Using Cluster Analysis for Medical Resource Decision Making

DAVID DILTS, PhD, JOSEPH KHAMALAH, MASc,
ANN PLOTKIN, OD, MSc

Escalating costs of health care delivery have in the recent past often made the health care industry investigate, adapt, and apply those management techniques relating to budgeting, resource control, and forecasting that have long been used in the manufacturing sector. A strategy that has contributed much in this direction is the definition and classification of a hospital's output into "products" or groups of patients that impose similar resource or cost demands on the hospital. Existing classification schemes have frequently employed cluster analysis in generating these groupings. Unfortunately, the myriad articles and books on clustering and classification contain few formalized selection methodologies for choosing a technique for solving a particular problem, hence they often leave the novice investigator at a loss. This paper reviews the literature on clustering, particularly as it has been applied in the medical resource-utilization domain, addresses the critical choices facing an investigator in the medical field using cluster analysis, and offers suggestions (using the example of clustering low-vision patients) for how such choices can be made. *Key words:* cluster analysis decisions; resource classification schemes; clustering methodology. (Med Decis **Making** **1995;15:333-347)**

Classifying patients into groups is not a new phenomenon. It is a concept that, although not always recognized as such, dates back to the beginning of medical science.[1] In fact, it can be said that the idea is based on the notion of a search for a. natural ordering of things, which is a basic characteristic of human beings.[2] Fairly recent additions to this concept, however, are 1) the wide-scale application of clustering and classification techniques to patients intra- and inter-institutionally for determining medical resource utilization and 2) the growing importance being attached to the reliability and validity aspects of classification procedures and the resulting schemes in general.[3] Certain critical decisions must be made in order to properly utilize cluster analysis. The purpose of this paper is to review these decisions and, using an illustrative example, demonstrate how such decisions can be made.

The goal of clustering and cluster analysis is to group **and distinguish** comparable units and to separate them from differing units. Towards this end, cluster analysis encompasses a wide range of statistical techniques.[4-7] In cluster analysis, one attempts "to group large numbers of persons, jobs, or objects into smaller numbers of mutually exclusive classes in which the members have similar characteristics."[8] The ultimate objective is to develop clusters whose configurations would be such that each entity in the analysis would be classified into only a single, unique cluster. The product of this analysis is referred to by a variety of terms, including types, taxons, groups, classes, categories, classifications, or clusters. By extension, therefore, cluster analysis is also referred to as typologic analysis, numerical taxonomy, pattern recognition, classification, or botryologv.[2] Although the foregoing implies that classification can (as is frequently the case in the literature) be used synonymously with clustering, clustering is associated with the concept of forming classes, whereas classification has been used in the sense of identifying or assigning individual objects to predetermined classes based on specific criteria.[9] The two address different, albeit related, tasks, since one (clustering) typically precedes the other (classification). Frequently, cluster analysis is used to determine groups so that subsequent assigning of new objects to these groups can be achieved.[9,10] Discriminant analysis, a classification technique, is widely used to achieve these subsequent assignments or identifications.

Cluster analysis has been applied for such varied objectives as finding a true typology, model fitting, prediction based on groups, hypothesis generation, hypothesis testing, data exploration, data reduction, and grouping similar entities. into homogeneous classes.[2,5,7,11-13] It, alongwith classification techniques, has been used extensively for medical resource classification. For example, table 1 summarizes various characteristics of resource grouping schemes in both

Received December 21, 1993, from the Department of Management Sciences (DD, JK) and the School of Optometry (AP), University of Waterloo Waterloo, Ontario, Canada. Revision accepted for publication September 29, 1994.

Address correspondence and reprint requests to Dr. Dilts: Department of Management Sciences, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

Table 1 • Summary of Patient Resource Utilization Schemes

Patient Resource Classification Scheme	Setting	Purpose of Scheme	Scheme's Cluster and Validation Method	Sample Size	Groups Obtained	Nature of Scheme*	Reference†
Professional Activity Study (PAS)[70]	Inpatient	Length of stay (LOS) benchmarks	Expert intuition	n/a‡	7,000	Post-classification	CPHA, 1976[70]
Patient Classification by Type of Care (PCTC)	Inpatient	Assessment classification, and placement	Expert intuition, discriminant analysis, and Bayesian procedures	585	5	Pre-classification	Bay et al., 1982[3] Bay et al., 1979[61]
Diagnosis-related Groups (DRGs)	Inpatient	Reimbursement and case mix management	Expert intuition and AUTOGRP	702,000	475	Post-classification	Fetter et al., 1980[45] Fetter et al., 1991[43]
Diagnosis clusters	Outpatient	Code diagnosis	Expert intuition	98,332	92	Post-classification	Scheneeweiss et al., 1983[62]
Reason for Visit Classification (RVC)	Outpatient	Code patient's reason for visit	Expert intuition and summary statistics	n/a§	7 modules§	Post-classification	Schneider et al., 1979[63]
Ambulatory Visit Groups (AVGs)	Outpatient	Reimbursement and link with DRGs	Expert intuition and AUTOGRP	10,145	570	Post-classification	Fetter et al., 1991[43]
Products of Ambulatory Care (PACs)	Outpatient	Reimbursement	Expert intuition, SAS, and AUTOGRP	10,000	24	Post-classification	Tenan et al., 1988”
Low Vision Patient Resource Groups (LVPRGs)	Outpatient	Generate patient iso-resource groups	Expert intuition, block clustering, and replication	99 by 2	5	Post-classification	Dilts et al., 1994”

*Post-classification is classification after the patient consumes health care resources and pre-classification is classification before the patient consumes the health care resources.

† For complete reference citations, see the reference list.

‡ This scheme was not derived from any formal sample. Instead it split patient data from participating hospitals into patient classes on the basis of primary diagnosis, secondary diagnosis, and age.

§ This scheme used no formal sample. It is based on data previously collected by the National Ambulatory Medical Care Survey and categorizes the reasons (provided by patients for visiting their physician) into seven modules, each of which can be further split into hundreds of specific codes.

inpatient (acute and long-term) and outpatient (ambulatory) health care settings. All, except for patient classification by type of care (PCTC), set out to “discover” the latent patient groups in the data. PCTC's aim is one of “placing” patients into predetermined groups, hence its use of discriminant analysis.

When reviewing methodologies used to build existing medical resource classification schemes, it becomes apparent that there is a standard set of decisions that must be made when doing medical clustering. These decisions have not received prominent attention in the proliferation of literature that has accompanied the wide-scale application of clustering techniques. The focus of our paper is on that portion of the literature concerning conventional clustering procedures, which facilitates an understanding of how clustering methods may be used to develop patient resource utilization classification schemes and what decisions are required to use such methods.

Below is an illustrative example of a patient resource classification need in a low-vision clinic. This example is used throughout the paper to demonstrate the decisions to be made when collecting and analyzing medical resource data using cluster analysis (see Dilts et al.” for more details of the setting and the problem). Next, the issues regarding the clustering process are discussed. The process begins with sampling, followed by choice of data attributes, where the variables, data scales, and dimensional analysis are all selected. The choice of a similarity measure is discussed, as is the determination of how to deal with missing data. The various types of clustering algorithms are then reviewed. Determining the optimal number of clusters, interpreting the resulting clusters, and validating them, complete the discussion of using cluster analysis in medical resource classification. Finally, the paper closes with a recapitulation of the decision choices.

