Diversity indices and multivariate analysis of community structure

Community structure: species richness

- Problem with species richness (*S*):
  - Should only compared if based on the same sample size (area of habitat; time; number of individuals)
  - Abundance of species is ignored

S = 7

S = 7

Microsoft ClipArts

Community structure: species richness

- Margalef's index (*d*) incorporates the number of individuals (*N*)
- Measure for the number of species present for a given number of individuals

$$d = (S\text{-}1)/\log N$$

d = 2.3

d = 1.8

Microsoft ClipArts

Community structure: diversity

- Diversity indices try to combine both species richness and the eveness or equitability of the distribution among those species
- *Shannon diversity index* (*H*) takes into account species richness and relative abundance (also called *Shannon-Wiener* or *Shannon-Weaver diversity index;* sometimes *H'*)

$$H = -\sum_i p_i \cdot \ln p_i$$

- p_i = proportion of biomass or abundance of the *i*th species
- The product $p_i \cdot \ln p_i$ is calculated for each species and summed for all of them
- Sometimes the logarithm to the base of 2 (log) is used: **comparability!**
- For a given species richness, *H* increases with equitability
- For a given equitability, *H* increases with species richness

Community structure: diversity

- Shannon diversity index (*H*):

$$H = -\sum_i p_i \cdot \ln p_i$$

H = 1.9

H = 0.6

Microsoft ClipArts

Community structure: equitability

- Pilou's evenness index (*J*):

$$J = H/H_{max} = H/ln\ S$$

J = 1

J = 0.5

Microsoft ClipArts

Multivariate statistics: Similarity coefficient

- **Univariate indices** (e.g. diversity) have the disadvantage that the species composition of the studied community is not considered
- **Multivariate statistics** considers abundance (or biomass) and species composition
- The **similarity (*S*) of samples** is calculated based on a data matrix containing samples in columns and species in rows

Loch Linnhe macrofauna {L} subset. *(a) Abundance (unstransformed) for some selected species and years. (b) The resulting Bray-Curtis similarities between every pair of samples.*

(a) Year:	64	68	71	73
(Sample:	1	2	3	4)
Species				
Echinoca.	*9*	*0*	*0*	*0*
Myrioche.	*19*	*0*	*0*	*3*
Labidopl.	*9*	*37*	*0*	*10*
Amaeana	*0*	*12*	*144*	*9*
Capitella	*0*	*128*	*344*	*2*
Mytilus	*0*	*0*	*0*	*0*

(b) Sample	1	2	3	4
1	–			
2	8	–		
3	0	42	–	
4	39	21	4	–

Clarck & Warwick (2011)

Multivariate statistics: Similarity coefficient

- The data matrix could be arranged in the following ways:

1. Absolute numbers (/biomass/cover)
   - ➢ two samples are considered perfectly similar if they contain the same species in exactly the same abundance

2. Relative abundance (/biomass/cover)
   - ➢ standardisation to give the percentage of total abundance (over all species)
   - ➢ essential if, for example, differing and unknown volumes of sediment and water are sampled

3. Reduction to simple presence or absence of each species
   - ➢ e.g. when quantitative counts are unreliable due to artefacts

Multivariate statistics: Similarity coefficient

- A **similarity coefficient *S*** is conventionally defined to take values in the range 0-100 %
- *S* = 100 % if two sample are totally similar
- *S* = 0 if two samples are totally dissimilar

Loch Linnhe macrofauna {L} subset. *(a) Abundance (unstransformed) for some selected species and years. (b) The resulting Bray-Curtis similarities between every pair of samples.*

(a) Year:	64	68	71	73
(Sample:	1	2	3	4)
Species				
Echinoca.	*9*	*0*	*0*	*0*
Myrioche.	*19*	*0*	*0*	*3*
Labidopl.	*9*	*37*	*0*	*10*
Amaeana	*0*	*12*	*144*	*9*
Capitella	*0*	*128*	*344*	*2*
Mytilus	*0*	*0*	*0*	*0*

