African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25)
www.abjournals.org

# COMPARATIVE EVALUATION OF SIX AGGLOMERATIVE HIERARCHICAL CLUSTERING METHODS WITH A ROBUST EXAMPLE

**Eric U. Oti[1*] and Michael O. Olusola[2]**

[1*]Department of Statistics, Federal Polytechnic, Ekowe, Bayelsa State.

[2]Department of Statistics, Nnamdi Azikiwe University, Awka, Anambra State.

[*]Corresponding Author's Email: eluchcollections@gmail.com; Tel.: +2348037979262

**Cite this article:**

Oti Eric U., Olusola Michael O. (2024), Comparative Evaluation of Six Agglomerative Hierarchical Clustering Methods with a Robust Example. African Journal of Mathematics and Statistics Studies 7(2), 1-25. DOI: 10.52589/AJMSS-QXPH8R1N

**Manuscript History**

Received: 27 Jan 2024

Accepted: 8 Mar 2024

Published: 1 Apr 2024



**ABSTRACT:** *The agglomerative hierarchical clustering methods are the most popular type of hierarchical clustering used to group objects in clusters based on their similarity. The methods are represented by a bottom-up approach where each object starts in its cluster and pairs of clusters are merged as it moves up the hierarchy. In this paper, we present six agglomerative hierarchical clustering methods namely: the single linkage method, complete linkage method, average linkage method, centroid method, median method, and Ward's method. We also evaluated how these methods work on a practical basis using a matrix of distance pairs of five points. It was observed that the single linkage method through its dendrogram produced the most similarity measure between $x_i$ and $x_j$, while Ward's method produced the highest distance measure between $x_i$ and $x_j$.*

**KEYWORDS:** Agglomerative methods; Dendrogram; Distance matrix; Objects; Similarities.


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

## INTRODUCTION

In carrying out a cluster analysis of an n-dimensional multivariate dataset, one may wish to compare two or more hierarchical clustering methods of the same set of objects. Agglomerative clustering procedures are the most widely used of the hierarchical methods. The methods start with every single object in a single cluster and repeat merging the closest pair of clusters according to some similarity criteria until all of the data are in one cluster; or perhaps the process continues similarly until n objects are formed (Dillon & Goldstein, 1984). Just as with many other multivariate methods, objects to be classified have numerical measurements on a set of attributes or variables. So, the analysis is carried out on the rows of an array or a matrix. The objects or rows of the matrix can be viewed as a vector in a multidimensional space in which the dimensionality of the space is the number of variables or columns (Murtagh *et al.*, 2008).

Hierarchical clustering methods are generally represented by a tree diagram called a dendrogram (Gordon, 1996) and it could be represented either as a bottom-up clustering (also known as agglomerative clustering) method or viewed as a top-down clustering (divisive clustering) method. Divisive hierarchical clustering starts with all objects in one cluster and carries out the splitting of large clusters into smaller pieces. Both agglomerative and divisive hierarchical clustering have some disadvantages such as (a) when data points that have been incorrectly grouped at the early stage cannot be reallocated and (b) different similarity measures for measuring the similarity between clusters may lead to different results (Gan *et al.*, 2007).

A dendrogram is an n-tree-like diagram in which each internal node is associated with a height satisfying the condition that: $h(A) \leq h(B) \iff A \subseteq B$ for all subset of data points $A$ and $B$ if $A \cap B \neq 0$ where $h(A)$ and $h(B)$ denotes the heights of A and B respectively. Hence, it suffices that the heights in the dendrogram satisfy the ultra-metric condition in Equation (1) below (Johnson, 1967) where:

$$h_{ij} \leq max\{h_{ik}, h_{jk}\} \, \forall \, i, j, k \in \{1, 2, \dots, n\} \qquad (1)$$

In fact, the ultra-metric condition is a necessary and also sufficient condition for a dendrogram (Gordon, 1987).

The purpose of this paper is to evaluate and compare six agglomerative hierarchical clustering methods concerning the measures of distance or similarities of objects.

The rest of this paper is organised as follows: section 2 discusses the distance or similarity of objects as regards the measure of how close or far objects are to each other, it is known as the proximity of objects. In section 3, six agglomerative hierarchical clustering methods are discussed. Section 4 gives an illustrative example of how these six methods work in practice. Section 5 is the conclusion of the paper.


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25)
www.abjournals.org

## MEASURES OF PROXIMITY

Clusters are considered as groups containing data objects that are similar to each other than data objects in different clusters. Thus, in attempting to identify clusters of observations which may be present in data is knowledge of how "close" individuals or objects are to each other, or how far apart they are from each other (Jain & Dubes, 1988; Xu & Wunsch, 2008). Many clustering investigations have as their starting point a one-mode matrix, the elements which reflect in some sense, a quantitative measure of closeness, commonly referred to as dissimilarity (distance) or similarity, with a general term being known as proximity. Two individuals or objects are "close" when their dissimilarity or distance is small or their similarity is large (Everitt *et al.*, 2011; Romesburg, 1984).

The term proximity is the generalization for both dissimilarity and similarity. A dissimilarity or distance function on a data set x is defined to satisfy the following condition of a metric space (Anderberg, 1973; Zhang & Srihari, 2003):

- $d(x,y) \geq 0 \; for \; all \; x \; and \; y$ (Non-negativity)
- $d(x,y) = d(y,x)$ (Symmetry)
- $d(x,y) \leq d(x,z) + d(z,y) \; for \; all \; x,y,z$ (Triangle inequality)
- $d(x,y) = 0 \; iff \; x = y$ (Reflexivity)

A metric space $(X, d)$ is a set $X$ with a metric $d$ defined on $X$, but if the triangle inequality is not satisfied, the function is called a semi-metric. A metric is a function that defines a concept of distance between any two points of the set Also, if a metric is an ultra-metric (Johnson, 1967) implies that it satisfies a stronger condition that states that: $d(x,y) \leq max\,\{d(x,y)\} \; for \; all \; x,y,z$ where $x,y,z$ are arbitrary data points.

In application, the choice of distance is important and the best choice is often achieved through the combination of experience, skill and sometimes luck. There are many distances in practice, but the Euclidean distance is probably the most common distance ever used for numerical data. If we have two data points $x$ and $y$ in n-dimensional space, the Euclidean distance between them is defined in Equation (2) below:

$$d_{euc}(x,y) = \left[\sum_{i=1}^{n} (x_i, y_i)^2\right]^{\frac{1}{2}} = [\![(x-y)^T(x-y)]\!]^{\frac{1}{2}} \qquad (2)$$

Where $x_i$ and $y_i$ respectively are the values of the $ith$ attribute of $x$ and $y$. The generalised case of the above distance is known as the Minkowski distance which is defined as

$$d_{mink}(x,y) = \left[\sum_{i=1}^{n} |x_i - y_i|^p\right]^{\frac{1}{p}} \qquad (3)$$

and p is called the order or the exponent of the above Minkowski distance. Note that if we take $p = 1, 2, and\ \infty$, we get the Manhattan (City block) distance, Euclidean distance, and the Maximum (Chebyshev) distance as well (Mao & Jain, 1996).


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

## AGGLOMERATIVE HIERARCHICAL CLUSTERING METHODS

Many agglomerative hierarchical clustering methods have been proposed at one time or another but we are interested in six of the agglomerative clustering methods namely: single linkage method, complete linkage method, average linkage method, centroid method, median method, and Ward's method. Amongst these methods mentioned, the single, complete, and average linkage methods are referred to as graph methods, while the centroid, median, and Ward's methods are referred to as geometric methods (Murtagh, 1983); in the graph methods, a cluster can be represented by a subgroup or interconnected points while in geometric methods, a cluster can be represented by a centre point.

The steps in the graph methods (algorithms) for grouping n objects can be summarised as follows (Johnson & Wichern, 2007):

1) Start with n clusters, each containing a matrix of distance (or similarities) $D = \{d_{ik}\}$

2) Search the distance matrix for the nearest (most similar) pair of clusters. Let the distance between the "most similar" cluster $x$ and $y$ be $d(x, y)$ or $d_{ik}$

3) Merge cluster $x$ and $y$. Label the newly formed cluster $(xy)$. Update the entries in the distance matrix by deleting the rows and columns corresponding to the cluster $x$ and $y$, by adding a row and column giving the distance between the cluster $(xy)$ and the remaining clusters

4) Repeat steps (2) and (3) for a total of $n - 1$ times (all objects will be in a single cluster after the algorithm terminates). Record the identity of clusters that are merged and the levels (distances or similarities) at which the mergers take place.