Two issues noted in the literature are worth pointing out initially. First, it is readily apparent, even in the most recent material, that there are certain germane pieces of work that are frequently used as authoritative references. They include those of Anderberg,[14] Sneath and Sokal,[15] Everitt,[12] Lorr,[16] and Romesburg,[13] among others. Because they are the authoritative sources we likewise often refer to them. Second, in spite of the proliferation of clustering algorithms in the recent past, many of the contemporary techniques are essentially adaptations of older, more frequently utilized, algorithms, such as single linkage, complete linkage, and ward's method. Algorithm modifications are often developed to make the new technique suitable for a given application area, to take into account some specified aspect, or to overcome a weakness of the original algorithm. The result is an explosion of the number of clustering algorithms in existence. This makes an exhaustive review of existing clustering techniques impracticable. However, as our intention is to discuss the decisions required to effectively utilize the major cluster analysis techniques, our primary attention will be on the few key foundation methods. Readers are encouraged to follow up on the "newer" techniques that seem promising for their specific situations.

An Illustrative Example

Let us assume, for illustrative purposes, that investigators are faced with the following scenario. The objective is to discover whether latent groupings exist among patients utilizing a low-vision clinic's resources. There is no pre-existing grouping of such specialty patients. It is envisaged that once groups are identified and proven robust, the clusters can aid in the prediction of the resources a given patient type will require in a typical visit to the clinic. The investigators' immediate task is to generate clusters of patients whose members demonstrate similar resource utilization patterns. They are, however, constrained by two sub-objectives, namely: **1)** that the groupings generated be reasonable in number, that is, not so numerous as to consist of a few patients each (i.e., with some clusters being rarely seen), and not so few as to be meaningless (i.e., with some clusters being too general); and 2) that the clusters generated be medically and resource meaningful. Not only should they be easily understood when they are described to a low-vision clinician, but also the clinician should be able to relate to these clusters and to identify a particular low-vision management process for patients in each cluster.

Low vision describes a condition afflicting about **1%** of the Canadian and U.S. populations wherein visual performance is diminished to such a level that affected individuals are unable to perform common age-related tasks in spite of conventional medical, surgical, and/or optometric therapy.[17,18] A low-vision clinic provides diagnostic, therapeutic; and patient care services that are geared towards lessening the functional impact of this visual impairment sufficiently that the ensuing disabilities are significantly reduced or eliminated altogether.[18]

A service encounter at a low-vision clinic requires a patient to complete a process that first attempts to identify and 'describe any outstanding visual disabilities through an extensive case history investigation. This is followed by an examination designed to expose and quantify all hidden or manifest visual impairments that have contributed to the reported disabilities." Most school-aged (and other interested) patients also undergo a high-technology assessment that determines their visual performances and adaptations in connection with computer access and closed-circuit television (CCTV) for educational and/or vocational purposes. Depending upon a variety of factors, patients may also use the services of electrodiagnostic assessment, ocular health, or spectacle therapy.

Further, the therapy given in the clinic consists of prescribing sight-enhancement (or sight-assistive) devices and/or recommending suitable task modifications. The devices prescribed may be optical in nature, such as spectacle microscopes or telescopes, or non-optical, such as reading stands or signature guides. The patient's progress over time is monitored through follow-up and reassessment visits.

It has been established that, on average, there is at least a six-month interval between the first contact and a patient's first visit to the clinic. Although it may be medically desirable for time to elapse between first contact and actual patient visit, there is a need to reduce this service delivery delay. Currently, the clinic informally divides patients into two broad categories (students and adults), and, in most cases, patients in a given category are booked for the same sets and lengths of services.

From the clinic's records, the investigators can collect a wide variety of data pertaining to each patient's demographic, diagnostic, and therapeutic characteristics in the study sample. It is important to bear in mind that any single surrogate measure, such as dollars, cannot be used to predict specific resource utilization. For example, determining the dollar cost for a patient visit does not inform the clinic as to which clinical services (ocular health, high-technology assessment, spectacle therapy, etc) were used. To use scheduling, planning, and control techniques that have been used successfully in other settings, it is vital that the use of specific clinic resources be determined in the clusters,

Using a technique such as multiple regression is inappropriate for this problem, for a variety of reasons.

First, the objective is to discover the interdependence of multiple variables with regard to utilization of a variety of clinical resources. Regression analysis focuses on determining the dependency relationship of a single dependent variable (such as time or cost) from a set of independent variables. Second, a single dependent variable would not supply the clinic decision makers with the appropriate information, as their interest is on how to determine which set of patient groups uses which set of resources. They are concerned not with how the groups affect a single dependent variable such as cost or time, but with how each cluster of patients may be related to different clusters of resources. Third, from a statistical perspective, standard stepwise regression analysis begins by finding the single independent variable that best predicts the dependent variable, then enters a second independent variable such that the two independent variables best predict the dependent variable, and so on. The problem with this approach is that, if there are three or more variables that are highly correlated with the initial variable, they will not appear in the regression equation as good predictors of the dependent variable. Cluster analysis circumvents this difficulty by grouping these highly correlated variables together when determining the set of clusters.

The Clustering Process

It is generally agreed that in order to generate subgroups from a given set of entities on some basis of similarity, a process involving a sequence of critical steps should be followed.[6,8,9] In essence, these steps entail answering the following questions: What kind of sample should be collected? What kinds of attributes (variables) should be measured or recorded? How should missing values, if any, be treated? What measure of similarity should be chosen to compare the entities? What cluster-search technique or techniques should be used? What is the optimal number of clusters? What benefit is derived from the clusters generated? How accurate are the clusters? These decisions are interdependent and none should be addressed in isolation. Each of these questions is treated as a decision point and dealt with in turn below.

SAMPLING

Subjects used in a cluster analysis typically represent a sample from a much larger population where the overriding consideration in sampling is to ensure that all types of individual variation in the population of interest are represented in the sample. An effective procedure should therefore be devised for screening the sample objects in order to select the relevant objects while culling out those that are inimical to the purpose of the analysis.

In spite of the aforementioned concern, there is little agreement in the clustering literature on how the sampling process should be conducted. Some authors hold the view that the process "should ensure that 'all data units have an equal opportunity to be selected as part of the sample."' Thus, any group believed to be present in the data should be represented in the sample in proportion to its relative size. This implies strict adherence to statistical approaches of random sampling and independence.

Maintaining a less stringent position, others feel that the statistical principles of random and independent selection should not be applied strictly since not all cluster analyses are concerned with hypothesis testing." They argue that under random selection, small or rare groups in the data tend to become lost in the larger groups, hence it is necessary to be selective about the data units in order to obtain any sizeable sample from such groups. They also feel that situations in which the selection of some units promotes the candidacy of others should not be avoided altogether. Instead, such situations should be exploited for evidence of association rather than neutralized in deference to independence.