(b) Sample	1	2	3	4
1	–			
2	*8*	–		
3	*0*	*42*	–	
4	*39*	*21*	*4*	–

Clarck & Warwick (2011)

Multivariate statistics: Similarity coefficient

- Similarity matrices are the basis of many multivariate methods, e.g. cluster and ordination analysis and associated significance tests

- Similarity matrices can be used to:

➢discriminate sites (or times) from each other

➢cluster sites into groups that have similar communities

➢allow a gradation of sites to be represented graphically

- Analogue to the described **sample similarity matrix** one can also calculate a **species similarity matrix; species similarity (*S'*)**

- **Bray-Curtis coefficient** or **Bray-Curtis similarity**:

$$S_{jk} = 100\left\{1 - \frac{\sum_{i=1}^{p}|y_{ij} - y_{ik}|}{\sum_{i=1}^{p}(y_{ij} + y_{ik})}\right\} = 100\frac{\sum_{i=1}^{p}2\min(y_{ij}, y_{ik})}{\sum_{i=1}^{p}(y_{ij} + y_{ik})}$$

S_{jk}: similarity between the jth and kth samples
y_{ij}: entry in the ith row and jth column of the data matrix (i.e. abundance of the ith species in the jth sample)
y_{ik}: entry in the ith row and kth column of the data matrix (i.e. abundance of the ith species in the kth sample)

Multivariate statistics: Similarity coefficient

- **Bray-Curtis coefficient** or **Bray-Curtis similarity**:

$$S_{jk} = 100\left\{1 - \frac{\sum_{i=1}^{p}|y_{ij} - y_{ik}|}{\sum_{i=1}^{p}(y_{ij} + y_{ik})}\right\} \qquad S_{14} = 100\left\{1 - \frac{9+16+1+9+2+0}{9+22+19+9+2+0}\right\} = 39.3$$

Loch Linnhe macrofauna {L} subset. *(a) Abundance (unstransformed) for some selected species and years. (b) The resulting Bray-Curtis similarities between every pair of samples.*

(a) Year:	64	68	71	73
(Sample:	1	2	3	4)
Species				
Echinoca.	9	0	0	0
Myrioche.	19	0	0	3
Labidopl.	9	37	0	10
Amaeana	0	12	144	9
Capitella	0	128	344	2
Mytilus	0	0	0	0

(b)

Sample	1	2	3	4
1	–			
2	8	–		
3	0	42	–	
4	39	21	4	–

$$= 100\frac{\sum_{i=1}^{p} 2\min(y_{ij}, y_{ik})}{\sum_{i=1}^{p}(y_{ij} + y_{ik})}$$

$$S_{14} = 100\left\{\frac{2[0+3+9+0+0+0]}{9+22+19+9+2+0}\right\} = 39.3$$

Clarck & Warwick (2011)

Multivariate statistics: Similarity coefficient

- Transformation of data:
  - root: √*y*
  - 4th root: √ √*y*
  - log: log (1+*y*)
  - presence/absence

Loch Linnhe macrofauna {L} subset. *(a) Abundance (unstransformed) for some selected species and years. (b) The resulting Bray-Curtis similarities between every pair of samples.*

(a) Year:	64	68	71	73
(Sample:	1	2	3	4)
Species				
Echinoca.	9	0	0	0
Myrioche.	19	0	0	3
Labidopl.	9	37	0	10
Amaeana	0	12	144	9
Capitella	0	128	344	2
Mytilus	0	0	0	0

(b)

Sample	1	2	3	4
1	–			
2	8	–		
3	0	42	–	
4	39	21	4	–

Loch Linnhe macrofauna {L} subset. *(a) √√-transformed abundance for the four years and six species of Table 2.1. (b) Resulting Bray-Curtis similarity matrix.*

