A preliminary comparison of the effectiveness of cluster analysis weighting procedures for within-group covariance structure.

Donoghue, John R.

Educational Testing Service, Princeton, N.J.

ETT-RR-95-35

Oct 95


Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

ABSTRACT

A Monte Carlo study compared the usefulness of six variable weighting methods for cluster analysis. Data were 100 bivariate observations from 2 subgroups, generated according to a finite normal mixture model. Subgroup size, within-group correlation, within-group variance, and distance between subgroup centroids were manipulated. Of the clustering methods examined, the flexible average algorithm with beta equal to -.15 or -.20 gave the best recovery. Of the remaining methods, that of J. H. Ward yielded the best recovery, followed closely by beta-flexible linkage and the EML algorithm of the Statistical Analysis System. In the absence of variable weights, negative within-group correlation resulted in much poorer recovery for all clustering algorithms. The ACE weighing method of D. Art, R. Gnanadesikan, and J. R. Kettenring proved preferable overall. Clustering with Mahalanobis distance based on the pooled within-group covariance matrix indicated that knowing the correct covariance method would yield improved recovery over the ACE method approximately 10% of the time. Two appendix figures provide weighting data. (Contains 8 figures, 7 tables, 2 appendix figures, and 40 references.) (Author/SLD)
A PRELIMINARY COMPARISON OF THE EFFECTIVENESS OF CLUSTER ANALYSIS WEIGHTING PROCEDURES FOR WITHIN-GROUP COVARIANCE STRUCTURE

John R. Donoghue
A Preliminary Comparison of the Effectiveness of Cluster Analysis Weighting

Procedures for Within-group Covariance Structure

John R. Donoghue

Educational Testing Service

This research was supported by Educational Testing Service Research Allocation Project 794-10.

An earlier version of this paper presented at the Educational Statisticians SIG at the annual meeting of the American Educational Research Association in New Orleans, LA, April 1994. The author would like to thank Carolyn Crone and Scott Wesley for their helpful comments.
Abstract

A Monte Carlo study compared the usefulness of six variable weighting methods for cluster analysis. Datasets were 100 bivariate observations from two subgroups, generated according to a finite normal mixture model. Subgroup size, within-group correlation, within-group variance, and distance between subgroup centroids were manipulated.

Of the clustering methods examined, the flexible average clustering algorithm with $\beta = -.15$ or -.20 gave the best recovery. Of the remaining methods, Ward's method yielded the best recovery, followed closely by beta-flexible linkage ($\beta = -.50$) and SAS's EML algorithm.

In the absence of variable weights, negative within-group correlation resulted in much poorer recovery for all clustering algorithms. The ACE weighting method of Art, Gnanadesikan, and Kettenring provided a net improvement in 17-24% of the datasets when used with better clustering algorithms. When used with the same clustering algorithms, De Soete's ultrametric weighting yielded improved recovery 16-22% of the time. However, although ultrametric weighting was more sensitive than ACE to negative within-subgroup correlation. Clustering based on principal components was less effective. Therefore, the ACE method is preferred overall.

There is still room for improvement, however. Clustering with Mahalanobis distance based on the pooled within-group covariance matrix indicated that knowing the correct covariance matrix would yield improved recovery (over ACE) approximately 10% of the time.
Variable Weighting in Cluster Analysis

Often in educational and psychological research, a researcher is confronted by a population which, for lack of a better term, seems too heterogeneous. In such circumstances, the investigator is likely to hypothesize that the population under study is actually composed of two or more relatively homogeneous subgroups. For example, several investigators have suggested that the umbrella term "learning disabilities" actually encompasses a variety of disorders, with different etiologies and strategies for treatment.

In such circumstances, one may appeal to clinical insight to describe possible subgroups. However, to bolster such insight, the investigator may also turn to the statistical method of cluster analysis to try to isolate relatively homogeneous subgroups within the more heterogeneous population. Indeed, cluster analysis often has been used to try to identify putative subtypes of learning disability.

Unfortunately, the applied investigator is faced with a myriad of clustering techniques, most of which present options and suboptions in the analysis. To guide these choices, numerous studies have examined various aspects of the clustering process, including comparison of clustering algorithms (e.g., Belbin, Faith, & Milligan, 1992; Blasfield, 1976; Donoghue, 1994b, 1995; Milligan, 1979, 1989a; Scheibler & Schneider, 1985), the effect of various types of "error" in the data to be clustered (Milligan, 1980), variable standardization (Milligan & Cooper, 1988; Barton, 1993), selection/weighting of irrelevant variables (De Soete, 1986, 1988; Milligan, 1989b; Donoghue, 1994a), procedures to determine the number of clusters (e.g., Milligan & Cooper, 1985), and procedures to compare clustering solutions (Milligan & Cooper, 1986).

Although these studies have obtained a variety of useful findings, one aspect of the clustering process has not received much attention: the within-group covariance structure. Recent work (Donoghue, 1994b) examined the case of two groups in two dimensions, and found a very large effect of within-group correlation. Specifically, within-group correlation which did not coincide with the direction of separation
in subgroup means was associated with lower recovery for all clustering methods; within-group correlation which did coincide with the direction of separation was associated with higher recovery. This result was interpreted in terms of the similarity measure used, the Euclidean distance.

A fundamental issue raised by Donoghue (1994b) is what the analyst should do to minimize the deleterious effect of within-group correlation. Donoghue suggested several possible alternatives to the Euclidean distance, but, in the absence of comparative data, could provide no guidance. The present study compares several of these variable weighting methods. In the remainder of this section, the finite mixture model which implicitly underlies this work will be introduced, and the various weighting strategies used in this study will be examined.

The Finite Mixture Model

This paper adopts the view of cluster analysis as the attempt to "unmix a mixture of distributions" (e.g., Titterington, Smith, & Makov, 1985; McLachlan & Basford, 1988). The finite mixture model states that the distribution function for the entire population, \( h(x) \), is given by:

\[
    h(x) = \sum_{g=1}^{G} \pi_g f_g(x)
\]

\( \pi_g > 0, \quad \sum_{g=1}^{G} \pi_g = 1 \) .

Subgroups are the homogeneous distributions \( f_g(x) \), which are mixed, and the \( \pi_g \) are termed the "mixing proportions." The conceptual framework of the finite mixture model of clustering underlies the present paper, and some of the results may not make sense in situations in which the finite mixture model does not apply.

---

\footnote{The general term "measure of similarity" will be used in this paper to denote both true measures of similarity, such as correlations and coefficients of concordance, and measures of dissimilarity, such as distances.}
In the discussion that follows, $X$ will be the full (entities by variables) data vector, and $x_i$ be the column vector of variable values associated with entity $i$ (i.e., the transpose of the $i$th row of matrix $X$).

**Variable Weighting Methods**

Prior to clustering, hierarchical clustering methods convert two-mode (variables by entities) multivariate data into a one-mode (entities by entities) univariate measure of similarity, which is then used as the basis of the clustering. The vast majority of clustering studies (applied and simulation) use $d_E$, Euclidean distance (or the squared distance) between entities $i$ and $j$ as similarity measure:

$$d_E(i,j) = \sqrt{\sum_{p=1}^{P} (x_{ip} - x_{jp})^2}$$

where $P$ is the number of variables used in the clustering. Basic geometry of vector spaces reveals that Euclidean distance is correct only if computed on an orthogonal basis. If the variables are correlated, the Euclidean distance can misrepresent the distance between two points. By ignoring the within-group correlations of variables, Donoghue (1994b) demonstrated that $d_E$ can seriously degrade cluster recovery.