To determine which group should be merged in agglomerative hierarchical clustering, we will first start with the single linkage method which is the simplest of all the methods. The single linkage (Sneath, 1957) utilises a minimum distance rule that starts by first, the two objects having the shortest (smallest) or largest similarity distance are merged; they constitute the first cluster. At the next stage, one of these two things can happen: Either a third object will join the already formed cluster of two, or the two closest un-clustered objects will be joined to form the second cluster. The decision rests on whether the distance from one of the un-clustered objects to the first cluster is shorter than the distances between the two closest un-clustered objects. The process continues until all objects belong to a single cluster. To find the minimum distance in $D = \{d_{ik}\}$, merge the corresponding objects: say $x$ and $y$ to get the cluster $(xy)$ and any other cluster $z$ are computed as $d_{(XY)Z} = min\{d_{XZ}, d_{YZ}\}$, where, $d_{XZ}$, and $d_{YZ}$ are the distance between the nearest neighbours of cluster $x$ and $y$ respectively.

The complete linkage method (McQuitty, 1960; Sokal & Sneath, 1963) uses the farthest neighbour distance to measure the dissimilarity between two groups. This method ensures that all items in a cluster are within the same maximum distance (or minimum similarity) to each other. To find the maximum distance $D = \{d_{ik}\}$, we merge the corresponding objects $x$ and $y$ to get a cluster $(xy)$. For step 3 of the graph algorithm, the distance between clusters $(xy)$ and any other cluster z is computed as $d_{(XY)Z} = max\{d_{XZ}, d_{YZ}\}$, where $d_{XZ}$ and $d_{YZ}$ are distances between the most distant members of the cluster $x$ and $z$ and cluster $y$ and z respectively.


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

The average linkage method is sometimes referred to as the unweighted pair group method using arithmetic averages (Jain and Dubes, 1988; Sokal and Michener, 1958). This method treats the distance between two clusters as the average distance between all pairs of items where one member of the pair belongs to each cluster. The distance can be computed as

$$d_{(XY)Z} = \frac{\sum_i \sum_k d_{ik}}{N_{(xy)} N_z} \tag{4}$$

where $d_{ik}$ is the distance between the object $i$ in the cluster $(xy)$ and the object $k$ in the cluster $z$. $N_{(xy)}$ and $N_z$ are the number of items in the cluster $(xy)$ and $z$ respectively.

The fourth is the centroid method where the distance between two clusters $i$ and $k$ is defined as the Euclidean distance between the mean vectors (often called centroids) of the two clusters stated as:

$$d(i,k) = d(\underline{x}_i \underline{x}_k) \tag{5}$$

where $\underline{x}_i$ and $\underline{x}_k$ are the mean vectors for the observation in $i$ and the observation in $k$ respectively, while $d(\underline{x}_i, \underline{x}_k)$ is the respective distance mean in the Euclidean space. The two clusters with the smallest distance between centroids are merged at each step; after the two clusters $i$ and $k$ are joined; the centroid of the new cluster $(ik)$ is calculated using the weighted average

$$\underline{x}_{ik} = \frac{n_i \underline{x}_i + n_k \underline{x}_k}{n_i + n_k} \tag{6}$$

The fifth is the median method which is also known as the "weighted pair group method using centroid" (Jain & Dubes, 1988). It was proposed by Gower (1967) to alleviate some disadvantages of the centroid method. In the centroid method, if the sizes of the two groups to be merged are quite different, then the centroid of the new group will be very close to that of the large group and may remain within that group (Everitt, 1993). In the centroid method, the centroid of the new group is independent of the size of the groups that form the new group. A disadvantage of this method is that it is not suitable for measures such as correlation coefficients, since interpretation in a geometrical sense is no longer possible (Lance & Williams, 1967). To avoid weighting the mean vector according to cluster size, we can use the median (midpoint) of the line joining $i$ and $k$ as the point for computing the next distance to other clusters $M_{ik} = \frac{1}{2}(\underline{x}_i + \underline{x}_k)$. The two clusters with the smallest distance between medians are merged at each step.

The sixth is Ward's method which was proposed by Ward and Hook (1963); it is a procedure-seeking method that forms the partition $p_k$, $p_{k-1}$, $p_{k-2}, \dots, p_1$ in a manner that minimises the loss of information associated with the merging. Usually, the loss of information is quantified in terms of an error sum of squares $(ESS)$ criterion. So, Ward's method is often referred to as the "minimum variance" method. Given a group c, the associated with c is given by: $ESS(c) = \sum_{x \in c} (x - \mu(c))(x - \mu(c))^T = \sum_{x \in c} xx^T - \frac{1}{|c|}\left(\sum_{x \in c} x\right)\left(\sum_{x \in c} x\right)^T = \sum_{x \in c} xx^T - |c|\mu(c)\mu(c)^T$ Where $\mu|c|$ is the mean of c, that is, $\mu|c| = \frac{1}{|c|}\sum_{x \in c} x$.


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25)


Suppose there are k groups $c_1, c_2, \dots, c_k$ in one level of the clustering. Then the information loss is represented by $ESS = \sum_{i=1}^{k} \quad ESS(c_i)$ which is the total within-group of the $ESS$. At each step of Ward's method, the union of every possible pair of groups is considered and two groups whose fusion results in the minimum increase in loss of information are merged. If the squared Euclidean distance is used to complete the dissimilarity matrix, then the dissimilarity matrix can be updated by the lance-William formula (Wishart, 1969) during the process of clustering as follows:

$$D\left(c_k, c_i \cup c_j\right) = \frac{|c_k|+|c_i|}{\sum_{ijk}} D(c_k, c_i) + \frac{|c_k|+|c_j|}{\sum_{ijk}} D\left(c_k, c_j\right) - \frac{|c_k|}{\sum_{ijk}} D\left(c_i, c_j\right) \qquad (7)$$

Initially, every single point forms a cluster and the total $ESS$ is $ESS_k = 0$.

## ILLUSTRATIVE EXAMPLE OF THE METHODS

Matrix of distance pair of five objects (Dillon and Goldstein, 1984) was used to illustrate the procedural steps of the agglomerative hierarchical clustering methods starting from the single linkage method:

$$D = d_{ik} = \begin{array}{c|c} & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} 0 & 1 & 5 & 6 & 8 \\ (1) & 0 & 3 & 8 & 7 \\ 5 & 3 & 0 & 4 & 6 \\ 6 & 8 & 4 & 0 & 2 \\ 8 & 7 & 6 & 2 & 0 \end{bmatrix} \end{array}$$

Treating each object as a cluster, we begin clustering by merging two closest items since $d_{ik} = d_{12} = 1$, object 2 and 1 are merged to form cluster $(12)$. To implement the next level of clustering, we need the distances between the cluster $(12)$ and the remaining objects 3, 4, and 5.

The nearest neighbor distances are: $d_{(12)3} = min\{d_{13}, d_{23}\} = min\{5,3\} = 3$. $d_{(12)4} = min\{d_{14}, d_{24}\} = min\{6,8\} = 6$. $d_{(12)5} = min\{d_{15}, d_{25}\} = min\{8, 7\} = 7$. Deleting the rows and columns of $D$ corresponding to object 1 and 2, and adding the remaining rows and columns for the cluster $(12)$; we obtain the new distance matrix as what we have below because a distance matrix will be symmetric and the lower echelon is shown because the upper echelon can be filled in by reflection:

$$D_2 = \begin{array}{c|c} & \begin{array}{cccc} (12) & 3 & 4 & 5 \end{array} \\ \begin{array}{c} (12) \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} 0 & & & \\ 3 & 0 & & \\ 6 & 4 & 0 & \\ 7 & 6 & 2 & 0 \end{bmatrix} \end{array}$$


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

The smallest distance between pair of clusters is now $d_{45} = 2$, so we merge object 4 and 5 to form cluster $(45)$. Calculating the nearest neighbor distances, we have:

$d_{(12)(45)} = min\{d_{(12)4}, d_{(12)5}\} = min\{6,7\} = 6.$ $d_{3(45)} = min\{d_{34}, d_{35}\} = min\{4,6\} = 4.$
The distance matrix becomes:

$$D_3 = \begin{array}{c} \\ (12) \\ 3 \\ (45) \end{array} \begin{array}{c} \begin{array}{ccc} (12) & 3 & (45) \end{array} \\ \begin{bmatrix} 0 & & \\ 3 & 0 & \\ 6 & 4 & 0 \end{bmatrix} \end{array}$$

The smallest distance between pairs of clusters is now $d_{(12)3} = 3$, so we merge cluster $(12)$ and 3 to get the next cluster $(123)$. Now the cluster left are cluster $(123)$ and $(45)$. Their nearest neighbor distance is $d_{(123)(45)} = min\{d_{(12)(45)}, d_{3(45)}\} = min\{6,4\} = 4$. The final distance matrix becomes:

$$D_4 = \begin{array}{c} \\ (123) \\ (45) \end{array} \begin{array}{c} \begin{array}{cc} (123) & (45) \end{array} \\ \begin{bmatrix} 0 & \\ 4 & 0 \end{bmatrix} \end{array}$$

Consequently, clusters $(123)$ and $(45)$ are merged to form a single cluster of all five objects $(12345)$ when the nearest neighbor distance reaches 4. The dendrogram is given below as:

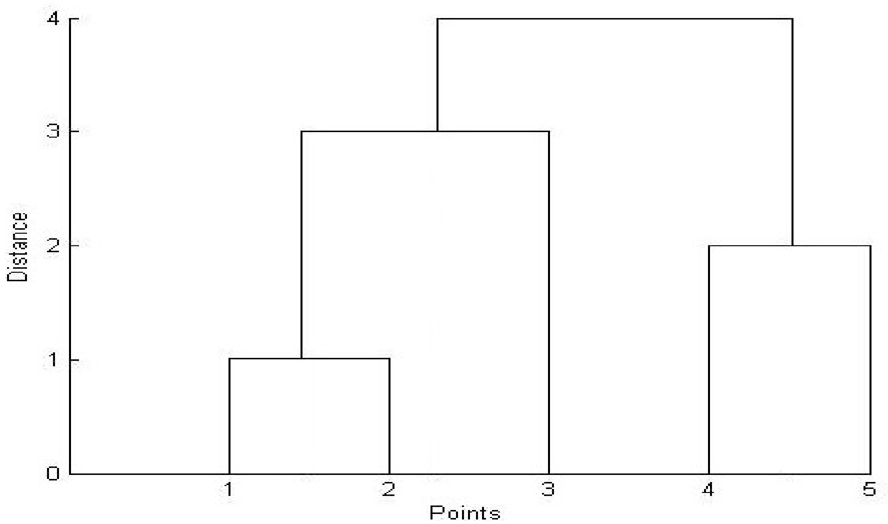



Figure 1: Single Linkage Dendrogram for distances between five objects.


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25)


For the complete linkage method using the same data, we have:

$$D = d_{ik} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & & \\ (1) & 0 & & & \\ 5 & 3 & 0 & & \\ 6 & 8 & 4 & 0 & \\ 8 & 7 & 6 & 2 & 0 \end{bmatrix} \end{array}$$

Treating each object as a cluster, clustering commences by merging the two closest items, since $d_{ik} = d_{12} = 1$. Object 1 and 2 are merged to form cluster (12) like the single linkage method which uses the minimum distance or most (maximum) similarity. At stage 2, we compute the maximum distance between clusters (12) and object 3, 4, and 5 as:

$d_{(12)3} = max\{d_{13}, d_{23}\} = max\{5,3\} = 5.$ $d_{(12)4} = max\{d_{14}, d_{24}\} = max\{6,8\} = 8.$ $d_{(12)5} = max\{d_{15}, d_{25}\} = max\{8,7\} = 8$ Which distance matrix becomes:

$$D_2 = \begin{array}{c} \\ (12) \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccc} (12) & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & \\ 5 & 0 & & \\ 8 & 4 & 0 & \\ 8 & 6 & 2 & 0 \end{bmatrix} \end{array}$$

The maximum similarity between pair of clusters 4 and 5 is $d_{45} = 2$. At stage 3, the distance matrix becomes:

$$D_3 = \begin{array}{c} \\ (12) \\ 3 \\ (45) \end{array} \begin{array}{c} \begin{array}{ccc} (12) & 3 & (45) \end{array} \\ \begin{bmatrix} 0 & & \\ 5 & 0 & \\ 8 & 6 & 0 \end{bmatrix} \end{array}$$

The maximum similarity between clusters (12) and 3 is $d_{(12)3} = 5$ to give cluster (123). Now the clusters left are cluster (123) and (45). The maximum distance between them is:

$d_{(123)(45)} = max\{d_{(12)(45)}, d_{3(45)}\} = max\{8,6\} = 8$. The final distance matrix becomes:

$$D_4 = \begin{array}{c} \\ (123) \\ (45) \end{array} \begin{array}{c} \begin{array}{cc} (123) & (45) \end{array} \\ \begin{bmatrix} 0 & \\ 8 & 0 \end{bmatrix} \end{array}$$


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25) 

At this final stage, clusters (123) and (45) are merged to form a single of all five objects (12345) when the maximum neighbor distance reaches 8. The dendrogram of the complete linkage is given in figure 2 below:

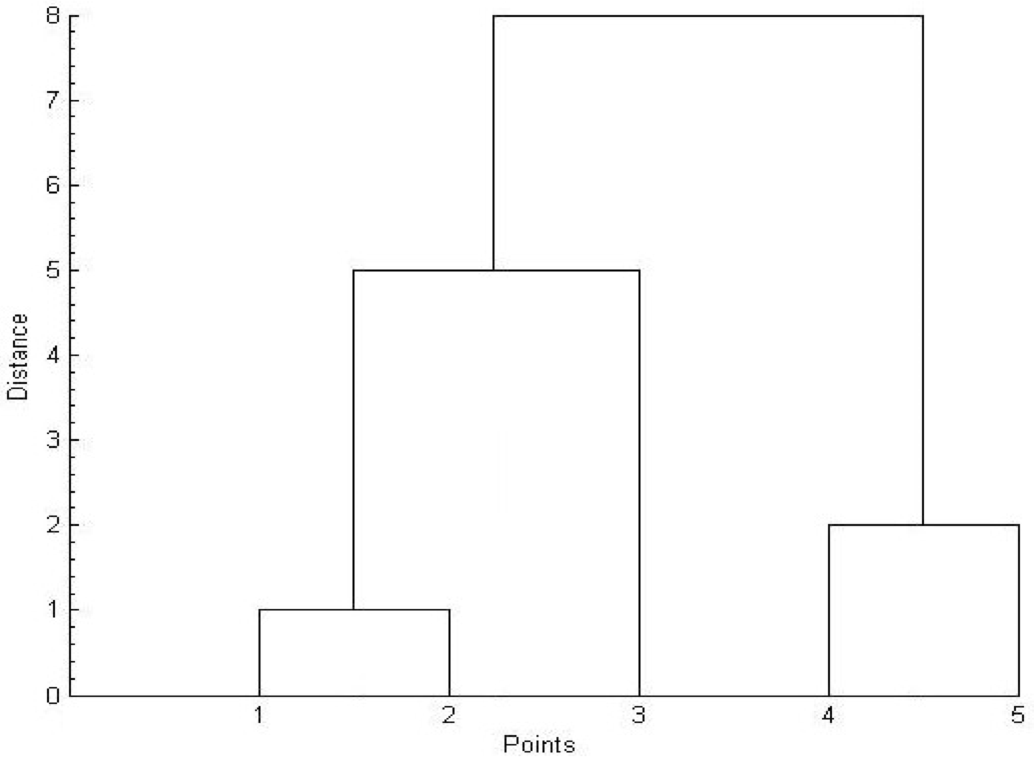



Figure 2: Complete Linkage Dendrogram for distance between five objects.

For the average linkage method:

$$D = d_{ik} = \begin{array}{c c} & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} 0 & & & & \\ (1) & 0 & & & \\ 5 & 3 & 0 & & \\ 6 & 8 & 4 & 0 & \\ 8 & 7 & 6 & 2 & 0 \end{bmatrix} \end{array}$$

Since min $d_{ik} = d_{12} = 1$, object 1 and 2 are merged to form cluster (12). To calculate the average distance we use Equation (4) which is:

$d_{(XY)Z} = \frac{\sum_i \sum_k d_{ik}}{N_{(xy)} N_z} = d_{(12)3} = \frac{1}{2} d_{13} + \frac{1}{2} d_{23} = \frac{5}{2} + \frac{3}{2} = 4$ . $d_{(12)4} = \frac{1}{2} d_{14} + \frac{1}{2} d_{24} = \frac{6}{2} + \frac{8}{2} = 7$. $d_{(12)5} = \frac{1}{2} d_{15} + \frac{1}{2} d_{25} = \frac{8}{2} + \frac{7}{2} = \frac{15}{2}$. The distance matrix becomes


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

$$D_2 = \begin{array}{c|cccc} & (12) & 3 & 4 & 5 \\ \hline (12) & 0 & & & \\ 3 & 4 & 0 & & \\ 4 & 7 & 4 & 0 & \\ 5 & \frac{15}{2} & 6 & 2 & 0 \end{array}$$

The minimum distance between pair of clusters is now $d_{45} = 2$. So object 4 and 5 are merged to form cluster (45) and the average distance is computed as: $d_{(12)(45)} = \frac{1}{2}d_{(12)4} + \frac{1}{2}d_{(12)5} = \frac{7}{2} + \frac{15}{4} = \frac{29}{4}$ $d_{3(45)} = \frac{1}{2}d_{34} + \frac{1}{2}d_{35} = \frac{4}{2} + \frac{6}{2} = 5$ then the distance matrix becomes:

$$D_3 = \begin{array}{c|ccc} & (12) & 3 & (45) \\ \hline (12) & 0 & & \\ 3 & 4 & 0 & \\ (45) & 29/4 & 5 & 0 \end{array}$$

Since the minimum distance between pair of clusters is now $d_{(12)3} = 4$. We merge cluster (12) and 3 to form cluster (123) and (45) which average distance is $d_{(123)(45)} = \frac{1}{2}d_{(12)(45)} + \frac{1}{2}d_{3(45)} = \frac{29}{8} + \frac{5}{2} = \frac{49}{8}$, so the final distance becomes

$$D_4 = \begin{array}{c|cc} & (123) & (45) \\ \hline (123) & 0 & \\ (45) & 49/8 & 0 \end{array}$$

At this final stage, cluster (123) and (45) are merged to form a single cluster (12345) when the average distance reaches $\frac{49}{8} = 6.125$.