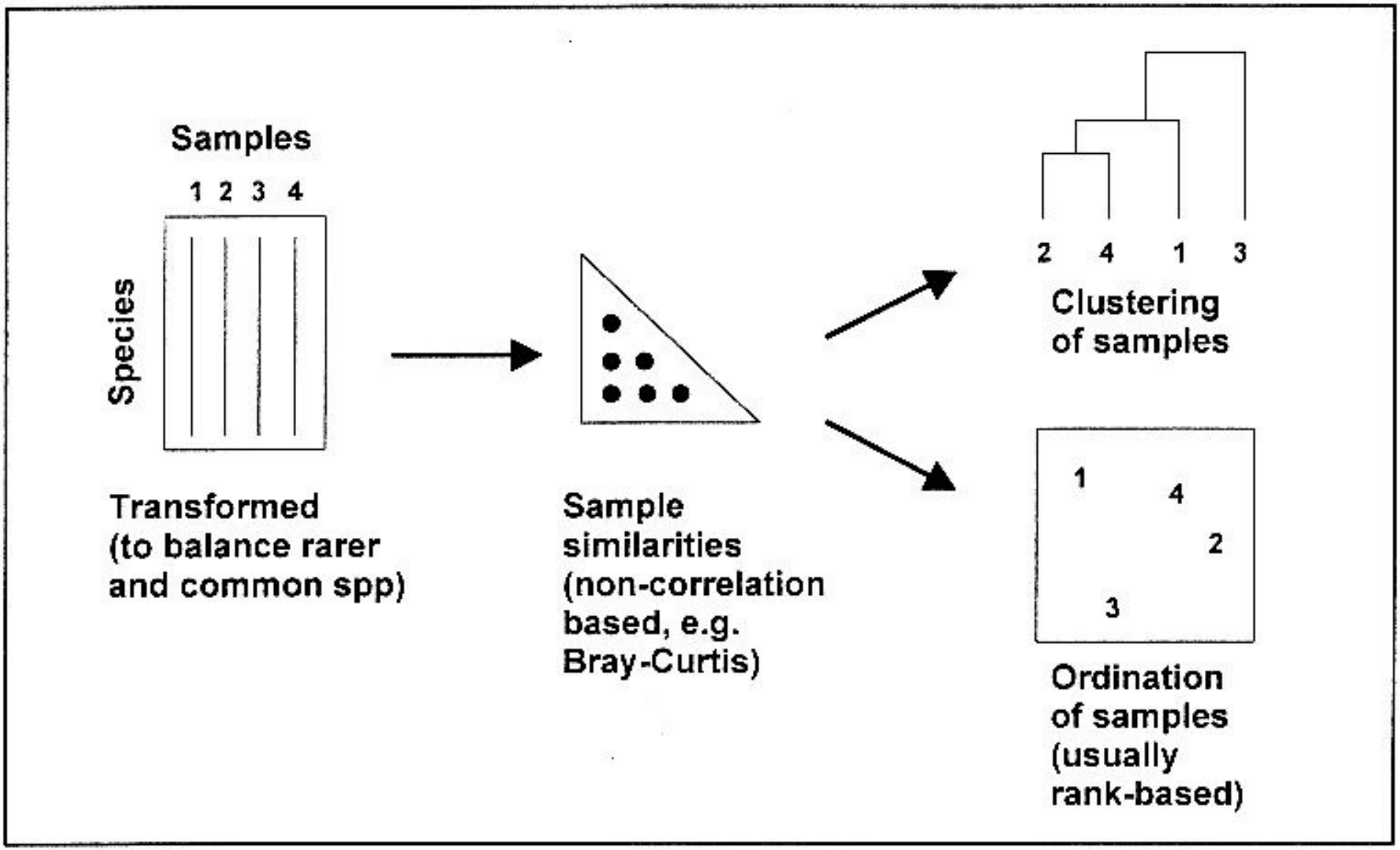
(a) Year:	64	68	71	73
(Sample:	1	2	3	4)
Species				
Echinoca.	1.7	0	0	0
Myrioche.	2.1	0	0	1.3
Labidopl.	1.7	2.5	0	1.8
Amaeana	0	1.9	3.5	1.7
Capitella	0	3.4	4.3	1.2
Mytilus	0	0	0	0

(b)

Sample	1	2	3	4
1	–			
2	26	–		
3	0	68	–	
4	52	68	42	–

Clarck & Warwick (2011)