Euclidean distance is widely used and is the default measure for most clustering software. However, most hierarchical clustering algorithms can make use of any one-mode measure, and other similarity measures are available. Two commonly mentioned possibilities are Mahalanobis distance $D_M$ and clustering based upon principal component scores. The problems with these obvious choices will be discussed, and then some possible alternatives will be discussed.

**Mahalanobis Distance**

Ideally, one would cluster based on the Mahalanobis distance measure, $D_M$:

$$D_M(i,j) = (x_i - x_j)' \Sigma^{-1} (x_i - x_j)$$

where $\Sigma$ is the covariance matrix.
in practice however, $D_M$ is difficult to compute because of the difficulty in estimating the appropriate $\Sigma$ matrix. The correct choice is the pooled within-group covariance matrix, $\Sigma_w$:

$$\Sigma_w = \sum_{g=1}^{G} n_g \Sigma_g$$  \hspace{1cm} (4)

Without knowing the subgroup structure, it is not possible to compute the $\Sigma_g$ matrices. An alternative is to use the overall covariance matrix, $\Sigma_T$:

$$\Sigma_T = \frac{1}{n} (X - \bar{X})(X - \bar{X})'$$  \hspace{1cm} (5)

However, the standard MANOVA decomposition reveals:

$$\Sigma_T = \Sigma_B + \Sigma_W$$  \hspace{1cm} (6)

The inclusion of $\Sigma_B$ can give $\Sigma_T$ very different properties from $\Sigma_w$. Hartigan (1975, p. 63) provides an example demonstrating that $D_M$ based on $\Sigma_T$ can result in worse cluster recovery than using $d_E$.

**Principal Components**

Another commonly suggested alternative is to cluster based upon principal component scores rather than the original variables. The chief advantage of this method is that principal component scores are orthogonal to one another; because $d_E$ implicitly assumes orthogonal variables, clustering based on the principal components seems to remove the objections to this measure. However, the difficulty with principal components is identical to that of $D_M$. The ideal is to perform principal components analysis based on $\Sigma_w$. In practice, however, $\Sigma_w$ is not known, and so component scores are computed based on $\Sigma_T$ (or the correlation matrix $R_T$). As is the case with $D_M$, the inclusion of $\Sigma_B$ can give $\Sigma_T$ very different properties from $\Sigma_w$. Rohlf (1970) discusses some of the difficulties with using principal components, and Chang (1983) gives a theoretical analysis of the difficulties involved, and provides a clear example of the
problems with this approach.

The ACE Algorithm

Art, Gnanadesikan, and Kettenring (1982) devised an ingenious method to estimate the pooled within-group covariance matrix without knowledge of the subgroup structure. They start with the known property that the covariance can be computed from inter-entity relations:

\[ \sum_{i} \sum_{j} (x_{ik} - x_{jk})(x_{il} - x_{jl}) \]

Analogous to the usual MANOVA decomposition in Equation (7), the inter-entity relations give a between-and within-groups decomposition:

\[ T = B^* + W^* \]

If \( \Delta \) is the matrix of differences between subgroup centroids, Art Gnanadesikan, and Kettenring show that \( W^{-1} \Delta \) has the same eigenvectors as \( W^{-1} \Delta \), and so \( D_M \) based on \( W^* \) is equivalent (up to a multiplicative constant) to \( D_M \) based on \( W \).

Reasoning that most of the smaller inter-entity distances are likely to be within-group distances, Art Gnanadesikan, and Kettenring use the inter-entity relations from the \( m \) smallest distances to form an approximate covariance estimate (ACE), \( W^* \):

\[ W^* = \frac{1}{n} \sum_{d < m} (x_i - x_j)'(x_i - x_j) \]

where the summation runs over the \( m \) smallest pairs. The distances are then recomputed as \( D_M \) based on \( W^* \), and the process repeated until the current estimate of \( W^* \) does not differ significantly from that of the previous iteration.

A modified version of ACE is computed by SAS's PROC ACECLUS (SAS Institute, 1988),
including the option to output a matrix of inter-entity similarities for further analysis. The present study
used the original algorithm given in Art, Gnanadesikan, and Kettenring (1982). To date, I am not aware
of any published studies (beyond the original article) of this approach, although work is underway at
Bellcore (J. Kettenring, personal communication, July 2, 1993).

**Ultrametric Weights**

De Soete (1986, 1988) developed an algorithm which determines weights to apply to variables in
computing a distance measure:

\[ d_u(i,j) = \sqrt{\sum_{p=1}^{P} w_p (x_{ip} - x_{jp})^2} \]  

\[ w_p > 0, \quad \sum_{p=1}^{P} w_p = 1 \]  

(10)

The weights are chosen so that the resultant distances maximally satisfy the "ultrametric inequality," which
states that any three points \(i, j,\) and \(k,\) the distances between the points should satisfy the relation:

\[ d_{ij} \leq \max (d_{ik}, d_{jk}) \]  

(11)

This is equivalent to requiring all sets of three points to lie on an acute isosceles (or equilateral) triangle.
Johnson (1967) and Milligan (1979) demonstrated the relationship between the ultrametric inequality and
many commonly used hierarchical clustering algorithms, and Milligan and Isaac (1980) give simulation
results which provide support for the utility of the conceptualization. Two studies (Milligan, 1989b,
Donoghue, 1994a) have examined De Soete's algorithm, and found that it greatly improved cluster
recovery when the data contained "error" dimensions, i.e., dimensions which contained no information
about subgroup membership. However, it is not known whether this method helps with the difficulties of
within-group covariance structure.
The present study sought to examine these weighting strategies, and compare their usefulness in recovering known subgroup structure in data with the presence of a variety of within-group covariance structures.

Method

This study used Monte Carlo methods to systematically investigate the utility of various variable weighting methods to account for the effects within-group covariance structures. Subgroups were generated according to a model of a finite mixture of normal distributions (e.g., McLachlan & Basford, 1988). Similar to Donoghue (1994b), the study will examine the limited case of two groups in two dimensions, but this case will be examined in some detail. While restricting the investigation to bivariate data somewhat limits its generalizability, this provides a minimal test of the weighting procedures; any procedure which does not function well with bivariate data seems unlikely to be useful for higher dimensional cases.

Design

Five aspects of the data were manipulated in data generation. The first three of these determined the within-group covariance matrices.

1) R1: The within-group correlation of subgroup 1 (3 levels) -- $r_1 = -0.7, 0.0, 0.7$.
2) R2: The within-group correlation of subgroup 2 (3 levels) -- $r_2 = -0.7, 0.0, 0.7$.
3) COV: The relationship of the within-group covariance matrices (5 levels). For the first three levels, the variances of each variable were equal ($\sigma^2_{11} = \sigma^2_{22}$) within a subgroup, and both of the variances of the second subgroup were equal to a constant times the variance within the first subgroup ($\sigma^2_{12} = k\sigma^2_{11}$, $\sigma^2_{22} = k\sigma^2_{21}$). For the last two conditions, the variances of the two
variables within a subgroup differed. The within-subgroup matrices were equal for one of the conditions, but differed for the other condition. The five levels of COV were:

A) $1:9 - \sigma^2_{11} = 1, \sigma^2_{21} = 1, \sigma^2_{12} = 9, \sigma^2_{22} = 1$;

B) $1:9 - \sigma^2_{11} = 1, \sigma^2_{21} = 1, \sigma^2_{12} = 9, \sigma^2_{22} = 9$;

C) $9:1 - \sigma^2_{11} = 9, \sigma^2_{21} = 9, \sigma^2_{12} = 1, \sigma^2_{22} = 1$;

D) $H0 - \sigma^2_{11} = 1, \sigma^2_{21} = 9, \sigma^2_{12} = 1, \sigma^2_{22} = 9$;

E) $H1 - \sigma^2_{11} = 1, \sigma^2_{21} = 9, \sigma^2_{12} = 9, \sigma^2_{22} = 1$.

