Using Hierarchical Cluster Analysis in Nursing Research

Jason W. Beckstead

This article is a pedagogical piece on hierarchical cluster analysis, a method for investigating the structure underlying data. Such methods are useful for finding similar groups of cases in data sets when it is not known a priori how many groups are present. The article is laid out as follows: First, a brief history and overview of the methods is presented; second, an illustrative example with a small hypothetical data set is used to clarify fundamental concepts; third, hierarchical cluster analysis is applied to a data set from the author's own program of research to illustrate one way in which the methods may be employed in nursing research; fourth, the limitations of the methods are discussed; and finally, a list of suggested readings, at varying levels of detail, are provided for the interested researcher.

Cluster analysis is a generic name for a wide variety of methods or procedures that can be used to create a classification. These procedures empirically form discrete groups or "clusters" of similar entities based on similarities among their measured features. Entities may be individuals, objects, or variables. This article presents a brief introduction to these methods with a focus on their application in nursing research.

Methods of cluster analysis may be used in nursing research to group hospitals or clinics based on their organizational and patient characteristics, patients based on their needs, diseases according to their symptoms, nurses according to their attitudes and beliefs, and countless other applications. The major stimulus for the development of cluster analysis was Principles of Numerical Taxonomy, published in 1963 by two biologists, Robert Sokal and Peter Sneath. Since then, there have been expansive developments in both theory and application of cluster analysis in psychology, sociology, and anthropology, as well as biology. Over the years these methods have taken on numerous names within the various disciplines such as profile analysis, numerical taxonomy, seriation, and clumping strategies, to name a few.

Jason W. Beckstead, Ph.D., College of Nursing, University of South Florida.
© 2002 Sage Publications
HOW CAN CLUSTER ANALYSIS BE APPLIED IN NURSING RESEARCH?

The primary reason for using cluster analysis is to find similar (homogeneous) groups of cases in a data set. Cluster analysis refers to a set of objective methods for quantifying the structural characteristics of a set of observations (Hair & Black, 2000, p. 173). Hierarchical methods are particularly well suited to situations in which the researcher cannot state up front how many groups are present in the data set. The value of exploratory cluster analysis to the researcher resides in the tendency for new arrangements of data to suggest relationships and principles previously unnoticed. In this sense, cluster analytic methods represent a kind of “retrospective” reasoning in science (Anderberg, 1973, p. 22). Aldenderfer and Blashfield (1984) suggested that cluster analysis may be useful in the pursuit of four principal goals: development of a typology or classification, investigation of useful conceptual schemes for grouping entities, hypothesis generation through data exploration, and hypothesis testing, or the attempt to determine whether classifications defined through other means are, in fact, present in a data set. Pursuit of these goals may further advances in nursing research.

The typical procedure for conducting a cluster analysis has the following steps: (a) Select a sample of entities. (b) Define a set of variables (i.e., features) on which to measure the entities. (c) Calculate measures of proximity between all the entities. (d) Group the entities together based on their proximity scores using a clustering algorithm. (e) Create a graphic depiction of the groupings that emerge in order to facilitate interpretation of the results. (f) Validate the resulting cluster solution using information about the entities that was not employed in the formation of the clusters.

Hierarchical cluster analysis differs from other methods of grouping people or variables, such as discriminant analysis or confirmatory factor analysis, in two important ways. First, the number of groups need not be known a priori. Second, there are situations in which clusters have subclasses within them, which in turn have subclasses of their own. Examples can be found in biological taxonomy where similar species are placed within a genus, which is within an order, within a class, within a phylum, and finally, within a kingdom. Similar structures appear in various diagnostic systems.