Multivariate statistics: Similarity coefficient



Clarck & Warwick (2011)

Multivariate statistics: Cluster analysis

- Cluster analysis (or classification) aims to find “natural groupings” of samples

- Samples within a group are more similar to each other than samples in different groups

- Cluster analysis is used in the following ways:

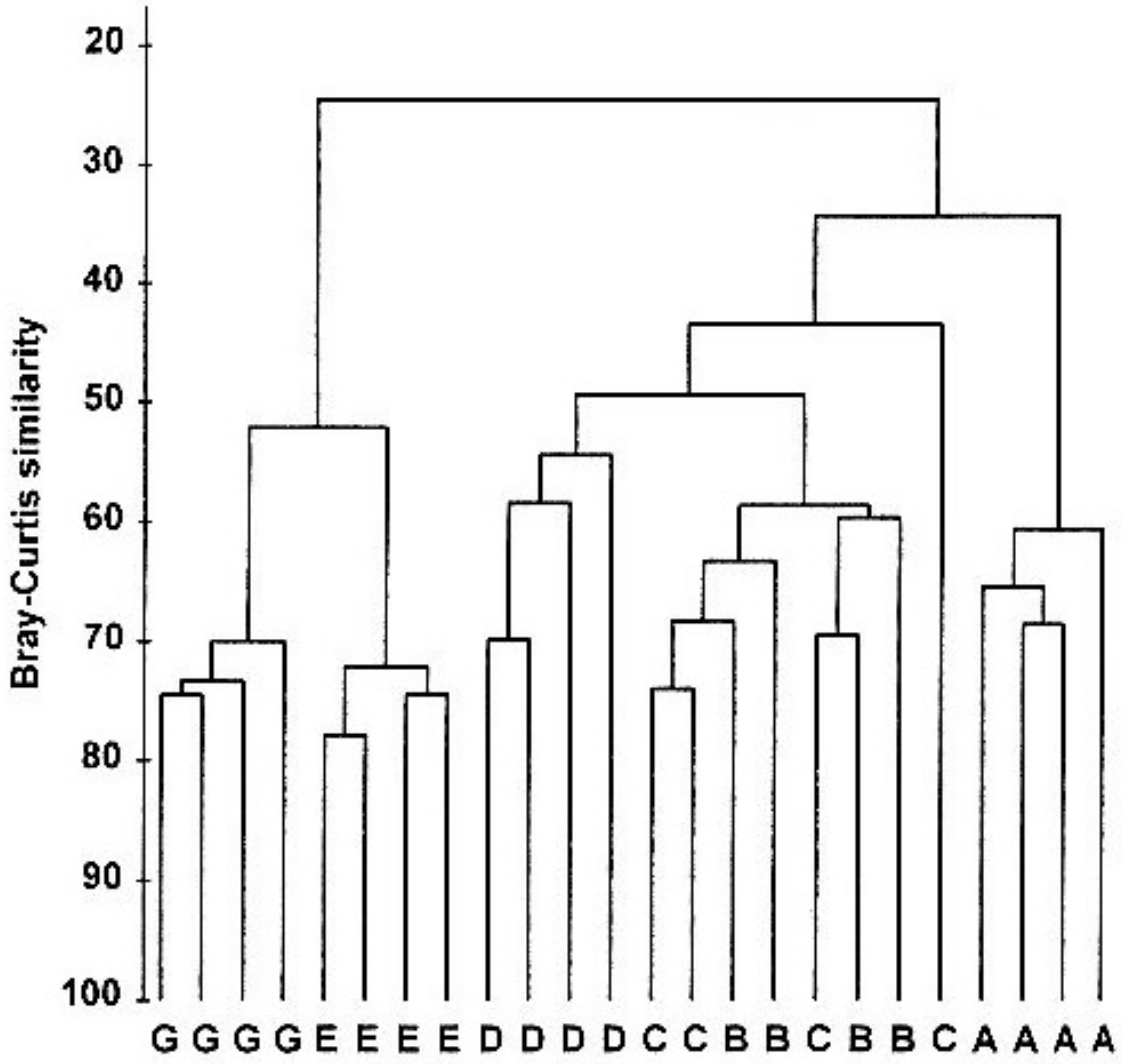
➢ Different sites (or times at the same site) can be seen to have differing community compositions by noting that replicate samples within a site form a cluster that is distinct from replicates within other sites