Turning to our example, the restrictive random sampling approach is not congruent with the objective of discovering any possible clusters. Investigators thus should follow the selective sampling approach to achieve a reasonable number of medically meaningful clusters. They should ensure not only that all the informal groupings used by the clinic's personnel be well represented in the sample, but also that the sample reflect the clinic's general patient population in terms of gender, age, eye condition, additional impairments, and other characteristics that may be used to distinguish between patients,

CHOICE OF VARIABLES

It is frequently stated in the literature that this step probably has the greatest influence on the clusters that are finally obtained from the analysis. Having determined the sample of entities, investigators must specify a domain that will constitute the basis of classification.[2] Lorr[16] notes that similarity is not a general quality of all elements; elements in a sample may be similar in one domain (e.g., political attitude) but quite disparate in another (e.g., food preference). Domains should not be mixed. In our example, the domain of interest is the utilization of the clinic's resources by the patients; hence only those items that may affect resource usage should be collected. Several additional decisions must be made when choosing attributes or variables.

Variables

Each of the elements of the sample must be consistently described in terms of the variables that measure the domain of interest. There is a need to include all variables that are broadly representative of the domain in such exploratory research. If relevant variables are left out of the analysis, there is a danger that otherwise distinctive groups will remain ambiguous.*“ On the other hand, including strong discriminators not relevant to the purpose at hand is undesirable because they may mask the sought-for groupings and give misleading results.

Over time, different approaches have been used to confront this issue. Zoologists use the “hypothesis of nonspecificity,” which assumes that the classification structure is dependent on many variables, any single one of which can be deleted or added without noticeable effect.*” This, in large part, explains why numerical taxonomy classifications often involve large numbers of variables. Behavioral and social scientists, statisticians, and engineers, on the other hand, emphasize parsimony and seek to minimize the number of variables measured. The latter approach places a premium on wise selection of variables both for relevance and for discriminating power. The lack of pretheory in the area of interest may force an investigator to use a combination of both approaches.” This would mean beginning with a large number of variables, most of which may later be found not to be relevant to the purpose of the investigation. After the irrelevant variables have been weeded out, a more parsimonious set of variables would remain for analysis.

Since the goal in our example was to identify isoresource groups (groups of patients demonstrating similar levels and patterns of the resource demands imposed on the clinic), the investigators originally identified all potentially relevant variables. Over 250 potential variables were collected in total. It was quickly discovered that the majority of these variables had either little variation or little relation to resource usage. The collection of variables was then refocused primarily on resource variables. First, all available data were scrutinized to determine whether they related to resource utilization. A variable was considered to be a resource measure if it identified a service, a facility, an outcome (i.e., purchase of a device), or the time expended by the clinic on the patient. Non-resource variables were discarded when it was determined that they could not be used to supplement the resource variables in distinguishing between the patients. Given the initial nature of the study, however, a large number of additional variables should be collected, which may be discarded after further analysis. In the example, the original set of 256 was reduced to 44 key variables.

Data Scales

Once the foregoing issues have been resolved, the data for analysis are constituted in a sample of N entities or objects described or measured on each of M variables. These are assembled in a data matrix of N rows and M columns. Before the next stage in the process can be attempted, the issue of what data scales to use must be addressed.

When conducting cluster analysis, an investigator is likely to encounter all four basic types of data scales, namely; nominal, ordinal, interval, and ratio.[13,14] Nominal scales distinguish between unordered classes, ordinal scales provide an ordering of the objects, interval scales augment the foregoing characteristics by also specifying the distance between objects, and ratio scales are simply interval scales with a true zero point. Variables on nominal and ordinal scales are usually referred to as categorical or qualitative variables, whereas those on interval and ratio scales are termed quantitative or numerical variables.

Some researchers contend that the only admissible descriptive statistics for nominal scales are mode, frequency counts, and the contingency coefficient C. Median and rank order correlations such as Spearman’s or Kendall’s tau are admissible for ordinal scales, whereas all parametric statistics, such as means, standard deviations, and correlations, may be computed only for the quantitative scales. On this basis, it is argued that since most social science measures are categorical in nature (or at best, simple approximations of interval scales), it may be inappropriate to compute indices such as correlations and distances between entities.[16]

This view originates from what was first suggested by Stevens” and later expounded by authors such as Siegel,[20] Blalock,[21] and others, who insist that parametric procedures are acceptable for quantitative variables, whereas nonparametric procedures are required for categorical scales. A dissenting position is, however, held by many other researchers, including Lord,[22] Burke,[23] Anderson,[24] McNemar,[25] and Gaito,[26] who feel that statistics apply to numbers rather than to things and that these “numbers do not remember where they came fram.” Consequently, although a knowledge of indiaes is needed, indices of similarity such as distances and correlations may be computed even for categorical scales.

This controversy about data scales accounts for the lack of unanimity in the literature regarding what similarity measures should be used where. In general, clustering techniques assume a homogeneity of scale types, whereas actual data sets usually involve mixed scales.[12] This problem can be overcome by choosing the dominant scale in the data and suitably transforming the variables into this one scale type to achieve homogeneity.[2,14]

Transformation techniques are amply covered in the literature.‘” It should be noted, though, that transformations do entail certain “costs.” If scales are ranked in a sequence reflecting decreasing complexity and information demands, the resulting order is: ratio, interval, ordinal, and nominal. Investigators are cautioned that transforming a scale upwards (to ratio) implies gathering of additional information, whereas transforming downwards (to nominal) means forfeiting some information.[2,14]

In the example at hand, the cost of transforming upwards is so prohibitive as to leave the investigators with no choice but to go downwards. It would have been extremely difficult to develop a scale that would have allowed many of the variables to be ratio or interval in nature. Hence, the resulting data set was converted to a completely categorical data set. Another option is to treat categorical data, where the variables have many levels, as binary data, with each level being regarded as a single binary variable.[16] The disadvantage of this approach is the rapid explosion in the number of variables; hence it was not selected.

Data Rationalization

In almost every classification study, the issue of how many variables are desirable must inevitably be addressed.[12] The tendency is to take more rather than fewer variables. This subsequently poses another problem, because, in clustering, the amount of computation increases dramatically with an increase in the number of variables. It is also difficult to avoid some element of redundancy when numerous variables are used. A key consideration here, however, is that it is much harder to describe and explain the final clusters on numerous variables, say 400, than it is on fewer variables, say 15.[16]

It has been suggested that the problem be tackled by first performing dimensional analysis (factor analysis) on the data and then using the more parsimonious principal component or factor scores as input variables for the clustering process.[12,16] This, however, presupposes that the existence of numerical variables in the data can be demonstrated.[27-29] It also behooves the investigator to interpret or infer meaning into the principal components or factors used-a necessary requirement if the final clusters are to be interpretable.

After sifting out those variables inimical to the purpose of the study, the remaining set may be so frugal as to make dimensional analysis unnecessary. Such is the case in our example, as 44 variables can easily be analyzed with existing statistical software. For a listing of the resulting variables and data scale, see the appendix.

DEALING WITH MISSING VALUES

It frequently happens that the data set being analyzed has "holes" in it, that is, it has missing values.2 A missing value results from any one of a number of causes: it may not have been recorded due to oversight or lack of time, it may simply not be available, it may be impossible to measure the desired feature, or it may have been recorded and lost.[2,13]

Several ways of handling missing values exist. In clustering, however, it is usually necessary to replace the missing value with some estimate, because most clustering techniques will either cease operating if them are missing values or delete all data on any entities with missing values.[29] The danger here is the tendency to get artificially higher measures of similarity from a data set with missing values in it than would have been the case if there had been no missing values.‘" There is a need therefore to replace the missing values with estimates.

