RONALD E. FRANK and PAUL E. GREEN*

Recent developments in numerical taxonomy appear to show applicability for many classification problems in marketing. This review article describes some of the approaches to numerical classification and presents illustrative marketing applications. Current limitations of the procedures are also discussed.

Numerical Taxonomy in Marketing Analysis: A Review Article

Marketing managers and researchers often comment on their difficulty in developing useful ways of classifying customers for formulating marketing policy. The source of the difficulty frequently stems from the abundance of alternative classification methods rather than from a lack of possibilities. Changes in our concepts of customer behavior have more often been associated with the generation of new measures of behavior than with the integration of existing measures. In 50 years, researchers have stopped focusing almost exclusively on customer socioeconomic characteristics as a basis for policy formulation and have begun considering a wide range of measures of sociological and psychological phenomena (such as personality, preferences, buying intentions, perceived risk, interpersonal influence) and an increasing number of measures of actual buying behavior (such as total consumption and brand loyalty).

Much of customer behavior has many factors—it is multidimensional. Researchers often sidestep its complexity by picking some unidimensional attribute as-

* Ronald E. Frank is associate professor of marketing and Paul E. Green is professor of marketing, both at Wharton School of Finance and Commerce, University of Pennsylvania. The original version of the article was presented at the Consumer Behavior Conference, Columbia University, May 1967.

Journal of Marketing Research, Vol. V (February 1968), 83–98

19

sumed to be an indicator of the more complex phenomena to be understood. For example, in studies of household brand loyalty (with respect to frequently purchased, branded food products), the researcher often finds variables used to measure brand loyalty such as the proportion of purchases spent on the most frequently purchased brand or the proportion spent on the brand that is of central interest to the researcher. For many purposes, however, these might be too limited a measure of loyalty since they fail to approximate a full description of a rather complex phenomenon. Customers do not typically buy a single brand or even two brands. Many households purchase three, four, or five brands of a product. In addition, the subset of brands chosen for consumption will vary from household to household.

What procedure could be used to study the clusters of brands that different households consume? All possible combinations of brands could be computed and households sorted into respective classes, but this approach presents a few problems. How many combinations are there in a market with only twelve brands? There are over four million if the number of partitions resulting from grouping twelve brands into two or more clusters is added.[1] Even worse, one may want to measure the similarity of brand purchasing behavior not only for the combination of brands but also for the relative proportion of money spent on each brand.

This kind of classification problem is not unique to brand loyalty. How are television programs classified for similarity of audience profiles? Here, too, practitioners often use a single category as the basis for classification, such as the modal audience group, (for example, teenagers loyal to "Rat Patrol"). How should market areas for choosing test markets be grouped? How can a potential purchaser compare the performance specifications of a wide range of computers? How should the readership characteristics of a number of alternative magazines be compared?

Almost every major analytical problem requires the classification of objects by several characteristics—whether customers, products, cities, television programs, or magazines. Seldom are explicit classification systems with some combination of attributes, such as those used for measuring a customer's social class or stage in life cycle, found. Such classification systems typically represent self-imposed taxonomies; that is, taxonomies the researcher believes to be relevant because of a theory or prior experience.[2] Although this approach can be useful, it has limitations. Regardless of the complexity of reality, it is difficult to classify objects by more than two or three characteristics at a time. If reality requires greater complexity, researchers are severely constrained by their conceptual limitations.

The difficulty of seeing through this often bewildering maze is not unique to marketing, (not to mention business problems) as indicated by Sokal, an entymologist:

> "Classification is one of the fundamental concerns of science. Facts and objects must be arranged in an orderly fashion before their unifying principles can be discovered and used as a basis for prediction. Many phenomena occur in such variety and profusion that unless some system is created among them, they would be unlikely to provide any useful information" [82].

A new technology, numerical taxonomy, has been developed, primarily in biology. It consists of a set of numerical procedures for classifying objects [83]. These taxonomic procedures may be called preclassification techniques since their purpose is to describe the natural groupings that occur in large masses of data. From these natural groupings (or clusters) the researcher can sometimes develop the requisite conceptual framework for classification.

Numerical taxonomy is still new, and to the authors' knowledge, only three articles in marketing have appeared [34, 50, 66]. This article introduces potential marketing applications of this set of techniques, giving some attention to their mathematical bases, current limitations, and assumptions. The following topics are discussed:

1. the nature of taxonomic procedures,
2. illustrative applications of taxonomic methods to marketing problems,
3. the assumptions and limitations of the procedures.

The authors feel that taxonomic methods will be used increasingly to describe complex marketing data. Hopefully, this article will alert more researchers to the potential of these methods and to some of the cautions associated with use.

THE NATURE OF TAXONOMIC PROCEDURES

Assume that there is a set of objects, such as people, products, advertisements, and marketing channels, each of which can be characterized by a measurement (or more generally, by an attribute score) on each of a set of characteristics. The researcher has no external criterion for grouping the objects into subsets of similar objects; instead, he wants to identify natural groupings in the

[1] The general formula [29] for finding all possible partitions of a given set of entities is

$$P(n,m) = \left[m^n - \sum_{i=1}^{m-1} m_{(m-i)}\, P(i) \right] / m!$$

where

m is number of partitions; $m \geq 2$
n is number of entities in set to be clustered; $n \geq m$
$P(m)$ is number of distinct partitions containing exactly m clusters
$m_{(m-i)}$ is $m(m-1)(m-2) \cdots (m-i+1)$.

[2] Taxonomies can be distinguished from classifications since they denote interconnections (usually a hierarchy) among characteristics of the objects—a less generic term than classifications. In practice, however, the terms are often used interchangeably.

data, after which more formal models might be developed.

More formally stated, the problem is: How should objects be assigned to groups so there will be as much likeness within groups and as much difference among groups as possible? From this question four others arise: (1) what proximity measure is to be used to summarize the likeness of profiles, (2) after these likeness measures have been computed, how should the objects be grouped, (3) after the objects have been grouped, what descriptive measures are appropriate for summarizing the characteristics of each group, (4) are the groups formed really different from each other (the inferential problem)?

There are numerous taxonomic procedures for achieving the major objective. The following discussion illustrates the logic of one of them, followed by a brief overview of other kinds of procedures that have been developed. The purpose is to show the relevance of these techniques for establishing multidimensional classification systems, not to provide a definitive methodological statement.

An Example

Suppose that the objects of interest are television programs and the characteristics are (assumed independent) measures of the socioeconomic profile of each program. Let us start with measures of two characteristics, number of teenagers (X_1) and number of adult men (X_2), for each of ten programs. Our problem is to find a way of grouping the programs by the similarity of their audience profiles. Figure 1 plots the programs in two dimensions.

