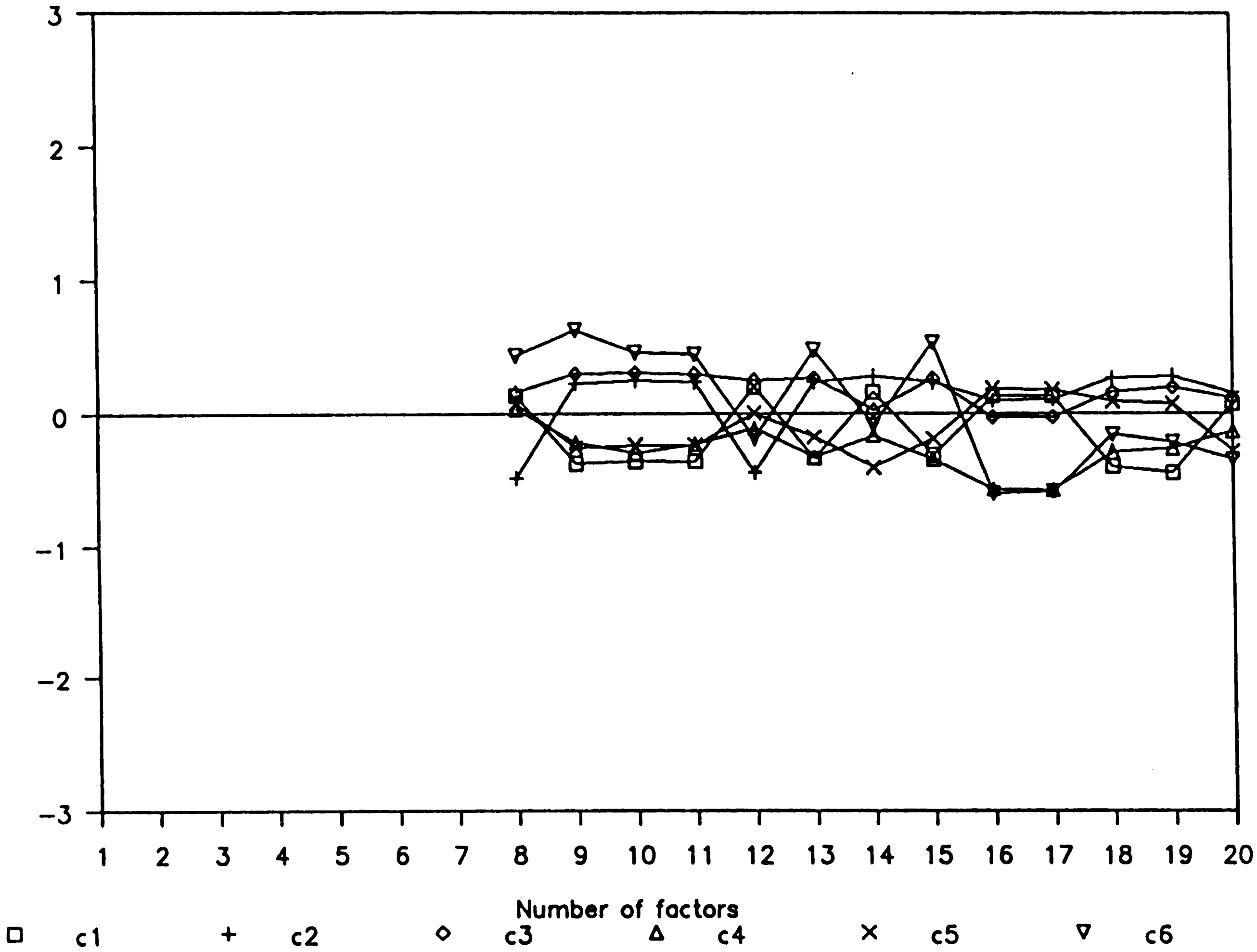


Figure 35. The "factor 8" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

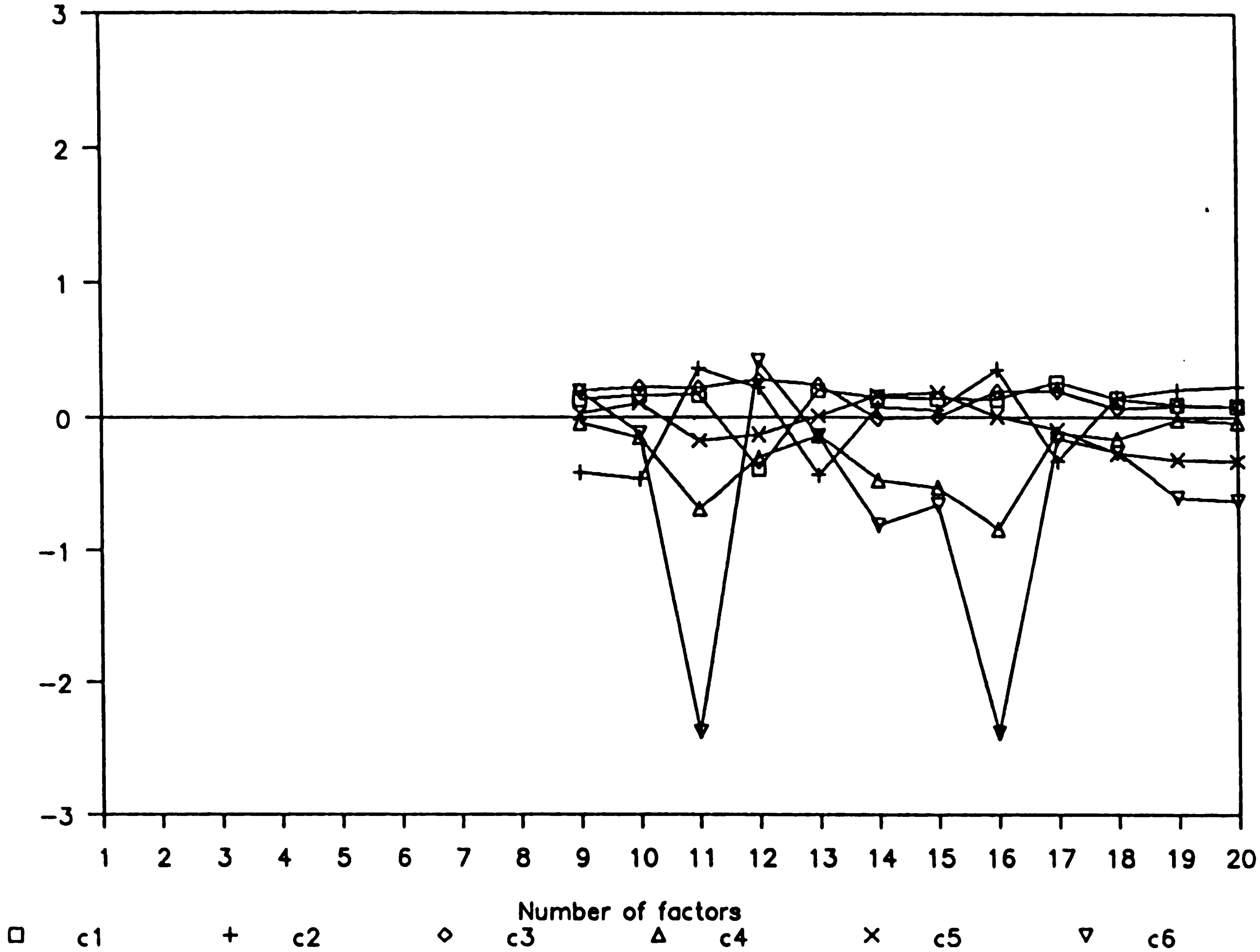


Figure 36. The "factor 9" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

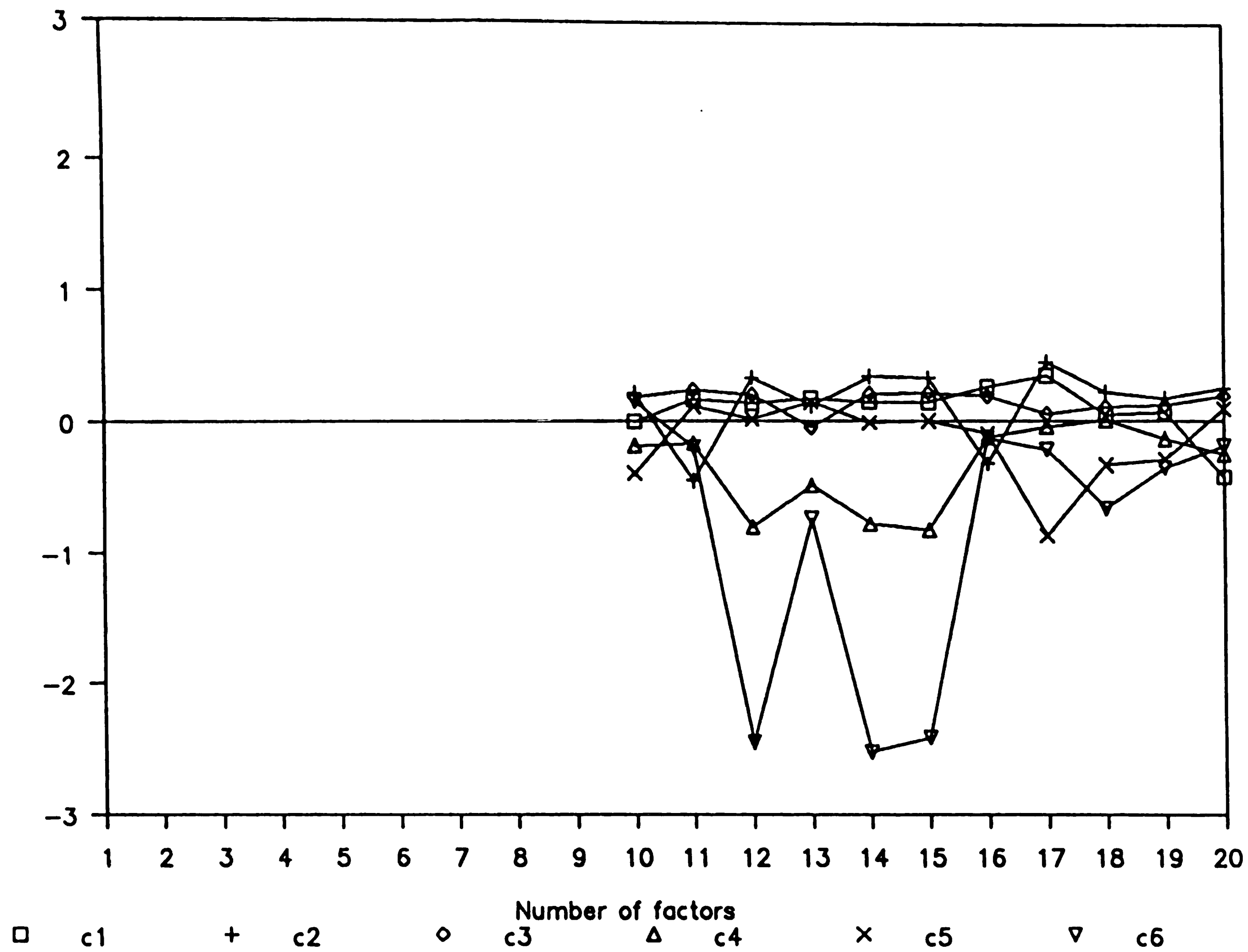


Figure 37. The "factor 10" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

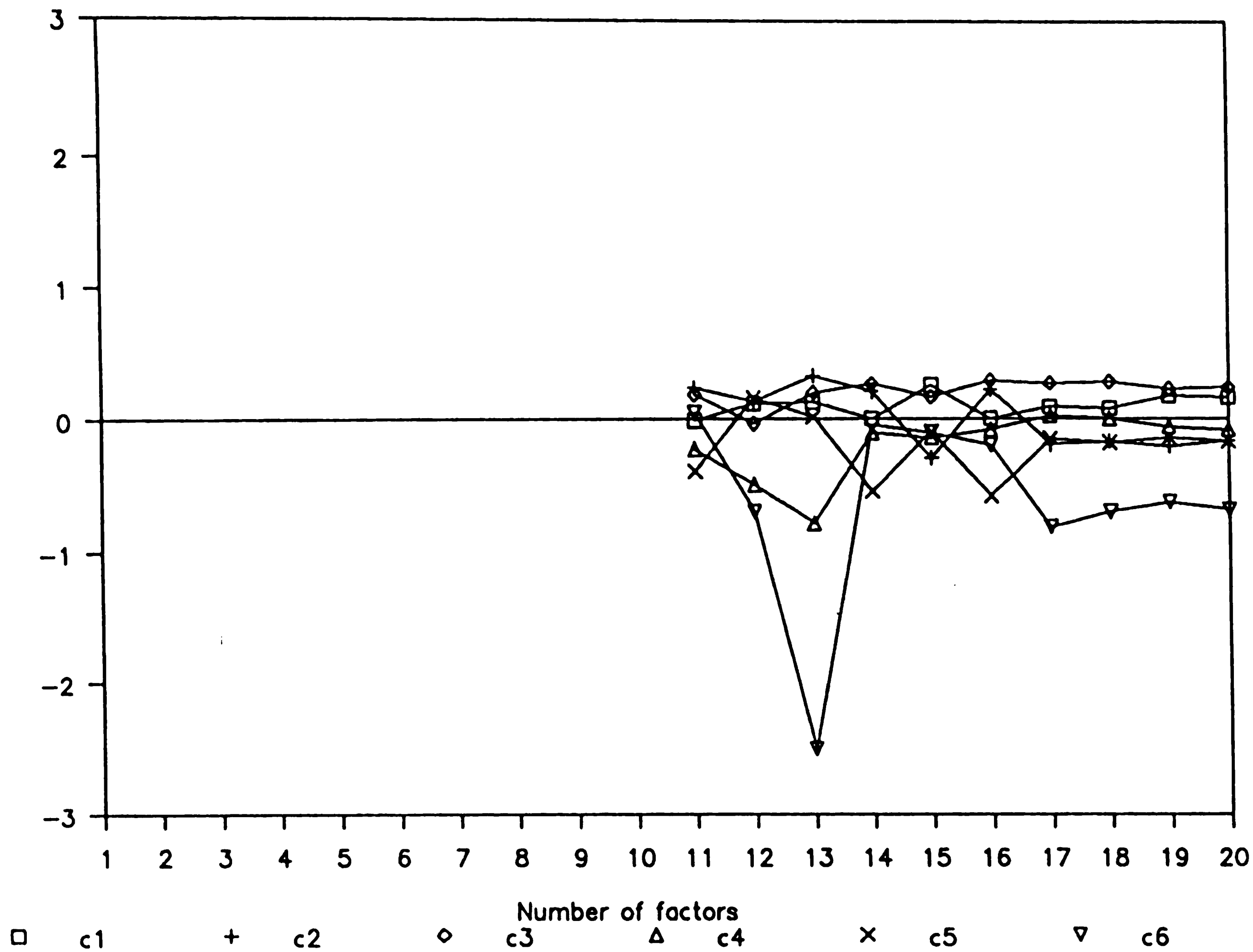


Figure 38. The "factor 11" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

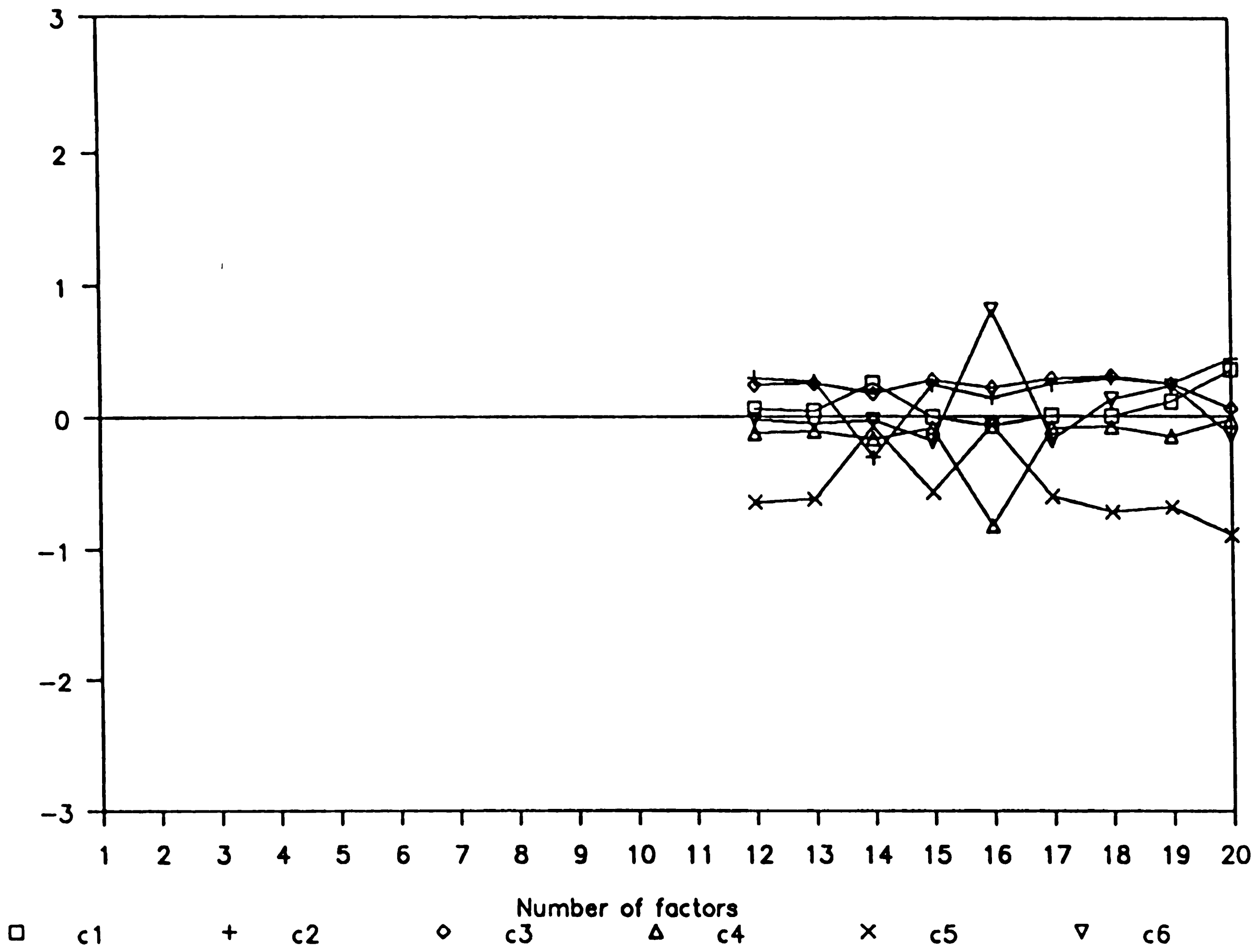


Figure 39. The "factor 12" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

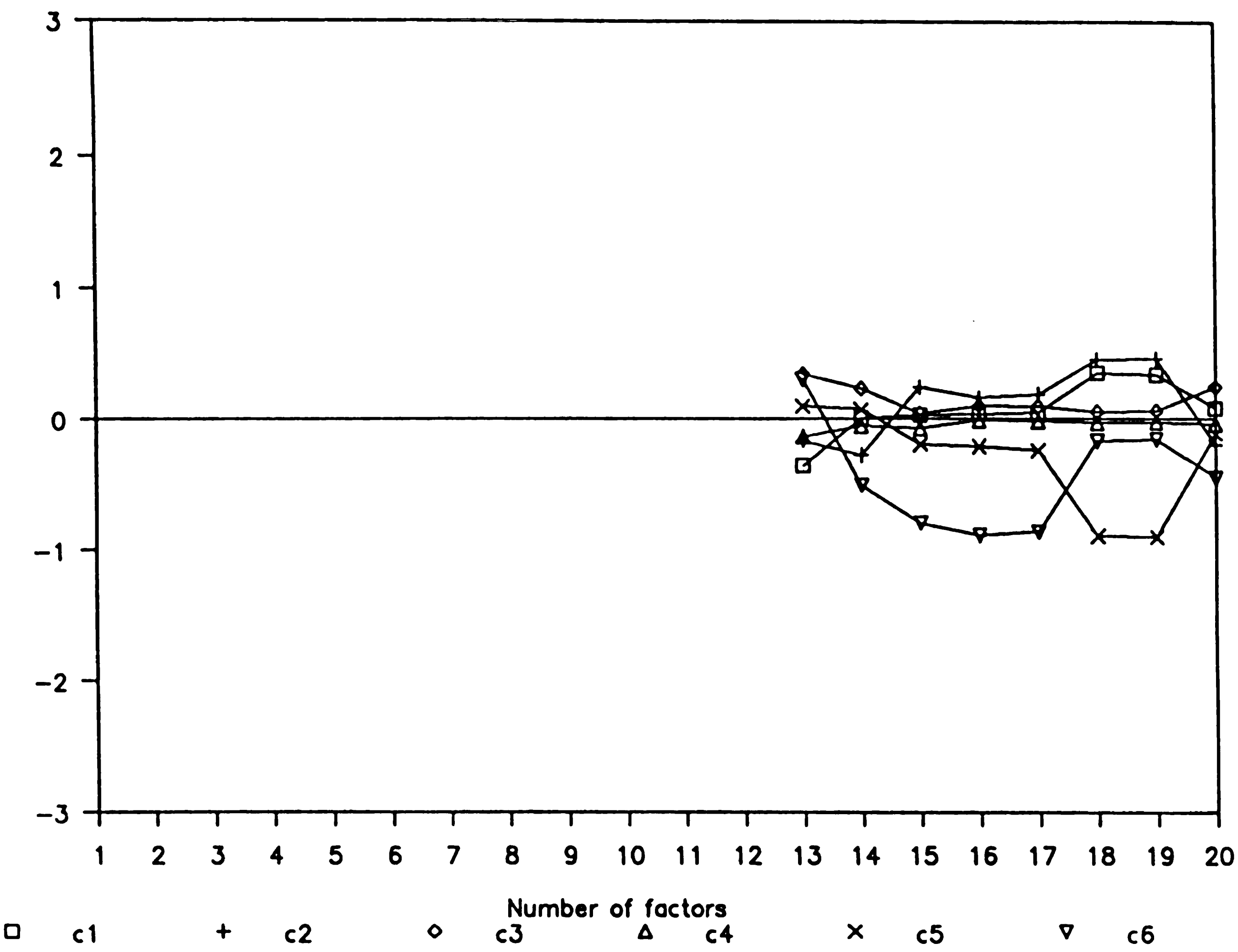


Figure 40. The "factor 13" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

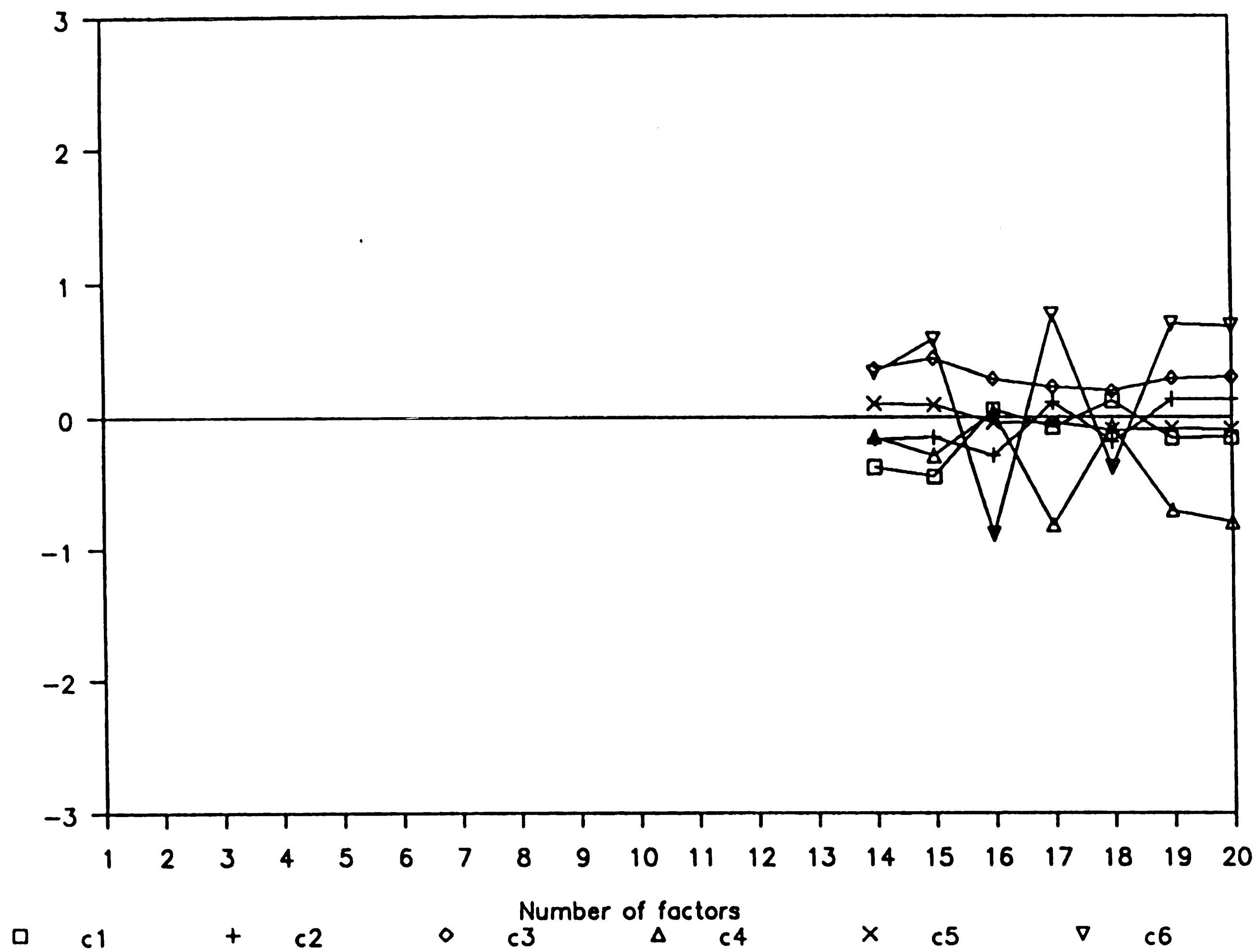


Figure 41. The "factor 14" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

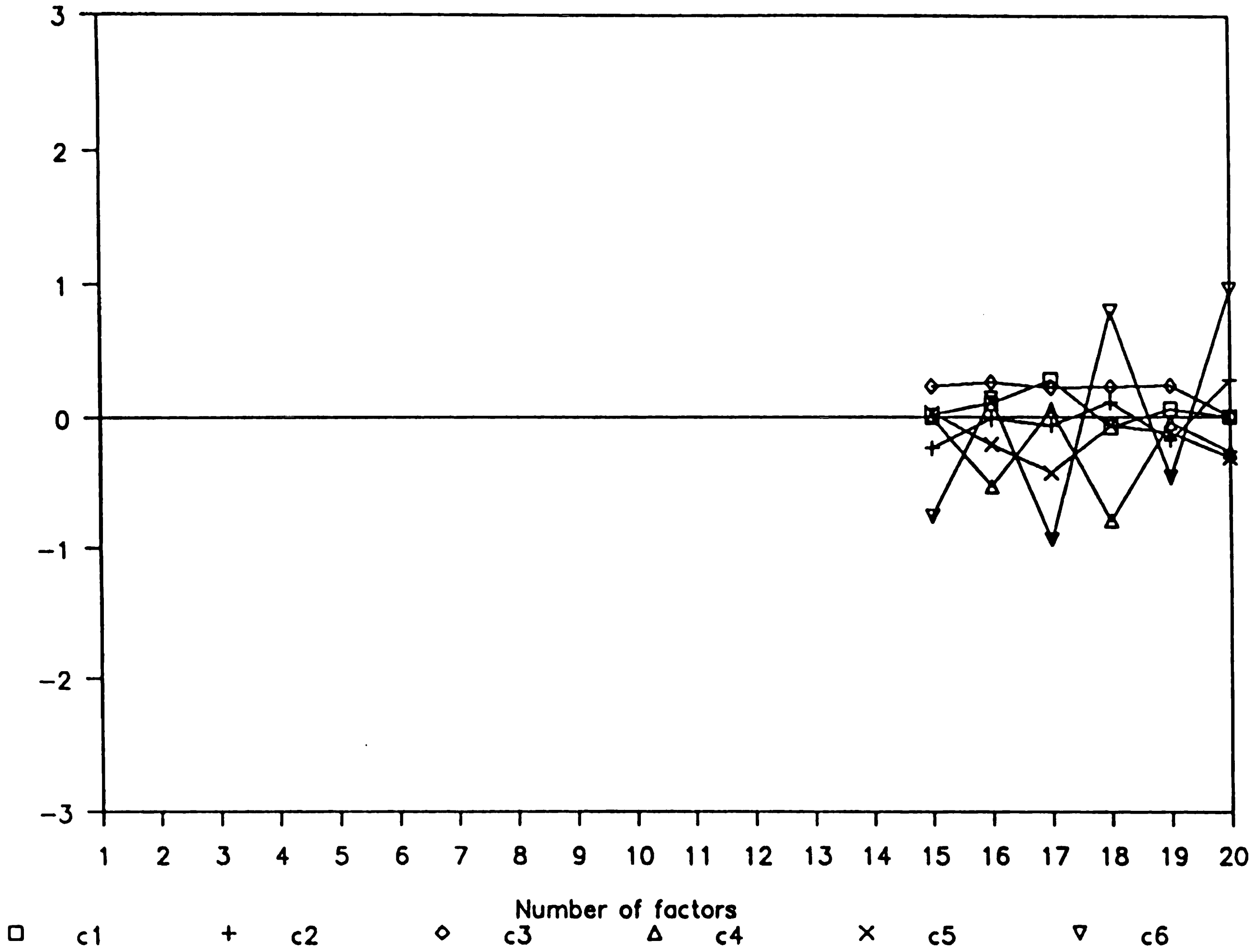


Figure 42. The "factor 15" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