In quantitative data scales a missing value can be replaced with the mean of values present on that variable.[2] Qualitative data scales, however, present a special problem, because it may not be possible to calculate the mean on a given variable. Even if the mean can be obtained, it is at best meaningless, especially on a nominal scale. It is therefore helpful to resort, for instance, to a technique that uses the patterns of related variables to estimate and replace missing values in the data.[30]

In our example, certain variables were deleted due to excessive missing values. For the remaining variables, either the missing data were estimated or additional information was gained by inspection of patient records directly. With the goal of determining whether clustering was even possible in the low-vision clinic setting, this handling of missing values was considered acceptable. It would not be acceptable if formal hypotheses were to be tested.

CHOICE OF SIMILARITY MEASURE

Most conventional clustering algorithms begin with the calculation of a matrix of similarities. Many diverse similarity measures are used in cluster analysis, and they are extensively covered in both statistical and clustering literature.[4,6,12-14,31,32] Some writers divide similarity measures into four groups, namely: correlation coefficients, distance measures, association measures, and probabilistic similarity measures.[4,12] Correlation coefficients are the similarity measures most commonly used in the social sciences.[14] Of these, the most widely known is Pearson's product-moment correlation coefficient. It is also pointed out that this measure is used only with quantitative data scales, and with binary data (where it is transformed into the phi coefficient).

Distance measures, on the other hand, can take on any non-negative value. The most common distance measures are Euclidean metric, Manhattan or city block metric and the "sup" distance (all three commonly referred to as the Minkowski metrics), and the Mahalanobis distance." The Calhoun distance and the Lance and Williams non-metric distance are other measures proposed in the literature. These, however, are also applicable only to quantitative variables.[14]

Association coefficients measure the relationship between two entities by the values they assume in a given set of variables common to both. In general, these measures take values in the range of 0 to 1, and in many cases the variables are of the "presence" or "absence" type (i.e., binary). A value of 0 indicates that the objects are not similar, whereas 1 indicates maximal similarity. Qualitative variables with many levels can be treated as if each level were a binary variable[2,12] A large number of these coefficients have been proposed in the literature. The most commonly used, however, are three: a simple matching coefficient, Jaccard's coefficient, and Gower's coefficient.[5]

Unlike the foregoing similarity measures, probabilistic similarity coefficients do not calculate similarity between two cases. Instead, they work directly on the raw data. When used in the formation of clusters, the information gain of the combination of two cases is evaluated, and that combination of cases that provides the least information gain is fused into a single cluster.[12] These coefficients can be used only with binary data.

The x^2 statistic and several other measures based on it, including the coefficient of contingency, C, have been suggested for use as similarity measures, but this is disputed by a number of researchers, who maintain that although they may be good tests of hypotheses, such statistics perform poorly as measures of association.[14,33]

Different indices used on the same data give different results of similarity between entities, and, as in several other areas of clustering literature, there is no firm agreement as to which index is the most suitable. As one sage puts it,

> . . . the choice of the correct measure to use would be much simpler if we had prior knowledge of the structure of the data, but this is essentially what we are trying to uncover, and so we encounter a basic circularity- if we knew the clusters we could define the appropriate metric, and if we knew the appropriate metric we could extract the clusters.[12]

What compounds this problem in a practical sense is that measures useful for qualitative variables have not received prominent attention in the clustering procedures of commonly available statistical packages. Thus, the problems associated with finding ideal distance and similarity measures, and a suitable software package to analyze such data, give those clustering algorithms that do not rely heavily on such choices a definite advantage[12]

The qualitative nature of the data in some studies, our example being a case in point, precludes the usage of most of the similarity measures discussed above. In such cases, the issues to be addressed are reduced to: Which one, among the available clustering techniques can handle qualitative data? and Which similarity measure does the technique use? The task becomes one of first identifying the technique and then calculating the similarity measure it requires from the data.

CHOICE OF CLUSTERING ALGORITHM

Several reviews of cluster analysis techniques have been completed.[5,6,12,14,16,30,34-38] The myriad clustering algorithms in existence today make a detailed discussion impossible in this paper. Instead, an overview of the major generic types of algorithms is given, with brief specific examples highlighted when possible. The focus of the review is on the conceptual rather than the mathematical foundation of the algorithms. We follow the recent convention of dividing the algorithms into "hierarchical" and "partitional" and group those that do not conveniently fit into these two under a third category, "other."

Hierarchical Algorithms

Hierarchical algorithms do not generate clusters in one step. Instead, they first split the data into a few broad clusters, each of which is further subdivided into smaller clusters, and the process repeats itself until terminal clusters that cannot be further subdivided are obtained. Techniques using this approach are of two basic types: agglomerative and divisive. Agglomerative techniques proceed by a series of successive fusions of the N objects into groups, whereas divisive algorithms progressively partition the set of N entities into successively smaller sets. Thus, agglomerative methods build an inverted tree or dendrogram structure from branches to the root, and divisive methods begin at the root and form a branching sequence.[16] A feature that has been both a weakness and a strength for hierarchical techniques is that once a fusion or division is made it is irrevocable.2

One of the primary advantages of hierarchical algorithms is that they can be used in situations where the optimal number of clusters is not known beforehand. All possible configurations are provided, and the investigator must select the most appropriate one.

Agglomerative methods begin by computing a similarity or distance matrix between the entities in the data set. They then proceed by successively fusing the objects or groups closest to one another until one group is finally achieved. Many specific agglomerative algorithms exist. A few of the most widely known are summarized in table 2. It is worth pointing out that their differences arise because of the different ways they define distance (or similarity).

In divisive methods, given that there are $2^{N-1} - 1$ ways of splitting a set of N entities into two subsets, considering all possible ways of splitting becomes prohibitive, especially as N becomes large. Restrictions must thus be imposed on the number of divisions considered, and these give rise to monothetic and polythetic divisive methods. Monothetic methods are based on the possession of a single binary attribute that is used to split a set of entities, whereas polythetic methods rely on several variables shared by the members of a cluster.[12,16] Table 3 presents a summary of