Assume that two clusters of five programs each are desired. A start is to compute Euclidean distances of every point from every other point with the usual formula:

$$\Delta_{jk} = [(X_{1j} - X_{1k})^2 + (X_{2j} - X_{2k})^2]^{1/2}.$$

Points 1 and 2 in Figure 1 appear to be closest together. The first cluster would then be formed by finding the midpoint between Points 1 and 2, the centroid of the point coordinates. Then the distance of each point from this average would be computed and the point closest to this average would be added (here, Point 3). Similarly, Point 4 and then Point 5 would be added, giving a cluster of five programs as desired.

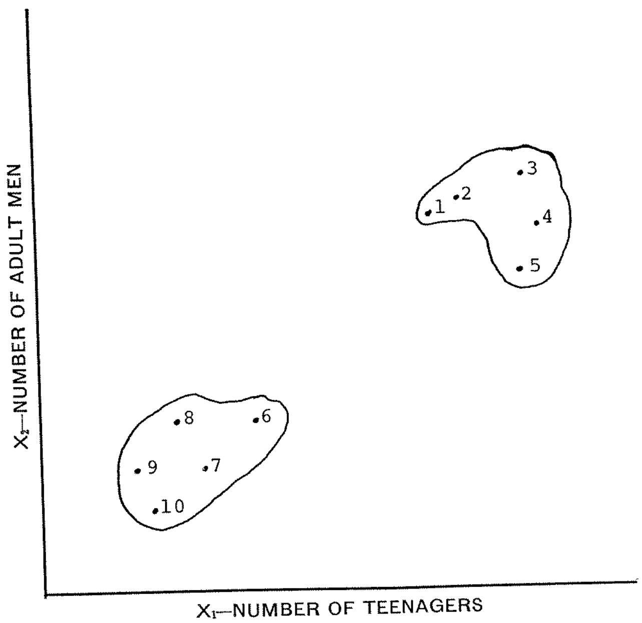
Generalizing to More than Two Dimensions

In the previous illustration, only two measurements were considered for each point (television program). It is relatively easy to follow the procedure visually.[3] In practice there may be many measurements for each program; hence, the graphical procedure must be supplemented by a computational technique that can deal with several characteristics.

Several computer routines are available for this type of taxonomic analysis often called cluster analysis. For example, one computer routine used involves these steps:

1. Each characteristic is first converted to a standardized variate with zero mean and unit standard deviation.
2. Euclidean distances are then computed for each of all possible pairs of points.
3. The pair with the smallest distance is chosen as the node of the first cluster, and the average of this pair is computed.
4. Additional points are added to this cluster (based on closeness to the last-computed average) until:
   a. Some prespecified number of points has been clustered.
   b. The point to be added to the cluster exceeds some prespecified distance-cutoff or threshold number.
5. The program then proceeds to the next pair of points which are closest together of all unclustered points, and the above process is repeated.
6. If desired, the program can be modified to allow points to be in more than one cluster.
7. The program can be further modified to shift points from cluster to cluster to obtain final clusters which are best in the sense of having the lowest average within-cluster distance summed over all clusters at a given stage in the clustering.

Figure 1
ILLUSTRATION OF TAXONOMIC TECHNIQUES (HYPOTHETICAL)



[3] The typical Euclidean distance measure can be easily generalized to more than two dimensions as:

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

OTHER CLUSTERING TECHNIQUES

Proximity Measures. This program is only one way to cluster points in multidimensional space. Other proximity measures and clustering techniques have been proposed by researchers in the biological and social sciences. With some simplification, the proximity measures can be categorized as:

1. distance measures,
2. correlation measures, and
3. similarity measures for attribute data.

The input data—nominal, ordinal, interval, ratio or mixed scales—often determines the proximity measure used to express pairwise relationships among the elements.

Distance Measures. One kind of clustering technique based on Euclidean distance has already been described. Two problems exist with regard to this kind of measure: (1) correlated characteristics and (2) noncomparability of the original units in which the characteristics are measured [69]. The second problem is usually "solved" by standardizing all characteristics to mean zero and unit standard deviation. Thus it is assumed that mean and variance among characteristics is not important in the grouping process.

The first problem can be handled two ways. A principal component analysis may be run on the characteristics and factor scores computed for the objects. Each component score may then be weighted by the square root of the eigenvalue associated with that component before computing the distance measure. A second approach uses the Mahalanobis [60] generalized distance in which the squared distances between objects is measured as a linear combination of the correlated measurements expressed in units of the estimated population dispersion of the composite measure. If the characteristics are uncorrelated and measurements are first standardized (mean zero and unit standard deviation), the square root of the Mahalanobis measure is equivalent to the Euclidean measure discussed.

In practice, distance measures of the kind just described are usually used when data are at least intervally scaled. Kendall [54], however, proposed a distance measure requiring only ordinally scaled measurements. Also, Restle [75] and others have shown that even nominally scaled data may be characterized in distance terms, in the sense of obeying the distance axioms. The resulting metric, however, may not be Euclidean.

Correlation Measures. Probably the most widely used proximity measure in clustering procedures involves the correlation coefficient.[4] Inverse factor analysis, the Q-technique, is a fairly widely used procedure in which objects replace tests in the computation of factor loadings. Clusters may then be formed by grouping subjects with similar factor loadings. Three problems are associated with this class of techniques. First, correlation removes the elevation and scatter of each object, thereby losing information. Second, in grouping objects by factor loadings, the analyst risks obtaining some objects that are split among clusters. Finally, the analyst must usually resort to an R-technique to interpret the clusters' characteristics according to their correlations with underlying factors.

Similarity Measures. Similarity measures are often used in clustering when the characteristics of each object are only nominally scaled, for example, dichotomous or multichotomous. The usual notion of distance seems less applicable here (although it is still possible to use multidimensional scaling techniques to "metricize" such data before clustering). Typically, however, the analyst tries to develop similarity coefficients based on attribute matching.

For example, if two objects are compared on each of eight attributes, the following might result:

Entity	*Attribute*							
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1	1	0	0	1	1	0	1	0
2	0	1	0	1	0	1	1	1

The fractional match coefficient would be:

$$S_{12} = \frac{M}{N} = 3/8,$$

where M denotes the number of attributes held in common (matching 1's or 0's) and N denotes the total number of attributes. If weak matches (non-possession of the attribute) are to be deemphasized, the Tanimoto [76] coefficient is appropriate:

$$\text{Tanimoto } S_{ij} = \frac{\text{No. of attributes which are 1 for both objects } i \text{ and } j}{\text{No. of attributes which are 1 for either } i \text{ or } j\text{, or both}}.$$

In this problem the coefficient would be $2/7$. Many other similarity measures have been developed that represent variations of the fractional match coefficient. (See [83].)