In generating the data, two other "nuisance variables" were manipulated, due to their consistently large effects in other studies:

4) PROB: Probability of each of the subgroups in the population ($p_1$ and $p_2$). This was manipulated by varying the sizes ($n_1 = N*p_1$, $n_2 = N*p_2$) of the subgroups (2 levels)—Equal sized groups ($n_1=50$, $n_2=50$) or unequal group sizes ($n_1=90$, $n_2=10$).

5) DIST: Separation of subgroups. This was defined in terms of $D_M$, the Mahalanobis distance (in the populations) between the subgroup centroids:

$$D_M(12) = \sqrt{(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)}$$  \hspace{1cm} (12)

This distance was based on the pooled within-group covariance matrix: $\Sigma = p_1\Sigma_1 + p_2\Sigma_2$ DIST had 3 levels—$D_M = 2$, $D_M = 4$, or $D_M = 6$.

6) WEIGHT: The method of weighting the variables to form the similarity measure (6 levels)—which are discussed in a separate section below.

7) METHOD: The method of cluster analysis (11 levels)—which are discussed in a separate section below.

The first five factors were fully crossed to yield 270 conditions. Twenty datasets were generated
for each condition according to the procedure given below, giving a total of 5400 base datasets. Although
more replications would be desirable, the results of previous work (e.g., Donoghue, 1994b) indicate that
this number of replications is likely to provide sufficient precision for the purposes of this study. Each
dataset was then weighted according to each of the weighting methods, and analyzed by each of the
hierarchical clustering methods, yielding a total of 356,400 cluster analyses. For each cluster analysis, the
solution for the correct number of subgroups (i.e., \( G = 2 \)) was used as the result for that clustering method.

Data generation, variable weighting, and clustering were performed using FORTRAN programs
written by the author. Eigenvalues and eigenvectors were computed using routines from EISPACK
(Smith, Boyle, Garbow, Ikebe, Klema, & Moler, 1974). Accuracy of these programs was ensured
through numerous comparisons of results of subroutines and final classifications with routines from SAS
and SPLUS.

Data Generation

Subgroups were generated according to a finite mixture of normal distributions. All datasets
consisted of 100 observations. The means were separated by both variables equally. To ensure that the
data generation procedures worked properly, the Mahalanobis distance between the two known groups was
computed for each dataset. Table 1 presents descriptive statistics for \( D_M \).

Insert Table 1 about here

The means and medians for each of the conditions are very close to the desired values, supporting the
validity of the data generation procedures.

Variable Weighting Methods

For analysis, each dataset was subjected to 6 variable weighting procedures. The procedures
either yield a set of weights to apply to the variables before forming the Euclidean distance, or they yield
an alternative similarity measure such as $D_m$.

1) Euclidean distance (no weighting)

2) PC-R: Principal component scores based on $R_T$

3) $D_m$-T: $D_m$ based on $S_T$

4) Ultrametric—d. computed using weights from De Soete’s (1986, 1988) algorithm to maximize agreement with the ultrametric inequality

5) ACE: $D_m$ based on $W^*$ from the ACE algorithm (Art, Gnanadesikan, & Kettenring, 1982). As in Art et al., the parameter $m$ (the number of inter-entity distances used to determine the pairs used to estimate $W^*$) was set at two-thirds of the number of within-group pairs: $m = 1633$ for $p_1 = 0.5$ and $m = 2700$ for $p_1 = 0.9$.

6) $D_m$-W: $D_m$ based on pooled within $S_w$.

A seventh method, principal component scores based on $S_T$, was originally included. However, because both component scores were retained, these scores are an orthogonal rotation of the original variables.

Because the Euclidean distance is invariant under orthogonal rotation, every one of the 59,400 solutions was identical to that computed from $d_e$, and so these results were not included in the analyses discussed below. Clearly, clustering based on principal components is not the solution to the problem of within-group correlation.

The first and last methods provide useful comparisons for the other methods. Clearly, a weighting method which results in worse cluster recovery is undesirable. On the other hand, the last method provides a useful upper bound on how good recovery could be in a given dataset.

Cluster Algorithms

Each dataset was analyzed 11 times, corresponding to different hierarchical clustering algorithms.

The clustering methods are:
1) Single linkage, distance

2) Complete linkage, distance

3) Ward's (1963) method, squared distance

4) EML-SAS's (1988) maximum likelihood hierarchical clustering procedure, which is a modification of Ward's method to alleviate the method's tendency to yield equal-sized clusters, squared distance.

5) Average flexible method, (Belbin, Faith, & Milligan, 1992), distance. Five levels of $\beta$ were examined, 0.0, -.10, -.15, -.20, -.25. Note that $\beta = 0$ is the usual average linkage method.

6) Flexible-beta method (Lance & Williams, 1967), distance. Two levels of $\beta$ were examined, -.25 and -.50.

The clustering methods were chosen because: a) they are widely used (single linkage, complete linkage, average linkage, and Ward's method); b) they have performed well in previous studies (average linkage, Ward's method, beta-flexible and flexible average); or c) they are designed to rectify a known weakness of another algorithm (EML). The values of $\beta$ used are based on the studies by Milligan (1989a), Belbin, Faith, and Milligan (1992), and Donoghue, 1994b, 1995). For a discussion of these algorithms, the reader is referred to standard introductions to cluster analysis (e.g., Everitt, 1993; Lorr, 1983). Milligan (1989a), Belbin, Faith, and Milligan (1992), and Donoghue (1995) contain discussions of the beta-flexible and flexible average methods, and the SAS documentation (SAS Institute, 1988) is the primary reference for the EML algorithm.

Outcome Measure

The outcome measure for the study was the Hubert and Arabie (1985) modification of Rand's (1971) statistic, which will be denoted HA-Rand. The index was computed between each cluster solution.
and the true subgroup membership used to generate the data. This index is based on examining pairs of entities, and determining whether they are classified into the same or different subgroups. A value of zero reflects chance agreement with the true membership, and 1.0 reflects perfect agreement. A study by Milligan and Cooper (1986) supports the accuracy of Hubert and Arabie's modification.

Analyses

To summarize the results, a factorial analysis of variance was conducted, as in Milligan (1980, 1981, 1989a) and Donoghue (1994b, 1995). The HA-Rand index served as the dependent variable. The independent variables were the design factors used to generate the data, weighting method, and the cluster algorithm used to analyze the data. The purpose of the ANOVA was to summarize the data and help to highlight the more important effects. Therefore, a measure of effect size was adopted in favor of traditional significance testing. Usually, $\eta^2$ would be used in this context:

$$\eta^2 = \frac{SS_{\text{effect}}}{SS_{\text{tot}}} \quad (13)$$

However, this index has a disadvantage in large designs, namely that the denominator contains not only error variance and systematic variance of interest, but also irrelevant systematic variance due to other effects. The larger the design becomes, the more apparent this effect becomes. In the present case, this is particularly noisome, because one of the factors, Mahalanobis distance between the subgroup means, has a very large effect and so serves to obscure the effects of other factors. Therefore, an alternate version, $\eta^2_{alt}$ (Tabachnick & Fidell, 1983, p. 47), was used:

$$\eta^2_{alt} = \frac{SS_{\text{effect}}}{SS_{\text{effect}} + SS_{\text{error}}} \quad (14)$$

Note that, unlike the other formulation, this version, $\eta^2_{alt}$, does not sum to 1.0. As in Donoghue (1994b), the practical criterion of $\eta^2_{alt} > .03$ was selected. Note that, given the large amount of data, any effect
which met the practical criterion was always highly significant (e.g., \( p < .0001 \)).