Table 2 • Hierarchical Agglomerative Algorithms

Method	Definition of Distance between Clusters for Fusion	Remarks	References*
Nearest neighbor (single linkage)	Distance between their nearest pair of members	May result in long straggly clusters	Anderberg, 1973[14] Everitt, Zupan, 1982[71] Lorr, 1983[16] Aldenderfer et al., 1984[5] Jain et al., Kaufman et al.,
Furthest neighbor (complete linkage)	Distance between their most remote pair of members	May result in small compact clusters	Anderberg, 1973[14] Everitt, 1980[12] Zupan, 1982[71] Lorr, 1983[16] Aldenderfer et al., 1984[5] Jain et al., 1988[6] Kaufman et al., 1990[2]
Centroid method	Distance between their centroids (variable means across the cluster)	Smaller groups could be lost in the larger groups	Fisher, 1969[72] Anderberg, 1973[14] Everitt, 1980[12] Zupan, 1982[71] Lorr, 1983[16] Jain et al., 1988[6] Kaufman et al., 1990[2]
Average linkage	Average distance between all pairs of their members	Could miss outliers	Anderberg, 1973[14] Everitt, 1980[12] Lorr, 1983[16] Aldenderfer et al., 1984[5] Jain et al., 1988[6] Kaufman et al., 1990[2]
Median method	Distance between cluster medians	Similar to centroid but avoids the tendency of small clusters to be lost in larger groups after fusion	Anderberg, 1973[14] Everitt, 1980[12] Lorr, 1983[16] Jain et al., 1988[6] Kaufman et al.. 1990[2]
Ward's method	Uses error sum of squares (ESS)	Fuses the two clusters whose fusion results in the minimum increase in ESS	Anderberg, 1973[14] Biljen, 1973[4] Everitt, 1980[12] Zupan, 1982[71] Lorr, 1983[16] Aldenderfer et al., 1984[5] Jain et al., 1988[6] Kaufman et al.,1990[2]

*For complete reference citations, see the reference list.

the main features of these methods.

The dissimilarity analysis method developed by MacNaughton-Smith et al.[39] is the most common of the polythetic methods used in divisive algorithms. In this method, a splinter group is accumulated by the sequential addition of the entity whose total dissimilarity with the remainder less its total dissimilarity with the splinter group is a maximum. The process is repeated on the two subgroups when this difference becomes negative.

Monothetic methods are applied in cases involving binary data. The data set is divided into those entities possessing, and those lacking, some specified attribute. Examples include the association analysis method[40,41] (and its several variants), and the automatic interaction detector (AID).[42] Although the latter was not initially designed to be a clustering technique, it has been found useful for clustering tasks and is often cited as a monothetic divisive method. A variation of AID, the AUTOGRP, has been used in such resource classification schemes in the medical field as DRGs[43] AVGs[43] and PACs.[44] AUTOGRP was designed to handle large amounts of data, a feature lacking in AID.[45]

Partitioning (Optimization, K-Means) Algorithms

Unlike hierarchical methods, partitioning methods permit the reallocation of entities, thus allowing a poor initial partition to be corrected at a later stage. They partition the data set in a way that optimizes a predetermined criterion (hence the alternative title of op-

Table 8 • Hierarchical Divisive Algorithms

Type of Method	Basis of Division between Clusters	Remarks	Examples	References*
Monothetic	A single binary attribute	Clusters are formed on the basis of having or not having a given attribute	Association analysis AID AUTOGRP	Anderberg, 1973[14] Everitt, 1980[12] Fetter et al., 1991[43] Lorr, 1983[16] Tenan et al., 1988[44]
Polythetic	Several attributes possessed by the cases in a cluster	A new cluster is accumulated by adding cases that are most dissimilar to existing clusters	Dissimilarity analysis	Lorr, 1983[16]

*For complete reference citations, see the reference list.

timization). Their general forms are largely similar-they all require 1) that the number of clusters (k) be determined beforehand; 2) a definition of the cluster mean or centroid (hence the alternative title of K-means methods); and 3) a distance/similarity measure (most use Euclidean distance1.[16,36]

It is worth noting that two issues serve to distinguish between the many variations of these methods, namely the approach they use to obtain an initial partition of the data and the clustering criterion that is optimized. Starting on the premise that the investigator would like to have the entities divided into ***k*** clusters, most of the partitioning techniques begin by finding ***k*** points in the n-dimensional space that act as initial estimates of the cluster center.[12] For example, MacQueen's method selects the first ***k*** points in the sample as the initial ***k*** cluster mean vectors[6] Beale's method[46] begins with a trial value of ***k*** larger than is considered necessary and sets up cluster centers regularly spaced at intervals of one standard deviation on each variable. The number of clusters is thereafter reduced until criterion based on the residual sum of squares is satisfied. Thornedike's method"' selects the ***k*** points furthest apart.

After these ***k*** starting points have been selected, entities are allocated to the cluster whose center they are nearest to in terms of the Euclidean distance. The estimate of the center may be updated after each additional entity joins the cluster, or only after all the entities have been allocated. Employing coarsening and refinement parameters at the analyst's disposal, these methods allow k to change during the clustering process by enforcing the division of a cluster with a large within-cluster sum of squares or by the fusion of two clusters with a small between-cluster sum of squares.[6,16]

Once an initial classification has been determined, a search is made, using a variety of mechanisms, for entities whose reallocation to another cluster would bring an improvement in the criterion being optimized in the classification. This procedure is continued until no further reallocation of a single entity can produce a better criterion value. A local optimum is then said to have been reached and this solution may be accepted, or alternatively, the procedure may be repeated using a different starting configuration.[5,12]

A general flaw with these methods is that it usually cannot be determined whether global maxima have been achieved. Despite this, the methods have been applied in a number of studies.[6,14] It is also worth noting that these methods require the investigator to control the clustering process through a number of parameters, for instance, desired number of clusters, level of optimization, and type of refinement criteria. Most of the available statistical packages do provide an investigator with the flexibility to achieve the desired combination of cluster initiation, reallocation of entities, and stopping rules from the numerous possibilities entailed by these methods. Although this may be desirable in situations where some idea of the final output is known beforehand, it is a definite drawback where the opposite is the case.

Other Algorithms

Some methods, frequently referred to as density search or mode-seeking analysis, were developed to overcome the chaining problem in single linkage methods, that is, the tendency of an algorithm to incorporate entities into existing clusters rather than to initiate new clusters.[12] These techniques adopt an approach that seeks regions of high density or modes in the data, each mode being taken to signify a different cluster.[6]

An example is the taxometric map method,[14] which initially forms clusters in a manner similar to single linkage, but adopts some criteria for judging when additions to the clusters should stop. One of the criteria used is that of subtracting the drop in average similarity on addition of an entity to the cluster from the new average. This value decreases smoothly until a discontinuity is reached. Another criterion is the ratio of the minimum similarity between any pair of points already in the cluster to the minimum similarity of the incoming entity and any point in the cluster. The chief drawback here is that the investigator must choose from numerous parameters to control the method.

Other examples include Girman and Levine's method[48] for detecting unimodal fuzzy sets, and the mixture methods of Pearson, Cohen, Bhattacharya and Hazel." Like their partitioning cousins, these algorithms also suffer from the problem of suboptimal solutions, since there may be more than one solution.

There are numerous other techniques that attempt to address some of the limitations of the more common algorithms. Rohlfs method[49] overcomes the problem of scaling and correlation between variables by adopting a sequential scheme for cluster formation in which distances from an object already in a cluster are measured in the local geometry of that cluster. Hartigan[50-52] describes a method that simultaneously clusters both the entities and the variables. This method sidesteps the issue of similarity / distance measures by seeking modal values that blocks of entities assume on blocks of variables. The blocks identify the latent clusters in the data.