One interesting distance-type measure which can also be used for attribute matching is the pattern similarity coefficient, r_p, proposed by Cattell, Coulter, and Tsujioka [16]. In interval-scaled data, the coefficient compares the computed distance with that expected by chance alone:

$$r_{p(jk)} = \frac{E_i - \sum_{i=1}^{n} d^2_{(jk)}}{E_i + \sum_{i=1}^{n} d^2_{(jk)}},$$

[4] If the characteristics are expressed in standard scores, the Euclidean distance between two objects is a monotone transformation of their correlation [18].

where i is the number of dimensions, $d^2{}_{(jk)}$ is the squared Euclidean distance in standard units between entities j and k, and E_i is twice the median chi-square value for i degrees of freedom. Cattell's coefficient has the convenient property of varying from $+1$ for complete agreement, 0 for no agreement, to -1 for inverse agreement.

The coefficient may also be adapted for dichotomous items as:

$$r'_p = \frac{E_i - d}{E_i + d},$$

where d represents the number of disagreements on d items.

Finally, some mention should be made of the mixed scale problem in which the characteristics are measured in different modes. One possibility is to degrade interval-scaled data into categories and use similarity coefficients. Another possibility is to upgrade nominally or ordinally scaled data. There seems to be no satisfactory solution to this problem although it is conceivable that some highly general measure of proximity, perhaps one derived from information theory, may be appropriate.

Clustering Routines

After the analyst has decided on some measure of pairwise proximity, he must still contend with the grouping process itself. A variety of approaches are possible. One major class of approaches to the clustering problem consists of hierarchical routines. For example, Edwards and Cavalli-Sforza [24] describe a clustering procedure (based on a least-squares technique) which first clusters the data into two groups. The procedure is repeated sequentially so that progressively smaller clusters are formed sequentially by splitting the original clusters. A hierarchical array is obtained. A variant of this procedure starts with clusters of one object each and builds new clusters hierarchically until one overall cluster results. This approach was described by Ward [93].

Other grouping routines use threshold or cutoff measures similar to the algorithm described earlier. Some procedures, for example, suggest selecting an object closest to the centroid of all the data to serve as a prime node around which other points are clustered until some threshold distance level is reached. An unclustered object farthest from the centroid of the first cluster may then be chosen as a new prime node. The process is continued, the third and subsequent prime nodes being selected on the basis of largest average distance from the centroids of clusters already formed.

Some grouping routines [24, 93] are highly metric since effectiveness measures involve the computation of within-cluster variance around the centroid of the cluster members. Others [83] use only the proximity between an unclustered object and some single member of the clustered set as a criterion for set inclusion.

In Q-technique, objects are often clustered by highest factor loadings, a simple approach; but it does not use all available information.

Finally, there is the possibility of clustering by systematic space-density search routines in which the n-dimensional space is cut into hypercubes and the computer program counts the number of cases falling into each region. Relatively little work, however, has been done on this taxonomic routine.

Descriptive Characteristics of the Groups

Even after objects are grouped, each cluster must be characterized by its representative profile. In some instances the cluster's centroid is used as a description of its members. In others the actual profile of the object closest to the group's centroid may be used. As in choice of proximity measure and choice of grouping routine, however, the criteria for describing each group are usually ad hoc, a main problem being that *cluster* is still not a precisely defined term. Some of these problems and the inferential problem will be reconsidered later in this article.

ILLUSTRATIVE MARKETING APPLICATIONS

Some appreciation for the versatility and unresolved problems of taxonomic methods can be gained from the following short review of studies conducted by the authors in the past two years.

Clustering Analysis in Test Marketing

One of the earliest pilot applications involved the use of cluster analysis in the grouping of cities (standard metropolitan areas) for test marketing purposes [34]. Data for each of 88 cities were available on 14 measured characteristics, such as population, number of retail stores, percent non-white. A clustering program using the Euclidean distance measure grouped the cities into homogeneous five-point clusters. Centroids of each cluster in 14-space and average distances of each point from the grand centroid and from the centroid of its own cluster were obtained. As an alternative for comparison purposes, the original data matrix was factored, and cluster analysis was performed on the resultant (standardized) factor scores.

The cluster analysis yielded some interesting findings. First, the cluster of five cities closest to the grand mean of all 88—Dayton, Columbus, Indianapolis, Syracuse, and New Haven—agreed well with various lists of typical cities prepared by such magazines as *Sales Management* and *Printers' Ink* indicating results consistent with industry judgment. This method also provides homogeneous groups of cities with centroids quite distant from the grand origin. Second, the combined procedure of factor analysis (and subsequent clustering of factor scores) indicated that two major dimensions, a city size construct and a demographic construct, explained most of the variance in the data.

This study was only a pilot effort. In practice, the marketing manager would use those city characteristics most relevant to his product line. The clusters could then serve as homogeneous blocks from which individual cities could be chosen to serve as treatment and control units, that is matched units for various experimental purposes.

Television Program Audience Profile Analysis

Grouping of television programs into clusters having similar audience profile, which was used to illustrate the nature of taxonomic procedures, comprises still another exploratory investigation currently in progress. American Research Bureau data for both day and evening programs in October, 1965, are the bases for this analysis. For each, program measures of the number of adult men and women in different age categories and the number of children and teenagers viewing the program are available. The primary objective is to group programs by viewer characteristics so that their grouping is a function of viewer reaction to content and casting—not to the effects of time of day, day of week, and lead-in programs.

The analysis is divided into two stages. The first is the adjustment of raw data for the effects of time of day, day of week, and lead-in programs. The adjustment is roughly analogous to making a cyclical adjustment in a time series analysis to ensure a cleaner set of data for studying trend movements. When variations in audience profile from program to program are caused primarily by the effect of program content and casting, the adjusted data are subjected to a taxonomic analysis. The first stage of the study is complete, and the taxonomic work is about to begin. (It will soon appear as a working paper [30].)

Patterns of Customer Brand Loyalty

At the beginning of this article the study of brand loyalty was used to illustrate the tendency for letting unidimensional measures represent customer behavior that may be multidimensional. In this study cluster analysis and Kruskal's algorithm [56] is used to characterize customer brand purchasing behavior. The objective is to develop more comprehensive classification systems for analyzing brand choice.

Chicago Tribune panel data for three product categories (carbonated beverages, regular coffee, and ready-to-eat cereals) for 1961 were used in the analysis. For each product category for each of 480 households, the percentage of units (based on weight) purchased by brand was computed.