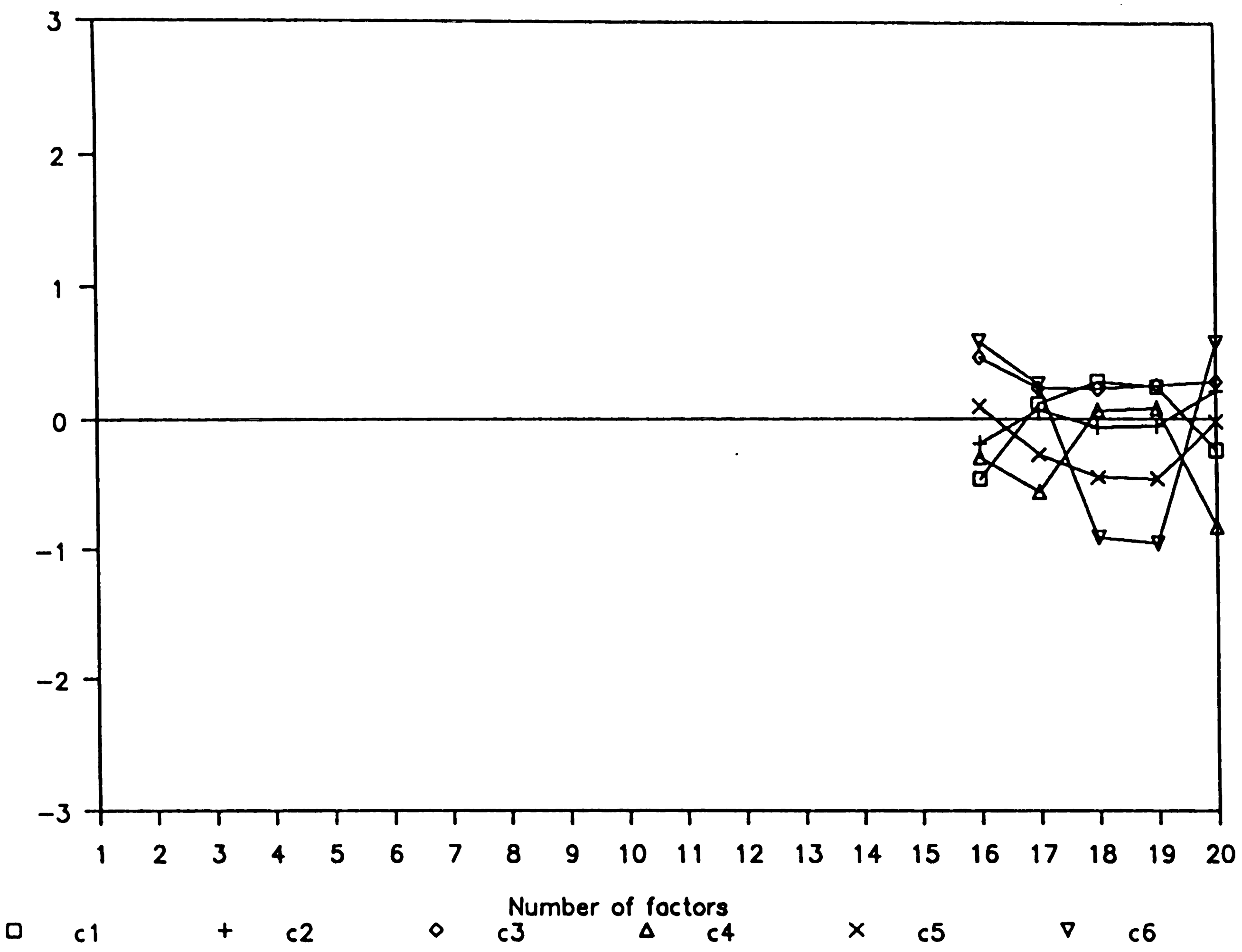


Figure 43. The "factor 16" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

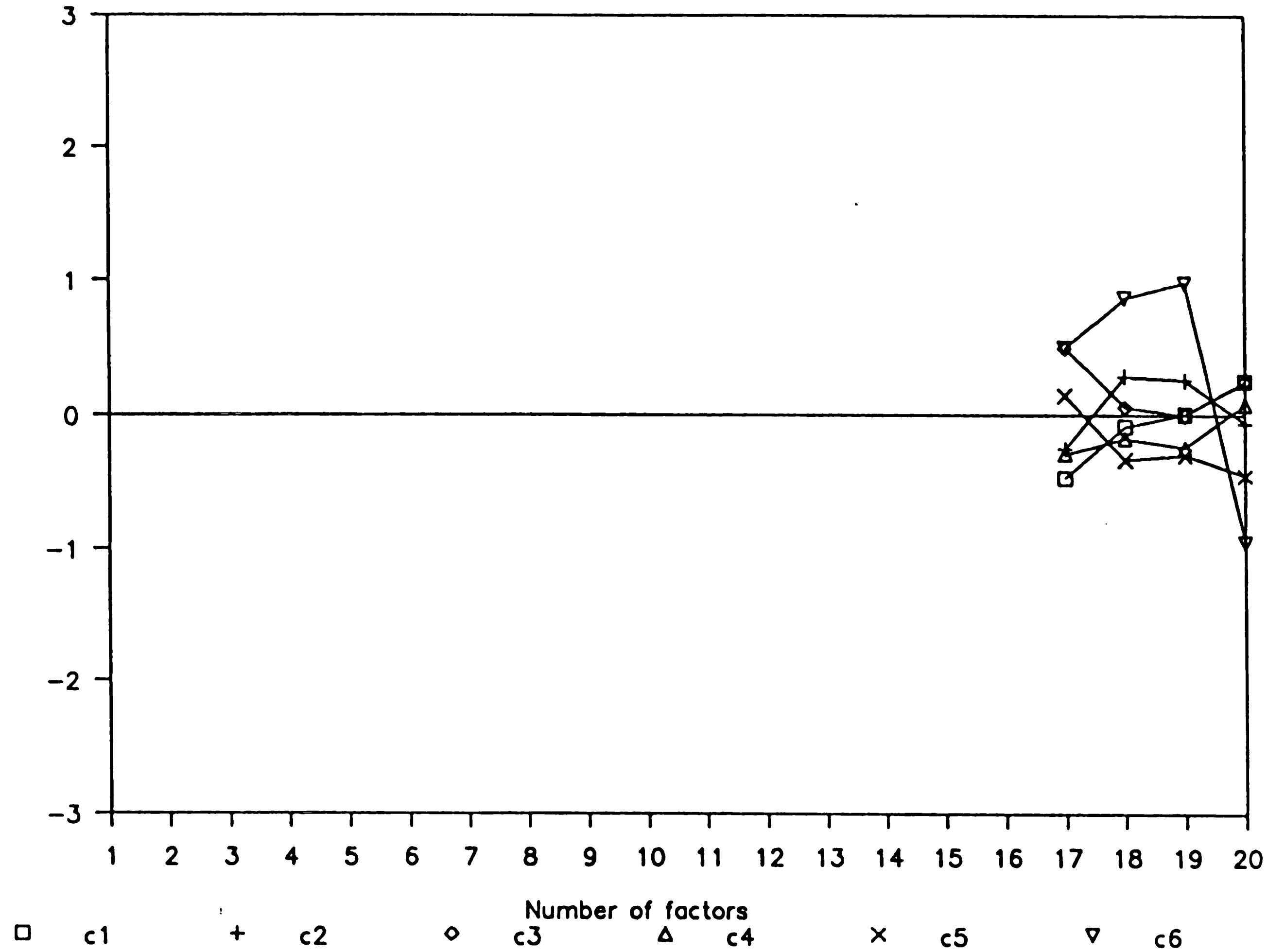


Figure 44. The "factor 17" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

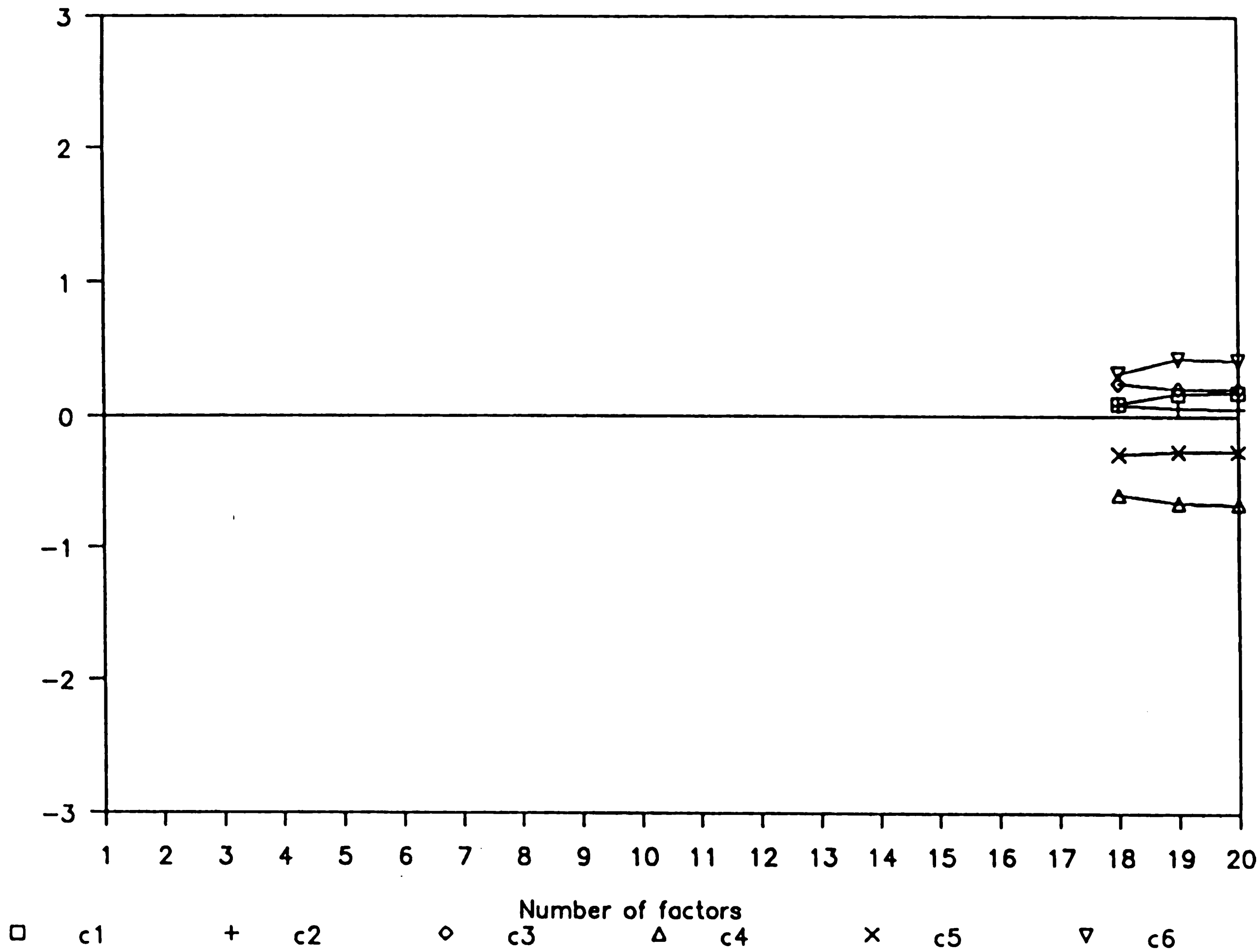


Figure 45. The "factor 18" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

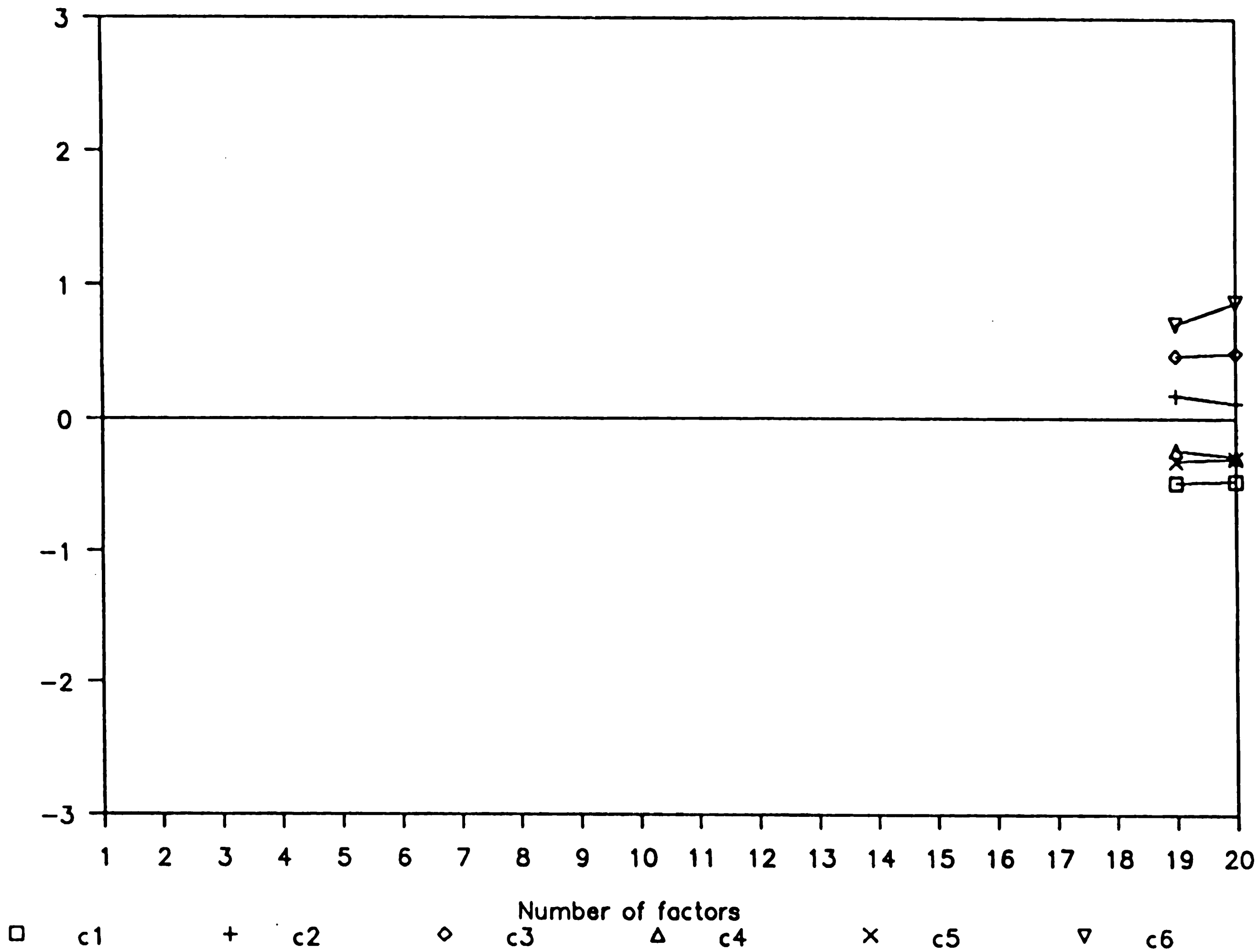


Figure 46. The "factor 19" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

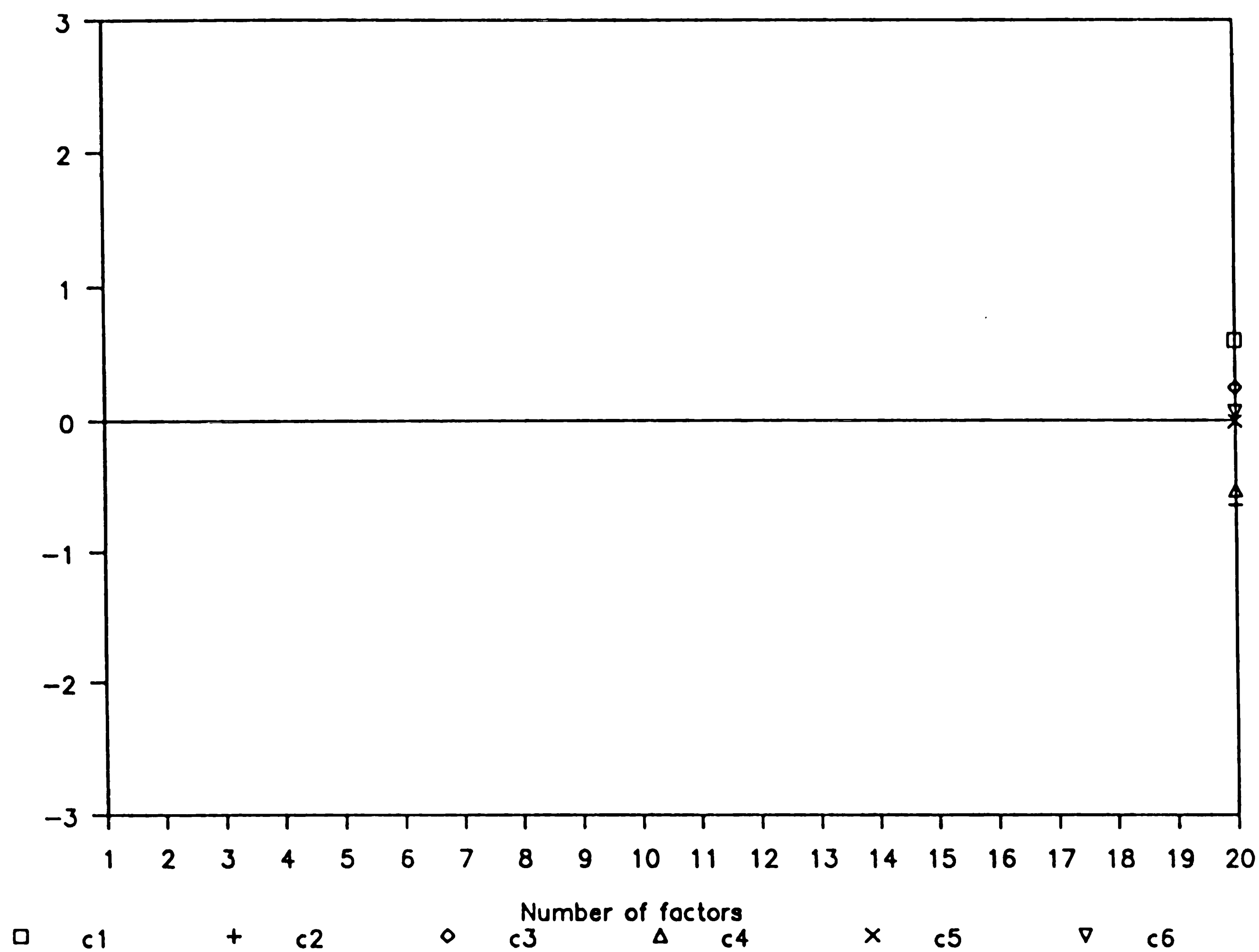


Figure 47. The "factor 20" factor score centroids for six clusters across the different factor solutions when clustering on raw data (c1: cluster 1, c2: cluster 2, c3: cluster 3, c4: cluster 4, c5: cluster 5, c6: cluster 6).

For example, Figure 28 shows the "factor 1" factor score centroids for each of the six clusters for each of the 19 factor solutions.