Methods of Clustering

The three most popular classes of clustering methods used in the social sciences are hierarchical agglomerative, iterative partitioning, and factor
analytic techniques. Hierarchical agglomerative clustering is a general term for those methods that attempt to group a sample of $N$ entities together in nested or concentric arrangements. The resulting clusters are hierarchical in that higher order clusters are formed from lower order clusters. Hierarchical methods use either agglomerative or divisive algorithms. These methods begin with calculation of an $N \times N$ matrix of proximities. The aim is to investigate the underlying structure of the data. In the agglomerative approaches presented here, entities that are most similar, or which are the least far apart, are grouped together to form clusters. The number of entities in the sample is reduced in a stepwise fashion. The initial $N \times N$ matrix is examined for the pair of entities that are the most similar. This pair is combined to form the first cluster. A new $(N - 1) \times (N - 1)$ matrix of proximities is calculated and examined for the pair of entities (one of which is now a cluster of two entities) that is the most similar. This pair is then combined, and the two-step process is repeated until all entities reside in a single cluster. In this way, cluster analysis may be understood as an iterative process. Hierarchical divisive methods are the logical opposite of agglomerative methods. All the entities belong to one cluster (i.e., $N = 1$) at the beginning of the procedure and then this cluster is cut into successively smaller chunks. The purpose of this article is to introduce hierarchical cluster analysis to nurse researchers using a small, illustrative example followed by an application of the methods to a data set from the author’s own area of research. First, a discussion of similarity measures is needed.

**Measures of Similarity or Proximity.**

There are many indices available to quantify the degree of similarity or closeness between entities. Correlation coefficients are probably the most familiar. Two examples are the Pearson coefficient $r$ and the Spearman correlation coefficient $r_s$, known as angular measures because of their geometric interpretations. Larger values on these measures indicate greater degrees of similarity.

There are other proximity indices, referred to as distance measures because larger values on such measures indicate that entities are farther apart, or more dissimilar. One of the most popular of these is the distance measure $d$, which is the generalized Pythagorean distance between two points in Euclidean space. Euclidean distance ($d$) is defined as the square root of the squared discrepancies between two entities summed over all variables (i.e., features) measured. For any two entities $A$ and $B$ and $k = 2$ features, say, $X_1$ and $X_2$, $d_{ab}$ is the length of the hypotenuse of a right triangle.
The square of the distance between the points representing A and B is obtained as follows: \( d_{ab}^2 = (X_{a1} - X_{b1})^2 + (X_{a2} - X_{b2})^2 \). The square root of this expression is the distance between the two entities. If the number of features being considered exceeds two (e.g., \( k > 2 \)), then the distance is given by \( d_{ab} = \left[ \sum(X_{aj} - X_{bj})^2 \right]^{1/2} \) where \( j \) ranges from 1 to \( k \). Using this approach, the distance between any two entities is expressed as a composite sum of their discrepancies across all measured variables making up their profiles. Proximity measures such as \( d \) may be calculated for any number of features considered in \( k \)-dimensional hyperspace. Distance, or dissimilarity measures, normally have no upper bounds and are scale-dependent. The examples below are based on the distance measure \( d \).

The term profile comes from the practice in applied testing of plotting an individual's scores on a battery of tests as a profile. As first demonstrated by Cronbach and Gleser (1953), the similarity between profiles is made up of three components: shape, the pattern of dips and rises across the variables; scatter, the dispersion of an individual's scores about their average; and elevation, the mean score of the individual over all the variables making up the profile. These components are represented, to varying degrees, in different measures of proximity. Consider two individuals whose profiles show the same shape and scatter but differ by a constant amount across all variables. A correlation coefficient calculated between their profiles would not reflect this difference in elevation. In comparison to Euclidean measures, correlation coefficients tend to be very sensitive to similarities in shape at the expense of magnitude and dispersion. Jaccard and Wood (1985) demonstrated how the three components of proximity may be isolated prior to the clustering procedure when working with Euclidean distances. By “centering” or subtracting each individual's mean score over all variables from each score within his or her profile, differences in elevation are removed. Differences in scatter may be removed by dividing each score in an individual's profile by his or her unique standard deviation.