Two different approaches were then taken. A Euclidean distance measure was used to group households that had relatively similar percentage distributions of brand purchasing behavior within a product. This is equivalent to studying brand loyalty for the bundle of brands households purchase. The results showed that with only one exception in the regular coffee market each cluster of households bought only one brand at a rate greater than the brand's overall market share. Although other brands were purchased, none was given this degree of favor. The only exceptions are the clusters containing several private brands. Households that purchase one private brand at a greater rate than its overall share are likely to purchase another with a similar degree of concentration. Customers who buy them may be less sensitive to differences in product characteristics, or the products themselves may be more similar.

A second approach organized the data by brand instead of by customer. This part of the analysis started with the transpose of the data matrix used, that is the data were organized by brand and within brand, by household. For each brand the percentage of purchases devoted to that brand by each of about 100 households was available. Euclidean distance measures characterized brand similarity by pattern of purchase requirements over households.

Results so far have provided few surprises and have raised more questions than can be answered here. For example in the cereal market, evidence appears that old standard brands (Kellogg's Corn Flakes, Cheerios, Wheaties) tend to serve segments which overlap, yet many health-oriented cereals (Special K, All Bran, Grape Nuts) tend to serve a somewhat different group of customers.

An Experimental Gaming Application

Another application of clustering was prompted by experimental data obtained in studying the relationship between risk taking (in a no-information-improvement context) and the propensity to acquire uncertainty-reducing information in an information-buying context [35]. Data were available for 42 men and women subjects on a variety of behavioral and personality variables.

Preliminary analysis using a variety of multivariate techniques showed little support for the study's primary hypotheses. Part of the problem was thought to be that different subjects were using different behavioral models; these differences became obscured in the process of data aggregation. Accordingly, each subject's behavior in the experiment was viewed as a point in task performance space, the axes of which were represented by the situational and personality variables comprising the experimental situation. Subjects were then clustered by their similarities to each other over the whole experiment.

This procedure produced various clusters of subjects—some supported the hypothesis and others suggested other kinds of behavioral models. The potentialities of this approach appear provocative in the examination of experimental gaming data generally. Perhaps even more interesting, however, is the application of this kind of approach to the design of behavioral experiments. Alternative explanatory models are the rule rather than the exception in experimental games. Before collecting any data the researcher could characterize the play of ideal subjects (those whose behavior corresponded to each alternative model) by points in experimental performance space. Levels of the experimental variables might then be chosen to maximize the discrimination among alternative models before the experiment is conducted.

Operational Characterization of Inter-Brand Competition

In another pilot study, cluster analysis helped to characterize inter-brand competition in the computer field [38]. Performance data were obtained for over 100 different computer models with installation date used to categorize them as first- or second-generation models. For each computer model, data were available on 12 measured characteristics, such as word length, execution time, digital storage, transfer rate, and 10 categorical characteristics, such as whether the computer possessed Boolean operations, table look-up, and indirect addressing.

The data's mixed character (continuous variables and dichotomous features data) required a different approach from that typically used in cluster analysis. First, the attribute data were metricized by a multidimensional scaling technique [56]. A two-dimensional representation revealed that each computer model could be characterized by the dimensions of capacity (number of different features) and orientation (scientific versus business), as based on the particular pattern of zeroes and ones.

The resultant clusters, developed by a hierarchical grouping technique, displayed interesting characteristics from the standpoint of intermodel competition. For example, a machine's cluster of features appears to be idiosyncratic to the particular manufacturer, that is, each manufacturer tends to build all his machines with a particular set of features. Each manufacturer's complex, however, may vary from that of his competitors. It is interesting that only IBM had a model in each of the major clusters. However, the time period comparison —first- versus second-generation computers—indicated a trend toward all models having a greater number of features.

The measured variables were then analyzed separately, yielding two main dimensions—speed and size of computers. Finally, the measured data were dichotomized about the median of each characteristic (taken separately) and submitted to a combination multidimensional scaling and cluster analysis.

Figure 2 shows a two-space configuration derived from applying a nonmetric program to proximity measures developed from the above steps. After adjusting for intercorrelation of the characteristics [39], similarity measures were developed by tabulating the number of (weighted) matches for all computer pairs. The higher this number, the more similar each pair was assumed to be with respect to all 22 performance characteristics. For $n = 55$, there are 1,485 interpoint proximities as input to the program; only their rank order is required.

The two-space configuration of Figure 2 shows the boundaries of clusters formed (by another means) on a more precise configuration obtained in four-space. Such compression of results (into two-space) seriously distorts the makeup of Cluster 8; otherwise the clusters are fairly compact. It is interesting to note that Cluster 5 is composed of small, fairly slow, business-oriented machines, but Cluster 7 is characterized by large, relatively fast, scientific machines.

The complete study on which Figure 2 is based revealed that four dimensions—speed, size, number of different features (qualitative characteristics), and orientation (scientific versus business)—appeared to adequately describe the computer market.

The possibilities of such performance-space analyses over time have potential for the study of product innovation and modification—particularly industrial products like electric motors and machine tools. In this approach a whole series of performance spaces could be viewed through time—their dimensions, number of points (models), and interrelationships among points could all be changing, reflecting changes in technology and inter-model competition. Such an approach would seem to indicate the data's fine structure better than the more traditional reliance on S-curves to describe product life cycles.