The graphs show that factor score centroids of the different clusters based on raw data do not differ significantly across the different factor solutions. For example, Figure 28 shows that the "factor 1" factor score centroid for each of the six clusters is relatively stable across the different factor solutions. In comparison, the factor score centroids of clusters formulated on the basis of factor scores differ significantly across the different factor solutions (see Figure 6). Figure 48 compares the cluster centroid stability across factor solutions (20, 19, ..., 2) for clusters based on factor scores and raw data. They reveal that the cluster centroids/membership is more stable when clustering is based on raw data.

The sum of squared distance between centroid points was calculated for each of the six clusters for each of the two clustering approaches (i.e., raw data and factor scores). Table 62 reports the sum of squared distance for each of the six clusters for 18 different factor scores centroids.

The sum of squared distances were used as input to a computer program (see discussion in Chapter III, page 63) to determine the best set of matched clusters between clusters formulated on factor scores and clusters formulated on raw data. The results are also shown in Table 62. The table shows which clusters are most similar. For example, cluster 1 (based on factor scores) is most similar to cluster 5 (based on raw data) for the "factor 1" factor score centroid pattern.

Within the best set of matched clusters for different factors, standard deviations of 18 different factor score centroids were

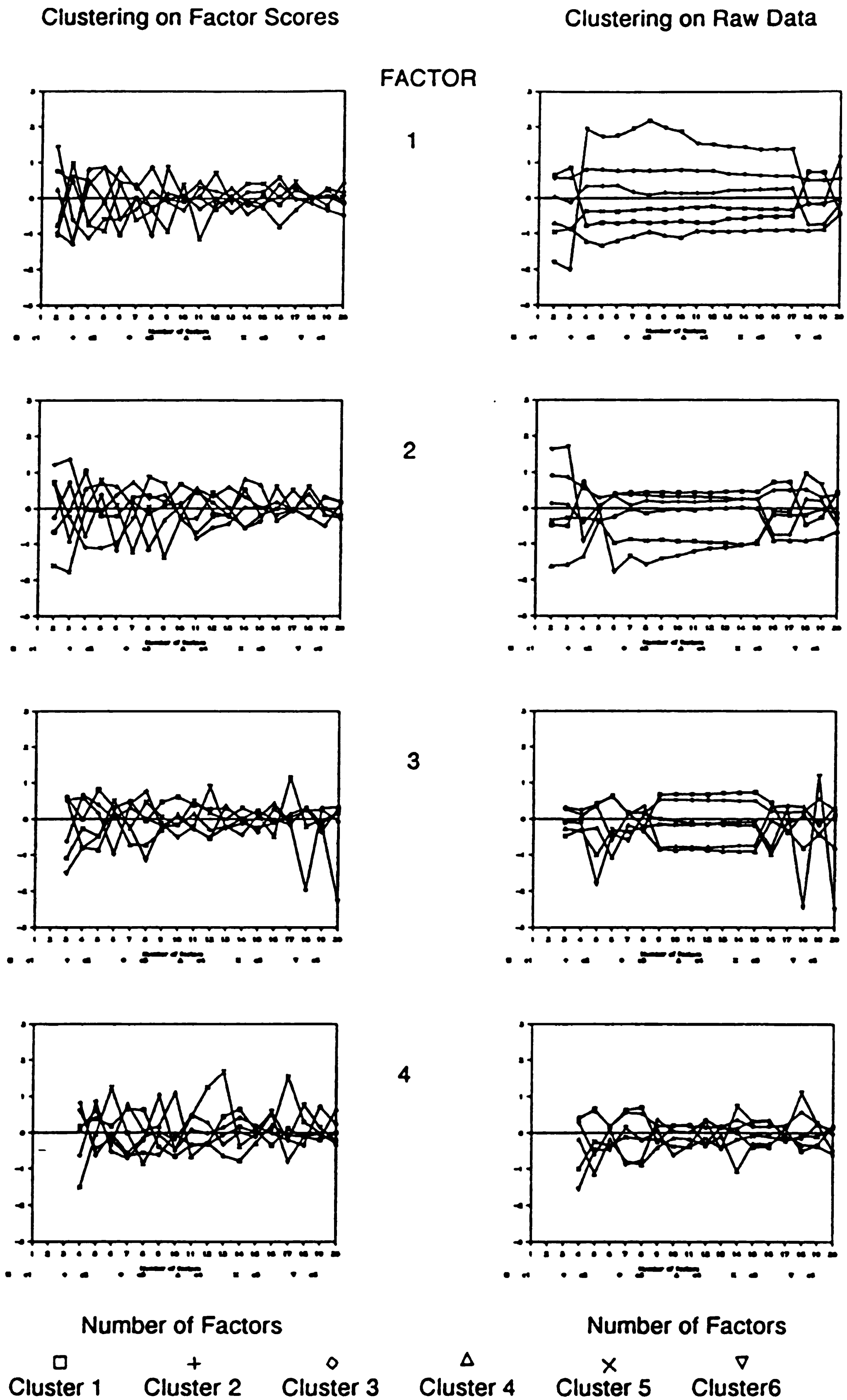


Figure 48. Comparisons of the stability of cluster centroids based on factor scores with clustering based on raw data.

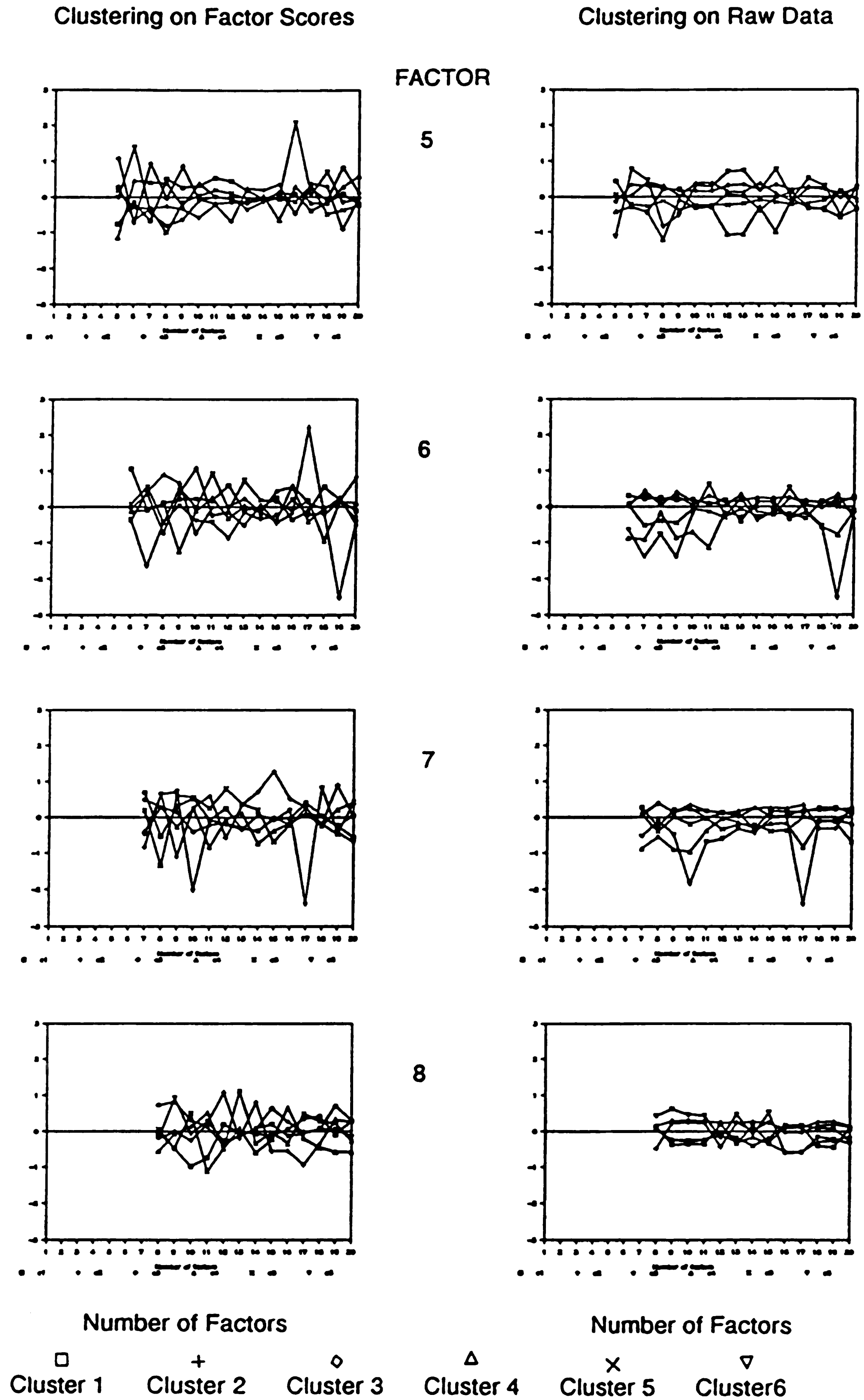


Figure 48 (Cont'd).

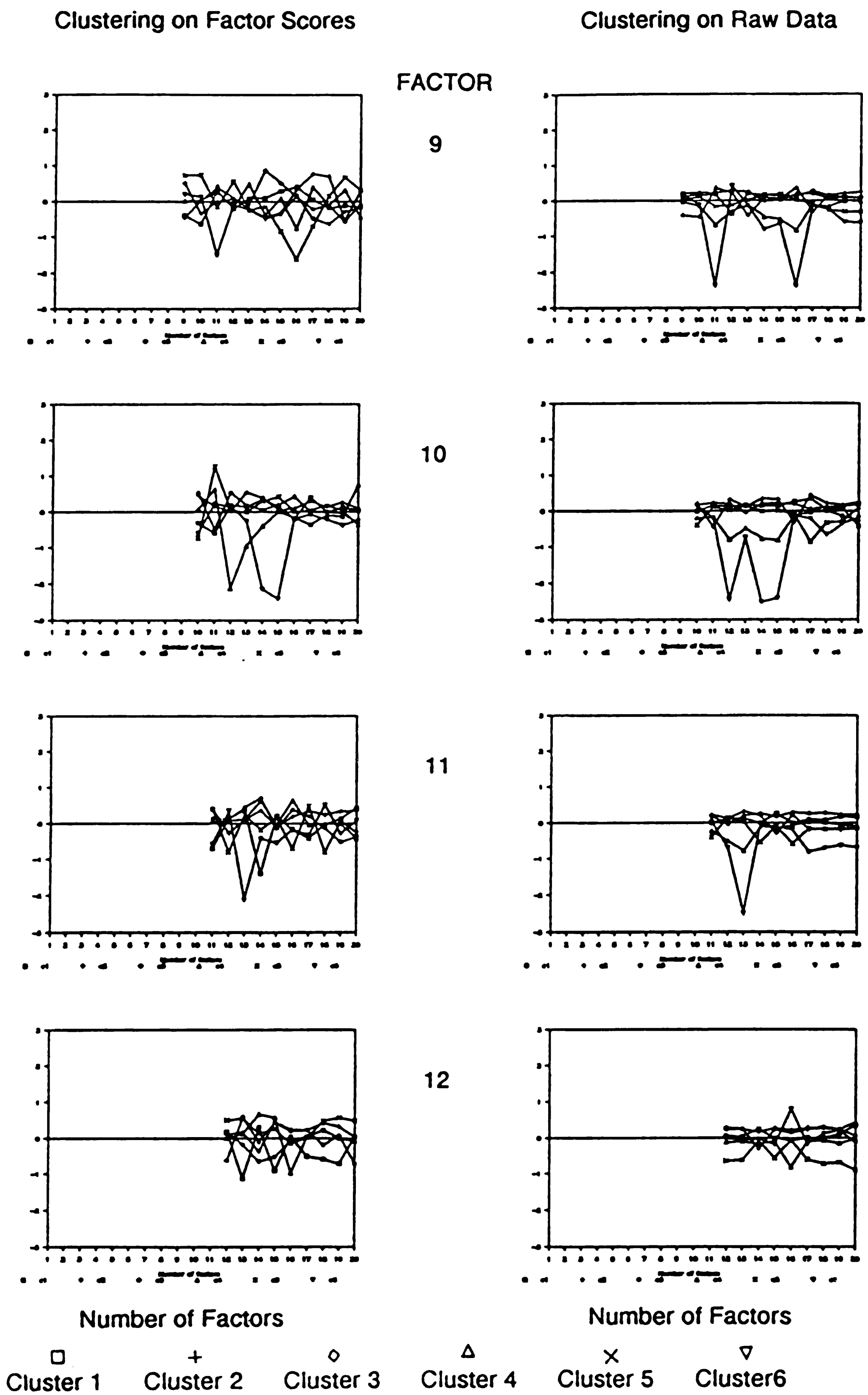


Figure 48 (Cont'd).