**AN ILLUSTRATIVE EXAMPLE WITH HYPOTHETICAL DATA**

Consider the following example. Five entities, let us refer to them as \( A, B, C, D, \) and \( E \), are measured on three variables, \( X_1, X_2, \) and \( X_3 \). The first step is the calculation of a proximity matrix containing the distance scores \( d_{ij} \) among the five entities (see Table 1). For example, the distance score, \( d_{ac} \), between the entities \( A \) and \( C \) is formed by \( [(19 - 14)^2 + (52 - 50)^2 + (20 - 24)^2]^1/2 = [25 + 4 + 16]^1/2 = 45^{1/2} = 6.708 \). The next step in the process is to
### Table 1: Illustration of Hierarchical Cluster Analysis of Hypothetical Data Using Between-Groups Averaging Algorithm Applied to Euclidean Distance Measures

<table>
<thead>
<tr>
<th>Raw Data Matrix</th>
<th>Step 1 Proximity Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>X₁</td>
<td>X₂</td>
</tr>
<tr>
<td>A</td>
<td>19</td>
</tr>
<tr>
<td>B</td>
<td>20</td>
</tr>
<tr>
<td>C</td>
<td>14</td>
</tr>
<tr>
<td>D</td>
<td>15</td>
</tr>
<tr>
<td>E</td>
<td>18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 2 Proximity Matrix</th>
<th>Step 3 Proximity Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>(AB)</td>
<td>C</td>
</tr>
<tr>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>7.096</td>
</tr>
<tr>
<td>D</td>
<td>6.113</td>
</tr>
<tr>
<td>E</td>
<td>5.611</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 4 Proximity Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>(ABE)</td>
</tr>
<tr>
<td>0.000</td>
</tr>
</tbody>
</table>

**Note:** Numbers in boldface represent smallest distance at each step.

Inspect these distance scores for the closest pair of entities (i.e., those with the smallest distance score). Inspection of Table 1 reveals this to be the pair A and B, which has a distance value \( d_{ab} = 1.000 \). These two entities are combined to form our first cluster \((AB)\), and the distance from this cluster to all the other elements of the proximity matrix is computed. When two entities have merged, they are joined together permanently and become the building block for later merges. This step raises the question of how best to combine the information on entities A and B so as to calculate their joint distance from the other entities. This issue is central in the development of cluster analytic theory and has received the most attention from cluster analysts since the inception of these methods. Many algorithms exist that may be grouped as linkage methods, centroid methods, or sums of squares methods; a comparison between them is beyond the scope of this article (see Milligan, 1980, for discussion). One of the most intuitive linkage algorithms, referred to as the between-groups averaging method, will be used to illustrate the analytical process. This method simply averages the values in the proximity matrix as to estimate the distances between a newly formed cluster and the other entities. For example, the distance from the cluster \((AB)\) to C is given by...
Figure 1. Dendrogram of hypothetical data clustered using between-group average algorithm applied to Euclidean distance measures.

averaging the distances between $A$ and $C$ (6.708) and between $B$ and $C$ (7.483), resulting in the distance $d_{abc} = 7.096$ between $(AB)$ and $C$. Similar calculations were performed to obtain the distances between the cluster $(AB)$ and entities $D$ and $E$, and the results are shown in Table 1 in the Step 2 proximity matrix. The next step is to reexamine the updated distance scores for the next smallest value. The pair $C$ and $D$ have the smallest distance $d_{cd} = 1.414$ and is combined to form the second cluster $(CD)$. The distances from
(CD) to the other entities and clusters are found by averaging their individual distances to E and (AB), as above, to produce the values in the Step 3 proximity matrix. The values in this matrix are examined to reveal that E is closest, or most similar, to cluster (AB). Subsequently, E is merged with cluster (AB) to form (ABE), which is 7.006 distant from cluster (CD). At this point, all five entities are combined into one cluster and the process is terminated.

