Cluster analysis
Cluster analysis

• Generic name of wide variety of multivariate procedures used to infer ‘classifications’.
  – Form clusters, = partitions of objects, = groups of similar objects, = homogeneous groups.
  – Objective, replicable.
  – Widely use in sciences:
    • All sciences built on classifications that structure their areas of inquiry.
    • Most common in psychology, social sciences, biology.
  – Principal uses:
    (1) Development of ‘typologies’ or classifications.
    (2) Investigation of useful conceptual schemes for grouping observations.
    (3) Inductive hypothesis generation, from data.
    (4) Empirical hypothesis testing of previously defined groups.
  – Phylogenetics: special case of cluster analysis.
Cluster analysis

- General caveats:
  1. Most CA methods are relatively simple procedures that are not supported by extensive bodies of statistical reasoning.
     - *Heuristics*: “rules of thumb”.
     - Plausible algorithms that can be used to create clusters of objects.
     - Contrast with eigenanalyses.
     - So clusters should not be *reified*.
  2. CA methods have evolved from many disciplines.
     - Each discipline has own biases and preferences:
       - Kinds of questions asked from data.
       - Types of data thought to be useful for classification.
       - Structure of classifications thought to be useful.
     - Methods in any one discipline often comprise a mixture of methods.
Cluster analysis

(3) Different clustering methods produce different solutions for the same data set.
   • Methods use different rules of group formation and evaluation.
   • Can be source of considerable confusion.
   • Q: What is the most ‘natural’ classification?

(4) Strategy of CA is structure-seeking, but its operation is structure-imposing.
   • Clustering methods are used to discover structure in data that is not readily apparent by visual inspection or other methods.
   • Clustering methods always place objects into groups, whether or not groups exist.
   • Q: Are groups ‘real’, or imposed on data by the method?
   • Leads to questions of statistical significance of clusters.
Similarity

- All CA methods based on some concept of similarity:
  - = Resemblance, = association, = proximity.
  - Concept tends to be vague.
  - The same clustering method used with different similarity measures will usually produce different results.
  - Similarity vs. dissimilarity.
- Many kinds of quantitative measures (coefficients).
  - Correlation coefficients.
  - Association coefficients.
  - Probabilistic similarity measures.
  - Feature-matching measures.
Dissimilarity

- **Distance metrics**: special class of (dis)similarity measures:
  - Objects represented as points in a coordinate space.
  - Distance between pair of points represents dissimilarity between them.
  - Excellent statistical properties.

- **Metric properties**:
  1. **Symmetry**: \( d(x, y) = d(y, x) \geq 0 \)
  2. **Triangle inequality**: \( d(x, y) \leq d(x, z) + d(y, z) \)
     - Length of any side of a triangle is less than or equal to the sum of the other two sides.
  3. **Distinguishability of nonidenticals**: if \( d(x, y) \neq 0 \), then \( x \neq y \)
  4. **Indistinguishability of identicals**:
     - For two identical objects: \( d(x, x') = 0 \)

- In particular, all metrics are *jointly monotonic*: all provide the same rankings of distances for a set of objects.
  - Non-metric measures (e.g., correlation coefficients, or nearest-neighbor distances) might not be jointly monotonic.
Additive vs ultrametric: triangle inequality

- Metric properties:
  
  (1) \( d(a, a) = 0, \ d(b, b) = 0, \ d(a, b) \geq 0 \)
  
  - Identical objects are indistinguishable.
  
  - Nonidentical objects might or might not be distinguishable.

  (2) If \( a \neq b \), then \( d(a, b) > 0 \)
  
  - If the objects differ in properties being measured, then their distance must be greater than zero.

  (3) \( d(a, b) = d(b, a) \)
  
  - Symmetry.
Additive vs ultrametric: triangle inequality

(4) \[ d(a, c) \leq d(a, b) + d(b, c) \]

- Triangle inequality.

- Any set of distances having these four metric properties will produce an additive tree.

- Properties can be superimposed on distances, even if they don’t possess the properties.
  - ‘Forces’ the distances into an additive tree.

- 4\textsuperscript{th} property can be relaxed:

(4) \[ d(a, c) \leq \max\{d(a, b), d(b, c)\} \]

- Any set of distances having these properties will produce an ultrametric tree.