The dendrogram of the average linkage method is given in figure 3 below

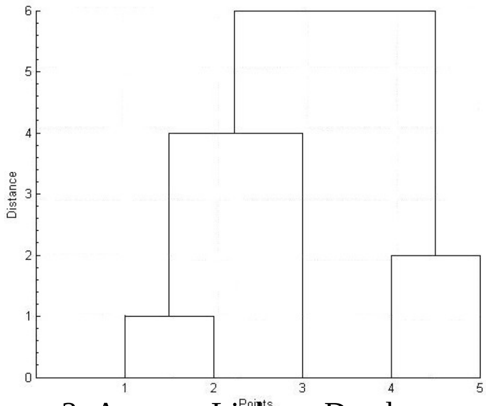



Figure 3: Average Linkage Dendrogram for distance between five objects


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25) www.abjournals.org

For the centroid method

$$D = d_{ik} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & & \\ (1) & 0 & & & \\ 5 & 3 & 0 & & \\ 6 & 8 & 4 & 0 & \\ 8 & 7 & 6 & 2 & 0 \end{bmatrix} \end{array}$$

Since min $d_{ik} = d_{12} = 1$, object 1 and 2 are merged to form cluster (12), then the distance is updated as: $d_{(12)3} = \frac{1}{2}\{d_{13} + d_{23}\} - \frac{1}{4}d_{12} = \frac{1}{2}\{5 + 3\} - \frac{1}{4}(1) = 3.75$

$d_{(12)4} = \frac{1}{2}\{d_{14} + d_{24}\} - \frac{1}{4}(1) = \frac{1}{2}\{6 + 8\} - \frac{1}{4}(1) = 6.75$

$d_{(12)5} = \frac{1}{2}\{d_{15} + d_{25}\} - \frac{1}{4}(1) = \frac{1}{2}\{8 + 7\} - \frac{1}{4}(1) = 7.25$ The distance matrix becomes

$$D_2 = \begin{array}{c} \\ (12) \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccc} (12) & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & \\ 3.75 & 0 & & \\ 6.75 & 4 & 0 & \\ 7.25 & 6 & 2 & 0 \end{bmatrix} \end{array}$$

The minimum distance between pair of clusters is $d_{45} = 2$, so we merge object 4 and 5 to form cluster (45). At the second stage, the distance is updated as:

$d_{(12)(45)} = \frac{1}{2}\{d_{(12)4} + d_{(12)5}\} - \frac{1}{4}d_{45} = \frac{1}{2}\{6.75 + 7.25\} - \frac{1}{4}(2) = 6.5$

$d_{3(45)} = \frac{1}{2}\{d_{34} + d_{35}\} - \frac{1}{4}d_{45} = \frac{1}{2}\{4 + 6\} - \frac{1}{4}(2) = 4.5$. Then the distance matrix becomes

$$D_3 = \begin{array}{c} \\ (12) \\ 3 \\ (45) \end{array} \begin{array}{c} \begin{array}{ccc} (12) & 3 & (45) \end{array} \\ \begin{bmatrix} 0 & & \\ 3.75 & 0 & \\ 6.5 & 4.5 & 0 \end{bmatrix} \end{array}$$

The minimum distance between pair of cluster is now $d_{(123)} = 3.75$, so cluster (12) and 3 are merged to form cluster (123). Now, the clusters left are (123) and (45). At this third stage, the distance is updated as:


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25)



$d_{(123)(45)} = \frac{2}{3}d_{(12)(45)} + \frac{1}{3}d_{3(45)} - \frac{2}{9}d_{(123)} = \frac{2}{3}(6.5) + \frac{1}{3}(4.5) - \frac{2}{9}(3.75) = 5$. And the final distance matrix becomes

$$D_4 = \begin{array}{c} \\ (123) \\ (45) \end{array} \begin{array}{c} \begin{array}{cc} (123) & (45) \end{array} \\ \begin{bmatrix} 0 & \\ 5 & 0 \end{bmatrix} \end{array}$$

The dendrogram of the centroid method is given below in figure 4

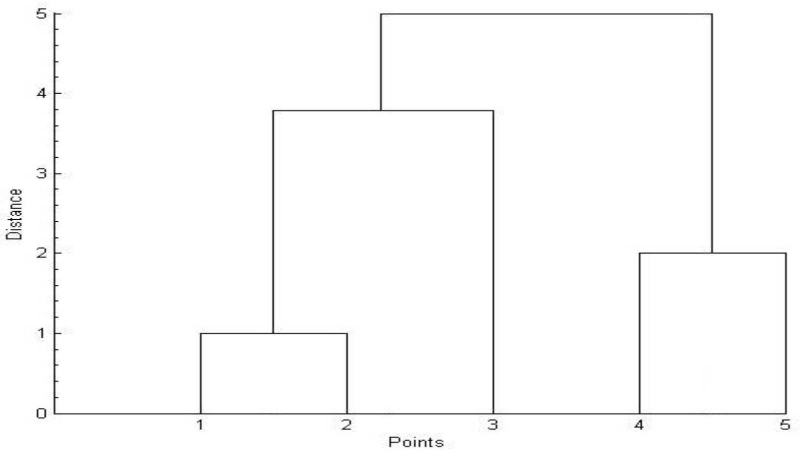



Figure 4: The centroid dendrogram for distance between five objects

For the median method

$$D = d_{ik} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & & \\ (1) & 0 & & & \\ 5 & 3 & 0 & & \\ 6 & 8 & 4 & 0 & \\ 8 & 7 & 6 & 2 & 0 \end{bmatrix} \end{array}$$

Since min $d_{ik} = d_{12} = 1$, object 1 and 2 are merged to form cluster (12). The median distance between newly formed groups and other groups are computed using Equation (8) below

$$D(c_k, c_i \cup c_j) = \frac{1}{2}D(c_k, c_i) + \frac{1}{2}D(c_k, c_j) - \frac{1}{4}D(c_i, c_j) \qquad (8)$$

$d_{(12)3} = \frac{1}{2}\{d_{13} + d_{23}\} - \frac{1}{4}d_{12} = \frac{1}{2}\{5 + 3\} - \frac{1}{4}(1) = 3.75$

$d_{(12)4} = \frac{1}{2}\{d_{14} + d_{24}\} - \frac{1}{4}d_{12} = \frac{1}{2}\{6 + 8\} - \frac{1}{4}(1) = 6.75$


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

$d_{(12)5} = \frac{1}{2}\{d_{15} + d_{25}\} - \frac{1}{4}d_{12} = \frac{1}{2}\{8 + 7\} - \frac{1}{4}(1) = 7.25$ The distance matrix becomes

$$D_2 = \begin{array}{c} \\ (12) \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccc} (12) & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & \\ 3.75 & 0 & & \\ 6.75 & 4 & 0 & \\ 7.25 & 6 & 2 & 0 \end{bmatrix} \end{array}$$

The minimum distance between pair of clusters is $d_{45} = 2$, so we merge objects 4 and 5 to form a cluster (45). The median distance becomes

$d_{(12)(45)} = \frac{1}{2}\{d_{(12)4} + d_{(12)5}\} - \frac{1}{4}d_{45} = \frac{1}{2}\{6.75 + 7.25\} - \frac{1}{4}(2) = 6.5$

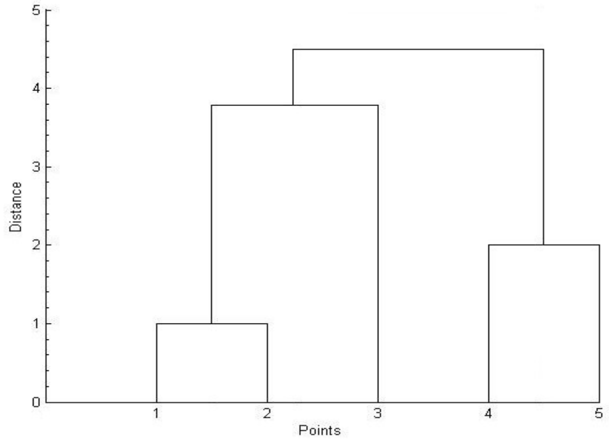
$d_{3(45)} = \frac{1}{2}\{d_{34} + d_{35}\} - \frac{1}{4}d_{45} = \frac{1}{2}\{4 + 6\} - \frac{1}{4}(2) = 4.5$. Then the distance matrix becomes

$$D_3 = \begin{array}{c} \\ (12) \\ 3 \\ (45) \end{array} \begin{array}{c} \begin{array}{ccc} (12) & 3 & (45) \end{array} \\ \begin{bmatrix} 0 & & \\ 3.75 & 0 & \\ 6.5 & 4.5 & 0 \end{bmatrix} \end{array}$$

The minimum distance between pair of cluster is now $d_{(123)} = 3.75$, so cluster (12) and 3 are merged to form cluster (123). The median distance becomes $d_{(123)(45)} = \frac{1}{2}\{d_{(12)(45)} + d_{3(45)}\} - \frac{1}{4}d_{123} = \frac{1}{2}\{6.5 + 4.5\} - \frac{1}{4}(3.75) = 4.56$ and the final matrix distance becomes

$$D_4 = \begin{array}{c} \\ (123) \\ (45) \end{array} \begin{array}{c} \begin{array}{cc} (123) & (45) \end{array} \\ \begin{bmatrix} 0 & \\ 4.56 & 0 \end{bmatrix} \end{array}$$

The dendrogram of the median method is given below in figure 5





Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25) 

Figure 5: The median dendrogram of distance between five objects.