This iterative process may be summarized graphically using a branching diagram known as a dendrogram. The entities being clustered are arranged along the horizontal axis perpendicular to the values of the proximity measure d (see Figure 1). Horizontal lines are placed at the values of d that correspond to the distances at which the entities (labeled at the base of the vertical lines) combine to form clusters. For example, A and B combine on the left side of the figure at a distance $d_{ab} = 1.000$. The next cluster (CD) is formed at $d_{cd} = 1.414$, shown on the right-hand side. Traditionally, the dendrogram appears as an inverted tree with branches at the bottom and the trunk at the top. Moving from the branches to the trunk depicts increasing aggregation of the entities into clusters. The hierarchical nature of the process becomes visible as E joins with cluster (AB) at a distance $d_{abE} = 5.611$. Using a dendrogram it is possible to “see” the groupings or degree of structure within a data set. The relative distances provide an intuitive feeling for the tightness or looseness of the groupings. Inspecting Figure 1, however, one may be left wondering if the correct result involves two clusters, (ABE) and (CD), or three clusters, (AB), (CD), and E. Or for that matter, just one cluster containing all five entities, or five clusters of one entity each!

In practice, interpretation is often largely subjective. One way of reducing subjectivity is to use stopping rules that take into account the mean and standard deviation among the $N-1$ fusion values. Using this information it is possible to calculate test statistics, such as $t$ with degrees of freedom equal to the number of entities minus 2, for each step and determine when to stop combining clusters (Mojena, 1977). Comparing these statistics to conventional critical values, it is possible to determine, in a probabilistic sense, the maximum number of significantly different clusters. In this example, the first step to produce a significant statistic was Step 3, ($t = 25.85$, the critical value is 3.18 for $\alpha = .05$), suggesting that we should stop the process prior to combining E with (AB), resulting in three distinct clusters: (AB), (CD), and E. Other stopping rules are based on inspecting a scree plot of the average proximity scores at successive steps in the process and looking for irregularities (Aldenderfer & Blashfield, 1984, pp. 53-58).
AN EXAMPLE WITH REAL DATA

The author is engaged in a program of research focusing on how nurses think about working alongside other nurses whose professional performance is impaired due to drug or alcohol problems. In a recent experiment, 126 nurses were asked to rate their impressions of fictitious coworkers impaired due to drugs or alcohol. Descriptions of coworkers were constructed from a factorial design that systematically combined information about the type of substance involved (alcohol, marijuana, or narcotics), known history of medication errors, and co-occurrence of depressive symptoms (see Beckstead, in press, for details). The participants gave their impressions of the fictitious coworkers using an 11-point rating scale (0 = not at all impaired to 10 = completely impaired). Each participant’s set of 24 responses formed their profile and was used to calculate proximities between all pairs of individuals. Cluster analysis was then used to identify groups of nurses who showed similar patterns in their impressions. Prior to clustering, the data were centered to remove differences in elevation. This meant that the resulting clusters would not be based on individual differences in mean ratings but on differences in the shape and scatter of each nurse’s responses.

A more complex algorithm, Ward’s (1963) sums of squares method, was employed using a cluster analysis program ClustanGraphics Version 4.08 (Wishart, 1999) to produce the cluster solution shown in Figure 2. Ward’s method determines cluster membership by minimizing a function based on the sum of squared deviations between the entities and the centers (or means) of each cluster. It was used here because it tends to produce very homogeneous clusters with maximal distances among them. ClustanGraphics prepares the dendrogram for printing in color or for export to other software for presentation. Many standard statistical software packages (e.g., SAS, SPSS, S-Plus, etc.) have procedures to conduct cluster analysis but their graphic displays are less sophisticated.