- Any distances can be ‘forced’ into an ultrametric tree.

- Ensures that clustering monotonically increases with distance.
Some commonly used distances

- **Euclidean distance:**

- **Manhattan distance:**

- **Mahalanobis distance:**

- **Hamming** distance between two strings of equal length: number of positions in which symbols are different (→ percent dissimilarity).
Distance matrix

- Distance or dissimilarity measures calculated for all possible pairs of objects.
- Arranged into a *distance or dissimilarity matrix*:

\[
D = \begin{bmatrix}
0 & d_{12} & d_{13} & \cdots & d_{1N} \\
 d_{21} & 0 & d_{23} & \cdots & d_{2N} \\
 d_{31} & d_{32} & 0 & \cdots & d_{3N} \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 d_{N1} & d_{N2} & d_{N3} & \cdots & 0 \\
\end{bmatrix}
\]

- Zeros on the diagonal, indicating self-identity.
- Off-diagonal elements represent pairwise distances.
- Matrix can be:
  - Symmetric, as with distance metrics.
  - Asymmetric, as with asymmetric dissimilarity measures, such as nearest-neighbor distances.
Clusters

- Purpose of cluster analysis is to find *clusters*.
- No standard, general definition of *cluster*.
  - Everitt (1980): “continuous region of a space containing a relatively high density of points, separated from other such regions by regions containing a relatively low density of points.”
  - Defines clusters in terms of the space rather than the points.
- However, clusters (whatever they are) have properties:
  - *Density*: property that defines cluster as a relatively thick swarm of points in a geometric space, compared to other areas of the space.
  - *Variance*, connectivity: degree of dispersion of points from center of the cluster (“tight”, “loose”); nearness of points to one another.
  - *Dimension*: average ‘radius’ of a cluster.
  - *Shape*: arrangement of points in the space; often assumed to be hypersphere.
  - *Separation*: degree to which clusters overlap or lie apart in the space.
Cluster analysis

- Seven major families of clustering methods:
  1. Hierarchical agglomerative
  2. Hierarchical divisive
  3. Iterative partitioning
  4. Density search
  5. Factor analytic
  6. Clumping
  7. Graph theoretic
(1) Hierarchical agglomerative methods

- Dominant method in biological sciences.
- Conceptually simple to understand.
  - Begin with set of $N$ objects (entities, observations).
  - Successively group into clusters.
  - Eventually end with all objects in one cluster.
- Begins with square symmetric distance matrix:
  - Iteratively:
    - Find pair of objects or clusters having least distance.
    - Merge into new cluster.
    - Recalculate distances between new cluster and all remaining objects and clusters.
  - Continues until all objects are in a single cluster.
  - For $N$ objects, requires $N–1$ steps.
(1) Hierarchical agglomerative methods

- Different methods differ in rules for recalculating distances (= linkage rules):
  - **Single linkage (nearest-neighbor)**: distance between clusters is minimum pairwise cross-distance of members.
  - **Complete linkage (furthest-neighbor)**: distance between clusters is maximum pairwise cross-distance of members.
  - **Average linkage (centroid)**:
    - Unweighted pair-group method (UPGMA): distance between clusters is unweighted mean of all cross-distances between clusters.
    - Weighted pair-group method (WPGMA): as in UPGMA, except that distances are weighted by cluster sizes.
  - **Ward’s method**: clusters formed in such a way as to minimize the within-cluster variance.
(1) Hierarchical agglomerative methods

- Example: *single-linkage* (nearest neighbor) cluster analysis
  
  - **Distance matrix:**

    |   | 1  | 2  | 3  | 4  | 5  |
    |---|----|----|----|----|----|
    | 1 | 0  | 2  | 6  | 10 | 9  |
    | 2 | 2  | 0  | 5  | 9  | 8  |
    | 3 | 6  | 5  | 0  | 4  | 5  |
    | 4 | 10 | 9  | 4  | 0  | 3  |
    | 5 | 9  | 8  | 5  | 3  | 0  |

  - Objects 1 & 2 fuse to form a cluster (12) at distance 2.
  - Recalculate distances from other objects to cluster (12):

    \[
    d_{(12)3} = \min \{d_{13}, d_{23}\} = d_{23} = 5
    \]

    \[
    d_{(12)4} = \min \{d_{14}, d_{24}\} = d_{24} = 9
    \]

    \[
    d_{(12)5} = \min \{d_{15}, d_{25}\} = d_{25} = 8
    \]
(1) Hierarchical agglomerative methods