For the Ward's method

$$D = d_{ik} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & & \\ (1) & 0 & & & \\ 5 & 3 & 0 & & \\ 6 & 8 & 4 & 0 & \\ 8 & 7 & 6 & 2 & 0 \end{bmatrix} \end{array}$$

Min $d_{ij} = d_{12} = 1$. When we compute the dissimilarity matrix for a data set $D = \{x_1, x_2, \dots, x_k\}$ using the squared Euclidean distance, then the entry $(i, j)$ of the dissimilarity matrix is $d_{ij}^2 = d(x_i, x_j) = (x_i - x_j)(x_i - x_j)^T = \sum_{i=1}(x_i - x_j)^2$ where d is the dimensionality of the data set D. Initially, $ESS_0 = 0, 1$ *and* 2 will merge at first stage and the increase in $ESS$ that result from the fusion (merger) of 1 and 2 is $\Delta ESS_{12} = \frac{1}{2}(1) = 0.5$ hence the $ESS$ becomes $ESS_1 = ESS_0 + \Delta ESS_{12} = 0 + 0.5 = 0.5$ using Equation (7) and the distance are updated as

$d_{(12)3} = \frac{2}{3}(d_{13} + d_{23}) - \frac{1}{3}d_{12} = \frac{2}{3}(5 + 3) - \frac{1}{3} = 5$

$d_{(12)4} = \frac{2}{3}(d_{14} + d_{24}) - \frac{1}{3}d_{12} = \frac{2}{3}(6 + 8) - \frac{1}{3} = 9$

$d_{(12)5} = \frac{2}{3}(d_{15} + d_{25}) - \frac{1}{3}d_{12} = \frac{2}{3}(8 + 7) - \frac{1}{3} = 9.67$ The distance matrix becomes

$$D_2 = \begin{array}{c} \\ (12) \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{cccc} (12) & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & \\ 5 & 0 & & \\ 8 & 4 & 0 & \\ 9.67 & 6 & 2 & 0 \end{bmatrix} \end{array}$$

At the second stage, 4 and 5 will merge and the resulting increase in $ESS$ is $\Delta ESS_{45} = \frac{1}{2}(2) = 1$. The total $ESS$ becomes $ESS_2 = ESS_1 + \Delta ESS_{45} = 0.5 + 1 = 1.5$, after 4 and 5 are merged, the distances are updated as $d_{(12)(45)} = \frac{1}{4}(d_{(12)4} + d_{(12)5}) - \frac{1}{4}d_{45} = \frac{3}{4}(8 + 9.67) - \frac{2}{4}(2) = 12.25$

$d_{3(45)} = \frac{2}{3}(d_{34} + d_{35}) - \frac{1}{3}d_{45} = \frac{2}{3}(4 + 6) - \frac{1}{3}(2) = 6$ The matrix becomes

(12) 3 (45)

 Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25) 

$$D_3 = \begin{array}{c} (12) \\ 3 \\ (45) \end{array} \begin{bmatrix} 0 & & \\ 5 & 0 & \\ 12.25 & 6 & 0 \end{bmatrix}$$

At the third stage, 1, 2 and 3 are merged and the resulting increase in $ESS$ is $\Delta ESS = \frac{1}{2}(5) = 2.5$, then the total $ESS$ becomes $ESS_3 = ESS_2 + ESS_{(12)3} = 1.5 + 2.5 = 4$ and the distance is updated as $d_{(123)(45)} = \frac{4}{5} d_{(12)(45)} + \frac{3}{5} d_{3(45)} - \frac{2}{5} d_{123} = \frac{4}{5}(12.25) + \frac{3}{5}(6) - \frac{2}{5}(5) = 11.4$ and the distance matrix becomes

$$D_4 = \begin{array}{c|cc} & (123) & (45) \\ \hline (123) & 0 & \\ (45) & 11.4 & 0 \end{array}$$

When all the data points are merged to a single cluster, the increase in $ESS$ will be $\Delta ESS_{(123)(45)} = \frac{1}{2}(11.4) = 5.7$ and the total $ESS$ will be $ESS_4 = ESS_3 + \Delta ESS_{(123)(45)} = 4 + 5.7 = 9.7$, the dendrogram of Ward's method is shown below in figure 6.

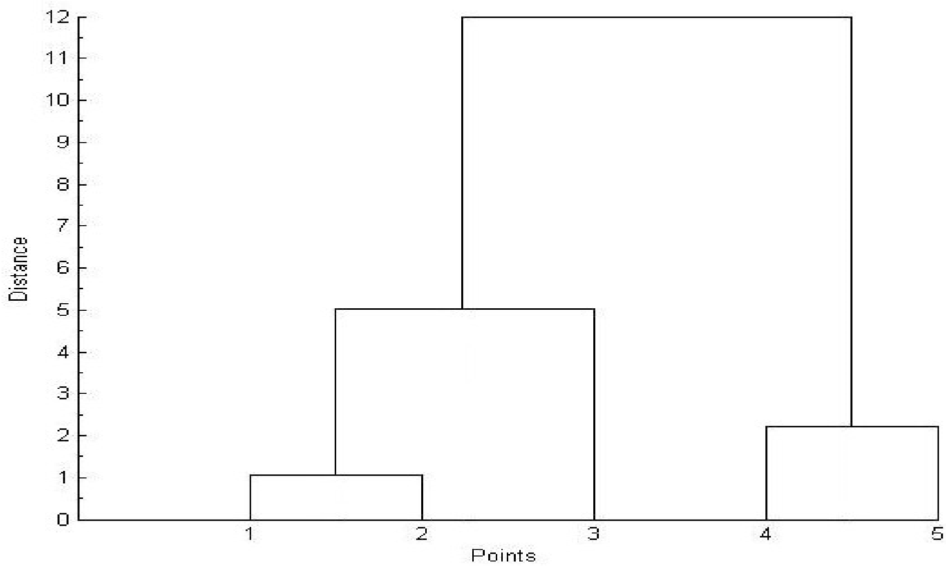



Figure 6: The Ward's Dendrogram of distance between five objects


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25)
www.abjournals.org

Matrix of distance pair of five objects (Dillon and Goldstein, 1984) was used to illustrate the procedural steps of the agglomerative hierarchical clustering methods starting from the single linkage method:

1 2 3 4 5

$D = d_{ik} =$

Treating each object as a cluster, we begin clustering by merging two closest items since $d_{ik} = d_{12} = 1$, object 2 and 1 are merged to form cluster $(12)$. To implement the next level of clustering, we need the distances between the cluster $(12)$ and the remaining objects 3, 4, and 5.

The nearest neighbour distances are: $d_{(12)3} = min\{d_{13}, d_{23}\} = min\{5,3\} = 3$. $d_{(12)4} = min\{d_{14}, d_{24}\} = min\{6,8\} = 6$. $d_{(12)5} = min\{d_{15}, d_{25}\} = min\{8, 7\} = 7$. Deleting the rows and columns of $D$ corresponding to objects 1 and 2, and adding the remaining rows and columns for the cluster $(12)$; we obtain the new distance matrix as what we have below because a distance matrix will be symmetric and the lower echelon is shown because the upper echelon can be filled in by reflection:

(12) 3 4 5

$D_2 =$

The smallest distance between pair of clusters is now $d_{45} = 2$, so we merge object 4 and 5 to form cluster $(45)$. Calculating the nearest neighbour distances, we have:

$d_{(12)(45)} = min\{d_{(12)4}, d_{(12)5}\} = min\{6,7\} = 6$. $d_{3(45)} = min\{d_{34}, d_{35}\} = min\{4,6\} = 4$. The distance matrix becomes:

(12) 3 (45)

$D_3 =$

The smallest distance between pairs of clusters is now $d_{(12)3} = 3$, so we merge cluster $(12)$ and 3 to get the next cluster $(123)$. Now the cluster left are cluster $(123)$ and $(45)$. Their nearest neighbour distance is $d_{(123)(45)} = min\{d_{(12)(45)}, d_{3(45)}\} = min\{6,4\} = 4$. The final distance matrix becomes:

(123) (45)