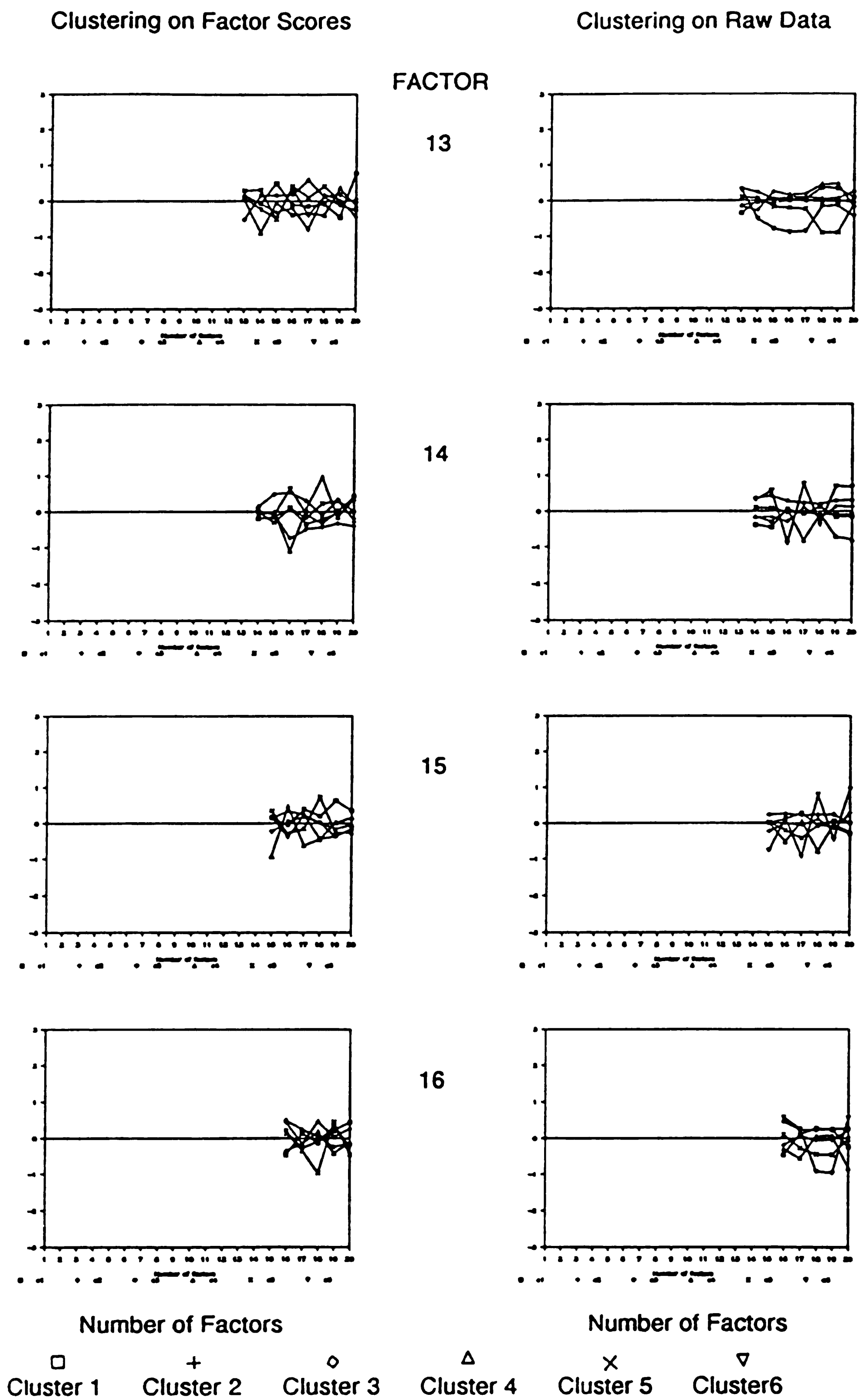


Figure 48 (Cont'd).

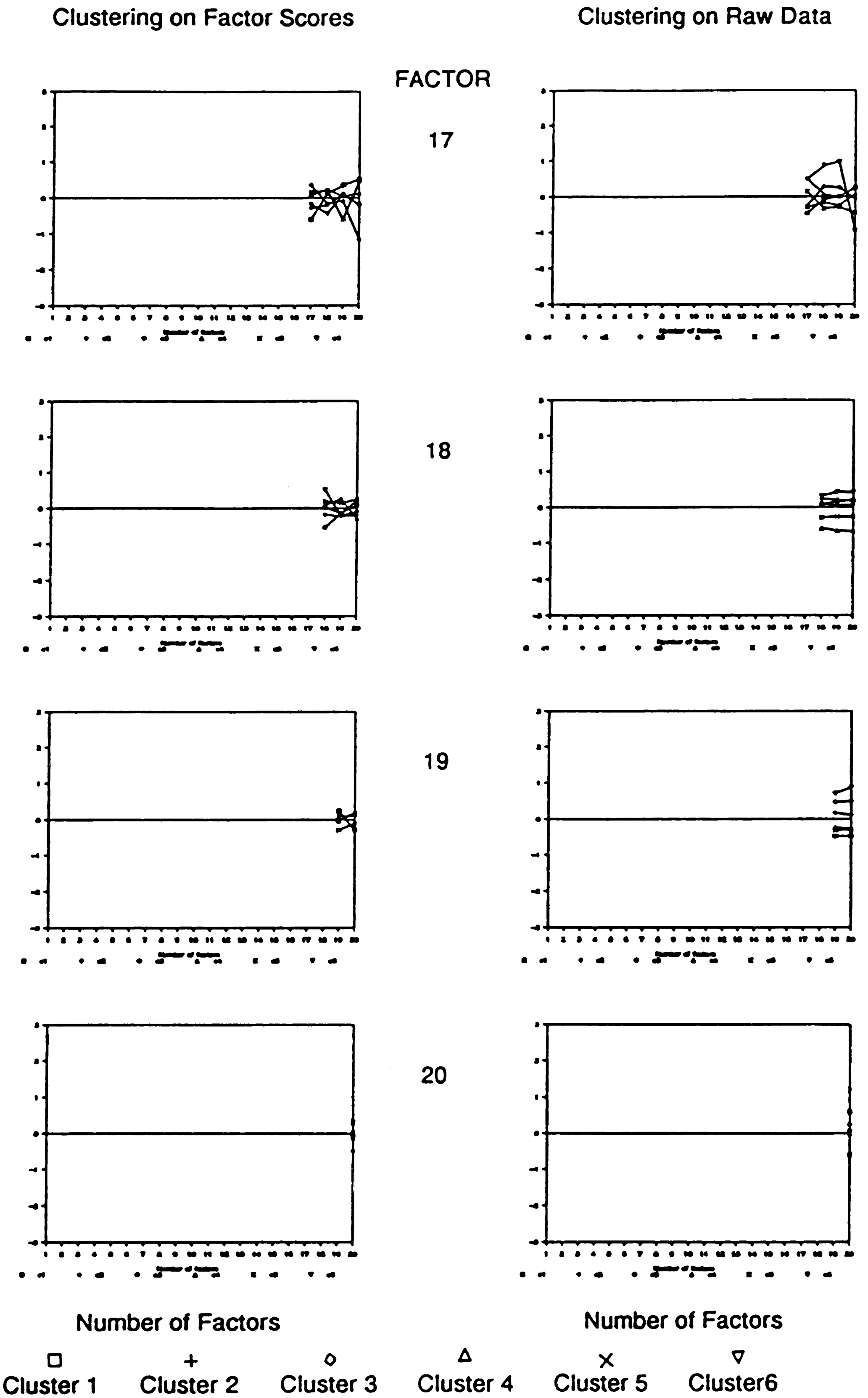


Figure 48 (Cont'd).

Table 62. Comparison of stability of factor score patterns between two approaches.

Cluster Order	Clustering on Factor Scores: Sum of Distance	Clustering on Factor Scores: Standard Deviation	Cluster Order	Clustering on Raw Data: Sum of Distance	Clustering on Raw Data: Standard Deviation	Comparison of Stability
Factor 1						
1	4.612	0.374	5	0.305	0.208	a
2	6.909	0.488	4	0.472	0.186	a
3	6.656	0.466	3	0.084	0.109	a
4	9.453	0.510	2	1.672	0.312	a
5	15.527	0.584	6	20.342	1.197	b
6	12.783	0.620	1	5.685	0.569	a
Factor 2						
1	5.455	0.384	5	1.137	0.367	a
2	4.477	0.426	3	0.290	0.182	a
3	11.008	0.447	2	3.120	0.333	a
4	10.812	0.547	1	4.333	0.401	a
5	11.919	0.670	6	14.028	1.040	b
6	7.999	0.599	4	1.952	0.552	a
Factor 3						
1	3.257	0.290	4	3.152	0.318	b
2	5.611	0.405	3	0.692	0.157	a
3	11.800	0.559	6	41.624	0.914	b
4	5.820	0.432	2	4.486	0.505	b
5	8.807	0.478	1	3.093	0.400	a
6	10.647	0.690	5	2.464	0.280	a
Factor 4						
1	2.926	0.388	5	0.973	0.233	a
2	6.696	0.406	4	4.364	0.330	a
3	10.353	0.500	6	7.702	0.588	b
4	8.567	0.514	1	3.276	0.375	a
5	17.068	0.829	2	4.177	0.378	a
6	4.331	0.413	3	0.858	0.179	a
Factor 5						
1	2.700	0.374	5	0.354	0.127	a
2	5.959	0.410	4	4.697	0.353	a
3	11.189	0.509	6	9.398	0.573	b
4	5.479	0.437	2	0.759	0.154	a
5	2.506	0.314	3	0.510	0.128	a
6	15.803	0.174	1	2.332	0.263	b
Factor 6						
1	3.188	0.342	5	0.915	0.184	a
2	16.430	0.708	6	15.796	0.786	b
3	7.591	0.465	2	1.667	0.222	a
4	5.346	0.407	4	3.180	0.369	a
5	8.322	0.464	3	0.378	0.102	a
6	14.872	0.678	1	1.175	0.198	a

Table 62 (Cont'd).

Cluster Order	Clustering on Factor Scores: Sum of Distance	Clustering on Factor Scores: Standard Deviation	Cluster Order	Clustering on Raw Data: Sum of Distance	Clustering on Raw Data: Standard Deviation	Comparison of Stability
Factor 7						
1	3.642	0.374	5	1.062	0.200	a
2	13.045	0.658	4	1.768	0.332	a
3	0.953	0.256	3	0.373	0.108	a
4	4.180	0.366	2	1.136	0.217	a
5	9.233	0.563	1	0.965	0.175	a
6	31.822	0.900	6	15.005	0.704	a
Factor 8						
1	1.975	0.400	5	0.566	0.189	a
2	6.710	0.543	6	3.072	0.428	a
3	1.867	0.354	4	0.338	0.161	a
4	4.038	0.412	3	0.263	0.117	a
5	10.315	0.550	1	2.084	0.256	a
6	2.966	0.298	2	1.481	0.250	a
Factor 9						
1	1.500	0.290	5	0.202	0.171	a
2	3.886	0.431	4	1.231	0.264	a
3	2.287	0.389	3	0.124	0.095	a
4	4.670	0.357	2	2.164	0.302	a
5	5.048	0.700	1	0.688	0.154	a
6	5.519	0.440	6	21.845	0.022	a
Factor 10						
1	1.449	0.281	5	1.334	0.287	b
2	0.662	0.172	3	0.155	0.080	a
3	2.549	0.311	2	2.172	0.266	a
4	10.712	0.741	6	16.980	0.996	b
5	6.052	0.474	1	0.382	0.187	a
6	9.541	0.840	4	1.128	0.304	a
Factor 11						
1	6.897	0.531	5	1.256	0.221	a
2	1.026	0.214	4	0.639	0.240	b
3	1.491	0.361	3	0.134	0.090	a
4	1.833	0.306	2	0.756	0.222	a
5	7.628	0.469	1	0.178	0.081	a
6	11.581	0.687	6	10.224	0.693	b
Factor 12						
1	7.001	0.487	6	2.201	0.292	a
2	1.176	0.234	5	1.126	0.266	b
3	3.188	0.422	4	1.112	0.229	a
4	3.832	0.442	2	0.672	0.193	a
5	0.619	0.247	3	0.062	0.069	a
6	0.671	0.281	1	0.188	0.127	a

Table 62 (Cont'd).

Cluster Order	Clustering on Factor Scores: Sum of Distance	Clustering on Factor Scores: Standard Deviation	Cluster Order	Clustering on Raw Data: Sum of Distance	Clustering on Raw Data: Standard Deviation	Comparison of Stability
Factor 13						
1	2.348	0.366	6	1.304	0.389	b
2	1.687	0.312	4	0.011	0.036	a
3	0.945	0.314	2	0.791	0.270	a
4	2.476	0.346	1	0.290	0.204	a
5	1.945	0.290	3	0.086	0.104	a
6	1.757	0.275	5	1.122	0.366	b
Factor 14						
1	0.476	0.238	5	0.394	0.194	a
2	0.561	0.237	3	0.044	0.076	a
3	0.647	0.223	2	0.369	0.165	a
4	4.965	0.575	6	7.415	0.592	b
5	0.841	0.222	1	0.394	0.194	a
6	1.302	0.280	4	1.830	0.338	b
Factor 15						
1	0.550	0.213	5	0.277	0.153	a
2	0.845	0.214	3	0.058	0.087	a
3	0.562	0.298	2	0.359	0.170	a
4	3.446	0.483	6	8.554	0.733	b
5	2.221	0.389	4	1.952	0.308	a
6	0.327	0.187	1	0.185	0.110	a
Factor 16						
1	0.604	0.311	3	0.056	0.089	a
2	1.002	0.242	4	1.289	0.354	b
3	0.200	0.266	2	0.151	0.134	a
4	0.630	0.214	1	0.583	0.287	b
5	3.447	0.490	6	3.817	0.698	b
6	0.736	0.237	5	0.353	0.222	a
Factor 17						
1	0.609	0.427	5	0.249	0.225	a
2	0.238	0.160	1	0.215	0.258	b
3	1.398	0.554	3	0.263	0.196	a
4	0.060	0.068	4	0.118	0.141	b
5	1.816	0.399	6	3.890	0.772	b
6	0.439	0.188	2	0.390	0.224	b
Factor Score 18						
1	0.071	0.092	6	0.011	0.048	a
2	0.003	0.020	5	0.000	0.009	a
3	0.289	0.310	4	0.004	0.033	a
4	0.346	0.233	3	0.002	0.019	a
5	0.014	0.041	2	0.000	0.012	a
6	0.527	0.320	1	0.005	0.037	a

Note: Two approaches are (I) clustering on factor scores and (II) clustering on raw data

[a]Clustering on factor scores has a larger standard deviation.