- Objects 4 & 5 fuse to form a cluster (45) at distance 3.
- Recalculate distances from other objects to cluster (45):

\[
d_{(12)(45)} = \min\{d_{14}, d_{15}, d_{24}, d_5\} = d_{25} = 8
\]

\[
d_{3(45)} = \min\{d_{34}, d_3\} = d_{34} = 4
\]

<table>
<thead>
<tr>
<th></th>
<th>(12)</th>
<th>(45)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(12)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>(45)</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>

- Objects 3 & (45) fuse to form a cluster (345) at distance 4:
- Recalculate distance between objects (12) and (345):

\[
d_{(12)(345)} = \min\{d_{(12)3}, d_{(12)(45)}\} = d_{(12)3} = 5
\]

<table>
<thead>
<tr>
<th></th>
<th>(12)</th>
<th>(345)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(12)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>(345)</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

- Objects (12) & (345) fuse to form a cluster (12345) at distance 5.
- Summarize with a dendrogram:
1) Hierarchical agglomerative methods

[Diagram of a single-linkage dendrogram with labels 1, 2, 3, 4, 5 and y-axes labeled 5.0, 4.0, 3.0, 2.0, 1.0.]

Single-linkage dendrogram
(1) Hierarchical agglomerative methods

- Lance & Williams (1967):
  - Showed that all forms of linkage rules are special cases of a general 4-parameter model ("the flexible clustering method"):
    \[ d_{hk} = \alpha_i d_{hi} + \alpha_j d_{hj} + \beta d_{ij} + \gamma |d_{hi} - d_{hj}| \]
  - Objects/clusters \( i \) and \( j \) fuse to form new cluster.
  - Model provides ‘rule’ for calculating distance between objects/clusters \( h \) and \( k \), given their distances to \( i \) and \( j \).
  - Model is general, but particular combinations of parameter values correspond to special cases: e.g.,
    - Single linkage: \( \alpha_i = \alpha_i = \frac{1}{2}, \beta = 0, \gamma = -\frac{1}{2} \)
    - Complete linkage: \( \alpha_i = \alpha_i = \frac{1}{2}, \beta = 0, \gamma = \frac{1}{2} \)
    - UPGMA: \( \alpha_i = \frac{n_i}{n_k} , \alpha_i = \frac{n_i}{n_k} , \beta = -\alpha_i \alpha_j , \gamma = 0 \)
Flexible clustering

- Parameter $\beta$ is a ‘space distorting’ parameter.
  - Example varying only $\beta$ for the same data set:
Flexible clustering

- Extended by Jambu (1978) to more general 5-parameter model:

Table 2. Clustering strategies obtainable from the general recurrence relation of Jambu (1978)

<table>
<thead>
<tr>
<th>Name (Reference)</th>
<th>$\alpha_i$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
<th>$\delta_i$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 Single-link (Florek et al., 1951; Sneath, 1957)</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
<td>$-\frac{1}{2}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C2 Complete-link (McQuitty, 1960)</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C3 Group average-link (Sokal &amp; Michener, 1958; McQuitty, 1967)</td>
<td>$\frac{n_i}{n_i + n_j}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C4 Weighted average-link (McQuitty, 1966, 1967)</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C5 Mean dissimilarity (Podani, 1989)</td>
<td>$\frac{n_i}{2}$</td>
<td>$\frac{n_i + n_j}{2}$</td>
<td>0</td>
<td>$-\frac{n_i}{2}$</td>
<td>$-\frac{n_j}{2}$</td>
</tr>
<tr>
<td>C6 Sum of squares (Jambu, 1978; Podani, 1979)</td>
<td>$\frac{n_i + n_k}{n_i}$</td>
<td>$\frac{n_i + n_j}{n_i}$</td>
<td>0</td>
<td>$-\frac{n_i}{n_i}$</td>
<td>$-\frac{n_k}{n_i}$</td>
</tr>
<tr>
<td>C7 Incremental sum squares (Ward, 1963; Anderson, 1966; Wishart, 1969)</td>
<td>$\frac{n_i + n_k}{n_i}$</td>
<td>$-\frac{n_k}{n_i}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C8 Centroid (Sokal &amp; Michener, 1958; Gower, 1967)</td>
<td>$\frac{n_i}{n_i + n_j}$</td>
<td>$-\frac{n_i n_j}{(n_i + n_j)^2}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C9 Median (Lance &amp; Williams, 1966a; Gower, 1967)</td>
<td>$\frac{1}{2}$</td>
<td>$-\frac{1}{4}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C10 Flexible (Lance &amp; Williams, 1966a)</td>
<td>$\frac{1}{2}(1 - \beta)$</td>
<td>$\beta$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: $n_i$ is the number of objects in class $C_i$; $n_+ = n_i + n_j + n_k$. 
Additive cluster methods