Results

The results of the ANOVA and values of \( \eta^2_{\text{in}} \) are summarized in Table 2. All of the main effects were identified as salient, with the exception of \( \text{Prob.} \) and \( \text{R2} \), the correlation within subgroup 2. Means for the main effects are presented in Table 3. The main effects are all modified by salient interactions, and so will not be discussed in detail. However, a few comments are in order, bearing in mind that the observations may well be modified by the interactions. First, the extremely large effect of \( D \), the distance between the subgroup centroids, is both noteworthy and expected; as the subgroups become more separated, the clustering task becomes easier and hence recovery improves. The smaller effect for \( \text{R2} \) (compared to \( \text{R1} \)) is also an expected artifact of the design; for one-half of the datasets, subgroup 2 comprised only 10% of the sample, and so had a smaller effect. Finally, for the weighting methods, Table 3 reveals that, overall, best recovery is obtained for \( D_{m-W} \), the Mahalanobis distance based on the pooled within-group covariance matrix. This is followed by distances based on the approximate covariance estimate, \( \text{ACE} \); distances computed from the ultrametric weights; and \( \text{PC-R} \), the principal components based on the total correlation matrix. Overall, \( D_{m-T} \), Mahalanobis distance based on the total covariance matrix, yielded worse recovery than simply using the Euclidean distance \( d_e \).

Weighting Method

The main effect of weighting method was modified by several salient interactions. Figure 1 portrays the \( \text{R1 by Weight} \) interaction. Three of the weighting methods, \( d_e \), \( \text{PC-R} \), and the ultrametric weights show large drops in recovery when \( \text{R1} = -.7 \). The other three methods based on \( D_m (D_{m-W}, \)
ACE, and $D_{MT}$) show a relatively flat profile; recovery using ACE actually decreases slightly as $R1$ increases. Although ultrametric weights and PC-R show lower recovery for $R1 = -.7$, both methods yield good recovery for $R1 = 0$ and $R1 = .7$, outperforming ACE.

Figure 2 plots the means of COV by Weight interaction. The effect appears to be due to the methods which are not based on Mahalanobis distance methods, $d_e$, PC-R and the ultrametric weights. The chief effect seems to be for $H0$, the condition in which $\sigma_{11} = \sigma_{21} = 1$ and $\sigma_{12} = \sigma_{22} = 3$. The ultrametric weights show a small decrease in recovery for this condition; PC-R has a substantial decrement in recovery, and $d_e$ demonstrates a huge decrement in recovery. The three methods also differ to a lesser extent on $H1$, but give virtually identical results for the other conditions.

The Prob. by Weight interaction is portrayed in Figure 3. $Dm-W$ is relatively unaffected by subgroup size. PC-R and ultrametric weights yield moderately lower recovery when $p_i = .9$, and $d_e$ yields substantially worse recovery. On the other hand, $D_{MT}$ and ACE yield better recovery for $p_i = .9$ than for $p_i = .5$. For ACE, the number of pairs used in the estimation, $m$, was larger for $p_i = .9$, so this result may simply reflect more stable estimation from more data.

Finally, Figure 4 displays the means for the Weight by Method interaction. Although there are some differences, five of the weighting methods give very similar profiles of recovery across the clustering algorithms. On the other hand, $D_{MT}$ differs wildly in effectiveness. It performs well for Ward's method.
and flexible average, but performs at or below the level of $d_e$ when used with other methods. Because these two factors (clustering algorithm and weighting method) are under control of the analyst, this interaction is of considerable interest. Therefore, it will be examined in more detail below.

---

Insert Figure 4 about here

---

**Clustering Methods**

Three additional two-way interactions involving clustering method were identified as salient: Distance by Method, Prob. by Method, and COV by Method. Figure 5 portrays the means of the Distance by Method interaction. In general, the ordering of the methods is similar for each distance, although complete linkage ranks ahead of average linkage for $D_m = 2$, but behind average linkage when the groups are further apart. The largest difference is due to the single linkage method, which shows a very small gain (relative to other methods) from $D_m = 2$ to $D_m = 4$, and a consequent large relative gain from $D_m = 4$ to $D_m = 6$. Otherwise, the effects of increasing subgroup separation are relatively stable across the other methods.

---

Insert Figure 5 about here

---

Figure 6 plots the means for the Prob. by Method interaction. The clustering methods differ substantially in their recovery of equal- versus unequal-sized groups. Consistent with other studies (e.g., Donoghue, 1994b), single linkage yields much better recovery for unequal-sized groups, as do, to a lesser extent, EML and complete linkage. On the other hand, Ward's method yields much better recovery for equal-sized groups. EML and Ward's method showed a similar pattern in Donoghue (1994b), but there EML yielded slightly better recovery for equal-sized groups.
Finally, the COV by Method interaction is portrayed in Figure 7. The 1:1, H0 and H1 conditions present similar profiles of recovery across the clustering algorithms. The 1:9 condition results in much higher recovery for flexible average, beta-flexible and Ward’s methods but lower recovery for average linkage. In general, the 9:1 condition yields a similar profile to the 1:9 condition, although the 1:9 condition results in worse recovery for single and average linkage. In spite of these differences, however, the clustering methods are, in general, similarly ordered for each condition of COV.

Other Effects

The two-way interaction of Distance by COV was identified as salient, as was the three-way interaction of Distance by Prob. by COV. Figures 8a and 8b plot the means of this three-way interaction. When \( p_1 = .5 \) as in Figure 8a, the Distance by COV interaction is fairly small. For each level of \( D_m \), recovery is slightly lower for the H0 condition of COV than for the other conditions, although the amount does differ somewhat by \( D_m \). However, when \( p_1 = .9 \) as in Figure 8b, there is a substantial interaction of Distance with COV. For \( D_m = 2 \), the 1:9 condition results in good recovery, while the recovery for the 9:1 condition falls near chance, and H0 and H1 yielded recovery equal to or better than that of the 1:1 condition. For \( D_m = 4 \), the 1:9 condition yields similar recovery to that of 1:1 and 9:1, and for \( D_m = 6 \), 1:9 results in somewhat lower recovery than the other two conditions.
Other Weight by Method Interactions

Inspection of Figure 4 indicated that the bulk of the Weight by Method interaction was due to the differing utility of $D_m$-$T$, the Mahalanobis distance between entities computed from $\Sigma_T$. Table 2 indicated that two three-way interactions involving Weight and Method (Distance by Weight by Method and Prob. by Weight by Method) were large but did not meet the criterion for salience. For completeness, plots of the means for these interactions are presented in the appendix. These interactions are largely due to the behavior of $D_m$-$T$; the interaction of $D_m$ with Method increases with subgroup separation, and the interaction is much larger for $p_1 = .5$ than for $p_1 = .9$.

What Works Best

As was noted above, the Weight by Method interaction is of particular interest. While the other factors in the design are aspects of the data which are outside of the analyst's control, the weighting method and clustering algorithm are decisions made by the analyst. The key questions are:

1) Which weighting method yields the best recovery?

2) Which clustering algorithm yields the best recovery?

The salient interaction between Weight and Method indicates that the answer to either of these questions may depend on the value of the other factor. Essentially, we would like to know if the interaction is ordinal (effect size changes but the order of the levels of one factor remains stable across levels of the other) or disordinal (the ordering of results for one factor depends upon the level of the other factor).