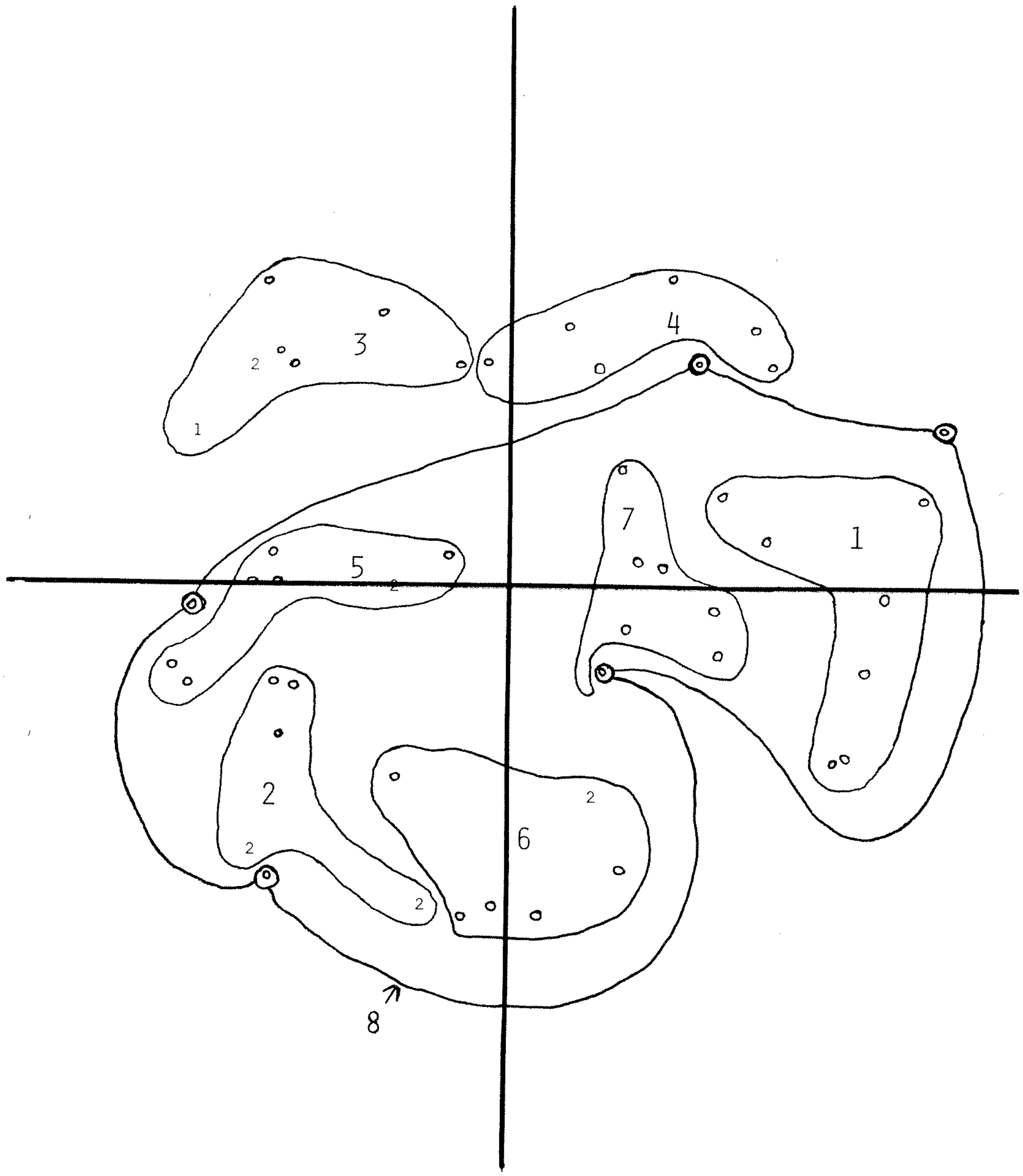
Physician's Media Reading Habits

In another study [41], numerical taxonomy was used to cluster reading profiles of both physicians and medical journals. The basic data consisted of zero-one matrixes in which each physician was classified as a light (zero) or heavy (one) reader of each of 19 medical journals. Each physician was also classified as one who lightly or heavily prescribed each of 29 therapeutic drug classes. Data were also available on the physician's specialty, age, and total weekly patient and prescription loads. The zero-one matrixes were again metricized by a multidimensional scaling program. Clusters of journals with similar, physician reading habits and clusters of physicians with similar journal profiles were developed.

Findings indicated that, within a given specialty, media reading profiles are not associated with such variables as physician age, total prescribing frequency, and product mix selection. However, the journal clusters provided an interesting output of the analysis by summarizing a diverse set of zero-one data. The marketing manager could use these clusters as a guide to media scheduling. For example, if he wishes to choose journals with high overlap of coverage, he can choose all journals within a given cluster. If, however, he wishes to emphasize diversity, he can choose one journal from each cluster.

From a methodological viewpoint, the interesting

Figure 2
TWO-SPACE CONFIGURATION OF COMPUTER MODELS IN "PERFORMANCE SPACE"



concept is the dual use of multidimensional scaling and cluster analysis. The first technique allows the researcher to make a concise description of the data—frequently interpretable in its own right—and the last allows him to organize the data into similar journal profiles that can then be subjected to further analysis.

Taxonomy in Psychometric Studies

Some mention should also be made of the usefulness of clustering procedures in psychometric studies involving perceptual and preference mapping. A recently completed study [36] involved the analysis of proximities data developed during a study of student perception of six graduate schools of business. Three modes of data collection—similarity triads, direct ratings, and the semantic differential—were used to collect proximity judgments.

In this study a hierarchical grouping method was used to develop clusters of respondents with similar perceptions of the six business schools. That is, although the main objective of this study was the development of perceptual maps, cluster analysis was useful in partitioning the respondents into homogeneous groups with similar perceptions.

The results indicated that a two-space solution adequately portrayed the respondents' perceptions. From other data collected in the study, the dimensions of the space could be characterized as prestige of school and quantitativeness of its curriculum. Not inconsequentially all three data collection methods yielded fairly similar perceptual maps, on an aggregate basis. Moreover, differences in perceptual mappings were not generally explainable by respondent personal data, such as undergraduate major, previous work experience, graduate major. Only one variable, home state of respondent, appeared to influence his perception of the business schools in any significant way.

A similar study [37] involved a multidimensional scaling of professional journals typically read by marketing academics and researchers. Perception and preference data were obtained for eight journals, and respondents were clustered on the basis of similarity of perception and preference.

Figure 3 shows the results of applying a nonmetric clustering routine to the perception data [37]. Note that this program is hierarchical. Respondents 4 and 7 are first clustered because they had the highest proximity measure of the group. Respondents 2 and 11 are next clustered at level two, and so on, until all points are eventually in one large cluster. On the left-hand side of Figure 3 one can see how the proximity measure declines as more disparate points are clustered.

The results of this study indicated that preferences and perception were independent over stimuli, that is respondents clustered by commonality of perception were unrelated to clusters formed by commonality of preference.

ASSUMPTIONS AND LIMITATIONS OF CLUSTERING METHODS

Cluster analysis is not a single, cohesive set of techniques but rather a variety of procedures, each having a kind of ad hoc flavor and certain advantages and disadvantages. Some of the limitations are shared by all these techniques to some degree, but specific procedures have both advantages and disadvantages.

General Problems in Cluster Analysis

All clustering techniques have certain general analytical inadequacies because the data are used to generate the groupings. Illustrative questions are:

1. How many clusters should be formed?
2. If, as is usually the case, the characteristics of the objects are measured in different units, how can equivalence among metrics be achieved?
3. If the objects' scores along several dimensions are intercorrelated, how should these interdependencies be handled?
4. Even if the number of clusters can be determined in some satisfactory way, how does the analyst decide on the appropriate boundaries for clusters, summary measures of the characteristics of each cluster, and their statistical significance?