Other methods under this category include an inverse of the "Q" factor analysis,[53] latent structure analysis?" Ling's,[55] Wong and Chiu's method[56] for detecting clusters in mixed-mode data, and many more. The voluminous and diverse literature on the conventional and other more specialized clustering algorithms in existence makes the choice of a clustering technique difficult for the novice investigator. It is worth noting that most of the numerous algorithms cited in the literature today either incorporate features of those algorithms discussed or are variations to adapt to the discipline in which the method is being applied.2 Table 4 summarizes the main features of two of these "other" methods.

Table 5 summarizes the clustering methods in the commonly available statistical packages. For similarity / distance measures, SYSTAT gives. the investigator a choice of Euclidean, gamma, and the Pearson correlation coefficient-implying that only quantitative data scales are admissible.[57,58] The SPSS package offers Euclidean, cosine of vectors of variables, city-block or Manhattan, and Chebychev distances." Like SYSTAT, this package also presupposes that the data are quantitative in scale. The clustering methods in SAS can use coordinate data, distance data, and correlation or covariance matrices.[50] It too, requires quantitative data scales. BMDP offers a wide choice of similarity / distance measures (such as the Jaccard, simple matching, Dice, Ochiai, and Fager coefficients) that are useful for binary data sets.[60] It also contains a clustering algorithm designed to handle categorical data.

In addition to these statistical packages, clustering routines are available in numerous other packages including NT-SYS (Numerical Taxonomy System of Multivariate Statistical Programs), OSIRIS, and CLUSTAN (Cluster Analysis Package). Computer programs for clustering routines are also available in most of the authoritative texts on cluster analysis.

Invariably all the medical classification schemes in the literature use expert intuition in the generation of patient groups. In addition, although their developers do not specifically mention it, most of the schemes have clustered variables that are basically binary in nature (that is, dealing with the presence or absence of some specified feature), and they have therefore tended to utilize algorithms that are designed to proceed through a series of binary splits of the data set (AID and AUTOGRP) .[43-45] Other schemes do not use any formal clustering algorithm because the problem they address is more an assignment than a clustering one; hence their resorting to discriminant analysis.[3,61] Still other schemes are less empirical in approach, relying almost exclusively on expert intuition rather than on formal algorithms in their classifications[62'63] (See table **1** for examples).

In our example, which involves a large number of categorical data (i.e., use of a type of clinical service, gender, types of assistive device purchased, etc.), the nature of the data calls for use of Hartigan's algorithm because of its ability to use categorical data. One advantage of this method is that it allows the investigator to avoid the question of similarity measures.

OPTIMAL NUMBER OF CLUSTERS

A very common problem in cluster analysis is the difficulty associated with determining the optimal number of clusters in a set of data.[12] How this is resolved depends on a number of factors, not least of which is the type of algorithm used. As noted earlier, hierarchical methods in general give a configuration of clusters from **1** to the number of variables. Some algorithms find a best-fitting structure for a given number of clusters. Others begin with a chosen number of clusters and then alter this number per the dictates of some given criteria.[5,12] Many attempts have been made to identify the optimal means of determining

Table 4 • Other Algorithms

Method	Formation of Clusters	Remarks	References*
Density search methods (example: taxometric map)	Regions of high density in the data represent clusters	Numerous parameters are needed to control the clustering	Lorr, 1983[16] Everitt, 1980[12] Aldenderfer et al., 1984[5] Kaufman et al., 1990[2]
Block clustering	Splits data into blocks (clusters) of cases (variables) with similar modal values	Suitable for categorical data Avoids use of distance (similarity) measures	Everitt, 1980[12] Hartigan, 1975,[51] 1992[52]

*For complete reference citations, see the reference list.

Table 5 • Summary of Available Clustering Methods in Selected Statistical Packages

Algorithm	Lowest Data Type	Statistical Package: BMDP Ver 7.0	SAS Ver 6.0	SPS Ver 5.0	SYSTAT Version 5.0 and Windows
Hierarchical agglomerative					
Single linkage	Quantitative	Yes	**Yes**	Yes	Yes
Complete linkage	Quantitative	Yes	Yes	Yes	Yes
Centroid method	Quantitative	Yes	Yes	Yes	Yes
Average linkage	Quantitative	Yes	Yes	Yes	Yes
Median method	Quantitative	Yes	Yes	Yes	**Yes**
Ward's method	Quantitative	Yes	Yes	Yes	Yes
Hierarchical divisive					
Monothetic	Quantitative	No	No	No	No
Polythetic	Quantitative	**NO**	No	No	No
Partitioning					
K-means	Quantitative	Yes	Yes	No	Yes
Other					
Density search	Quantitative	No	Yes	No	No
Block clustering	Categorical	Yes	No	No	No

the number of clusters.[2,13,47,64-69] A completely satisfactory solution has yet to be discovered.

In the example at hand, the investigators were required to rely on expert intuition in determining what configuration of clusters made the most medical sense, as was done in all the medical classifications discussed in table **1.** This resulted in a selection of five potential clusters.

INTERPRETING AND VALIDATING CLUSTERS

Producing a set of clusters is not an end in itself; some use must be made of the results. It has been #maintained that clusters may not only be summary descriptive statistics about the data but also can serve as an aid to reasoning from the data.[14] Most cluster analyses have the latter objective. Viewed as a proposition about the organization of the data, the clusters may give rise to novel interpretations of what is already known, and shed light on previously unnoticed regularities and relations in the data. Alternatively, clusters may be statistically significant but contain falsehoods or sheer irrelevancies. Care should thus be taken in determining and interpreting the clusters.

The results of a classification study need to be evaluated against a background of validating criteria. This is necessary to determine how useful the results are and whether the research goal has been achieved.[13] Frequently, ad hoc methods based on the application area are used to justify the clusters generated.[6] This seems to have been the case in most of the medical classification schemes that have been developed. In spite of this, many objective validation techniques exist. They include: agreement with existing classifications, replication, cophonetic correlation, agreement with expert intuition, agreement of classification with one derived using a different data matrix, agreement of different multivariate methods, demonstration of stability and robustness, significance tests, Monte Carlo procedures, and internal consistency checking.[5,6,7,13] Space constraints do not allow a discussion of these here. It is worth pointing out, however, that invariably all medical classifications have used expert intuition in their development, and some additionally have used agreement of other multivariate methods in the validation of the patient groups obtained.

In our example, five unique clusters were generated (see table 6). As can be seen, the number of relevant variables is considerably reduced from the 256 originally collected and from the 44 variables analyzed. These clusters explain approximately 74% of the variance in the sample, a reasonable amount in such an exploratory study.

One of the greatest difficulties in cluster analysis is identifying for each cluster a "label" that would be meaningful to a user. In our example, the first cluster appears to represent nongeriatric females for reassessment. The second cluster, educationally/vocationally active young males, was one of the surprising results of the analysis. The clinic was aware that young patients typically utilized high-technology assessment but was unaware of the dominance of males in this group. Group three, new geriatric patients with additional impairments, and group four, additionally impaired geriatric females for initial consultation, were groups which the clinic was aware of but not as distinctive groups. Other, the label for group five, represented those patients who could not be grouped into the other four clusters. The results of the clustering process clearly demonstrated that there are significant differences in the amounts and types of resources used by these groups of patients.