- Trees based on flexible clustering are *ultrametric*:
  - All terminal values are equally distant from the root.

- *Additive* clustering methods relax the ultrametric assumption.
  - E.g., neighbor-joining, least-squares.
Measuring agreement of dendrogram with original distance matrix

- Hierarchical agglomerative dendrograms reflect only information in original distance matrix.
  - Constrained by a clustering algorithm: set of rules for joining objects into clusters.
  - Process of clustering *distorts* the pairwise distances among objects.

- *Cophenetic (=patristic) distance*: distance between two objects as measured on the dendrogram.

- Methods for comparing original and cophenetic distances:
  - Cophenetic correlation coefficient
  - Gower’s coefficient

- Etc.
(2) Hierarchical divisive methods

- Similar to hierarchical agglomerative methods, but in opposite direction:
  - Begin with set of $n$ objects.
  - Successively subdivide objects into subsets.
  - End with individual objects.
- Forms *polythetic* clusters: groups of objects for which no single variable value is *necessary* for cluster membership, but certain subsets of variables are *sufficient* for membership.
- Used primarily with binary data.
- Rather than cluster-joining rules, use cluster-dividing rules.
  - E.g., based on chi-square statistic or information statistic.
- *Association analysis*: version of divisive clustering used in ecology.
(3) Iterative partitioning

- Attempt to partition data into a set of clusters, assuming that the number of clusters \((k)\) is known.
  - Produce single-rank clusters that are not nested (not hierarchical).
- Utilize raw data rather than a distance matrix.
- \(K\)-means method (e.g., CLUSTAN):
  1. Begin with initial ‘guess’ by partitioning data points into specified number of clusters. Compute centroids of clusters.
  2. Allocate each point to the cluster that has the nearest centroid.
  3. Compute new centroids of clusters.
     - Repeat steps 2 and 3 until no points change clusters.
- Not guaranteed to find globally optimal solution, which is computationally impossible.
  - E.g., for 15 objects and 3 clusters, there are 217,945,728,000 unique possible partitions.
  - Many heuristic shortcuts developed.
(3) Iterative partitioning

- If number of clusters \((k)\) not known, need *clustering criterion*: 
  - Measures the suitability of a partition, relative to number of clusters.
  - Many different criteria used; e.g., sum-of-squares, maximum likelihood, minimum entropy.

- Then use iterative procedure:
  - Set \(k = 2\), find best partition.
  - Set \(k = 3\), find best partition.
  - Etc., up to some specified maximum number of clusters.
  - Use clustering criterion to compare partitions and estimate ‘best’ value of \(k\).

- Almost all criteria used tend to ‘discover’ spherical clusters of approximately equal radius.
- Strauss (2001): \(k\)NN method to identify non-circular clusters.
(3) Iterative partitioning
(4) Density search methods

• = *Kernel methods*: identify regions of ‘high’ density of points in space relative to surrounding regions.
  – Search space for natural modes.

• Two basic strategies:
  (1) Variant of single-linkage clustering: *minimum-spanning tree (MST)*.
    – Prune longest edges (branches) to form clusters.
(4) Density search methods

(2) Methods based on *mixtures* of multivariate probability distributions.