In some of the illustrative applications described here, the number of clusters was decided in advance. Increasing the number of clusters will tend to reduce the average within-cluster distance but, obviously, one must stop short of ending with each point being a cluster.

In addition, all data including variables originally interval-scaled were standardized to zero mean and unit standard deviation. Although this step enables the analyst to work with common metrics, it is assumed that central tendency and variability among dimensions are not important.

The problem of dealing with intercorrelated characteristics was pointed out in the test marketing illustration. In this study an alternative procedure was used in which the set of characteristics was first reduced to independent constructs by a principal component analysis before the cluster analysis. This procedure can lead to different clusters from those obtained by the first procedure that ignored the intercorrelations among charac-

Figure 3

ILLUSTRATION OF HIERARCHICAL CLUSTERING ROUTINE

Proximity measure	*Subject number* 09		10		01		08		06		02		11		03		04		07		05		12
1.8253	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X	X	X	.	.	.	.
1.7745	.	.	.	.	.	.	.	.	.	.	X	X	X	.	.	.	X	X	X	.	.	.	.
1.7111	.	.	.	.	.	.	.	.	.	.	X	X	X	.	X	X	X	X	X	.	.	.	.
1.6871	.	.	.	.	.	.	.	.	.	.	X	X	X	.	X	X	X	X	X	.	X	X	X
1.5961	.	.	.	.	.	.	.	.	X	X	X	X	X	.	X	X	X	X	X	.	X	X	X
1.4724	.	.	.	.	X	X	X	.	X	X	X	X	X	.	X	X	X	X	X	.	X	X	X
1.3715	.	.	.	.	X	X	X	.	X	X	X	X	X	.	X	X	X	X	X	X	X	X	X
1.2901	.	.	.	.	X	X	X	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.1388	.	.	.	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.0558	X	X	X	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.8077	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

teristics. Finally, the researcher might wish to use the Mahalanobis generalized distance measure discussed by Morrison [66].

Appropriate boundaries and descriptive statistics of clusters are usually determined by the specific technique used—in many instances by a generalized distance function, the computation of centroids, and the use of a preset number of points or cutoff distances. Even so, it is fair to say that good measures of cluster compactness are not available. In the test marketing illustration each dimension included in the analysis was given (manifest) equal weight in determining similarity. In a given situation one might choose to give a single dimension or some subset of dimensions more weight than others in defining proximity measures. Cluster analysis can be easily modified to take into account unequal weights, but this approach still largely varies with circumstances.

Still less is known about the inferential characteristics of clustering techniques. Unlike other multivariate techniques, such as discriminant analysis and principal component analysis, clustering techniques are much less structured, and little investigation has been made to date of their statistical proprieties.

Limitations of Specific Proximity Measures

In earlier sections of this article, the characteristics of specific proximity measures—distance measures, correlation techniques, similarity coefficients—were briefly described. Each measure suffers from certain specific limitations.

Distance measures are usually restricted to instances in which the objects' characteristics to be measured can be expressed as interval-scaled variables. This represents a limitation on the kind of variable meaningfully handled although Kendall's nonparametric measure (mentioned earlier) could be used to handle data that are scaled only ordinally and the researcher could develop non-Euclidean metrics.

In addition, the Euclidean measure suffers from the disadvantage that two objects may be viewed as different solely because their values on one variable differ markedly. Finally, it should be reiterated that the researcher would, in general, obtain different results by using original versus standardized data for the characteristics of the objects being clustered by this method.

Correlative techniques, such as Q-factor analysis, have an even more serious limitation because one must standardize over objects, thus losing mean and scatter information. That is, in this technique, each object is given the same mean and variance.

A second disadvantage is that rotation of factor axes (to get purer loadings) lends a certain arbitrariness to the procedure. Finally, also mentioned earlier, in this procedure objects may be split on factors, leading to uncertainty of the placement of an object into a specific group.

Similarity measures are flexible since they can be adapted to handle nominal, ordinal, and interval-scaled data. Furthermore, it can be shown that similarity measures can be metricized by multidimensional-scaling procedures. Moreover, similarity measures are generally less sensitive to the impact of a single characteristic on the resultant dissimilarity of two objects than are the Euclidean distance measures.

However, similarity measures have their set of limitations. First, if a group is to be formed on the basis of overall matches, two objects may not be grouped even if they match well on some subset of characteristics. Conversely, an object may be in a group because it is similar to different members of the group on different subsets of characteristics.

Second, if a large number of characteristics are involved, objects which match may do so for accidental reasons, reflecting the noise in the data; and third, if some variables are dichotomous and others are multichotomous, the two-state attributes will tend to be more heavily weighted in the similarity measures. For example, if one attribute were broken down into 100 states, we would rarely find matches. Hence this attribute would receive little importance in the overall similarity measure.

Finally, if continuous data are discretized in order to use similarity measures, valuable information can be lost. The analyst is thus plagued with the problem of deciding both the kinds of attributes to include in the analysis and the number of states to be associated with each.

Choosing Appropriate Techniques

Numerical taxonomy invites some ambivalence by the analyst wanting to use the techniques. On one hand, the procedures are designed to cope with a relevant aspect of marketing description—the orderly classification of multivariate phenomena. On the other hand, the varying character of various proximity measures and clustering techniques—and the basic lack of structure at either the descriptive statistic or inferential statistic level—suggests that the analyst be cautious in applying them.

Until more structure is introduced, it seems prudent to conduct analyses in parallel where alternative proximity measures and grouping procedures are used [40]. Moreover, sensitivity analyses on synthetic data might be helpful in exploring the various idiosyncracies of alternative techniques. If the data are well clustered to begin with, similar results over alternative techniques will usually be obtained—but how often will these pleasant states of affairs exist? Though the authors believe numerical taxonomy can be useful in marketing analysis, they would urge prudence in its application and the systematic study of similarities and differences among alternative procedures. (The references may help to facilitate this study.)