Because these are explicitly ordinal questions, Cliff's (1993) distribution free method of comparing two distributions was used. Ordinal comparisons were performed using a modified version of Cliff's (1992) program PAIRDEL1, for paired observations. Two types of ordinal comparisons were made. The first estimates the probability that a randomly sampled observation from one distribution is higher (i.e., better recovery) than a randomly sampled observation from the other distribution. This results in one of
three decisions for each pair of methods: a) Method A is higher (better recovery) than Method B; b) Method B is higher than Method A; or c) the methods do not significantly differ. These pairwise comparisons were made on HA-Rand index recovery values. Shaffer's (1986) modification to the Bonferroni procedure was used to maintain familywise Type I error at $\alpha=0.05$. Finally, the pairwise relations were converted into ranks based upon the number of methods which were significantly higher than a given method versus the number which were significantly below it.

The second type of ordinal comparison is based on Cliff's index $d_m$, which is the proportion of datasets for which Method A yielded higher recovery than Method B minus the proportion of datasets for which Method B yielded higher recovery than Method A; $d_m$ is the net proportion of datasets for which Method A yielded higher recovery. In the present context, it may be interpreted as the probability (for a randomly chosen dataset) of getting better recovery using Method A. Negative values of $d_m$ indicate lower recovery for Method A.

Comparison of Weighting Methods

To determine which weighting methods work better, pairwise comparisons of the weighting methods were computed within each clustering algorithm, and ranks were formed based on the pairwise comparisons. For each clustering algorithm, the initial family size was $6 \times 2 = 15$. The results of these comparisons are summarized in Table 4.

In spite of the strong interaction, the ranks of weighting methods are very consistent across clustering algorithms; only $d_b$ and $D_{W\cdot T}$ show reversals across methods. Clustering using $D_{W\cdot W}$ (Mahalanobis distance based on the pooled within-groups covariance matrix) yielded best cluster recovery. Of the remaining methods, which do not require knowledge of the cluster structure, the ACE algorithm
Variable Weighting in Cluster Analysis

(Art et al., 1982) is best, followed by the ultrametric weighting algorithm of De Soete (1986, 1988).

As was mentioned in the introduction, the (squared) Euclidean distance $d_E$ is the default measure of similarity for most implementations of clustering procedures (e.g., statistical packages). Table 5 provides $d_\omega$ values for each of the weighting methods compared to $d_E$. A negative value indicates that the method indexed at the top of the column gave superior recovery to that of $d_E$. For example, using $d_E$ instead of ACE would yield lower recovery 8-24% (depending upon the clustering algorithm used) of the time when analyzing datasets similar to those used here. Table 5 also gives $d_\omega$ values for each of the remaining methods compared to the best method, ACE. In comparing the two best methods which do not require knowledge of the subgroup structure, ACE and ultrametric, the $d_\omega$ value varies somewhat by clustering algorithm, but is always greater than 0; for the type of data considered in this study, ACE provides better recovery than ultrametric weights.

The last two columns of Table 5 compare cluster recovery using the two best weighting methods (ACE and ultrametric weights) to cluster recovery using $D_M-W$. Thus, knowing the subgroup structure and clustering based on $D_M-W$ instead of ACE would result in improved recovery 10-18% of the time, unless the analyst used the single linkage algorithm.

Comparison of Clustering Algorithms

A similar procedure was used to rank the clustering algorithms. Within each weight method, all pairwise ordinal comparisons of clustering methods were performed and the pairwise relations were converted into ranks. Shaffer's modification to the Bonferroni was again used, with the initial family size of 11 take 2 = 35. The results of these analyses are summarized in Table 6.
The ranks of the clustering methods are more variable than those of the weights. However, the flexible average method of Belbin, Faith, and Milligan (1992) clearly outperforms the other algorithms. For the conditions examined in this study, $\beta = -.15$ or $\beta = -.20$ gave the best recovery. These values coincide with Belbin et al.'s recommendation that $\beta \geq -.2$, although they are slightly lower than the final value they suggest, $\beta = -.10$. However, for the best weighting conditions, $\beta = -.10$ yielded very similar results to the best values.

Although there are some minor differences, the results for the remaining methods are similar to those obtained in other studies (e.g., Donoghue, 1994a,b; Milligan 1989a,b; Belbin, Faith, & Milligan, 1992). The results for the lowest of the methods, complete linkage and single linkage, are quite consistent. Based on theoretical considerations, such as its relationship to the minimum spanning tree of graph theory, some authors continue to recommend single linkage (e.g., IMS Panel on Discrimination, Classification, and Clustering, 1989). Yet this method uniformly produced the lowest recovery of the methods examined, a result that is routinely found in comparisons clustering algorithms (e.g., Milligan & Cooper, 1987, and the references therein).

Table 7 presents $d_{\infty}$ values for comparing each clustering method to one of the best (although not widely available) clustering methods, flexible average with $\beta = -.15$, and comparing each clustering method to one of the best of the widely available algorithms, Ward's method. In the present study, using the flexible average method instead of single linkage gives $d_{\infty} = .441$ for $d_E$, .590 for ACE, and .524 for ultrametric weights. Using single linkage gives worse recovery about one-half of the time! Even compared to Ward's method, which is very widely available, $d_{\infty}$ is over .30 for all weighting methods.
Limitations and Future Work

The following is a very brief discussion of some related areas for future research. It is not intended to be exhaustive by any means. In order to keep this section reasonably brief, procedures which differ from the present paper's focus on distance-based measures are not discussed. Examples of other approaches not considered here include non-distance similarity coefficients (e.g., q-correlations) and algorithms which combine specific variable weighting or multidimensional scaling models with specific clustering algorithms.

While restricting the present investigation to bivariate data somewhat limits its generalizability, it was felt that this provides minimal test of the weighting procedures; any procedure which does not function well with bivariate data seems unlikely to be useful for higher dimensional cases. Having established here that variable weighting methods can provide substantial benefit in cluster recovery, further work is needed to determine the extent to which the results obtained here generalize to higher numbers of subgroups and higher dimensional clustering problems.

Certain aspects of the methods used, such as how to choose the number of distances $m$ to use in computing the ACE estimate of $\Sigma$, need to be addressed. Currently, there is little to guide the applied researcher. J. Kettenring (personal communication, July 2, 1993) indicates that some work along these lines is underway at Bellcore. Obviously, more work would be useful. A related issue is the relationship of the original formulation of the method, used here, to the somewhat modified version which is implemented in SAS's PROC ACECLUS (1988).

The relationship between the methods used here and procedures aimed at variable standardization...
Variable Weighting in Cluster Analysis

(Milligan & Cooper, 1988; Barton, 1993) should be examined. Variable standardization methods seek to deal with problems associated with combining variables measured on very different scales (such as SAT and GPA) into a single similarity coefficient. Hence, variable standardization procedures tend to treat variables sequentially, as opposed to the multivariate focus of the procedures examined in the present study.

Of the clustering methods examined in this study, the flexible average clustering method (Belbin, Faith, & Milligan, 1992) gave the best recovery of the methods in this study. The optimal value of $\beta$ was found to be $-0.15$ to $-0.20$, which is slightly lower than the value of $\beta = -0.10$ suggested by Belbin et al. (1992). A companion paper (Donoghue, 1995) presents further results comparing the flexible average and beta-flexible methods. Of the remaining methods, Ward's method and beta-flexible linkage ($\beta = -0.50$) yielded the best recovery, followed closely by SAS's EML algorithm.

The relationship of the weighting methods identified here to the closely allied problem of variable selection (Milligan, 1989b; Donoghue, 1994a) remains to be addressed. Both of these papers found that the ultrametric weights were useful in ameliorating the degradation of cluster recovery caused by including irrelevant variables in the cluster analysis. Donoghue (1994a) also compared two additional variable selection procedures, and in their original paper, Art et al. (1982) presented preliminary evidence that the ACE weighting method may be useful with the variable selection problem. More work is needed to understand the interrelationship of the various variable weighting techniques proposed to solve different problems in clustering.

