**Cluster diagrams**

**QUESTION 2003**

A physical geographer is contrasting beach profiles at ten sites around a coral atoll. The following data were collected from each site:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Site | mean sediment size (mm) | beach slope (degrees) | mean wave height (cm) | mean water depth (cm) | mean current speed (m.s-1) |
| 1 | 0.9 | 10.1 | 251 | 307 | 1 |
| 2 | 0.8 | 9.0 | 235 | 278 | 1.5 |
| 3 | 4 | 7.6 | 283 | 326 | 0.9 |
| 4 | 3 | 8.2 | 205 | 251 | 0.9 |
| 5 | 2 | 8.5 | 185 | 310 | 1.7 |
| 6 | 0.05 | 1.2 | 68 | 106 | 0.5 |
| 7 | 0.1 | 2.3 | 84 | 124 | 0.2 |
| 8 | 0.08 | 3.6 | 93 | 151 | 0.4 |
| 9 | 0.09 | 4.1 | 87 | 149 | 0.6 |
| 10 | 0.1 | 3.7 | 62 | 179 | 0.1 |

The following are dendrograms using untransformed data, a correlation coefficient as the similarity index, and the complete linkage method of tree construction.

1. What would be the value of this similarity coefficient if two samples were identical? (1 mark)
2. Should this data have been transformed or standardised? (4 marks)
3. Is a distance-based similarity index (e.g., Euclidean distance) preferable to a correlation coefficient for this sort of data? (5 marks)
4. To prevent erosion, the resort on the island is developing a management plan for its beaches. How would you explain the results of this cluster analysis to the resort owners? How would these results guide management decisions? (10 marks)

**QUESTION 5: 2006**

# An ecologist surveyed three rivers for the presence of fish. The following table is the abundance of the four species per 100 m of river length at each of the three rivers.

|  |  |  |  |
| --- | --- | --- | --- |
| Species | River 1 | River 2 | River 3 |
| *Galaxias vulgaris* | 5 | 10 | 0 |
| *Galaxias anomolus* | 4 | 12 | 0 |
| *Galaxias* sp. | 3 | 5 | 0 |
| *Salmo trutta* | 1 | 0 | 25 |

# Produce a cluster diagram to represent the relationships among samples. Use single linkage and Euclidean distances (formula below). Show all the steps in your working. (10 marks)

b) What is the biological interpretation of the cluster diagram? (4 marks)

c) If you found out that river 2 was twice as wide as rivers 1 and 3, how might you standardise the data? (3 marks)

d) Is Euclidean distance the best choice of similarity/dissimilarity index for this sort of data? If not, why not? (3 marks)

**QUESTION 9: 2002**

A botanist collected the following physical data from ten of her study sites in eastern Australia.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Site | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Temperature (°C) | 35 | 32 | 30 | 28 | 27 | 15 | 25 | 23 | 26 | 22 |
| Hours of sunlight | 2.5 | 3.5 | 3.6 | 2.5 | 1 | 5 | 4 | 3.5 | 4 | 1.5 |
| Soil moisture (%) | 5 | 4 | 1 | 8 | 4 | 6 | 3 | 10 | 2 | 7 |
| Altitude (m) | 150 | 130 | 125 | 95 | 10 | 50 | 25 | 75 | 5 | 100 |
| Rainfall (mm) | 650 | 630 | 550 | 585 | 635 | 800 | 850 | 750 | 600 | 550 |

Two cluster analyses were performed to group sites. The data were untransformed and the similarity indices used were Euclidean distance, and the Pearson correlation coefficient (r).

0

50

100

150

Euclidean distance

1

2

3

4

5

6

7

8

9

10

0.00

0.01

0.02

0.03

1-r

1

2

3

4

5

6

7

8

9

10

1. Using the diagram produced using Euclidean distances, describe how site 7 is related to the other sites based on the physical variables measured (5 marks).
2. Which cluster diagram groups sites more effectively? Why? (5 marks)
3. Which distance measure is more appropriate for these data? Why? (5 marks)
4. Based on the cluster analysis produced used Euclidean distances, the botanist proceeded to design an experiment to test the effects of climate on the abundance of a rare plant. She divided her sites into two groups; group A with sites 6,7 and 8, and group B with the remaining sites. The abundance data were analysed using ANOVA with climate as a factor (two levels: group A and group B). Comment on this approach. (5 marks).