- Usually *multivariate normal* (e.g., NORMIX).
- *Mixture* = collection of samples representing different populations of objects.
- Based on statistical model that assumes that different groups (clusters) have distributions that are identical except for centroid (and perhaps variance).
- Sensitive to suboptimal solutions.
(5) Factor analytic methods

- Based on Q-mode eigenanalysis of $n \times n$ correlation matrix of observations rather than $p \times p$ matrix of variables.
  - Objects assigned to clusters based on eigenvector loadings.
- Primarily used in social sciences: very controversial.
(6) Clumping methods

- Unique in that permit creation of overlapping clusters.
  - Don’t produce hierarchical clusters.
  - Objects are permitted to be members of two or more clusters.
  - Originally developed by linguistics researchers.
    - Words can have more than one meaning.

- More recently, development of *fuzzy clustering* methods.
(7) Graph theoretic methods

- Includes methods based on well-developed theorems and axioms of graph theory.
  - *Random graph hypothesis*: all rank-order distance matrices are equally likely.
  - RGH used as null hypothesis to test for presence of clusters in a distance matrix.
  - Methods provide alternatives to heuristic strategies of other clustering methods.
Phylogenetics

• Cluster analysis methods commonly used in evolutionary biology.
  – E.g., UPGMA, additive tree methods.

• Used because trees comprise fundamental evolutionary models.
Trees describe evolutionary relationships

von Baer, 1837

Darwin, 1859

Haeckel, 1870
The phylogenetic model

- Two components: *anagenesis* and *cladogenesis*.
- Models provide process-level explanations:
  Anagenesis: natural selection, etc.
  Cladogenesis: allopatric speciation, etc.
Two classes of phylogenetic methods

(1) **Algorithmic:**
- Formulate an algorithm (set of rules) for generating a tree from a data matrix.
- Play out the algorithm (exact or iterative).
- Global optimality criterion: maximum-likelihood estimation.
- Local optimality criterion:
  - Additive and ultrametric clustering methods.
- **Problems:**
  - The rules for generating the tree might not have any biological/evolutionary basis.
  - Even if the optimality criterion is global, any particular solution might be only locally optimal.

(2) **Combinatoric:**
- Formulate a quantitative criterion for choosing the “best” tree.
- Compare all possible trees (or a reasonable subset), giving each tree a score based on the criterion.
- Choose the tree (or possibly set of trees) having the best score.
- **Problem:** the number of possible trees is enormous.
  - Heuristics are possible (e.g., branch and bound).
<table>
<thead>
<tr>
<th>Taxa</th>
<th>Number of rooted trees</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>105</td>
</tr>
<tr>
<td>6</td>
<td>945</td>
</tr>
<tr>
<td>7</td>
<td>10,395</td>
</tr>
<tr>
<td>8</td>
<td>135,135</td>
</tr>
<tr>
<td>9</td>
<td>2,027,025</td>
</tr>
<tr>
<td>10</td>
<td>34,459,425</td>
</tr>
<tr>
<td>11</td>
<td>654,729,075</td>
</tr>
<tr>
<td>12</td>
<td>13,749,310,575</td>
</tr>
<tr>
<td>13</td>
<td>316,234,143,225</td>
</tr>
<tr>
<td>14</td>
<td>7,905,853,580,625</td>
</tr>
<tr>
<td>15</td>
<td>213,458,046,676,875</td>
</tr>
<tr>
<td>16</td>
<td>6,190,283,353,629,370</td>
</tr>
<tr>
<td>17</td>
<td>191,898,783,962,510,000</td>
</tr>
<tr>
<td>18</td>
<td>6,332,659,870,762,850,000</td>
</tr>
<tr>
<td>19</td>
<td>221,643,095,476,699,000,000</td>
</tr>
<tr>
<td>20</td>
<td>8,200,794,532,637,890,000,000</td>
</tr>
</tbody>
</table>
Phylogenetic methods

- **Algorithmic methods:**
  - Hierarchical agglomerative cluster analysis (UPGMA).
  - Maximum-likelihood methods.
    - Based on specific evolutionary models.