Conclusion

This study was meant to constitute a first step in understanding the utility of various variable weighting algorithms. The results of Donoghue (1994b) clearly indicate that cluster analyses based upon
the Euclidean distance can yield suboptimal results in the presence of within-group correlation, but could not offer clear suggestions for alternatives.

Variable weighting methods were found to have a large effect on cluster recovery. Of the methods which do not require a priori knowledge of the subgroup structure, the ACE method of Art, Gnanadesikan, and Kettenring (1982) yielded the best results. When used with the better clustering algorithms, this method provided a net improvement in 17% (EML) to 24% (beta-flexible, $\beta = -.50$) of the datasets examined. When used with the same clustering methods, the next best weighting method, ultrametric weights, yielded improved recovery 16-21% of the time. However, ultrametric weighting was found to be more sensitive than ACE to within-subgroup correlation. Therefore, ACE is the method which is preferred overall. There is still plenty of room for improvement, however. Comparisons of ACE with $D_m$ based on the pooled within-group covariance matrix indicated that knowing the correct covariance matrix would yield improved recovery (over ACE) approximately 10% of the time.

Finally, it should be noted that the procedures examined and the results obtained here only apply to situations in which a mixture model formulation of the clustering problem makes sense. In some educational applications, such as clustering test items based on multidimensional IRT parameters (Miller & Hirsch, 1992), it is far from obvious whether the mixture model formulation of the clustering problem is valid. The application of the present work to such situations is problematic.

This study is intended to be a first step in understanding variable weighting methods. It attempts to provide some information to those educational investigators faced with doing empirical cluster analyses, allowing them to make a better informed choice of variable weighting methods and clustering algorithms.
Variable Weighting in Cluster Analysis

References


Variable Weighting in Cluster Analysis


Table 1
Descriptive Statistics for Observed $D_m$

<table>
<thead>
<tr>
<th>$D_m$</th>
<th>Mean</th>
<th>Std.</th>
<th>Min.</th>
<th>$Q_1$</th>
<th>Med.</th>
<th>$Q_3$</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.086</td>
<td>0.454</td>
<td>0.372</td>
<td>1.835</td>
<td>2.051</td>
<td>2.300</td>
<td>4.611</td>
</tr>
<tr>
<td>4</td>
<td>4.106</td>
<td>0.558</td>
<td>2.127</td>
<td>3.775</td>
<td>4.063</td>
<td>4.357</td>
<td>7.185</td>
</tr>
<tr>
<td>6</td>
<td>6.143</td>
<td>0.730</td>
<td>3.511</td>
<td>5.691</td>
<td>6.067</td>
<td>6.500</td>
<td>10.690</td>
</tr>
</tbody>
</table>
Table 2
Main Effects and Salient Interactions ($\eta^2_{sh} \geq .03$) for ANOVA of HA-Rand Index

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>DF</th>
<th>SS</th>
<th>$\eta^2_{sh}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>2.</td>
<td>23826.77</td>
<td>.641</td>
</tr>
<tr>
<td>P</td>
<td>1.</td>
<td>59.61</td>
<td>.004</td>
</tr>
<tr>
<td>C</td>
<td>4.</td>
<td>648.85</td>
<td>.046</td>
</tr>
<tr>
<td>R1</td>
<td>2.</td>
<td>425.51</td>
<td>.031</td>
</tr>
<tr>
<td>R2</td>
<td>2.</td>
<td>97.03</td>
<td>.007</td>
</tr>
<tr>
<td>M</td>
<td>10.</td>
<td>4166.77</td>
<td>.238</td>
</tr>
<tr>
<td>W</td>
<td>5.</td>
<td>1452.93</td>
<td>.098</td>
</tr>
<tr>
<td>D*C</td>
<td>10.</td>
<td>540.93</td>
<td>.039</td>
</tr>
<tr>
<td>D*M</td>
<td>20.</td>
<td>702.38</td>
<td>.050</td>
</tr>
<tr>
<td>P*M</td>
<td>10.</td>
<td>723.06</td>
<td>.051</td>
</tr>
<tr>
<td>C*M</td>
<td>40.</td>
<td>436.27</td>
<td>.032</td>
</tr>
<tr>
<td>C*R1</td>
<td>8.</td>
<td>337.65</td>
<td>.025</td>
</tr>
<tr>
<td>D*W</td>
<td>10.</td>
<td>332.11</td>
<td>.024</td>
</tr>
<tr>
<td>P*W</td>
<td>5.</td>
<td>671.72</td>
<td>.048</td>
</tr>
<tr>
<td>C*W</td>
<td>20.</td>
<td>834.55</td>
<td>.059</td>
</tr>
<tr>
<td>R1*W</td>
<td>10.</td>
<td>568.05</td>
<td>.041</td>
</tr>
<tr>
<td>R2*W</td>
<td>10.</td>
<td>83.78</td>
<td>.006</td>
</tr>
<tr>
<td>M*W</td>
<td>50.</td>
<td>631.73</td>
<td>.045</td>
</tr>
<tr>
<td>D<em>P</em>C</td>
<td>8.</td>
<td>414.63</td>
<td>.030</td>
</tr>
<tr>
<td>D<em>M</em>W</td>
<td>100.</td>
<td>328.51</td>
<td>.024</td>
</tr>
<tr>
<td>P<em>M</em>W</td>
<td>50.</td>
<td>312.40</td>
<td>.023</td>
</tr>
<tr>
<td>Error</td>
<td>338580.</td>
<td>13353.61</td>
<td>----</td>
</tr>
</tbody>
</table>

Abbreviations: D - Mahalanobis distance between subgroup centroids, P - probability of subgroup membership, C - covariance matrix condition, R1 - correlation in larger group (when $p_1 \neq p_2$), R2 - correlation in smaller group (when $p_1 \neq p_2$), M - method of clustering, W - weighting method.
### Table 3
HA-Rand Index Means and (Std. Dev.) for Selected Effects

<table>
<thead>
<tr>
<th></th>
<th>Prob.</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>p₀ = 0.5</th>
<th>p₁ = 0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dₘ</td>
<td></td>
<td>.241 (.249)</td>
<td>.637 (.362)</td>
<td>.867 (.283)</td>
<td>.595 (.396)</td>
<td>.569 (.399)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>1:1</th>
<th>1:9</th>
<th>9:1</th>
<th>H₀</th>
<th>H₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>COV</td>
<td></td>
<td>.594 (.405)</td>
<td>.636 (.354)</td>
<td>.584 (.441)</td>
<td>.505 (.410)</td>
<td>.590 (.395)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>-.7</th>
<th>0.0</th>
<th>0.7</th>
<th>-.7</th>
<th>0.0</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>R₁</td>
<td></td>
<td>.533 (.411)</td>
<td>.607 (.389)</td>
<td>.606 (.387)</td>
<td>.559 (.408)</td>
<td>.591 (.394)</td>
<td>.596 (.389)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Euclid</th>
<th>Dₘ-T</th>
<th>Dₘ-W</th>
<th>PC-R</th>
<th>ACE</th>
<th>Ultrametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td></td>
<td>.512 (.418)</td>
<td>.492 (.400)</td>
<td>.677 (.352)</td>
<td>.583 (.402)</td>
<td>.627 (.381)</td>
<td>.599 (.397)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Recovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Link (Si)</td>
<td>.318 (.414)</td>
</tr>
<tr>
<td>Complete Link (Cm)</td>
<td>.456 (.390)</td>
</tr>
<tr>
<td>Ward’s Method (Wₐ)</td>
<td>.623 (.385)</td>
</tr>
<tr>
<td>EML</td>
<td>.609 (.389)</td>
</tr>
<tr>
<td>Average Linkage (Avg.)</td>
<td>.515 (.422)</td>
</tr>
<tr>
<td>Flexible Average, β = -.10 (Av10)</td>
<td>.663 (.365)</td>
</tr>
<tr>
<td>Flexible Average, β = -.15 (Av15)</td>
<td>.682 (.353)</td>
</tr>
<tr>
<td>Flexible Average, β = -.20 (Av20)</td>
<td>.682 (.353)</td>
</tr>
<tr>
<td>Flexible Average, β = -.25 (Av25)</td>
<td>.673 (.359)</td>
</tr>
<tr>
<td>Beta Flexible, β = -.25 (Fx25)</td>
<td>.568 (.388)</td>
</tr>
<tr>
<td>Beta Flexible, β = -.50 (Fx50)</td>
<td>.609 (.381)</td>
</tr>
</tbody>
</table>
Variable Weighting in Cluster Analysis