$D_4 =$

Consequently, clusters $(123)$ and $(45)$ are merged to form a single cluster of all five objects $(12345)$ when the nearest neighbour distance reaches 4. The dendrogram is given below as


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25)
www.abjournals.org

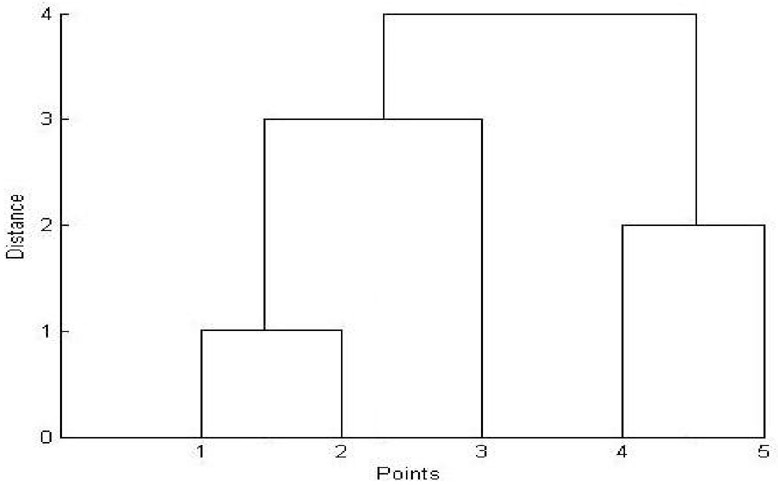



Figure 1: Single Linkage Dendrogram for distances between five objects.

For the complete linkage method using the same data, we have:

1 2 3 4 5

$D = d_{ik} =$

Treating each object as a cluster, clustering commences by merging the two closest items, since $d_{ik} = d_{12} = 1$. Objects 1 and 2 are merged to form cluster (12) like the single linkage method which uses the minimum distance or most (maximum) similarity. At stage 2, we compute the maximum distance between clusters (12) and objects 3, 4, and 5 as

$d_{(12)3} = max\{d_{13}, d_{23}\} = max\{5,3\} = 5.$ $d_{(12)4} = max\{d_{14}, d_{24}\} = max\{6,8\} = 8.$ $d_{(12)5} = max\{d_{15}, d_{25}\} = max\{8,7\} = 8$ Which distance matrix becomes:

(12) 3 4 5

$D_2 =$

The maximum similarity between pair of clusters 4 and 5 is $d_{45} = 2$. At stage 3, the distance matrix becomes:

$$D_3 = \begin{array}{c} \begin{matrix}(12) & 3 & (45)\end{matrix} \\ \begin{bmatrix}0 & & \\ 5 & 0 & \\ 8 & 6 & 0\end{bmatrix}\end{array}$$

The maximum similarity between clusters (12) and 3 is $d_{(12)3} = 5$ to give cluster (123). Now the clusters left are clusters (123) and (45). The maximum distance between them is:

$d_{(123)(45)} = max\{d_{(12)(45)}, d_{3(45)}\} = max\{8,6\} = 8$. The final distance matrix becomes:


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

(123) (45)

$D_4 =$

At this final stage, clusters (123) and (45) are merged to form a single of all five objects (12345) when the maximum neighbour distance reaches 8. The dendrogram of the complete linkage is given in Figure 2 below:

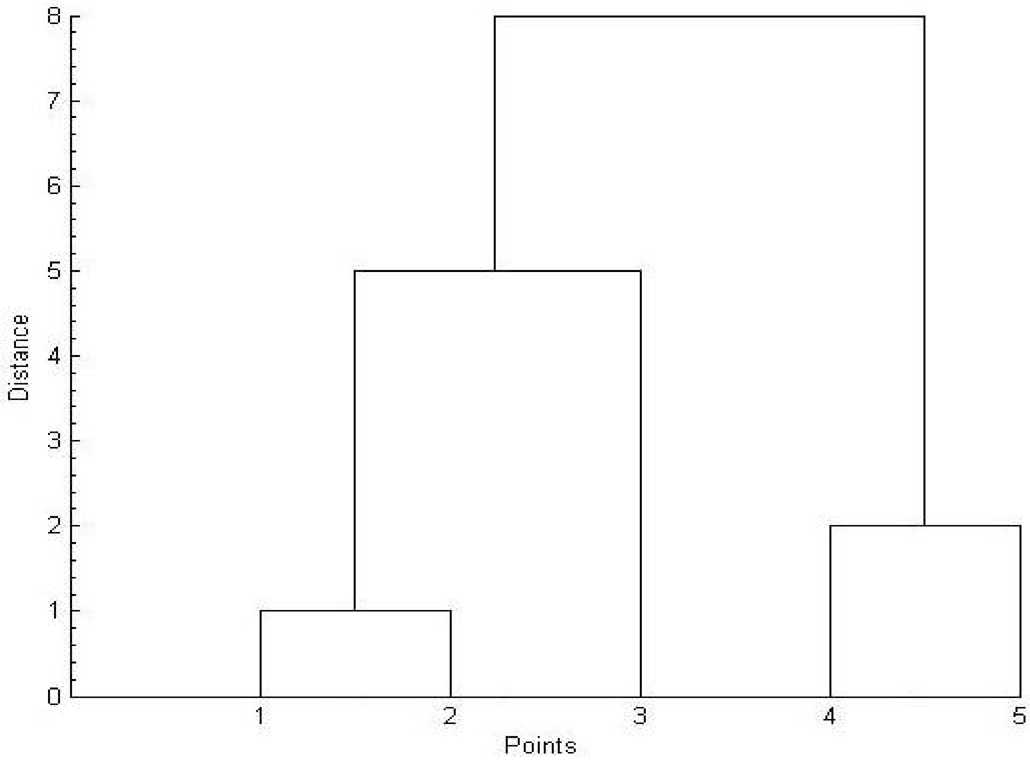



Figure 2: Complete Linkage Dendrogram for distance between five objects.

For the average linkage method:

1 2 3 4 5

$D = d_{ik} =$

Since min $d_{ik} = d_{12} = 1$, object 1 and 2 are merged to form cluster (12). To calculate the average distance we use Equation (4) which is:

$d_{(XY)Z} = \frac{\sum_i \quad \sum_k \quad d_{ik}}{N_{(xy)}N_z} \;=\; d_{(12)3} = \frac{1}{2}d_{13} + \frac{1}{2}\,d_{23} = \frac{5}{2} + \frac{3}{2} = 4\,.\qquad d_{(12)4} = \frac{1}{2}d_{14} + \frac{1}{2}d_{24} = \frac{6}{2} + \frac{8}{2} = 7.\; d_{(12)5} = \frac{1}{2}d_{15} + \frac{1}{2}d_{25} = \frac{8}{2} + \frac{7}{2} = \frac{15}{2}.$ The distance matrix becomes


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

$$D_2 = \begin{array}{c} \\ (12) \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{cccc} (12) & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & \\ 4 & 0 & & \\ 7 & 4 & 0 & \\ \frac{15}{2} & 6 & 2 & 0 \end{bmatrix} \end{array}$$

The minimum distance between pairs of clusters is now $d_{45} = 2$. So object 4 and 5 are merged to form cluster (45) and the average distance is computed as: $d_{(12)(45)} = \frac{1}{2}d_{(12)4} + \frac{1}{2}d_{(12)5} = \frac{7}{2} + \frac{15}{4} = \frac{29}{4}$ $d_{3(45)} = \frac{1}{2}d_{34} + \frac{1}{2}d_{35} = \frac{4}{2} + \frac{6}{2} = 5$ then the distance matrix becomes:

(12) 3 (45)

$D_3 =$

Since the minimum distance between pairs of clusters is now $d_{(12)3} = 4$. We merge clusters (12) and 3 to form clusters (123) and (45) which average distance is $d_{(123)(45)} = \frac{1}{2}d_{(12)(45)} + \frac{1}{2}d_{3(45)} = \frac{29}{8} + \frac{5}{2} = \frac{49}{8}$, so the final distance becomes

(123) (45)

$D_4 =$

At this final stage, clusters (123) and (45) are merged to form a single cluster (12345) when the average distance reaches $\frac{49}{8} = 6.125$.

