Worksheet 6 - Multivariate analysis

> **Multivariate analysis references**
>
> - Fowler *et al.* (1998) -Chpts 18
> - Holmes *et al.* (2006) - not covered
> - Quinn & Keough (2002) - Chpts 15 & 18

Question 1 - Dissimilarity

The following data are the abundances of 3 species of gastropods in 5 quadrats (ranging from high shore marsh, Quadrat 1, to low shore marsh, Quadrat 5) in a saltmarsh.

> **Format of gastropod.csv data file**
>
> | | Salinator | Ophicardelus | Marinula |
> |------|-----------|--------------|----------|
> | Q1 | 4 | 0 | 1 |
> | Q2 | 9 | 3 | 0 |
> | Q3 | 9 | 4 | 1 |
> | Q4 | 6 | 2 | 0 |
> | Q5 | 0 | 1 | 1 |
>
> **Salinator**   Number of Salinator gastropods - variable
>
> **Ophicardelus**   Number of Ophicardelus gastropods - variable
>
> **Marinula**   Number of Marinula gastropods - variable
>
> **Q1-Q5**   Quadrats - these are the samples (sites)
>
>

Q1-1. By hand, calculate the **Bray-Curtis(Czekanowski) dissimilarity coefficient** and **Euclidean distance** between all pairs of quadrats.

- **a.** Fill in the matrix below. Note that the lower left-hand section of a dissimilarity, correlation, distance, etc matrix is mirrored in the upper right-hand section and thus can be represented by a triangular matrix. To save space and assist comparisons, fill the lower left section with Bray-Curtis dissimilarity values and the upper right with Euclidean distances.

	Q1	Q2	Q3	Q4	Q5
Q1	0.00	☐	☐	☐	☐
Q2	☐	0.00	☐	☐	☐
Q3	☐	☐	0.00	☐	☐
Q4	☐	☐	☐	0.00	☐
Q5	☐	☐	☐	☐	0.00

b. Do the two measures correspond? In what ways are they similar/different?

Open the gastropod data file. Note the format of the file, with variables in columns and samples/sites in rows.
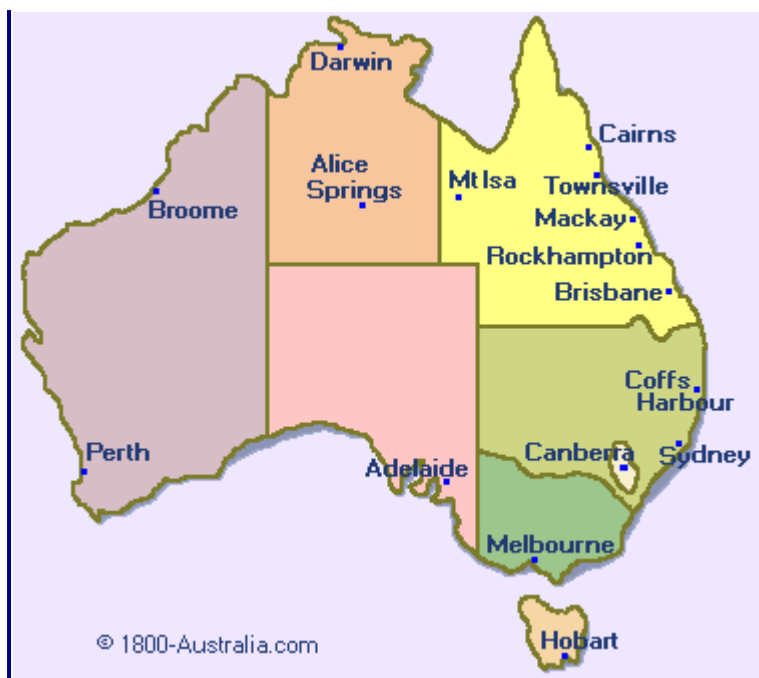
Q1-1. Now lets use R to calculate separately the **Bray-Curtis(Czekanowski) dissimilarity coefficients and Euclidean distances**.