Table 4
Ranks of Weighting Methods Based on Ordinal Comparisons
(by Clustering Method)

<table>
<thead>
<tr>
<th>Weighting Method</th>
<th>Si</th>
<th>Cm</th>
<th>Wa</th>
<th>EML</th>
<th>Avg.</th>
<th>Av10</th>
<th>Av15</th>
<th>Av20</th>
<th>Av25</th>
<th>Fx25</th>
<th>Fx50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dm-W</td>
<td>1.5⁺</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ACE</td>
<td>1.5⁺</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2⁺</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ultrametric</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2⁺</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>PC-R</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4⁺</td>
<td>4⁺</td>
<td>4⁺</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Dm-T</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5⁺</td>
<td>4⁺</td>
<td>4⁺</td>
<td>4⁺</td>
<td>6</td>
</tr>
<tr>
<td>Euclid</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5⁺</td>
</tr>
</tbody>
</table>

⁺ Entries with common superscripts do not significantly differ from one another.
Variable Weighting in Cluster Analysis

Table 5
*$d_{uv}$ Values for Selected Comparisons of Weighting Methods
(by Clustering Method)

<table>
<thead>
<tr>
<th></th>
<th>Euclid</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dm-T</td>
<td>PC-R</td>
<td>Ultra.</td>
<td>ACE</td>
<td>Dm-W</td>
<td>Dm-T</td>
<td>PC-R</td>
<td>Ultra.</td>
<td>ACE</td>
</tr>
<tr>
<td>Si</td>
<td>.085</td>
<td>-.031</td>
<td>-.053</td>
<td>-.081</td>
<td>-.069</td>
<td>.169</td>
<td>.050</td>
<td>.026</td>
<td>-.007</td>
</tr>
<tr>
<td>Cm</td>
<td>.275</td>
<td>-.121</td>
<td>-.182</td>
<td>-.234</td>
<td>-.417</td>
<td>.486</td>
<td>.123</td>
<td>.053</td>
<td>.184</td>
</tr>
<tr>
<td>Wa</td>
<td>-.070</td>
<td>-.133</td>
<td>-.168</td>
<td>-.230</td>
<td>-.331</td>
<td>.166</td>
<td>.104</td>
<td>.067</td>
<td>.112</td>
</tr>
<tr>
<td>EML</td>
<td>.165</td>
<td>-.147</td>
<td>-.160</td>
<td>-.168</td>
<td>-.272</td>
<td>.342</td>
<td>.046</td>
<td>.025</td>
<td>.109</td>
</tr>
<tr>
<td>Avg.</td>
<td>.213</td>
<td>-.097</td>
<td>-.120</td>
<td>-.132</td>
<td>-.224</td>
<td>.350</td>
<td>.043</td>
<td>.010</td>
<td>.096</td>
</tr>
<tr>
<td>Av10</td>
<td>-.034</td>
<td>-.148</td>
<td>-.162</td>
<td>-.160</td>
<td>-.280</td>
<td>.140</td>
<td>.025</td>
<td>.016</td>
<td>.128</td>
</tr>
<tr>
<td>Av15</td>
<td>-.113</td>
<td>-.160</td>
<td>-.163</td>
<td>-.184</td>
<td>-.296</td>
<td>.086</td>
<td>.048</td>
<td>.033</td>
<td>.130</td>
</tr>
<tr>
<td>Av20</td>
<td>-.127</td>
<td>-.155</td>
<td>-.176</td>
<td>-.210</td>
<td>-.306</td>
<td>.091</td>
<td>.069</td>
<td>.052</td>
<td>.116</td>
</tr>
<tr>
<td>Av25</td>
<td>-.121</td>
<td>-.139</td>
<td>-.171</td>
<td>-.222</td>
<td>-.303</td>
<td>.104</td>
<td>.086</td>
<td>.061</td>
<td>.103</td>
</tr>
<tr>
<td>Fx25</td>
<td>.114</td>
<td>-.154</td>
<td>-.211</td>
<td>-.239</td>
<td>-.393</td>
<td>.348</td>
<td>.114</td>
<td>.066</td>
<td>.161</td>
</tr>
<tr>
<td>Fx50</td>
<td>-.010*</td>
<td>-.144</td>
<td>-.174</td>
<td>-.244</td>
<td>-.366</td>
<td>.246</td>
<td>.119</td>
<td>.072</td>
<td>.118</td>
</tr>
</tbody>
</table>

* Does not significantly differ from 0.0, p = .05.
### Table 6
Ranks of Clustering Algorithms Based on Ordinal Comparisons (by Weighting Method)

<table>
<thead>
<tr>
<th></th>
<th>Euclid</th>
<th>Dm-T</th>
<th>Dm-W</th>
<th>PC-R</th>
<th>ACE</th>
<th>Ultrametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av20</td>
<td>$2^{ab}$</td>
<td>1.5$^f$</td>
<td>1$^h$</td>
<td>2</td>
<td>1$^n$</td>
<td>1.5$^r$</td>
</tr>
<tr>
<td>Av15</td>
<td>1$^a$</td>
<td>1.5$^f$</td>
<td>2$^{hi}$</td>
<td>1</td>
<td>3$^o$</td>
<td>1.5$^r$</td>
</tr>
<tr>
<td>Av10</td>
<td>3$^{bc}$</td>
<td>4</td>
<td>4</td>
<td>3.5$^l$</td>
<td>5$^p$</td>
<td>4$^s$</td>
</tr>
<tr>
<td>Av25</td>
<td>4$^c$</td>
<td>3</td>
<td>3$^i$</td>
<td>3.5$^l$</td>
<td>2$^{no}$</td>
<td>3$^s$</td>
</tr>
<tr>
<td>Wa</td>
<td>6$^{de}$</td>
<td>5</td>
<td>5.5$^j$</td>
<td>6.5$^m$</td>
<td>5$^p$</td>
<td>6$^l$</td>
</tr>
<tr>
<td>Fx50</td>
<td>7$^e$</td>
<td>6</td>
<td>5.5$^j$</td>
<td>6.5$^m$</td>
<td>5$^p$</td>
<td>6$^l$</td>
</tr>
<tr>
<td>EML</td>
<td>5$^d$</td>
<td>7</td>
<td>7.5$^k$</td>
<td>5</td>
<td>7</td>
<td>6$^l$</td>
</tr>
<tr>
<td>Fx25</td>
<td>8</td>
<td>8</td>
<td>7.5$^k$</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Avg.</td>
<td>9</td>
<td>9.5$^g$</td>
<td>10</td>
<td>9</td>
<td>9.5$^q$</td>
<td>9</td>
</tr>
<tr>
<td>Cm</td>
<td>10</td>
<td>9.5$^g$</td>
<td>9</td>
<td>10</td>
<td>9.5$^q$</td>
<td>10</td>
</tr>
<tr>
<td>Si</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