[b]Clustering on raw data has a larger standard deviation.

calculated for each cluster for the two clustering approaches. The results are reported in Table 62. The higher the standard deviation, the more unstable the cluster membership. Overall, the results indicate that the approach of clustering on raw data was better than the approach of clustering on factor scores in terms of cluster membership stability.

CHAPTER V

CONCLUSIONS

The primary purpose of this study was to examine the impact of factor analyses on cluster membership when clustering is based on factor scores. Although many researchers have utilized factor analysis as a prelude to clustering, very few have examined the potential effects of alternative factor solutions (number of factors) on clustering results. The study had three objectives: (1) to assess the effect of different factor solutions (number of factors) on cluster membership, (2) to ascertain the effect of factor rotation on cluster membership, and (3) to compare clustering on factor scores with clustering on raw data. This chapter presents a summary of the study, major conclusions, a discussion of study limitations, and recommendations regarding the combined use of factor analysis and cluster analysis.

Summary of the Study

The study utilized the importance ratings of 20 different campground attributes/facilities collected in a study of the 1988 Michigan Campvention. Respondents ranked the importance of these

a
b

a
a
a
a
A
a
I
m

w
a
A
T
c
c

i
s
a
a
o
s
cl
pr

attributes/facilities on a five-point scale ("1" being crucial and "5" being not important).

Nineteen (20, 19, 18, ..., 2) different principal component analyses with varimax rotation were performed on these data. Cluster analysis was performed on the factor scores from the "20 factor" factor analysis. A six-cluster solution was selected. Cluster analyses were also performed on the factor scores from the other 18 factor analyses. A six-cluster solution was derived for each of the other 18 factor analyses. The stability of cluster membership was compared across the 18 different factor-cluster analyses using an entropy (information) measure.

Nineteen different principal component analyses without rotation were performed on the attributes/facilities data. Cluster analyses were again performed on the factor scores from each of these factor analyses. A six-cluster solution was decided for each factor-cluster analysis. The cluster memberships derived from the nonrotated factor scores were compared (using membership crosstabulation) with the memberships of clusters based on rotated factor scores.

Cluster analysis was performed to group respondents based on the importance they assigned to the 20 different campground attributes. A six-cluster solution was selected. Nineteen principal component analyses with varimax rotation were performed on the 20 campground attributes. Factor score centroids were calculated and graphed for each of the six clusters across different factor solutions. The sum of squared distance for each cluster on each factor was computed for both clustering on raw data and clustering on factor scores. A computer program was utilized to determine the best set of matched clusters between two clustering approaches. The standard deviations of factor score centroids for each cluster across different factor solutions were calculated and used as the basis for comparing the stability of cluster membership derived from clustering on raw data with the stability of cluster membership derived from clustering on factor scores.

Major Conclusions

Three major conclusions were drawn from the analyses. First, when factor analysis is used in conjunction with cluster analysis, the factor solution (number of factors) selected has an effect on the cluster membership. Different factor solutions generate different factor scores, which result in different similarity measures. Different similarity measures lead to different cluster solutions. As a result, cluster membership is very unstable across clustering solutions based on factor scores.

Second, whether or not the initial factors are rotated does not affect cluster membership. Because the original relationship between variables does not change when the initial factors are rotated, the distance measure between cases for each variable in the clustering procedure will not be changed. The difference between clustering on rotated factor scores and clustering on nonrotated factor scores is that clusters will be labeled differently.

Third, clustering on raw data rather than factor scores results in more stable cluster membership. Because factor analysis is used to reduce observed variables into fewer dimensions by means of a linear combination of the observed data, a certain amount of information (percentage of variance explained) will be lost depending on the number of factors selected. Thus, when clustering on the different factor scores, the loss of information will result in significant changes in cluster membership as compared to the cluster membership derived from clustering on raw data (no information is lost).

Although this study identified that alternative factor solutions will affect cluster membership, it does not mean that results of previous studies using factor analysis in conjunction with cluster analysis are methodological and statistical incorrect. However, this study raises some significant concerns about the impact of alternative factor analyses on cluster analysis. These concerns should be incorporated into future studies which utilize combined factor analysis and cluster analysis.

Study Limitations

The study had five major limitations. First, the number of cases that could be analyzed by the clustering software was limited. Not all of the 424 respondents (cases) who rated all 20 campground attributes could be clustered. This required selection of a subsample of 212 cases. As a result, some of the formulated clusters had fewer than 10 members. Calculation of chi-square statistics to compare cluster membership differences was not possible because one or more of the cells in the cluster crosstabulation tables had less than five members.

Second, although considerable thought was given to identify relevant campground attributes, there is no assurance that they represent complete list of all the relevant attributes sought. The problem of identifying relevant attributes is not unique to this study, but is rather inherent in classification, especially attributes and/or benefits sought segmentation studies.

Third, only Ward's method with the squared Euclidean distance was used. Other clustering techniques are available that have different characteristics and procedures. These clustering techniques often yield different clustering results because different similarity measures (for hierarchical clustering methods) and different partitioning rules (for nonhierarchical clustering methods) are used.

Fourth, although the entropy (information) measure was used to assess the stability of cluster membership, no statistical test was used to reject or accept the hypotheses.

Fifth, because the similarity of the six clusters formulated on raw data and clustering on factor scores is uncertain, a computer program was used to identify the best set of matched clusters based on the criterion of minimum total difference of the sum of squared distance. However, there might be more appropriate ways to select the matched clusters.

Recommendations Regarding the Use of Factor Analysis and Cluster Analysis

Six major recommendations are offered regarding the use of factor analysis and cluster analysis. First, when factor analysis is performed as a preliminary step to cluster analysis, they should not be treated as distinct analyses. The findings show that alternative factor solutions will affect the clustering results (cluster membership). Researchers who use factor scores as the basis for clustering should examine the impact of alternative factor solutions on the clustering results. Decisions regarding the number of factors should be based on both the factor analysis criteria (eigenvalues greater than one, percentage of variance explained, scree test) and the impact on the cluster solution.

Second, researchers may first perform cluster analysis based on raw data for classification (segmentation) purposes, and then use factor analysis as a means of describing clusters. Selection of variables (raw data) to be used in cluster analysis should have theoretical support. Also, when many variables are included in the study, researchers should consider alternative methods (e.g., multiple discriminant analysis) to determine which variables can contribute the most to the correct group classification.

Third, the findings indicate that the entropy (information) measure can be used as an indicator of cluster stability. The entropy (information) measure has been commonly used in the fields of marketing, management, finance, accounting, biology, communication, and geography. It has rarely been used in the field of recreation. The results of this study show that the entropy (information) measure provides a good indicator with which to assess the uncertainty of cluster memberships. The information measure can also be used to assess the stability of derived clusters over time.

Fourth, the assessment of the impacts of alternative factor analyses on the clustering results should be repeated with different clustering data, similarity measures, and other clustering techniques that produce different clustering results. Fifth, although a specially designed computer program was used to assess the similarity of clustering results formulated on raw data and factor scores, alternatives to solve the problem of cluster matching should be examined in the future. Finally, the entropy (information) measure was used to assess the stability of cluster membership derived from clustering on raw data and factor scores; however, researchers should investigate appropriate statistical tests to use with the entropy (information) measure.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aczél, J., & Daróczy, Z. (1975). On measures of information and their characterizations. New York: Academic Press.

Aldenderfer, M. S., & Blashfield, R. K. (1984). Cluster analysis. Newbury Park, CA: Sage Publications.

Allen, L. R. (1982). The relationship between Murrary's personality needs and leisure interests. Journal of Leisure Research, 14(1), 63-76.

Anderberg, M. R. (1973). Cluster analysis for applications. New York: Academic Press.

Arbuckle, J., & Friendly, M. L. (1977). On rotating to smooth functions. Psychometrika, 42, 127-140.

Archer, C. O., & Jennrich, R. I. (1973). Standard errors for rotated factor loadings. Psychometrika, 38, 581-592.

Armstrong, J. S., & Soelberg, P. (1968). On the interpretation of factor analysis. Psychological Bulletin, 70(5), 361-364.

Attaran, M., & Guseman, D. (1988). An investigation into the nature of structural changes within the service sector in the U.S. Journal of the Market Research Society, 30(3), 387-396.

Attaran, M., & Zwick, M. (1987). Entropy and other measures of industrial diversification. Quarterly Journal of Business and Economics, 26(4), 17-34.

Bartholomew, D. J. (1985). Foundations of factor analysis: Some practical implications. British Journal of Mathematical and Statistical Psychology, 38(1), 1-10.

Bartko, J. J., Strauss, J. S., & Carpenter, W. T. (1971). An evaluation of taxometric techniques for psychiatric data. Classification Society Bulletin, 2(1), 1-27.

Bartlett, M. S. (1937). The statistical conception of mental factors. British Journal of Psychology, 28, 97-104.

Bartlett, M. S. (1950,June). Tests of significance in factor analysis, British Journal of Statistical Psychology, 3, 77-85.

Bartlett, M. S. (1951,march). A further note on tests of significance in factor analysis. British Journal of Statistical Psychology, 4, 1-2.

Bayne, C. K., Beauchamp, J. J., Begovich, G. L., & Kane, V. E. (1980). Monte Carlo comparisons of selected clustering procedures. Pattern Recognition, 12, 51-62.

Beale, E. M. L. (1969). Euclidean cluster analysis. Bulletin of the International Statistical Institute, 43, 92-94.

Beard, J. G., & Ragheb, M. G. (1983). Measuring leisure motivation. Journal of Leisure Research, 15(3), 219-228.

Beecher, M. D. (1989). Signaling systems for individual recognition: An information theory approach. Animal Behaviour, 38(2), 248-261.

Bieber, S. L., & Smith, D. V. (1986). Multivariate analysis of sensory data: A comparison of methods. Chemical Senses, 11(1), 19-47.

Bishara, H. I. (1984). Aggregate dividend decision making in Canadian life insurance companies. Akron Business and Economic Review, 15(2), 6-14.

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. (1978). The literature on cluster analysis. Multivariate Behavioral Research, 13, 271-295.

Bobko, P., & Schemmer, F. M. (1984). Eigenvalue shrinkage in principal components based factor analysis. Applied Psychological Measurement, 8(4), 439-451.

Boggis, J. G., & Held, I. (1971). Cluster analysis-a new tool in electricity usage studies. Journal of the Market Research Society, 13(2), 49-66.

Browne, M. W. (1968a). A comparison of factor analytic techniques. Psychometrika, 33, 267-334.

Browne, M. W. (1968b). A note on lower bounds for the number of common factors. Psychometrika, 33(2), 233-236.

Calantone, R. J., & Gross, A. C. (1980). The impact of segment dynamics on retail bank advertising strategies. In D. W. Scotton & R. L. Zallocco (Ed.), Readings in market segmentation (pp. 126-142). Chicago, IL: American Marketing Association.

Calantone, R. J., & Johar, J. S. (1984). Seasonal segmentation of the tourism market using a benefit segmentation framework. Journal of Travel Research 23(2), 14-24.

Calantone, R., Schewe, C., & Allen, C. T. (1980). Targeting specific advertising messages at tourist segments. In Hawkings, Shafer, and Rovelstad (Eds.), Tourism marketing and management issues (pp. 149-160). Washington, DC: George Washington University Press.

Carroll, J. B. (1953). An analytic solution for approximating simple structure in factor analysis. Psychometrika, 18, 23-38.

Cattell, R. B. (1966). The scree test for the number of factors. Multivariate Behavioral Research, 1(2), 245-276.

Collins, L. M., Cliff, N., & Cudeck, R. A. (1983). Patterns of crime in a birth cohort. Multivariate Behavioral Research, 18(3), 235-258.

Comrey, A. L. (1973). A first course in factor analysis. New York: Academic Press.

Connelly, N. A. (1987). Critical factors and their threshold for camper satisfaction at two campgrounds. Journal of Leisure Research, 19(3), 159-173.

Coovert, M. D., & McNelis, K. (1988). Determining the number of common factors in factor analysis: A review and program. Educational and Psychological Measurement, 48(3), 678-692.

Crask, M. R. (1980). Segmenting the vacation market: Identifying the vacation preferences, demographics, and magazine readership of each group. Journal of Travel Research, 20(2), 29-34.

Davis, D., Allen, J., & Cosenza, R. M. (1988). Segmenting local residents by their attitudes, interests, and opinions toward tourism. Journal of Travel Research, 27(2), 2-8.

Day, E., Fox, R. J., & Huszagh, S. M. (1988). Segmenting the global market for industrial goods: Issues and implications. International Marketing Review, 5(3), 14-27.

Day, G. S., & Heeler, R. M. (1971). Using cluster analysis to improve marketing experiments. Journal of Marketing Research, 8, 340-347.

DeSarbo, W. S., Carroll, J. D., & Clark, L. A. (1984). Synthesized clustering: A method for amalgamating alternative clustering bases with differential weighting of variables. Psychometrika, 49, 57-78.

Devall, B., & Garry, J. (1981). Who hates whom in the great outdoors: The impact of recreational specialization and technologies of play. Leisure Science, 4(4), 399-418.

Dielman, T. E., Cattell, R. B., & Wagner, A. (1972). Evidence on the simple structure and factor invariance achieved by five rotational methods on four types of data. Multivariate Behavioral Research, 7, 223-231.

Ditton, R. B., Goodale, T. L., & Jonsen, P. K. (1975). A cluster analysis of activity, frequency, and environment variables to identify water-based recreation types. Journal of Leisure Research, 7(4), 282-295.

Donderi, D. C. (1988). Information measurement of distinctiveness and similarity. Perception and Psychophysics, 44(6), 576-584.

Dreger, R. M., Fuller, J., & Lemoine, R. L. (1988). Clustering seven data sets by means of some or all of seven clustering methods. Multivariate Behavioral Research, 23(2), 203-230.

Driver, H. E., & Kroeber, A. L. (1932). Quantitative expression of cultural relationships. Archaeology and Ethnology, 31, 211-256.

Dubes, R., & Jain, A. K. (1979). Validity studies in clustering methodologies. Pattern Recognition, 11, 235-254.

Dulewicz, S. V., & Keenay, G. A. (1979). A practically oriented and objective method for classifying and assigning senior jobs. Journal of Occupational Psychology, 52(3), 155-166.

Edelbrock, C. (1979). Comparing the accuracy of hierarchical clustering algorithms: The problem of classifying everybody. Multivariate Behavioral Research, 14, 367-384.

Edelbrock, C., & McLaughlin, B. (1980). Hierarchical cluster analysis using intraclass correlations: A mixture model study. Multivariate Behavioral Research, 15, 299-318.

Ellis, G. D., & Rademacher, C. (1987). Development of a typology of common adolescent free time activities: A validation and extension of Kleiber, Larson, and Csikszentminalyi. Journal of Leisure Research, 19(4), 284-292.

Everitt, B. S. (1974). Cluster analysis. London: Heinemann Educational Books Ltd.

Everitt, B. S. (1979). Unresolved problems in cluster analysis. Biometrics, 35(1), 169-181.

Frank, R. E., & Green, P. E. (1968). Numerical taxonomy in marketing analysis: A review article. Journal of Marketing Research, 5(1), 83-98.

Funkhouser, G. R. (1983). A note on the reliability of certain clustering algorithms. Journal of Marketing Research, 20(1), 99-102.