➢ When it is established that sites can be distinguished from one another, sites can be partitioned into groups with similar community structure

➢ Cluster analysis of the species similarities matrix can be used to define species assemblages, i.e. groups of species that tend to co-occur

Multivariate statistics: Cluster analysis

- There are many different clustering methods, but the most commonly used are **hierarchical agglomerative clustering methods**
- Similarity matrix as starting point and successively fuses the samples into groups and the groups into larger groups
- The result is presented in a tree diagram or **dendrogram**, with the x axis representing the full set of samples and the y axis defining the similarity level



Clarck & Warwick (2011)

Multivariate statistics: Cluster analysis

***Loch Linnhe macrofauna {L} subset**. Abundance array after √√-transform, the resulting Bray-Curtis similarity matrix and the successively fused similarity matrices from a hierarchical clustering, using group average linking.*

Year:	64	68	71	73
Sample:	1	2	3	4
Species				
Echinoca.	1.7	0	0	0
Myrioche.	2.1	0	0	1.3
Labidopl.	1.7	2.5	0	1.8
Amaeana	0	1.9	3.5	1.7
Capitella	0	3.4	4.3	1.2
Mytilus	0	0	0	0

→

Sample	1	2	3	4
1	–			
2	25.6	–		
3	0.0	67.9	–	
4	52.2	68.1	42.0	–

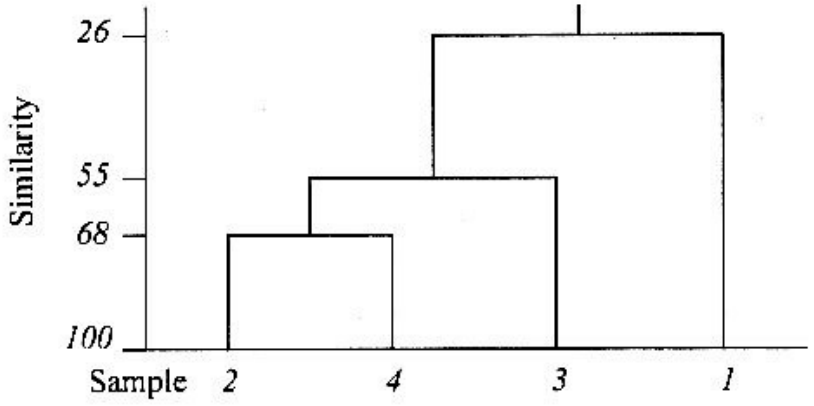
→

Sample	1	2&4	3
1	–		
2&4	38.9	–	
3	0.0	55.0	–

→

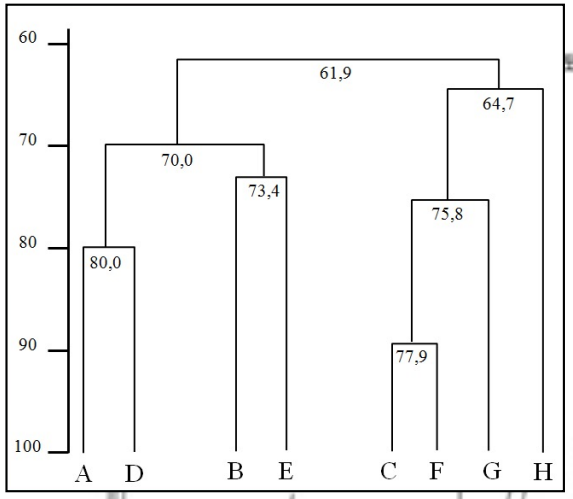
Sample	1	2&3&4
1	–	
2&3&4	25.9	–

- **Single linkage**: *S* (1, 2&4) is the maximum of *S* (1, 2) and *S* (1, 4) = 52.2 %
- **Complete linkage**: *S* (1, 2&4) is the minimum of *S* (1, 2) and *S* (1, 4) = 25.6 %
- **Group-average link**: *S* (1, 2&4) is the average of *S* (1, 2) and *S* (1, 4) = 38.9 %