Applying the t-test stopping rule described in the hypothetical example above, the maximum number of distinct clusters was determined to be five. The dendrogram in Figure 2 summarizes the clustering process and reveals five distinct clusters of approximately 25 participants (participant numbers were omitted from the dendrogram for clarity). The five clusters appear as shaded groups at the bottom of the figure. The program saves information that indicates cluster membership at specified points in the process for use in subsequent analysis. To gain some understanding of the resulting cluster solution, cluster membership was treated as a between-subjects factor in a 5
Figure 2. Dendrogram showing five relatively distinct clusters ($n = 21, 24, 29, 28,$ and 29, from left to right) formed within a sample of 126 nurses. The clusters were generated using Ward's (1963) algorithm applied to squared Euclidean distances. Clusters are based on similarities in impression ratings given to 24 fictitious nurses impaired by drugs or alcohol. Data were centered to remove interrespondent differences in elevation prior to calculating distance measures.

$\times (4 \times 3 \times 2)$, Type of Substance Involved $\times$ History of Medication Errors $\times$ Co-Occurrence of Depressive Symptoms factorial ANOVA performed on the responses. Reporting the full ANOVA is not germane to the discussion at hand. However, a summary of the results will be used as an aid in
substantive interpretation of the clustering solution and to provide validating support for the distinctiveness of the five clusters.

The five clusters of nurses differed significantly on several of the main effects and interactions making up the full ANOVA model. All interactions with cluster membership were significant at $p < .05$, with the exception of the (Cluster × Depression) and (Cluster × Medication Error × Depression) effects. Differences in elevation (i.e., grand means for each cluster) were not expected due to the centering technique employed. Thus, the patterns of main effects and interactions revealed how the clusters of nurses differed in the way they weighted information when forming their impressions of impaired coworkers. Gross distinctions among clusters rested on the simple main effects of type of substance and history of medication error. Inspection of effect size estimates ($\eta^2$ values) revealed that nurses in Clusters 3, 4, and 5 weighted information regarding substance use substantially as they formed impressions ($\eta^2 = .243$, .232, and .213, respectively). This factor did not have as much of an influence among nurses in Clusters 1 and 2 ($\eta^2 = .014$ and .091, respectively). Regarding the History of Medication Error factor, nurses in Cluster 2 weighted this factor quite heavily ($\eta^2 = .428$), whereas those in Clusters 1, 3, 4, and 5 did not ($\eta^2 = .093$, .056, .072, and .076, respectively).

Other more subtle differences involving interactions were observed. Clusters 2 and 5 showed significant (Type of Substance × History of Medication Error) interactions, whereas Clusters 1, 3, and 4 did not. The nature of this interaction was significantly different for these two clusters: For the nurses in Cluster 2, history of medication error made less of a difference when either narcotics or marijuana was involved compared to situations where alcohol or no substances were involved. History of medication error had its strongest impact in situations involving only marijuana for nurses in Cluster 5. Clusters 3, 4, and 5 showed significant (Type of Substance × Depression) interactions, whereas Clusters 1 and 2 did not. The nature of this interaction differed significantly among the three clusters: For the nurses in Cluster 3, depressive symptoms made no difference when narcotics were involved but did have an effect in the other three conditions. Depressive symptoms made no difference when the type of substance involved was marijuana for nurses in Cluster 4. For nurses in Cluster 5, depressive symptoms made more of a difference when no substances were involved as compared to situations where marijuana, alcohol, or narcotics were involved; the effect of depression did not differ among the three types of substances.