To validate the clusters, the investigators in our example used two methods. First, expert intuition of low-vision clinicians was used both to determine the medical meaningfulness of the clusters generated and to validate them. Second, a replication study of a second sample was completed to determine whether the clusters initially obtained were spurious. Results in both cases showed that the generated clusters were representative and robust.

Conclusion

From the foregoing, it can be concluded that clustering is a complicated process marked by several stages where critical decisions must be made. Each and every one of these choices impacts the final clusters generated. Unfortunately, the clustering literature does not provide firm rules-much is left to the judgement and skills of the investigator.

Table 6 • Characteristics of Low-vision Patient Resource Groups

Characteristic	Initial Sample Values	Replication Sample Values Different from Initial Sample
Group I/A		
Gender	Mainly female	—
Age	Mainly less than 55 years old	—
Impairments	About equal with/without additional impairment	Most with no additional impairment
Primary reason	Investigate effective aids	Reassessment
Services	Few booked or used training or CCTV assessment	—
Letters requested	One letter	
Other features	—	Glare problems; not slated for follow-up
Group 2/B		
Gender	Predominantly male	—
Age	Predominately less than 25 years old	—
Impairments	No additional Impairment	—
Patient goal	Educational/vocational	—
Patient type	Student	—
Patient category	New and repeats	—
Services	Used high-technology and CCTV assessment	—
Letters requested	Two letters	—
Other features		Nystagmus and optic atrophy; glare problems
Group 3/C		
Gender	Approximately equal male/female	Predominantly female
Age	All more than 55 years old	—
Impairments	Most have additional impairment	—
Primary reason	Reading	—
Services	Booked and used training and spectacle therapy	—
	Booked but did not use CCTV assessment	Booked and used CCTV assessment
Eye condition	ARMD	—
Devices provided	Presoription reading glasses	—
Other features		Glare problems; predominantly new patients
Group 4/D		
Gender	Predominantly female	
Age	Most more than 55 years old	
Impairments	Most with additional impairment	No additional impairment
Patient category	New	Repeat
Primary reason	Daily living skills, TV	—
Services	Booked and used training and spectacle therapy	Did not use training or spectacle therapy
Devices provided	Prescription spectacles	n/a
Other features	—	No letter, spectacle therapy, computers, or CCTV assessment; no other impairment; glare problems
Group 5/E		
Gender	Mostly female	Predominantly male
Age	—	Predominately more than 55 year old
Services	Did not use spectacle therapy	Did not use CCTV assessment
Impairments	Most with no additional impairment	—
Services	—	Booked and used spectacle therapy, devices
Other features	Have glare problems	n/a
Eye condition	—	ARMD
Patient goals	—	Investigate effective aids

Table 7 • Summary of Clustering Decisions

Decision Area	Choice
Sampling	If hypothesis testing, all data should have the same opportunity for inclusion If small or rare groups are of interest, use selective sampling
Data variables	
Variables	If a pre-theory exists, collect all relevant variables If no pre-theory exists, use a wide selection of variables
Data scales	For all data, assure that the proper scale (nominal, ordinal, interval, or ratio) is known and used to achieve homogeneity For mixed data sets, transform other scales into the dominant scale paying due attention to the "costs" entailed.
Dimensional analysis	For quantitative data, performed dimensional analysis if the number of variables is too large For qualitative data, sift out those variables that are inimical to the purpose of the study
Similarity measure	For quantitative data, select a suitable similarity (distance) measure and transform data into a matrix of similarity (distance) This step may be skipped for qualitative data
Missing values	If data have "holes," select an appropriate method of plugging them, for instance, means for quantitative and suitable estimates for qualitative data
Clustering algorithm	If a high degree of control is desired over the actual clustering by means of a number of parameters, use a partitioning algorithm For qualitative data, use "other" algorithm
Number of clusters	If the optimal number of clusters is not known beforehand, and the algorithm cannot determine the "best" configuration, use an appropriate method, for instance expect opinion, to establish which configuration is the most logical
Interpretation and validation	Assure that the resulting clusters make sense through expert endorsement Use an appropriate technique(s) to assure that the clusters obtained are not spurious

This paper has discussed the issues regarding the clustering process using a low-vision clinic setting for illustration. The process begins with sampling, followed by choice of data variables, where the variables, data scales, scale transformations, and dimensional analysis are all selected. The choice of a similarity measure is considered, as in the determination of how to deal with missing data. The various types of clustering algorithms are then reviewed. Determining the optimal number of clusters and interpreting and validating the resulting clusters are also covered in the discussion on using cluster analysis in medical resource classification. The paper ends with a recapitulation of the steps presented.

A summary of the steps is included in table 7. While there are no formal rules for making clustering decisions, table 7 provides a summary of major decision points and possible rules that could be used when making such choices.

References

1. Temkin 0. The history of classification in medical science. In Katz M, Cole J, Burton W (eds). The Role and Methodology of Classification is Psychiatry and Psychopathology. Rockville, MD: US. National Institute of Mental Health, 1968. DHEW Publication No. (HSM) 72-9015.
2. Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis. New York: John Wiley and Sons, 1990.
3. Bay KS, Leatt P, Stinson SM. A patient classification system for long-term care. Med Care. 1982;20:468-88.
4. Biljen EJ. Cluster Analysis: Survey and Evaluation of Techniques. Groningen, The Netherlands: Tilburg University Press, 1973.
5. Aldenderfer MS, Blashfield RK. Cluster Analysis. Beverly Hills, CA: Sage, 1984.
6. Jain AK, Dubes RC. Algorithms for Clustering Data. Engelwood Cliffs, NJ: Prentice Hall, 1988.
7. Wilson R, Spann M. A new approach to clustering. Pattern Recognition. 1990;23:1413-25.
8. Ward JH, Hook ME. Application of an hierarchical grouping procedure to a problem of grouping profiles. Educ Psycho1 Meas. 196323:69-81.
9. Bock HH (ed). Classification and Related Methods of Data Analysis. Amsterdam: The Netherlands; North Holland, 1988.
10. Gordon AD. Classification. London, U.K.: Chapman and Hall, 1981.
11. Dilts D, Khamalah J, Plotkin A. Clinical low vision resource usage prediction. Optometry and Vision Sci. 1994;71:422-36.
12. Everitt B. Cluster Analysis. New York: Halstead Press, 1980.
13. Romesburg HC. Cluster Analysis for Researchers. Belmont, CA: Lifetime Learning Publications, 1984.
14. Anderberg MR. Cluster Analysis for Applications. New York: Academic Press, 1973.
15. Sneath PHA, Sokal RR. Numerical Taxonomy. San Francisco, CA: W. H. Freeman, 1973.
16. Lorr M. Cluster Analysis for Social Scientists: Techniques for Analyzing and Simplifying Complex Blocks of Data. San Francisco, CA: Jossey-Bass, 1983.
17. Pace RJ. Low Vision: A Clinical Manual. Waterloo, Ontario, Canada: Centre for Sight Enhancement, University of Waterloo, 1990.
18. Strong JG, Pace RJ, Plotkin AD. Low vision services: a model for sequential intervention and rehabilitation. Can J Public Health. 1988;79:50-4.
19. Stevens SS. On the theory of scales of measurement. Science. 1946;103:677-80.
20. Siegel S. Nonparametrlc Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.
21. Blalock HM Jr. Social Statistics. New York: McGraw-Hill, 1979.
22. Lord FM. On statistical treatment of football numbers. Am Psychologist. 1953;8:750-1.
23. Burke DJ. Additive scales and statistics. Psycho1 Rev. 1953;60: 73-5.
24. Anderson NH. Scales and statistics: parametric and nonparametric. Psycho1 Bull. 1961;58:305-516.