REFERENCES

1. G. H. Ball, "Data Analysis in the Social Sciences: What About the Details?" *Proceedings Fall Joint Computer Conference*, 1965, 533–59.
2. B. M. Bass, "Iterative Inverse Factor Analysis: A Rapid Method for Clustering Persons," *Psychometrika*, 22 (March 1957), 105.
3. J. F. Bennett and W. L. Hays, "Multidimensional Unfolding: Determining the Dimensionality of Ranked Preference Data," *Psychometrika*, 25 (March 1960), 27–43.
4. A. Birnbaum and A. E. Maxwell, "Classification Procedures Based on Bayes' Formula," *Applied Statistics*, 9 (November 1961), 152–68.
5. Jack Block, "The Difference Between Q and R," *Psychological Review*, 62 (1955), 356–8.
6. ———, Louis Levine and Quinn McNemar, "Testing for the Existence of Psychometric Patterns," *Journal of Abnormal Social Psychology*, 46 (July 1951), 356–9.
7. R. E. Bonner, "Some Clustering Techniques," *IBM Journal of Research and Development*, 8 (January 1964), 22–33.
8. A. D. Booth, "An Application of the Method of Steepest Descent to the Solution of Simultaneous Non-linear Equations," *Quarterly Journal of Mech. Applied Mathematics*, 2 (December 1949), 460–8.
9. G. E. P. Box, "The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples," *Biometrics*, 10 (March 1954), 16–60.
10. S. H. Brooks, "A Discussion of Random Methods of Seeking Maxima," *Journal of Operations Research Society*, 6 (1958), 244–51.
11. ———, "A Comparison of Maximum Seeking Methods," *Journal of Operations Research Society*, 7 (1959), 430–57.
12. Cyril L. Burt, "Correlations Between Persons," *British Journal of Psychology*, 28 (July 1937), 59–96.
13. H. Cartwright, *Structural Models: An Introduction to the Theory of Directed Graphs*, New York: John Wiley & Sons, Inc., 1963.
14. Raymond B. Cattell, "r_p and Other Coefficients of Pattern Similarity," *Psychometrika*, 14 (December 1949), 279–98.
15. ———, "On the Disuse and Misuse of R, P, Q and O Techniques in Clinical Psychology," *Journal of Clinical Psychology*, 7 (1951), 203–14.
16. ———, M. A. Coulter and B. Tsujioka, "The Taxonometric Recognition of Types and Functional Emergents," in R. B. Cattell, ed., *Handbook of Multivariate Experimental Psychology*, Chicago: Rand McNally and Co., 1966, 288–329.
17. W. W. Cooley and Paul R. Lohnes, *Multivariate Procedures for the Behavioral Sciences*, New York: John Wiley & Sons, Inc., 1963.
18. C. H. Coombs, *A Theory of Data*, New York: John Wiley & Sons, Inc., 1964.
19. ———, "A Method for the Study of Interstimulus Similarity," *Psychometrika*, 19 (September 1954), 183–94.
20. Douglas R. Cox, "Note on Grouping," *Journal of American Statistical Association*, 52 (December 1957), 543–47.
21. Lee J. Cronbach and Goldine C. Gleser, "Assessing Similarity Between Profiles," *Psychological Bulletin*, 50 (November 1953), 456–73.
22. Frank M. duMas, "A Quick Method of Analyzing the Similarity of Profiles," *Journal of Clinical Psychology*, 2 (January 1946), 80–3.
23. ———, "On the Interpretation of Personality Profiles," *Journal of Clinical Psychology*, 3 (1947), 57–65.
24. A. W. F. Edwards and L. L. Cavalli-Sforza, "A Method for Cluster Analysis," *Biometrics*, 52 (June 1965), 362–75.
25. G. A. Ferguson, "The Factorial Interpretation of Test Difficulty," *Psychometrika*, 6 (October 1941), 323–29.
26. R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *American Eugenics*, 7 (1963), 179–88.
27. W. D. Fisher, "On Grouping for Maximum Homogeneity," *Journal of American Statistical Association*, 53 (December 1958), 789–98.
28. Claude Flament, *Applications of Graph Theory to Group Structure*, New Jersey: Prentice-Hall, Inc., 1963.
29. J. J. Fortier and H. Solomon, "Clustering Procedures," Unpublished paper, International Symposium on Multivariate Analysis, University of Dayton, June 1965.
30. Ronald E. Frank, "Television Program Audience Similarities: A Taxonomic Analysis," University of Pennsylvania, December 1967, mimeographed.
31. Eugene L. Gaier and Marilyn C. Lee, "Pattern Analysis: The Configural Approach to Predictive Measurement," *Psychological Bulletin*, 50 (March 1953), 140–8.
32. J. A. Gengerelli, "A Method for Detecting Subgroups in a Population and Specifying Their Membership," *Journal of Psychology*, 55 (1953), 140–48.
33. L. A. Goodman and W. H. Kruskal, "Measures of Association for Cross Classifications," *Journal of American Statistical Association*, 59 (September 1964), 732–64.
34. Paul E. Green, Ronald E. Frank and Patrick J. Robinson, "Cluster Analysis in Test Market Selection," *Management Science*, 13 (April 1967), 387–400.
35. ———, "A Behavioral Experiment in Risk Taking and Information Seeking," Working paper, University of Pennsylvania, January 1967.
36. Paul E. Green and P. J. Robinson, "Perceptual Structure of Graduate Business Schools—An Application of Multidimensional Scaling," Working paper, June 1967.
37. ———, "Perceptual and Preference Mapping of Professional Journals," Working paper, May 1967.
38. ——— and F. J. Carmone, "Structural Characteristics of the Computer Market—An Application of Cluster and Reduced Space Analysis," Working paper, May 1967.
39. ———, "WAGS: An IBM 7040 Computer Program for Obtaining Weighted Agreement Scores for Multidimensional Scaling," Working paper, May 1967.
40. ———, "Cross Techniques Study—Computer Model Clustering," Working paper, August 1967.
41. ———, "A Reduced Space and Cluster Analysis of Physicians' Media Reading Habits," Working paper, September 1967.
42. H. H. Harman, *Modern Factor Analysis*, Chicago: University of Chicago Press, 1960.
43. C. W. Harris, "Characteristics of Two Measures of Profile Similarity," *Psychometrika*, 20 (1955), 289–97.
44. G. C. Helmstadter, "An Empirical Comparison of Methods for Estimating Profile Similarity," *Educational and Psychological Measurement*, 17 (1957), 71–82.
45. J. L. Hodges, Jr., "Discriminatory Analysis I: Survey of Discriminatory Analysis," USAF School of Aviation Medicine, Randolph, Texas, 1950.
46. Karl J. Holzinger, "Factoring Test Scores and Implications for the Method of Averages," *Psychometrika*, 9 (December 1944), 257–62.
47. Paul Horst, *Matrix Algebra for Social Scientists*, New York: Holt, Rinehart and Winston, 1963.
48. ———, "Pattern Analysis and Configural Scoring," *Journal of Clinical Psychology*, 10 (January 1954), 1–11.
49. K. J. Jones, *The Multivariate Statistical Analyzer*, Cambridge, Mass.: Harvard Cooperative Society, 1964.
50. J. Joyce and C. Charron, "Classifying Market Survey Respondents," *Applied Statistics*, 15 (November 1966), 191–215.
51. H. F. Kaiser, "Formulas for Component Scores," *Psychometrika*, 27 (March 1962), 83–7.
52. M. G. Kendall, *The Advanced Theory of Statistics*, Vol. 1, New York: Hafner Publishing Company, 1958.
53. ———, *Rank Correlation Methods*, London: Griffin Publishing Company, 1948.