These findings suffice to illustrate that significant and interpretable patterns of impression formation among nurses were obviated via cluster analysis.
Validation of the clustering results was further supported by differences among the clusters with respect to variables external to the clustering process. The clusters showed significant differences on measures of attitude toward substance abuse using the Substance Abuse Attitude Survey (SAAS), a scale for assessing attitudes among clinicians (physicians, medical students, psychologists, social workers, and nurses) toward various aspects of alcohol and drug abuse (Chappel, Veach, & Krug, 1985). The instrument as used assessed permissive-restrictive attitudes regarding substance use. Significant differences were found among the five clusters on their attitudes, $F(4, 121) = 4.465, p < .01$. Ranked from most to least permissive on their attitudes were Cluster 2, Cluster 1, Cluster 3, Cluster 4, and Cluster 5, respectively. Follow-up analysis using Games-Howell $t$ tests revealed that Cluster 5 was significantly less permissive than either Clusters 1 or 2 ($p$'s < .05). The five clusters of nurses also showed significant differences when asked how likely they would be to report fictitious coworkers to supervisors, $F(4, 121) = 6.143, p < .001$. Follow-up tests revealed that nurses in Cluster 5 were significantly more likely to report impairment to a supervisor than were the nurses in any of the other clusters ($p$'s < .05).

ASSUMPTIONS AND LIMITATIONS OF HIERARCHICAL CLUSTERING

These methods have been described as being descriptive, atheoretical, or noninferential. Cluster analytic methods have no statistical basis on which to draw statistical inferences from a sample to a population and are primarily used as exploratory techniques. These methods have strong mathematical properties but not statistical foundations. The requirements of normality, linearity, and homoscedasticity that are so important in other techniques have little bearing on cluster analysis (Hair & Black, 2000, p. 173). Andersen (1973, p. 11) comments that hierarchical cluster analysis is generally not involved with hypothesis testing, and therefore the necessity of imposing principles such as random and independent selection of cases should be carefully examined. The methods will always create clusters regardless of the "true" existence of any structure in the data. Cluster analytic methods are sensitive to the inclusion of irrelevant variables that can degrade the reproducibility of group assignments (Hair & Black, 2000, p. 161). Consequently, researchers are strongly encouraged to always validate their results. Monte Carlo modeling has shown that the choice of clustering algorithm is more important than the choice of proximity measure with Ward's (1963)
method and the averaging methods performing best at recovering the structure of the data under various conditions (Milligan, 1980).

SUMMARY AND CONCLUSION

The goal of this article has been to introduce cluster analysis to nurse researchers using illustrative examples. Cluster analysis proved useful in discovering substantive differences in the way nurses formed impressions of impaired coworkers. Evidence for distinct types or groups of similarly thinking nurses emerged. These cognitive differences among clusters of nurses appeared to be related to the attitudes the nurses held toward substance abuse. The clusters also showed differences in how likely they were to report a fictitious impaired coworker to a supervisor.

Nurse researchers are encouraged to explore the utility of hierarchical clustering methods in various domains of research. As nurse researchers become more sophisticated in developing and testing nursing theory, they must make use of all available analysis strategies that facilitate the discovery of relationships and principles previously unnoticed. Cluster analytic methods have the ability to rearrange the data so as to sharpen the researcher’s faculties for pattern recognition and discovery. Adding these methods to their repertoire of critical thinking skills will serve nurse researchers well.

There are various sources, written at differing levels of detail, on hierarchical cluster analysis. A brief introduction to cluster analytic procedures is provided by Dunn-Rankin (1983, pp. 121-161). A clear and nontechnical presentation of cluster analysis may be found in Hair and Black (2000). Another excellent and quite readable treatment of the most popular techniques is given in Aldenderfer and Blashfield (1984). Nunnally and Bernstein (1994) discussed the methods for clustering subjects under the heading of profile analysis (see pp. 597-652). An in-depth, mathematically rigorous exposé on cluster analytic methods is offered by Anderberg (1973). There are also some excellent Web sites dealing with cluster analysis (e.g., http://www.cs.mcgill.ca/~viper/cs644/cluster.html). For another interesting application of hierarchical cluster analysis in nursing research, see the Nursing Interventions Classification (Iowa Intervention Project, 1993) that combines cluster analysis and traditional Q-sort methods to construct a taxonomy of over 350 nursing interventions.
REFERENCES