- **Combinatoric methods:**
  - Parsimony trees.
    - Use branch-and-bound for heuristic shortcuts.
  - Bayesian trees.
    - Based on specific evolutionary models.
    - Use Markov Chain Monte Carlo (MCMC) methods for heuristic shortcuts.
Other methods for modeling distance matrices

**Distance matrix**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>1.41</td>
<td>2.45</td>
<td>2.45</td>
<td>2.83</td>
</tr>
<tr>
<td>2</td>
<td>1.41</td>
<td>0.00</td>
<td>2.45</td>
<td>3.16</td>
<td>2.45</td>
</tr>
<tr>
<td>3</td>
<td>2.45</td>
<td>2.45</td>
<td>0.00</td>
<td>1.41</td>
<td>1.41</td>
</tr>
<tr>
<td>4</td>
<td>2.45</td>
<td>3.16</td>
<td>1.41</td>
<td>0.00</td>
<td>2.45</td>
</tr>
<tr>
<td>5</td>
<td>2.83</td>
<td>2.45</td>
<td>1.41</td>
<td>2.45</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Correlation matrix**

<table>
<thead>
<tr>
<th></th>
<th>X₁</th>
<th>X₂</th>
<th>X₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>X₁</td>
<td>1.00</td>
<td>0.16</td>
<td>0.79</td>
</tr>
<tr>
<td>X₂</td>
<td>0.16</td>
<td>1.00</td>
<td>0.37</td>
</tr>
<tr>
<td>X₃</td>
<td>0.79</td>
<td>0.37</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Principal coordinates analysis (PCoA)

- *Eigenanalysis* method:
  - Decomposes information in the distance matrix into a set of orthogonal axes:
    - Axes = principal coordinates.
    - Orthogonal = statistically independent.
  - *Principal coordinates*:
    - PCo1 accounts for the maximum *dispersal* in the distance matrix.
    - PCo2 accounts for the maximum residual dispersal in the distance matrix, independent of PCo1.
    - PCo3 accounts for the maximum residual dispersal in the distance matrix, independent of both PCo1 and PCo2.
    - Etc.
  - For an $N \times N$ distance matrix, there are $N$ principal coordinates.
    - The full set of $N$ principal coordinates accounts for all of the information in the distance matrix.
  - Observations (taxa) are projected onto the axes to give projection scores, which can be plotted with scattergrams.
Principal coordinates analysis (PCoA)

Distance matrix

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>1.41</td>
<td>2.45</td>
<td>2.45</td>
<td>2.83</td>
</tr>
<tr>
<td>2</td>
<td>1.41</td>
<td>0.00</td>
<td>2.45</td>
<td>3.16</td>
<td>2.45</td>
</tr>
<tr>
<td>3</td>
<td>2.45</td>
<td>2.45</td>
<td>0.00</td>
<td>1.41</td>
<td>1.41</td>
</tr>
<tr>
<td>4</td>
<td>2.45</td>
<td>3.16</td>
<td>1.41</td>
<td>0.00</td>
<td>2.45</td>
</tr>
<tr>
<td>5</td>
<td>2.83</td>
<td>2.45</td>
<td>1.41</td>
<td>2.45</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Additive

PCo1 (60.6%) vs. PCo2 (34.5%)
Multidimensional scaling (MDS)

- Not an eigenanalysis method:
  - Information in distance matrix is not decomposed.
- Points representing observations (taxa) are ‘squeezed’ into 1D, 2D, or 3D… space.
  - Placed into space so that interpoint distances reconstruct original distances as much as possible.
    - Metric and nonmetric (monotonic) versions.
  - Axes are arbitrary, and points can be arbitrarily rotated.
    - Often rotated to principal-component solution.
  - Amount of distortion measured as ‘stress’.
- Observations (taxa) are projected onto the axes to give projection scores, which can be plotted with scattergrams.
Multidimensional scaling (MDS)

Distance matrix

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>1.41</td>
<td>2.45</td>
<td>2.45</td>
<td>2.83</td>
</tr>
<tr>
<td>2</td>
<td>1.41</td>
<td>0.00</td>
<td>2.45</td>
<td>3.16</td>
<td>2.45</td>
</tr>
<tr>
<td>3</td>
<td>2.45</td>
<td>2.45</td>
<td>0.00</td>
<td>1.41</td>
<td>1.41</td>
</tr>
<tr>
<td>4</td>
<td>2.45</td>
<td>3.16</td>
<td>1.41</td>
<td>0.00</td>
<td>2.45</td>
</tr>
<tr>
<td>5</td>
<td>2.83</td>
<td>2.45</td>
<td>1.41</td>
<td>2.45</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Additive

Stress = 0.27