Entries with common superscripts do not significantly differ from one another.
Table 7

d_0 Values for Selected Comparisons of Clustering Methods
(by Weighting Method)

<table>
<thead>
<tr>
<th></th>
<th>Si</th>
<th>Cm</th>
<th>Wa</th>
<th>EML</th>
<th>Avg.</th>
<th>Av10</th>
<th>Av15</th>
<th>Av20</th>
<th>Av25</th>
<th>Fx25</th>
<th>Fx50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclid</td>
<td>.440</td>
<td>.394</td>
<td>.142</td>
<td>.131</td>
<td>.247</td>
<td>.030</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Dm-T</td>
<td>.723</td>
<td>.689</td>
<td>.159</td>
<td>.385</td>
<td>.566</td>
<td>.107</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Dm-W</td>
<td>.613</td>
<td>.295</td>
<td>.088</td>
<td>.118</td>
<td>.274</td>
<td>.034</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>PC-R</td>
<td>.536</td>
<td>.411</td>
<td>.136</td>
<td>.123</td>
<td>.273</td>
<td>.035</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>ACE</td>
<td>.591</td>
<td>.326</td>
<td>.076</td>
<td>.116</td>
<td>.296</td>
<td>.055</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Ultrametric</td>
<td>.521</td>
<td>.359</td>
<td>.114</td>
<td>.114</td>
<td>.262</td>
<td>.021</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
</tbody>
</table>

Comparison with Ward's Method

<table>
<thead>
<tr>
<th></th>
<th>Si</th>
<th>Cm</th>
<th>----</th>
<th>EML</th>
<th>Avg.</th>
<th>Av10</th>
<th>Av15</th>
<th>Av20</th>
<th>Av25</th>
<th>Fx25</th>
<th>Fx50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclid</td>
<td>.319</td>
<td>.270</td>
<td>----</td>
<td>.005*</td>
<td>.122</td>
<td>-.109</td>
<td>-.142</td>
<td>-.126</td>
<td>-.110</td>
<td>.086</td>
<td>-.005*</td>
</tr>
<tr>
<td>Dm-T</td>
<td>.543</td>
<td>.546</td>
<td>----</td>
<td>.251</td>
<td>.431</td>
<td>-.044</td>
<td>-.159</td>
<td>-.173</td>
<td>-.148</td>
<td>.285</td>
<td>.077</td>
</tr>
<tr>
<td>Dm-W</td>
<td>.545</td>
<td>.212</td>
<td>----</td>
<td>.026</td>
<td>.200</td>
<td>-.054</td>
<td>-.088</td>
<td>-.087</td>
<td>-.065</td>
<td>.028</td>
<td>-.012*</td>
</tr>
<tr>
<td>PC-R</td>
<td>.386</td>
<td>.273</td>
<td>----</td>
<td>-.008*</td>
<td>.143</td>
<td>-.102</td>
<td>-.136</td>
<td>-.124</td>
<td>-.104</td>
<td>.071</td>
<td>-.010*</td>
</tr>
<tr>
<td>ACE</td>
<td>.541</td>
<td>.274</td>
<td>----</td>
<td>.058</td>
<td>.246</td>
<td>-.011*</td>
<td>-.076</td>
<td>-.095</td>
<td>.076</td>
<td>-.015*</td>
<td></td>
</tr>
<tr>
<td>Ultrametric</td>
<td>.416</td>
<td>.259</td>
<td>----</td>
<td>-.008*</td>
<td>.158</td>
<td>-.087</td>
<td>-.114</td>
<td>-.116</td>
<td>-.110</td>
<td>.059</td>
<td>-.005*</td>
</tr>
</tbody>
</table>

* Does not significantly differ from 0.0, p = .05.
Figure Captions

Figure 1. Plot of mean cluster recovery (HA-Rand) for two-way interaction of correlation in the larger group (R1) with variable weighting method (Weight).

Figure 2. Plot of mean cluster recovery (HA-Rand) for two-way interaction of covariance condition (COV) with variable weighting method (Weight).

Figure 3. Plot of mean cluster recovery (HA-Rand) for two-way interaction of subgroup size (probability of subgroup 1 membership - \( p_1 \)) with variable weighting method (Weight).

Figure 4. Plot of mean cluster recovery (HA-Rand) for two-way interaction of variable weighting method (Weight) with clustering method (Method).

Figure 5. Plot of mean cluster recovery (HA-Rand) for two-way interaction of Mahalanobis distance between the centroids of the subgroups (Distance) with clustering method (Method).

Figure 6. Plot of mean cluster recovery (HA-Rand) for two-way interaction of subgroup size (probability of subgroup 1 membership - \( p_1 \)) with clustering method (Method).

Figure 7. Plot of mean cluster recovery (HA-Rand) for two-way interaction of covariance condition (COV) with clustering method (Method).

Figure 8. Plot of mean cluster recovery (HA-Rand) for three-way interaction of Mahalanobis distance between the centroids of the subgroups (Distance) with covariance condition (COV) with subgroup size (probability of subgroup 1 membership - \( p_1 \)).
   a) Distance by COV interaction for equal-sized groups (\( p_1 = .5 \)).
   b) Distance by COV interaction for unequal-sized groups (\( p_1 = .9 \)).
R1 by Weight Interaction

Subgroup 1 Correlation
- $r_1 = -0.7$
- $r_1 = 0.0$
- $r_1 = 0.7$

Weighting Method
- Euclid
- Dm-T
- Dm-W
- PC-R
- ACE
- Ultra.
COV by Weight Interaction

Covariance Condition

- 1 : 1
- 1 : 9
- 9 : 1
- H0
- H1

Weighting Method

- Euclid
- Dm-T
- Dm-W
- PC-R
- ACE
- Ultra.
Prob. by Weight Interaction

Recovery

Subgroup Size

- p1=0.5
- p1=0.9

Euclid  Dm-T  Dm-W  PC-R  ACE  Ultra.

Weighting Method
Dist. by Method Interaction

Clustering Method

Mahalanobis Distance

- - - - Dm = 2
- - - - - - - - - - Dm = 4
- - - - - - - - - - - - - Dm = 6
Prob. by Method Interaction

Recovery

Subgroup Size
- p1=0.5
- p1=0.9

Clustering Method

Si Cm Wa EML Avg. Av10 Av15 Av20 Av25 Fx25 Fx50
COV by Method Interaction

Recovery

Si  Cm  Wa  EML  Avg.  Av10  Av15  Av20  Av25  Fx25  Fx50  Clustering Method

Covariance Condition
- 1:1
- 9:1
- H0
- H1

Clustering Method
Dist. by COV Interaction

a) $p_1 = 0.5$

Dist. by COV Interaction

b) $p_1 = 0.9$
Appendix

Additional Figures
Captions for Additional Figures

Figure A1. Plot of mean cluster recovery (HA-Rand) for three-way interaction of Mahalanobis distance between the centroids of the subgroups (Distance) with variable weighting method (Weight) with clustering method (Method).
   a) Weight by Method interaction for $D_M = 2.0$.
   b) Weight by Method interaction for $D_M = 4.0$.
   c) Weight by Method interaction for $D_M = 6.0$.

Figure A2. Plot of mean cluster recovery (HA-Rand) for three-way interaction of subgroup size (probability of subgroup 1 membership—$p_1$) with variable weighting method (Weight) with clustering method (Method).
   a) Weight by Method interaction for equal-sized groups ($p_1 = .5$).
   b) Weight by Method interaction for unequal-sized groups ($p_1 = .9$).
Weight by Method Interaction

a) $p_1 = 0.5$

b) $p_1 = 0.9$