Furse, D. H., Punj, G. N., & Stewart, D. W. (1984). A typology of individual search strategies among purchasers of new automobiles. Journal of Consumer Research, 10(4), 417-431.

Garrison, C., & Paulson, A. (1973). An entropy measure of the geographic concentration of economic activity. Economic Geography, 49, 319-324.

Gartner, W. B. (1990). What are we talking about when we talk about entrepreneurship? Journal of Business Venturing, 5, 15-28.

Gau, G. W. (1978). A taxonomic model for the risk-rating of residential mortgages. Journal of Business, 51(4), 687-706.

Gnanadesikan, R., & Wilk, M. B. (1969). Data analytic methods in multivariate statistical analysis. In P. R. Krishnaiah (Ed.), Multivariate analysis II (pp. 593-638). New York: Academic Press.

Goodrich, J. N. (1980). Benefit segmentation of U.S. international travelers: An empirical study with American Express. In Hawkins, Shafer, and Rovelstad (Eds.), Tourism marketing and management issues (pp. 133-147). Washington, DC: George Washington University Press.

Gorman, B. S. (1983). The complementary use of cluster and factor analysis methods. Journal of Experimental Education, 51(4), 165-168.

Gorsuch, R. L. (1983). Factor analysis. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Green, D. W., Sommers, M. S., & Kernan, J. B. (1973). Personality and implicit behavior patterns. Journal of Marketing Research, 10, 63-69.

Green, P. E., Frank, R. E., & Robinson, P. J. (1967). Cluster analysis in test market selection. Management Science, 13(8), 387-400.

Green, P. E., & Rao, V. R. (1969). A note on proximity measures and cluster analysis. Journal of Marketing Research, 6(3), 359-364.

Hair, J. F., Anderson, R. E., & Tatham, R. L. (1987). Multivariate data analysis with readings. New York: Macmillan.

Hakstian, A. R. (1976). Two-matrix orthogonal rotation procedures. Psychometrika, 41, 267-272.

Hakstian, A. R., & Muller, V. J. (1973). Some note on the number of factors problem. Multivariate Behavioral Research, 8(4), 461-475.

Hamer, R., & Cunningham, J. (1981). Cluster analyzing profile data confounded with interrater differences: A comparison of profile association measures. Applied Psychological Measurement, 5, 63-72.

Harraigan, K. R. (1985). An application of clustering for strategic group analysis. Strategic Management Journal, 6(1), 55-73.

Harris, M. L., & Harris, C. W. (1971). A factor analytic interpretation strategy. Educational and Psychological Measurement, 31, 589-606.

Hautaluoma, J., & Brown, P. J. (1979). Attributes of the deer hunting experience: A cluster-analytic study. Journal of Leisure Research, 10(4), 271-287.

Hawes, D. K. (1988). Travel-related lifestyle profile of older women. Journal of Travel Research, 27(2), 22-32.

Heeler, R. M., Whipple, T. W., & Hustad, T. P. (1977). Maximum likelihood factor analysis of attitude data. Journal of Marketing Research, 14(1), 42-51.

Hegwood, J. L. (1987). Experience preferences of participants in different types of river recreation groups. Journal of Leisure Research, 19(1), 1-12.

Henderson, K. A., & Stalnaker, D. (1988). The relationship between barriers to recreation and gender-role personality trait for women. Journal of Leisure Research, 20(1), 69-80.

Hollender, J. W. (1977). Motivational dimensions of the camping experience. Journal of Leisure Research, 9(2), 133-141.

Hooper, M. (1985). A multivariate approach to the measurement and analysis of social identity. Psychological Report, 57(1), 315-325.

Horn, J. L. (1965a). An empirical comparison of methods for estimating factor scores. Educational and Psychological Measurement, 25(2), 313-322.

Horn, J. L. (1965b). A rationale and test for the number of factors in factor analysis. Psychometrika, 30(2), 179-185.

Horst, P. (1965). Factor analysis of data matrices. New York: Holt, Rinehart & Winston.

Humphrey, A. B., Buechner, J. S., & Velicer, W. F. (1987). Differentiating geographic areas by socioeconomic characteristics. Northeast Journal of Business & Economics, 13(2), 47-64.

Huszagh, S. M., Fox, R. J., & Day, E. (1985). Global marketing: An empirical investigation. Columbia Journal of World Business, 20(4), 31-43.

Jones, D. S. (1979). Elementary information theory. Oxford: Clarendon Press.

Jones, F. L. (1968). Social area analysis: Some theoretical and methodological comments illustrated with Australian data. British Journal of Sociology, 19, 424-444.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. Psychometrika, 43, 443-477.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Kaiser, H. F. (1970,December). A second generation little jiffy. Psychometrika, 35, 401-415.

Kaiser, H. F., & Rice, J. (1974,Spring). Little jiffy mark IV. Educational and Psychological Measurement, 34, 111-117.

Kass, R. A., & Tinsley, H. E. A. (1979). Factor analysis. Journal of Leisure Research, 11(2), 120-138.

Kiel, G. C., & Layton, R. A. (1981). Dimensions of consumer information seeking behavior. Journal of Marketing Research, 18(2), 233-239.

Kikuchi, H. (1986). Segmenting Michigan's sport fishing market: Evaluation of two approaches. Unpublished doctoral dissertation, Michigan State University.

Kim, J. O., & Mueller, C. W. (1989). Factor analysis: Statistical methods and practical issues. Newbury Park, CA: Sage Publications.

Kim, L., & Lim, Y. (1988). Environment, generic strategies, and performance in a rapidly developing country: A taxonomic approach. Academy of Management Journal, 31(4), 802-827.

Knopp, T. B., & Merriam, L. C. (1979). Toward a more direct measure of river use preferences. Journal of Leisure Research, 11(4), 317-326.

Kotler, P. (1984). Marketing management: Analysis, planning, and control. Englewood Cliffs, New Jersey: Prentice-Hall.

Krazanowski, W. J., & Lai, Y. T. (1988). A criterion for determining the number of groups in a data set using sum-of-square clustering. Biometrics, 44(1), 23-34.

Krippendorff, K. (1986). Information theory: Structural models for qualitative data. Newburg Park, CA: Sage Publications.

Krzystofiak, F., Newman, J. M., & Anderson, G. (1979). A quantified approach to measurement of job content: Procedures and payoffs. Personnel Psychology, 32(2), 341-358.

Lathrop, R. G. (1987). The reliability of inverse scree tests for cluster analysis. Educational and Psychological Measurement, 47(4), 953-959.

Lesser, J. A. (1988). Entropy and the prediction of consumer behavior. Behavioral Science, 33(4), 282-291.

Lessig, V. P., & Tollefson, J. D. (1971). Market segmentation through numerical taxonomy. Journal of Marketing Research, 8, 480-487.

Lindell, M. K., & St. Clair, J. B. (1980). Tukknife: A jacknife supplement to canned statistical packages. Educational and Psychological Measurement, 40, 751-754.

Lounsbury, J. W., & Hoopes, L. L. (1988). Five-year stability of leisure activity and motivation factors. Journal of Leisure Research, 20(2), 118-134.

Love, J. (1987). Commodity concentration and export instability: The choice of concentration measure and analytical framework. Journal of Developing Areas, 21(1), 63-74.

Mahoney, E. M., Oh, I. K., & Ou, S. J. (1989). A study of the National Campers and Hikers Association's 1988 Michigan Campvention. Dept. of Parks and Recreation Resources, Michigan State University.

Manfredo, M. J., Driver, B. L., & Brown, P. J. (1983). A test of concepts inherent in experience based setting management for outdoor recreation areas. Journal of Leisure Research, 15(3), 263-283.

Mark, J. H. (1980). Identifying neighborhoods for preservation and renewal: Comment. Growth and Change, 11(4), 47-48.

Marriott, F. H. C. (1971). Practical problems in a method of cluster analysis. Biometrics, 27(3), 501-514.

Mazanec, J. A. (1984). How to detect travel market segments: A clustering approach. Journal of Travel Research, 23(1), 17-21.

McDonald, R. P., & Burr, E. J. (1967). A comparison of four methods of constructing factor scores. Psychometrika, 32(4), 381-401.

McIntyre, R. M., & Blashfield, R. K. (1980). A nearest-centroid technique for evaluating the minimum-variance clustering procedure. Multivariate Behavioral Research, 15(2), 225-238.

Meade, N. (1987). Strategic positioning in the UK car market. European Journal of Marketing, 21(5), 43-56.

Milligan, G. W. (1980). An examination of the effects of six types of error perturbation on fifteen clustering algorithms. Psychometrika, 45, 325-342.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. Psychometrika, 50(2), 159-179.

Mojena, R. (1977). Hierarchical grouping methods and stopping rules: An evaluation. The Computer Journal, 20(4), 359-363.

Moojjaart, A. (1985). Factor analysis for non-normal variables. Psychometrika, 50, 323-342.

Norusis, M. J. (1988). Spss/pc+ advanced statistics V2.0. Chicago, IL: SPSS Inc.

Oh, I. K. (1990). Evaluation of the effectiveness of a camping refund offer and the relationships among campers' characteristics. Unpublished doctoral dissertation, Michigan State University.

Overall, J. E. (1964). Note on multivariate methods for profile analysis. Psychological Bulletin, 61(3), 195-198.

Perreault, W. D., Darden, D. K., & Darden, W. R. (1977). A psychographic classification of vacation life styles. Journal of Leisure Research, 9(3), 208-225.

Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. Journal of Marketing Research, 20, 134-48.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association, 66, 846-850.

Rescorla, L. (1988). Cluster analytic identification of autistic preschoolers. Journal of Autism and Developmental Disorders, 18, 475-492.

Rohlf, F. J. (1970). Adaptive hierarchical clustering schemes. Systematic zoology, 19, 58-82.

Sampson, P., & Pergentino, de Fmendes de Almeida. (1979). A note on selecting the appropriate factor analytic solution from several available. European Research, 7(5), 212-217.

Saunders, D. R. (1961). The rationale for an oblimax method of transformation in factor analysis. Psychometrika, 26, 317-324.

Saunders, J. A. (1985). Cluster analysis for market segmentation. European Journal of Marketing, 14, 422-435.

Schaninger, C. M., & Buss, W. C. (1986). Removing response-style effects in attribute-determinance ratings to identify market segments. Journal of Business Research, 14(3), 237-252.

Sethi, S. P. (1971). Comparative cluster analysis for world markets. Journal of Marketing Research, 8, 348-354.

Shannon, C. E. (1948a). A mathematical theory of communication. Bell System Technology Journal, 27, 379-423.

Shannon, C. E. (1948b). A mathematical theory of communication. Bell System Technology Journal, 27, 623-656.

Shoemaker, S. (1989). Segmentation of the senior pleasure travel market. Journal of Travel Research, 27(3), 14-21.

Shutty, M. S., & DeGood, D. E. (1987). Cluster analyses of responses of low-back pain patients to the SCL-90: Comparison of empirical versus rationally derived subscales. Rehabilitation Psychology, 32(3), 133-144.

Skinner, H. A. (1978). Differentiating the contribution of elevation, scatter, and shape in profile similarity. Educational and Psychological Measurement, 38(2), 297-308.

Skinner, H. A. (1979). Dimensions and clusters: A hybrid approach to classification. Applied Psychological Measurement, 3(3), 327-341.

Smith, S. L. J. (1989). Tourism analysis: A handbook. New York: John Wiley & Sons.

Sneath, P., & Sokal, R. (1973). Numerical taxonomy. San Francisco: W. H. Freeman.

Sokal, R., & Rohlf, F. (1962). The comparison of dendrograms by objective methods. Taxon, 11, 33-40.

Sokal, R., & Sneath, P. (1963). Principles of numerical taxonomy. San Francisco: W. H. Freeman.

Sorce, P., Tyler, P. R., & Loomis, L. M. (1989). Lifestyles of older Americans. The Journal of Service Marketing, 3(4), 37-47.

Stanley, T. J., Powell, T., & Danko, W. D. (1987). Trust marketing: Courting a segmented market. Trusts & Estates, 126(11), 14-20.

Starr, M. K. (1980). Some new fundamental considerations of variety-seeking behavior. Behavioral Science, 25(3), 171-179.

Stewart, D. W. (1981). The application and misapplication of factor analysis in marketing research. Journal of Marketing Research, 18(1), 51-62

Stynes, D. J. (1983). Marketing Tourism. Leisure today. Journal of Physical Education, Recreation and Dance, 54(4), 21-23.

Stynes, D. J., & Mahoney, E. M. (1980). Michigan downhill ski marketing study: Segmenting active skiers. (Research Report No. 391). East Lansing: Michigan State University, Agricultural Experiment Station.

Swinyard, W. R., & Struman, K. D. (1986). Market segmentation: Finding the heart of your restaurant's market. Cornell Hotel & Restaurant Administration Quarterly, 27(1), 89-96.

Tatham, R. L., & Dornoff, R. J. (1971). Marketing segmentation for outdoor recreation. Journal of Leisure Research, 3(1), 5-16.

Thorndike, R. L. (1953). Who belongs in a family? Psychometrika, 18, 267-276.

Thurstone, L. L. (1935). The vectors of mind. Chicago: University of Chicago Press.

Tinsley, H. E. A., & Johnson, T. L. (1984). A preliminary taxonomy of leisure activities. Journal of Leisure Research, 16(3), 234-244.

Tinsley, H. E. A., & Kass, R. A. (1979). The latent structure of the need satisfying properties of leisure activities. Journal of Leisure Research, 11(4), 278-291.

Tucker, L. R. (1971). Relations of factor score estimates to their use. Psychometrika, 36(4), 427-436.

Tucker, L. R., Koopman, R. F., & Linn, R. C. (1969). Evaluation of factor-analytic research procedures by means of simulated correlation matrices. Psychometrika, 34, 421-459.

Velicer, W. F. (1976a). Determining the number of components from the matrix of partial correlations. Psychometrika, 41, 321-327.

Velicer, W. F. (1976b). The relation between factor score estimates, image scores, and principal component scores. Educational and Psychological Measurement, 36, 149-159.

Wahlers, R. G., & Etzel, M. J. (1985). Vacation preference as a manifestation of optimal stimulation and lifestyle experience. Journal of Leisure Research, 17(4), 283-295.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58, 236-244.

Williams, W. (1971). Principles of clustering. Annual Review of Ecology and Systematics, 2, 303-326.

Wind, Y. (1978). Issues and advances in segmentation research. Journal of Marketing Research, 15, 317-337.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. Multivariate Behavioral Research, 5, 329-350.

Wolfe, J. H. (1978). Comparative cluster analysis of patterns of vocational interest. Multivariate Behavioral Research, 13(1), 33-44.

Woodside, A. G., & Motes, W. H. (1981). Sensitivity of market segments to separate advertising strategies. Journal of Marketing, 45, 63-73.

Zubin, J. (1938). A technique for measuring like-mindedness. Journal of Abnormal Social Psychology, 33, 508-516.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. Psychological Bulletin, 99, 432-442.

APPENDIX A
Pretrip Questionnaire

Appendix A: Pretrip Questionnaire

1988 MICHIGAN CAMPVENTION STUDY

Michigan State University, Michigan Division of State Parks, Michigan Association of Private Campground Owners, and the National Campers and Hikers Association are conducting a comprehensive study of persons who attend the 1988 MICHIGAN CAMPVENTION being held at Highland Recreation Area. The study will provide information which will be useful in decisions regarding future campventions.