25. McNemar Q. Psychological Statistics. New York: John Wiley and Sons, 1962.
26. Gaito J. Measurement scales and statistics: resurgence of an old misconception. Psycho1 Bull. 1980;87:564-7.
27. Cureton EE, D'Agostino RB. Factor Analysis: An Applied Approach. Hillsdale, NJ: Lawrence Erlbaum, 1983.
28. Hair FJ, Anderson RE, Tatham RL, Grablowsky BJ. Multivariate Data Analysis with Readings. New York: Macmillan, 1984.
29. Norusis MJ. SPSS/PC+ Advanced Statistics V2.0 for the IBM PC/XT/AT and PS/2. Chicago, IL: SPSS Inc., 1988.
30. Frane J. Description and estimation of missing data. In Dixon WJ (ed). BMDP Statistical Software Manual Volume 2. Berkeley. CA: University of California Press, 1988, pp. 813-29.
31. Clifford DHT, Stephenson W. An Introduction to Numerical Classification. New York: Academic Press, 1975.
32. Skinner HA. Differentiating the contribution of elevation, scatter, and shape in profile similarity. Educ Psycho1 Meas. 1978;38:297-308.
33. Goodman IA, Kruskal WH. Measures of association for cross classifications. J Am Statis Assoc. 1954;49:732-64.
34. Bijman J. Cluster Analysis. Tilberg, The Netherlands: Tilberg University Press, 1973.
35. Duran BS, Ode11 PL. Cluster Analysis: A Survey. Berlin: Springer-Verlag, 1974.
36 Hartigan JA. Clustering Algorithms. New York: John Wiley and Sons, 1975.
37. Trahanias P, Skordalaskis E. An efficient clustering method. Pattern Recognition. 1989;22:449-53.
38. Sridhar V, Murty MN. Clustering algorithms for library comparisons. Pattern Recognition. 1991;24:815-23.
39. MacNaughton-Smith P, Williams WT, Dale MB, Mockett LG. Dissimilarity analysis. Nature. 1964;202:1034~5.
40. Lambert JM, Williams, WT. Multivariate methods in plant ecology. IV: Nodal analysis. J Ecology. 1962;54:775-802.
41. Lambert JM, Williams WT. Multivariate methods in plant ecology. IV Comparison of information analysis and associative analysis. J Ecologv. 1966;54:635-64.
42. Sonquist AJ, Morgan JN. The Detection of Interaction Effects. Ann Arbor, MI: Institute of Social Research, University of Michigan, 1964.
43. Fetter RB, Brand DA, Gamache DE (eds). DRGs: Their Design and Development, Ann Arbor, MI: Health Administration Press, 1991.
44. Tenan PM, Fillmore HH, Caress B, et al. PACs: classifying ambulatory care patients and services for clinical and financial management. J Ambulatory Cam Manage. 1988;11(3):36-53.
45. Peter RB, Youngsoo S, Freeman JL, Averill RF, Thompson JD. Case mix definition by diagnosis-related groups. Med Care. 1980;18(2) suppl:l-53.
46. Beale EML. Cluster Analysis. London: Scientific Control Systems Ltd., 1969.
47. Thornedike RL. Who belongs in a family? Psychometrika. 1953;18:267-76.
48. Gitman I, Levine MD. An algorithm for detecting unimodal fuzzy sets and its application as a clustering technique. IEEE Trans Comput. 1970;19:583-93.
49. Rohlf FJ. Adaptive hierarchical clustering schemes. Comput J. 1970;16:93-5.
50. Hartigan JA. Direct clustering of a data matrix. J Am Statist Assoc. 1972;67:123-9.
51. Hartigan JA. Clustering Algorithms. New York: John Wiley and Sons, 1975.
52. Hartigan JA. Block clustering. In Dixon WJ (ed). BMDP Statistical Software Manual. Volume 2. Berkeley, CA: University of California Press, 1992, pp. 891-8.
53. Cartel1 RB. Factor Analysis. New York: Harper, 1952.
54. Lazarsfeld PL, Henry NW. Latent Structure Analysis. Boston, MA: Houghton Miffin, 1968.
55. Ling RF. On the theory and construction of k-clusters. Comput J. 1972;15:326-32.
56. Wong AKC, Chiu DKY. Synthesizing statistical knowledge from incomplete mixed-mode data. IEEE Trans Pattern Anal Machine Intell. 1987;PAMI-9:459-65.
57. Wilkinson L. SYSTAT: The System for Statistics. Evanston, IL: SYSTAT Inc., 1990.
58. Wilkinson L. SYSTAT for Windows: Statistics, Version 5 Edition. Evanston, IL: SYSTAT Inc., 1992.
59: SAS Institute. SAS/STAT User's Guide, Version 6. Fourth Edition. Volume 1. Cary, NC: SAS Institute Inc., 1990.
60. Dixon WJ (ed). BMDP Statistical Software Manual. Volumes 1 and 2. Berkeley, CA: University of California Press, 1992.
61. Bay KS, Overton P, Harrison F. Patient Classification by Types of Care: A Research Report on the Development and Validation of the PCTC System. Edmonton, Alberta, Canada: Division of Health Administration, University of Alberta, 1979.
62. Scheneeweiss R, Rosenblatt RA, Cherkin DC, Kirkwood CR, Hart G. Diagnosis clusters: a new tool for analyzing the content of ambulatory care. Med Care. 1983;21:105-22.
63. Schneider D, Appleton L, McLemore T. A reason for visit classification for ambulatory care. Vital Health Stat. 1979;Series 2(8): l-7.
64. Engelman L, Hartigan JA. Percentage points off a test for clusters. J Am Statist Assoc. 1969;64:1647-8.
65. Marriot FHC. Practical problems in a method of cluster analysis. Biometrics. 1971;27:501-14.
66. Wolfe JH. Comparative Cluster Analysis of Patterns of Vocational Interest. Technical Bulletin STB 72-3. San Diego, CA: Naval Personnel and Training Research Laboratory, 1971.
67. Gower JC. Goodness-of-fit criteria for classification and other patterned structures. Pmc 8th International Conference on Numerical Taxonomy. 1975, pp. 38-62.
68. Chan KP, Cheung YS. Clustering of clusters. Pattern Recognition. 1992;25:211-7.
69. Lennington RK, Flake RH. Statistical evaluation of a family of clustering methods. Proc 8th International Conference on Numerical Taxonomy. 1975, pp. 75-97.
70. Commission on Professional and Hospital Activities. Length of Stay in. PAS Hospitals, by Diagnosis. Ann Arbor, MI: CPHA, 1976.
71. Zupan, J. Clustering of Large Data Sets. Chichester, U.K.: John Wiley and Sons, 1982.
72. Fisher WD. Clustering and Aggregation in Economics. Baltimore, MD: The Johns Hopkins Press, 1969.