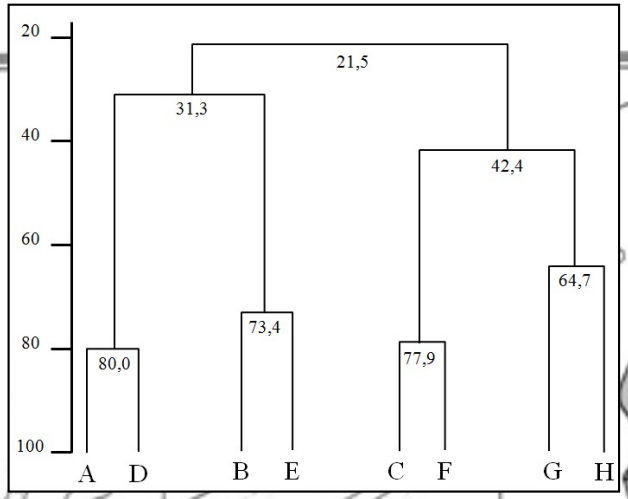


Clarck & Warwick (2011)

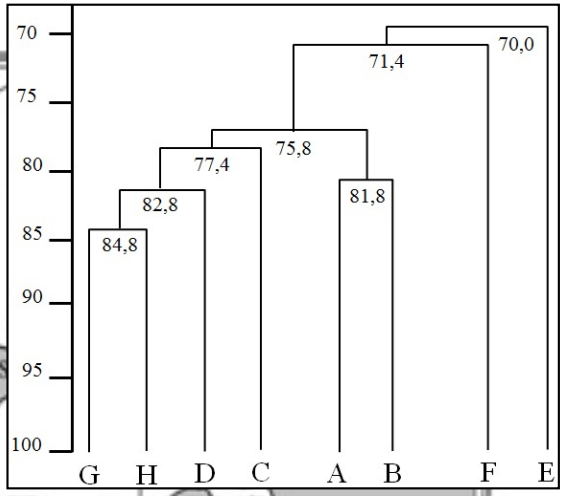
Multivariate statistics: Cluster analysis of North Sea benthos data



Untransformed data; single linkage

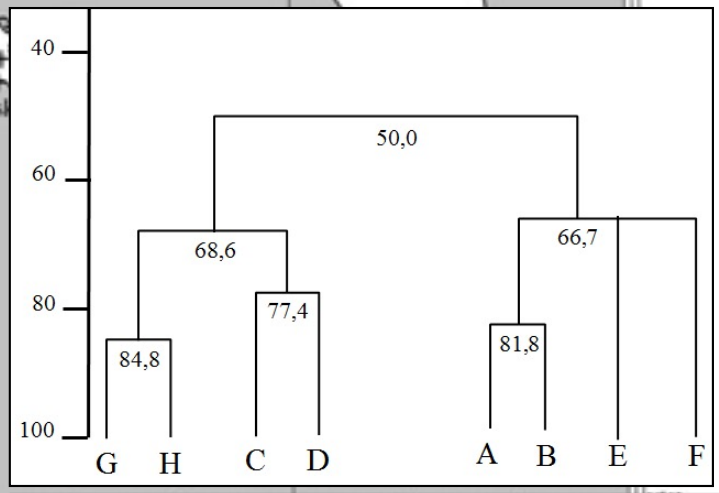


Untransformed data; complete linkage



Transformed data; single linkage

A = 230/1-2011
B = 230/2-2011
C = 120-2011
D = ZG03-2011
E = 230/1-2012
F = 230/2-2012
G = 120-2012
H =ZG03-2012



Transformed data; complete linkage

References

Clarke KR, Warwick RM (2001) Changes in marine communities: an approach to statistical analysis and interpretation. PRIMER-E, Plymouth