We will also be sending you another brief questionnaire after you return home from your trip to gather information on your satisfaction with the 1988 Campvention and camping in Michigan.

If you are planning to attend the 1988 Campvention PLEASE COMPLETE the following questionnaire and RETURN it to us in the attached postage paid envelope. PLEASE take the time to complete the questionnaire. Without your help the study will not be successful. We guarantee that your response will remain strictly confidential.

---

1. DATE YOU COMPLETED this QUESTIONNAIRE ________/_____/_____(MONTH/DAY/YEAR)

2. Will the 1988 Michigan CAMPVENTION be the FIRST National Campers and Hikers CAMPVENTION you have attended?

____ Yes (the 1988 will be my FIRST CAMPVENTION) (GO TO QUESTION 4)

____ No → Did you attend the 1987 IOWA CAMPVENTION? ____ Yes ____ No (GO TO QUESTION 4)

3. ON YOUR 1987 IOWA CAMPVENTION TRIP how many nights did you spend on:

3a) On your entire CAMPVENTION TRIP (This includes nights at the Campvention, nights in Iowa before and after the Campvention, and nights in other states traveling to and from the Campvention)
________ Number of total nights away from home

3b) At the Iowa CAMPVENTION SITE: _____Number of nights

3c) At campgrounds in Iowa (OTHER THAN THE CAMPVENTION SITE): ______ Number of nights at other campgrounds

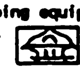
3d) At campgrounds OUTSIDE Iowa: ______ Number of nights
*** (3a should equal the SUM of 3b, 3c, and 3d) ***

********************** 1988 MICHIGAN CAMPVENTION QUESTIONS **********************

4. At the 1988 MICHIGAN CAMPVENTION how many persons including yourself will STAY ON YOUR SITE with you ? ______

4a) What are the AGES of the persons who will stay on your site? YOURSELF _____ , Person 2 _____ , Person 3 ____, Person 4 _____ , Person 5 _____ , Person 6_____, Person 7 ______.

5. On your MICHIGAN CAMPVENTION TRIP what type of camping equipment will you utilize?

____ Tent  ____ Camping Trailer  ____ Travel Trailer 

____ Motor Home  ____ Van/Bus Conversion  ____ 5th wheeler 

____ Other

6. On your MICHIGAN CAMPVENTION TRIP how many TOTAL NIGHTS AWAY FROM HOME will you spend? This includes: nights at the Campvention, nights in Michigan before and after the Campvention and, nights in other states traveling to and from the Campvention.

____ Total 1988 CAMPVENTION TRIP nights

7. How many nights are you planning to camp at the (Michigan) CAMPVENTION SITE located in Highland Recreation Area? (The CAMPVENTION will last 7 nights)

_____ Number of nights at the Michigan CAMPVENTION SITE

8. Other than the nights at the CAMPVENTION SITE are you planning to camp additional nights in MICHIGAN either before or after the Campvention?

___ No (GO TO QUESTION 15)

___ Yes → How many ADDITIONAL nights (not counting nights at the CAMPVENTION SITE) are you planning to camp in Michigan ? _____ Number of additional nights (GO TO QUESTION 9)

9. Will you likely SELECT THE CAMPGROUND(s), or have you already selected the campground(s) (OTHER THAN CAMPVENTION SITE) you will stay at in Michigan BEFORE LEAVING HOME on the trip?

____ No (GO TO QUESTION 11) ____ Yes (GO TO QUESTION 9a)

9a) Have you already selected the campground(s) (OTHER THAN CAMPVENTION SITE) you will stay at in Michigan ?

___ No

___ Yes → How many Michigan campgrounds have you already selected ? ________ number of campgrounds

10. WILL you make, or have you already made, reservations at these campgrounds (OTHER THAN CAMPVENTION SITE) before leaving home on the trip?

____ No

____ Yes → 10a) Have you ALREADY made reservations at campgrounds in Michigan ? ____ No ____ Yes

11. In your registration package there is an offer for a $1.00 OFF REFUND for each night you spend camping at a Michigan State Park or campground which is a member of Michigan Association of Private Campground Owners (MAPCO). The refund offer will not apply to other Michigan campgrounds OR nights at the campvention site.

WILL YOU LIKELY TAKE ADVANTAGE OF THIS OFFER?

___ No Why not?____

___ Yes ____

12. What PRIMARY SOURCE(S) information will you rely on most to select the campground(s) OTHER THAN THE CAMPVENTION SITE you will stay at in Michigan? (Please check all that apply)

___ Rand McNally Camping Directory
___ Woodalls Camping Directory
___ Michigan Campground Directory
___ Trailer Life
___ Highway signs
___ Campground brochures
___ Recommendations from other campers at CAMPVENTION
___ Recommendations from campers you meet in Michigan campgrounds
___ Recommendations of friends & relatives
___ Past camping experience in Michigan
___ Other (specify) ____________________

13. CIRCLE THE NUMBERS (1-6) on the map at the right TO SHOW THE REGIONS of Michigan YOU PLAN TO CAMP IN while on YOUR 1988 CAMPVENTION TRIP. CIRCLE the numbers of ALL REGIONS you are planning to camp in.

*ONLY CIRCLE REGION 1 IF YOU PLAN TO CAMP AT CAMPGROUNDS (OTHER THAN CAMPVENTION SITE) in this region.

14. Have you already written or called, or do you plan to write or call, for additional Michigan travel/recreational information?

___ No
___ Yes → 14a) Which Organization(s) have you written or called, or will you write or call for more information?

___ Michigan Travel Bureau
___ Michigan Dept. of Natural Resources
___ East Michigan Tourism Organization
___ Southeast Michigan Tourism Organization
___ West Michigan Tourism Association
___ Southwest Michigan Tourism Association
___ Upper Peninsula Tourism Association
___ Other (Specify) ____________

15. Please rate the IMPORTANCE of the following CAMPGROUND ATTRIBUTES AND/OR FACILITIES WHEN SELECTING A CAMPGROUND?

CAMPGROUND ATTRIBUTES	Crucial	Very Important	Important	Somewhat Important	Not Important
Large sites	___	___	___	___	___
Shaded Sites	___	___	___	___	___
Cleanliness	___	___	___	___	___
Quietness	___	___	___	___	___
Site Privacy	___	___	___	___	___
Security	___	___	___	___	___
Hospitality of campground staff	___	___	___	___	___
Low Price	___	___	___	___	___
Flush toilets	___	___	___	___	___
Electricity	___	___	___	___	___
Showers	___	___	___	___	___
Laundromat	___	___	___	___	___
Campground store	___	___	___	___	___
Water hookups	___	___	___	___	___
Sewer hookups	___	___	___	___	___
Natural surrounding	___	___	___	___	___
Situated on a lake/stream	___	___	___	___	___
Hiking trails	___	___	___	___	___
Pool	___	___	___	___	___
Playgrounds	___	___	___	___	___

16. Do you USUALLY prefer to camp in public or private (commercial) campgrounds?

_____ Public campground _____ Private (commercial) campground _____ No preference

17. Who is USUALLY MOST INFLUENTIAL in deciding which campgrounds you stay at?

_____ Myself _____ My spouse _____ Children _____ Family (Group) decision _____ Other

18. Approximately how many nights did you camp LAST YEAR (1987)? _____ (If you didn't camp, write "0" on the line)

19. How many of these nights were OUTSIDE THE STATE WHERE YOU LIVE? _____ (If none, write "0" on the line)

20. How many states (not including your home state) did you camp in during 1987? _____ (If no other states write "0")

21. Do you USUALLY camp BEFORE Memorial Day ? ____ No ____ Yes

22. Do you USUALLY camp AFTER Labor Day? ____ No ____ Yes

23. Have you EVER camped in Michigan ? ____ No ____ Yes → When was the last year you camped in MICHIGAN? 19____

24. Based on your impressions, experience, information from others, or travel/camping literature, please complete the following perception of MICHIGAN CAMPGROUNDS which include public and private campgrounds.

Michigan campgrounds:	Strongly Agree	Agree	Disagree	Strongly Disagree	No Impression
are very large (number of campsites)	___	___	___	___	___
are inexpensive	___	___	___	___	___
are crowded	___	___	___	___	___
have hospitable campground staff	___	___	___	___	___
offer many (in-campground) recreation facilities	___	___	___	___	___
provide large campsites	___	___	___	___	___
are clean	___	___	___	___	___
are quiet	___	___	___	___	___
are family oriented	___	___	___	___	___
offer modern hookups (electric,sewer,water)	___	___	___	___	___
are secluded	___	___	___	___	___
provide modern restroom/shower facilities	___	___	___	___	___
are safe/secure	___	___	___	___	___
are well maintained	___	___	___	___	___

25. Are you a RESIDENT of MICHIGAN? ___ Yes ___ No → 25a) Have you EVER LIVED in Michigan ? ___ Yes ___ No

25b) Do you have family/friends LIVING in Michigan? ___ Yes ___ No

25c) Will you VISIT them on Your Campvention trip? ___ Yes ___ No

26. What is the zip code of YOUR PERMANENT RESIDENCE? ______________

27. Are YOU male or female? ____ Female ____ Male

28. Are YOU retired? ____ No ____ Yes

29. Are you currently: ____ Single ____ Divorced/widowed ____ Seperated

____ Married → Is your spouse retired? ____ Yes ____ No

30. Do you have children LIVING AT HOME WITH YOU ?

____ No ____ Yes → What are their ages ? Child 1 ____ Child 2 ____ Child 3 ____ Child 4 ____ Child 5 ____

APPENDIX B
Differences in The Importance Ratings of Different Campground Attributes Between The Two Subsamples

Appendix B: Differences in the importance ratings of different campground attributes between two subsamples.

Table 63. Differences in the importance ratings of different campground attributes between two subsamples.

Campground Attributes	Subsample I Mean	Subsample I Standard Deviation	Subsample II Mean	Subsample II Standard Deviation	Significance
Large sites	3.11	1.01	3.08	.96	.730
Shade sites	3.20	.92	3.10	.94	.250
Cleanliness	1.88	.74	1.82	.62	.435
Quietness	2.76	.89	2.74	.92	.872
Privacy	3.33	.99	3.33	.96	1.000
Security	2.16	.88	2.18	.89	.784
Hospitality	2.50	.94	2.42	.82	.057
Low price	2.90	1.00	3.02	.93	.193
Flush toilets	3.33	1.17	3.20	1.24	.085
Electricity	2.75	1.09	2.56	.99	.062
Shower	3.00	1.13	2.96	1.21	.083
Laundromat	3.92	.99	4.05	.89	.164
Store	3.81	.97	3.76	.98	.583
Water hookups	3.10	1.23	2.98	1.14	.289
Sewer hookups	3.74	1.18	3.63	1.18	.325
Natural Surr.	3.22	1.06	3.18	1.00	.707
Lake/stream	4.03	1.03	3.95	.95	.378
Hiking trail	4.00	1.02	4.17	.92	.081
Swimming pool	3.98	1.09	3.87	1.12	.333
Playgrounds	4.44	.97	4.46	.94	.839

Note: Significant at .05 level.

APPENDIX C
Comparisons of Factoring Results Between Subsamples

Appendix C: Comparisons Of Factoring Results Between Two Subsamples.

Table 64. Comparisons of factoring results between two subsamples.

Subsample I			Subsample II		
Factor	Eigenvalue	Percent[a]	Factor	Eigenvalue	Percent[a]
1	5.601	28.0	1	4.087	20.4
2	1.938	9.7	2	2.244	11.2
3	1.699	8.5	3	1.878	9.4
4	1.329	6.6	4	1.402	7.0
5	1.168	5.8	5	1.149	5.7
6	1.091	5.5	6	1.110	5.5
7	1.020	5.1	7	1.024	5.1
8	.802	4.0	8	.959	4.8
9	.677	3.4	9	.831	4.2
10	.619	3.1	10	.745	3.7
11	.574	2.9	11	.668	3.3
12	.546	2.7	12	.629	3.1
13	.505	2.5	13	.616	3.1
14	.476	2.4	14	.520	2.6
15	.440	2.2	15	.489	2.4
16	.385	1.9	16	.402	2.0
17	.326	1.6	17	.387	1.9
18	.294	1.5	18	.324	1.6
19	.278	1.4	19	.293	1.5
20	.231	1.2	20	.243	1.2

[a]Percent of variance explained.

APPENDIX D
The computer program for finding sets of matched clusters

Appendix D: The computer program for finding sets of matched clusters.

```
#include    <stdio.h>

#define          Max   6
#define            One '\x01'
#define          Maximum 99999999
#define          TRUE  1
#define          FALSE -1

int   index[7], Best_Choice[7];
float table[7][7], Min;
float   t1[7], t2[7];

main(argc, argv)
int   argc;
char  *argv[];
{
int   i,j,k, depth;
unsigned a,b,c, mask;
float    sum;
FILE   *fp;


if(argc >1)
    if( (fp = fopen( *++argv,"r")) == NULL ) printf("error\n");
Read_data(fp);
Min = Maximum;

for( i = 1; i <= Max; i++ ){
      depth = 1;
      mask = One << i;
      index[depth] = i;
      sum = table[1][i];
      Comb_Search( mask, depth, sum );
}

PrintResult();

} /*  End of Program */


Comb_Search( mask, depth, sum )
unsigned mask;
int depth;
float sum;
{
   int i, j, k;
   float T_sum;
   unsigned T_mask;
```

```
   depth ++;
   for( i = 1; i <=Max; i++ ){
      if(  ( mask & ( One << i )) == 0  ){
       T_mask = mask | (One << i);
       index[depth] = i;
       T_sum = sum + table[depth][i];
       if( depth < Max )
           Comb_Search( T_mask, depth, T_sum );
       else {
           if( T_sum <= Min ){
            Min = T_sum;
            for( k = 1; k <= Max; k++ )
                  Best_Choice[k] = index[k];
           }  /* end if */
       }  /* end else */
      }  /* end if */
   }  /* end for */

}  /* end Comb_Search */


Read_data(p)
FILE *p;
{

   char  c;
   float a, b, diff;
   int   i,j,k, count;
   int   flag, terminate, start;

   for(i=1;i<3;i++){
      for(j=1;j<Max+1;j++){
        count = 0;
        a = 0;
        flag = FALSE;
        start = FALSE;
        terminate = FALSE;
        do{
           c = getc(p);
           switch(c) {
            case '0':
            case '1':
            case '2':
            case '3':
            case '4':
            case '5':
            case '6':
            case '7':
            case '8':
            case '9':
                  if(start != TRUE) start = TRUE;
                  if(flag == TRUE)
                     count++;
```

```
                    a = a*10+(c-'0');
                    break;
              case '.':
                    flag = TRUE;
                    break;
              default:
                    if(start==TRUE) terminate = TRUE;
                    break;
              } /* end switch */
           }while(terminate!=TRUE );

           for(k=1;k<count+1;k++) a = a/10;
           printf("%10.6f\n",a);
           if( i== 1) t1[j] = a;
           else t2[j] = a;
         }
     }

     for(i=1;i<=Max;i++){
        for(j=1;j<=Max;j++){
         diff = t1[i] - t2[j];
         if( diff < 0 ) diff = 0-diff;
         table[i][j] = diff;
         printf("%10.6f",table[i][j]);
        }
        printf("\n");
     }
}


PrintResult()
{
  int i;
  printf("Minimum is %10.6f\n", Min);
  for( i=1; i<7; i++) printf("%4d",Best_Choice[i]);
  printf("\n");
}
^Z
```