The dendrogram of the average linkage method is given in Figure 3 below

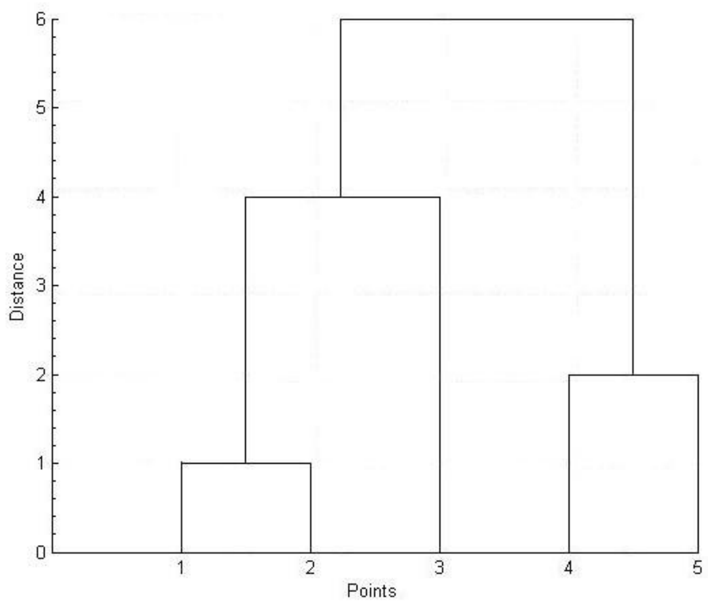



Figure 3: Average Linkage Dendrogram for the distance between five objects


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25)


For the centroid method

1 2 3 4 5

$D = d_{ik} =$

Since min $d_{ik} = d_{12} = 1$, object 1 and 2 are merged to form cluster (12), then the distance is updated as: $d_{(12)3} = \frac{1}{2}\{d_{13} + d_{23}\} - \frac{1}{4}d_{12} = \frac{1}{2}\{5 + 3\} - \frac{1}{4}(1) = 3.75$

$d_{(12)4} = \frac{1}{2}\{d_{14} + d_{24}\} - \frac{1}{4}(1) = \frac{1}{2}\{6 + 8\} - \frac{1}{4}(1) = 6.75$

$d_{(12)5} = \frac{1}{2}\{d_{15} + d_{25}\} - \frac{1}{4}(1) = \frac{1}{2}\{8 + 7\} - \frac{1}{4}(1) = 7.25$ The distance matrix becomes

(12) 3 4 5

$D_2 =$

The minimum distance between pairs of clusters is $d_{45} = 2$, so we merge objects 4 and 5 to form cluster (45). In the second stage, the distance is updated as:

$d_{(12)(45)} = \frac{1}{2}\{d_{(12)4} + d_{(12)5}\} - \frac{1}{4}d_{45} = \frac{1}{2}\{6.75 + 7.25\} - \frac{1}{4}(2) = 6.5$

$d_{3(45)} = \frac{1}{2}\{d_{34} + d_{35}\} - \frac{1}{4}d_{45} = \frac{1}{2}\{4 + 6\} - \frac{1}{4}(2) = 4.5$. Then the distance matrix becomes

(12) 3 (45)

$D_3 =$

The minimum distance between pairs of clusters is now $d_{(123)} = 3.75$, so clusters (12) and 3 are merged to form cluster (123). Now, the clusters left are (123) and (45). At this third stage, the distance is updated as:

$d_{(123)(45)} = \frac{2}{3}d_{(12)(45)} + \frac{1}{3}d_{3(45)} - \frac{2}{9}d_{(123)} = \frac{2}{3}(6.5) + \frac{1}{3}(4.5) - \frac{2}{9}(3.75) = 5$. And the final distance matrix becomes

(123) (45)

$D_4 =$

The dendrogram of the centroid method is given below in Figure 4


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

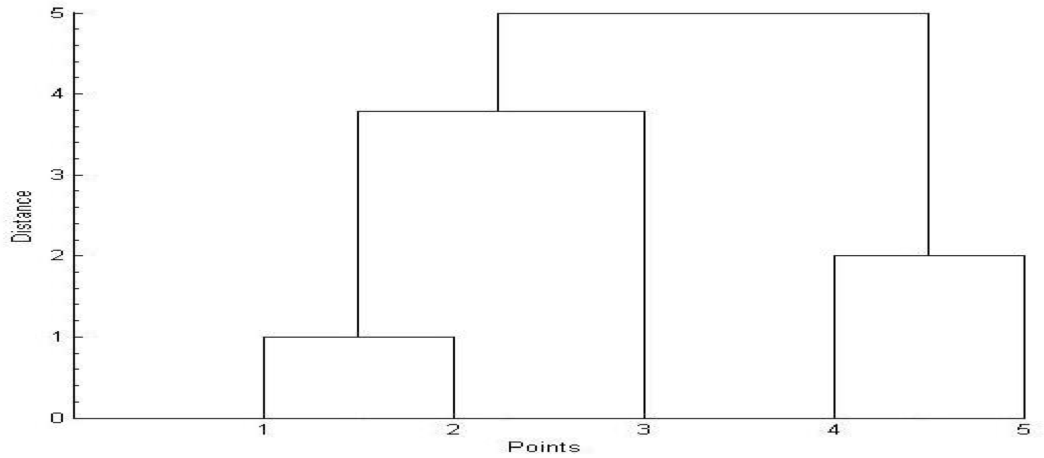



Figure 4: The centroid dendrogram for the distance between five objects

For the median method

1 2 3 4 5

$D = d_{ik} =$

Since min $d_{ik} = d_{12} = 1$, object 1 and 2 are merged to form cluster (12). The median distance between newly formed groups and other groups is computed using Equation (8) below

$$D(c_k, c_i \cup c_j) = \frac{1}{2}D(c_k, c_i) + \frac{1}{2}D(c_k, c_j) - \frac{1}{4}D(c_i, c_j) \quad (8)$$

$d_{(12)3} = \frac{1}{2}\{d_{13} + d_{23}\} - \frac{1}{4}d_{12} = \frac{1}{2}\{5 + 3\} - \frac{1}{4}(1) = 3.75$

$d_{(12)4} = \frac{1}{2}\{d_{14} + d_{24}\} - \frac{1}{4}d_{12} = \frac{1}{2}\{6 + 8\} - \frac{1}{4}(1) = 6.75$

$d_{(12)5} = \frac{1}{2}\{d_{15} + d_{25}\} - \frac{1}{4}d_{12} = \frac{1}{2}\{8 + 7\} - \frac{1}{4}(1) = 7.25$ The distance matrix becomes

(12) 3 4 5

$D_2 =$

The minimum distance between pairs of clusters is $d_{45} = 2$, so we merge objects 4 and 5 to form a cluster (45). The median distance becomes

$d_{(12)(45)} = \frac{1}{2}\{d_{(12)4} + d_{(12)5}\} - \frac{1}{4}d_{45} = \frac{1}{2}\{6.75 + 7.25\} - \frac{1}{4}(2) = 6.5$

$d_{3(45)} = \frac{1}{2}\{d_{34} + d_{35}\} - \frac{1}{4}d_{45} = \frac{1}{2}\{4 + 6\} - \frac{1}{4}(2) = 4.5.$ Then the distance matrix becomes

(12) 3 (45)

$D_3 =$


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

The minimum distance between the pair of cluster is now $d_{(123)} = 3.75$, so clusters (12) and 3 are merged to form cluster (123). The median distance becomes $d_{(123)(45)} = \frac{1}{2}\{d_{(12)(45)} + d_{3(45)}\} - \frac{1}{4}d_{123} = \frac{1}{2}\{6.5 + 4.5\} - \frac{1}{4}(3.75) = 4.56$ and the final matrix distance becomes

(123) (45)

$D_4 =$

The dendrogram of the median method is given below in Figure 5

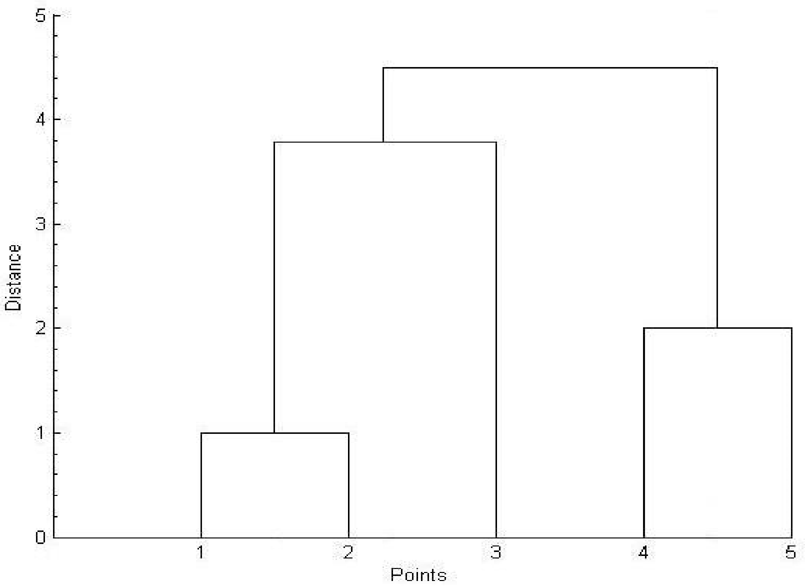



Figure 5: The median dendrogram of the distance between five objects.