54. ———, "Discrimination and Classification," London: CEIR Ltd., 1965.
55. J. B. Kruskal, "Nonmetric Multidimensional Scaling: A Numerical Scaling Method," *Psychometrika*, 29 (June 1964), 115–30.
56. ———, "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis," *Psychometrika*, 29 (March 1964), 1–28.
57. E. O. Laumann and L. Guttman, "The Relative Association Contiguity of Occupations in an Urban Setting," *American Sociological Review*, 31 (April 1966), 169–78.
58. J. C. Lingoes, "A Taxonometric Optimization Procedure: An IBM 7090 Classification Program," *Behavioral Science*, 8 (October 1963), 370.
59. ———, "An IBM 7090 program for Guttman-Lingoes Smallest Space Analysis," Computer Center, University of Michigan, 1965.
60. P. C. Mahalanobis, "On the Generalized Distance in Statistics," *Proceedings National Institute of Science*, Vol. 12, India, 1936, 49–58.
61. F. Massarik and P. Ratoosh, *Mathematical Explorations in Behavioral Science*, Homewood, Ill.: Richard D. Irwin, Inc., 1965.
62. P. McNaughton-Smith, *et al.*, "Dissimilarity Analysis: A New Technique of Hierarchical Subdivision," *Nature*, 202 (June 1964), 1033–4.
63. Louis L. McQuitty, "Hierarchical Syndrome Analysis," *Educational and Psychological Measurement*, 20 (1960), 293–304.
64. ——, "Typal Analysis," *Educational and Psychological Measurement*, 20 (1960), 293–304.
65. ——, "Best Classifying Every Individual at Every Level," *Educational and Psychological Measurement*, 23 (July 1963), 337–46.
66. Donald G. Morrison, "Measurement Problems in Cluster Analysis," *Management Science*, 13 (August 1967), B-755–80.
67. Jum Nunnally, "The Analysis of Profile Data," *Psychological Bulletin*, 59 (July 1962), 311–19.
68. Charles E. Osgood and George J. Suci, "A Measure of Relation Determined by Both Mean Difference and Profile Information," *Psychological Bulletin*, 49 (May 1952), 251–62.
69. J. E. Overall, "Note on Multivariate Methods of Profile Analysis," *Psychological Bulletin*, 61 (March 1964), 195–8.
70. K. Pearson, "On the Dissection of Assymetrical Frequency Curves," *Contributions to the Mathematical Theory of Evolution, Phil. Trans. of Royal Society*, 1894.
71. ———, "On the Coefficient of Racial Likeness," *Biometrika*, 18 (July 1926), 105–17.
72. R. G. Pettit, "Clustering Program: Continuous Variables," Advanced Systems Development Division, IBM, Yorktown Heights, New York, 1964.
73. C. R. Rao, "Tests of Significance in Multivariate Analysis," *Biometrika*, 35 (May 1948), 58–79.
74. ———, "The Utilization of Multiple Measurements in Problems of Biological Classification," *Journal of Royal Statistical Society*, Section B, 10 (1948), 159–203.
75. F. Restle, *Psychology of Judgment and Choice*, New York: John Wiley & Sons, Inc., 1961.
76. D. J. Rogers and T. T. Tanimoto, "A Computer Program for Classifying Plants," *Science*, 132 (October 1960), 1115–22.
77. P. J. Rulon, "Distinctions Between Discriminant and Regression Analysis and a Geometric Interpretation of the Discriminant Function," *Harvard Educational Review*, 21 (June 1951), 80–90.
78. R. N. Shepard, "The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Functions: I and II," *Psychometrika*, 27 (June 1962, September 1962), 125–40, 219–46.
79. ———, "Analysis of Proximities as a Technique for the Study of Information Processing in Man," *Human Factors*, 5 (February 1963), 33–48.
80. G. G. Simpson, "Numerical Taxonomy and Biological Classification," *Science*, 144 (May 1964), 712–13.
81. P. H. A. Sneath, "The Application of Computers to Taxonomy," *Journal of General Micro-Biology*, 17 (August 1957), 201–27.
82. R. R. Sokal, "Numerical Taxonomy," *Scientific American*, 215 (December 1966), 106–16.
83. ——— and P. H. A. Sneath, *Principles of Numerical Taxonomy*, San Francisco: Freeman & Company, 1963.
84. William Stephenson, "Some Observations on Q Technique," *Psychological Bulletin*, Vol. 49 (September 1952), 483–98.
85. S. A. Stouffer, *et al., Measurement and Prediction*, Princeton, N. J.: Princeton University Press, 1950.
86. Robert L. Thorndike, "Who Belongs in the Family?", *Psychom trika*, 18 (December 1953), 267–76.
87. Warren S. Torgerson, "Multidimensional Scaling: Theory and Method," *Psychometrika*, 17 (December 1952), 401–19.
88. ———, "Multidimensional Scaling of Similarity," *Psychometrika*, 30 (December 1965), 379–93.
89. Fred T. Tyler, "Some Examples of Multivariate Analysis in Educational and Psychological Research," *Psychometrika*, 17 (September 1952), 289–96.
90. Robert C. Tyron, *Cluster Analysis*, Edwards Bros., 1939.
91. ———, "Cumulative Communality Cluster Analysis," *Educational and Psychological Measurement*, 18 (March 1958), 3–35.
92. J. W. Tukey, "The Future of Data Analysis," *Annals of Mathematical Statistics*, 33 (March 1962), 1–67.
93. J. H. Ward, "Hierarchical Grouping to Optimize an Objective Function," *Journal of American Statistical Association*, 58 (March 1963), 236–44.
94. Joe E. Ward, Jr., and Marion E. Hook, "Application of an Hierarchical Grouping Procedure to a Problem of Grouping Profiles," *Educational and Psychological Measurement*, 23 (1963), 69–82.
95. Harold Webster, "A Note on Profile Similarity," *Psychological Bulletin*, 49 (September 1952), 538–9.
96. Joseph Zubin, "A Technique for Measuring Like-Mindedness," *Journal of Abnormal Social Psychology*, 33 (October 1938), 508–16.