For Ward's method:

1 2 3 4 5

$D = d_{ik} =$

$Min\, d_{ij} = d_{12} = 1$. When we compute the dissimilarity matrix for a data set $D = \{x_1, x_2, \dots, x_k\}$ using the squared Euclidean distance, then the entry $(i, j)$ of the dissimilarity matrix is $d_{ij}^2 = d(x_i, x_j) = (x_i - x_j)(x_i - x_j)^T = \sum_{i=1} (x_i - x_j)^2$ where d is the dimensionality of the data set D. Initially, $ESS_0 = 0, 1$ *and* 2 will merge at the first stage and the increase in $ESS$ that result from the fusion (merger) of 1 and 2 is $\Delta ESS_{12} = \frac{1}{2}(1) = 0.5$ hence the $ESS$ becomes $ESS_1 = ESS_0 + \Delta ESS_{12} = 0 + 0.5 = 0.5$ using Equation (7) and the distance are updated as

$$d_{(12)3} = \frac{2}{3}(d_{13} + d_{23}) - \frac{1}{3}d_{12} = \frac{2}{3}(5 + 3) - \frac{1}{3} = 5$$

$$d_{(12)4} = \frac{2}{3}(d_{14} + d_{24}) - \frac{1}{3}d_{12} = \frac{2}{3}(6 + 8) - \frac{1}{3} = 9$$


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 1-25) 

$d_{(12)5} = \frac{2}{3}(d_{15} + d_{25}) - \frac{1}{3}d_{12} = \frac{2}{3}(8 + 7) - \frac{1}{3} = 9.67$ The distance matrix becomes

(12) 3 4 5

$D_2 =$

In the second stage, 4 and 5 will merge and the resulting increase in $ESS$ is $\Delta ESS_{45} = \frac{1}{2}(2) = 1$. The total $ESS$ becomes $ESS_2 = ESS_1 + \Delta ESS_{45} = 0.5 + 1 = 1.5$, after 4 and 5 are merged, the distances are updated as $d_{(12)(45)} = \frac{1}{4}\left(d_{(12)4} + d_{(12)5}\right) - \frac{1}{4}d_{45} = \frac{3}{4}(8 + 9.67) - \frac{2}{4}(2) = 12.25$

$d_{3(45)} = \frac{2}{3}(d_{34} + d_{35}) - \frac{1}{3}d_{45} = \frac{2}{3}(4 + 6) - \frac{1}{3}(2) = 6$ The matrix becomes

(12) 3 (45)

$D_3 =$

At the third stage, 1, 2 and 3 are merged and the resulting increase in $ESS$ is $\Delta ESS = \frac{1}{2}(5) = 2.5$, then the total $ESS$ becomes $ESS_3 = ESS_2 + ESS_{(12)3} = 1.5 + 2.5 = 4$ and the distance is updated as $d_{(123)(45)} = \frac{4}{5}d_{(12)(45)} + \frac{3}{5}d_{3(45)} - \frac{2}{5}d_{123} = \frac{4}{5}(12.25) + \frac{3}{5}(6) - \frac{2}{5}(5) = 11.4$ and the distance matrix becomes

(123) (45)

$D_4 =$

When all the data points are merged to a single cluster, the increase in $ESS$ will be $\Delta ESS_{(123)(45)} = \frac{1}{2}(11.4) = 5.7$ and the total $ESS$ will be $ESS_4 = ESS_3 + \Delta ESS_{(123)(45)} = 4 + 5.7 = 9.7$, the dendrogram of Ward's method is shown below in Figure 6.

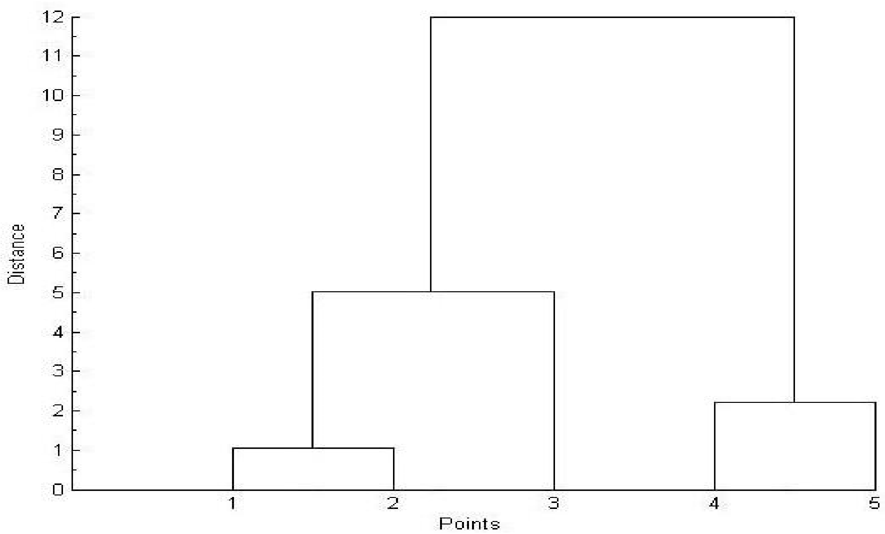



Figure 6: The Ward's Dendrogram of distance between five objects


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

## CONCLUSION

In this paper, we have presented six agglomerative hierarchical clustering methods and also evaluated how these methods work on a matrix of distance pairs of five objects.

The dendrograms of the agglomerative hierarchical clustering methods show all the steps in the hierarchical procedure which includes the distances (or similarities) at which clusters are merged. The dendrograms displayed information in respective tables in the form of a tree diagram. It was observed that in each dendrogram; points (observations) in the x-axis, that is, $1, 2, \ldots, 5$ made up of one single cluster (12345). The y-axis indicates the distances or heights $(h_{ij})$ of the internal nodes for each pair of data points.

Small values of $(h_{ij})$ indicate a high similarity and large values of $(h_{ij})$ indicate a high distance between $x_i$ and $x_j$. It was observed that the single linkage method via its dendrogram produced the most similarity measure between $x_i$ and $x_j$, while Ward's method provided the highest distance measure between $x_i$ and $x_j$.

## REFERENCES

Anderberg, M. R. (1973). *Cluster Analysis for Applications*. Academic Press, New York.

Dillon, W. R., & Goldstein M. (1984). *Multivariate Analysis Methods and Applications*. John Wiley and Sons Inc.

Everitt, B. S. (1993). *Cluster Analysis*, 3rd Edition. New York, Toronto: Halsted Press.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis*, 5th Edition. John Wiley and Sons.

Gan, G., Ma, C., & Wu, J. (2007). *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM Series.

Gordon, A. (1987). A review of the hierarchical classification. *Journal of the Royal Statistical Society*. A (General). 150(2): 119-137.

Gordon, A. (1996). *Hierarchical classification. In Arabie, P., Hubert, L., and Soete, G., editors, clustering and classification*, pages 65-121. River Edge, NJ: World Scientific.

Gower, J. C. (1967). Multivariate Analysis and Multidimensional Geometry. J*ournal of the Royal Statistical Society*. Series D (The Statistician) 17(1): 13-28.

Jain, A., & Dubes, R. (1988). *Algorithms for Clustering Data*. Eaglewood Cliffs, NJ: Prentice-Hall.

Johnson, S. C. (1967). Hierarchical clustering scheme. *Psychometrica*, 32, 241-254.

Johnson, R. A., & Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis*. 5th Edition, Eaglewood Cliffs. NJ: Prentice Hall.

Lance, G. N., & Williams, W. T. (1967). A general theory of classificatory sorting strategies: Hierarchical systems. *The Computer Journal*, 9(4): 373-380.

Mao, J., & Jain, A. (1996). A self-organizing network for hyper-ellipsoidal clustering (HEC). *IEEE Transactions on Neural Networks*.7(1): 16-29.

McQuitty, L. (1957). Elementary linkage analysis for isolating orthogonal and oblique types and relevancies. *Educational and Psychological Measurement*, 17: 207-222.

Murtagh, F. (1983). A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, 26(4): 315-323.


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N

Murtagh, F., Downs, G., & Contreras, P. (2008). Hierarchical clustering of massive, high dimensional data sets by exploiting ultra-metric embedding. *SIAM Journal on Scientific Computing*, 30(2): 707-730.

Romesburg, H. C. (1984). *Cluster Analysis for Researchers*. London: Wadswoth.

Sneath, P. H. A. (1957). The application of computers to taxonomy. *Journal of General Microbiology*, 17, 201-226.

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin*, 38, 1409-1438.

Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of Numerical Taxonomy*. San Francisco: Freeman.

Ward, Jr., J., & Hook, M. (1963). Application of a hierarchical grouping procedure to a problem of group profiles. *Educational and psychological measurement*, 23(1): 69-81.

Wishart, D. (1969). 256 Note: *An algorithm for hierarchical classifications Biometrics*, 25(1): 165-170.

Xu, R., & Wunch, D. C. (2009). *Clustering, IEEE Computational Intelligence Society*, Series Press. John Wiley and Sons.

Zhang, B., & Srihari, S. (2003). *Properties of Binary Vectors Dissimilarity Measures*. Technical report, CEDAR, Department of Computer Science and Engineering, University of Buffalo, New York.


Article DOI: 10.52589/AJMSS-QXPH8R1N
DOI URL: https://doi.org/10.52589/AJMSS-QXPH8R1N