Question 2 - Multidimensional scaling

The following example is designed to help you appreciate the link between distance measures and ordination space (MDS). The data set consists of distances (km) between major Australia cities (as the crow flies), and is in the form of a triangular matrix.

Format of austcities.csv data file

	Canberra	Sydney	Melbourne	..
Canberra	0	NA	NA	..
Sydney	246	0	NA	..
Melbourne	467	713	0	..
Adelaide	958	1160	653	..
Perth	3090	3290	2720	..
..	..	..	..	..

© 1800-Australia.com

Open the austcities data file. Note the format of the file, it is a triangular distance matrix.

While the file is a distance matrix, at this stage R is unaware of it, we must manually make it aware (a round about way of saying that we must type a command to force R to treat the data set as a distance matrix. We do this by typing the following command in the top window (Script Window) of R Commander (actually we could alternatively type it on the command prompt of RGui).

```
aust.cities.dis <- as.dist(aust.cities)
```

where `aust.cities` is the name you gave the Australian cities data set when you imported it, `as.dist` is the name of the function that forces a data set to be considered a distance matrix, and `aust.cities.dis` is a new name that I am defining to store the distance matrix.

If you chose to enter the statement in the R Commander command log window, once you have completed typing the above statement, press the **Submit** button in the top right hand corner of R Commander. This will case the statement to be evaluated and the task to be completed.

We are now ready to perform the MDS for the purpose of examining the ordination plot.

Q2-1. Perform an **MDS** with 2 dimensions on the city distances matrix.

    a. What was the final stress value? _____ ☐

    b. What does this stress value suggest about the success of the MDS?

    c. The Sheppard diagram (plot) represents the relationship between the original distances (y-axis) and the new MDS distances (x-axis). Does this and the stress value indicate that the patterns present in the original distance matrix (crow flies distances between cities) are adequately reproduced from the 2 new dimensions? ☐

    d. The final ordination plot summarizes the relationship between the cities. Does this ordination plot approximate the true geographical arrangement of the cities? _____
    ☐

    e. In this case, what might the two new MDS dimensions (variables) represent? (hint think of the ordination plot as a map)

Question 3 - Multidimensional scaling

A fish ecologist investigated differences in fish community composition associated with flow regulation in rivers of four major New South Wales catchements (Murray, Darling, North Coast and South Coast). A subset of the community data is presented in the file riverfish.csv.

Format of riverfish.csv data file

	H.comp	G.mac	M.flu	..
Darling Unreg	0	0	62	..
Darling Reg	0	0	42	..
Murray Unreg	0	0	0	..
Murray Reg	0	0	2	..
North Coast Reg	1657	0	0	..
..	..	..	..	..

Open the riverfish data file.

Q3-1. Examine the whole data set. Before proceeding with any formal analysis, lets explore the patterns in the data visually.

   a. Based solely on *Hypseleotris compressa* which sites are most similar?

   b. What do the abundances of *Nematalosa.erebi* suggest about the sites?

Of course, we could explore the patterns amongst sites according to each separate fish species. However, in the full investigation, the fish ecologist had 51 species of freshwater fish, each of which yield slightly different patterns. Furthermore, the fish ecologist was not interested in the patterns of any one species of fish. What he was interested in was whether (and how) the fish communities differed between catchments and flow regimes. Consequently, an MDS was used to explore the patterns amongst sites based on all the fish species.

Q3-2. Calculate the **Bray-Curtis(Czekanowski) dissimilarity coefficients** amongst the sites using all six fish species (note we are only using a subset of the entire data set).

   a. Which sites are most similar?

Q3-3. Perform an **MDS** with 2 dimensions on the dissimilarity matrix.

  a. What was the final stress value? ☐

  b. What does this stress value suggest about the success of the MDS?

  c. The Sheppard diagram (plot) represents the relationship between the original distances (y-axis) and the new MDS distances (x-axis). How would you describe the shape of this curve, and base on this is metric or non-metric MDS more appropriate?

  d. The final ordination plot summarizes the relationship between the sites. What would you conclude from this?

Question 4 - Multidimensional scaling

Ludwig and Reynolds (1988) described a data set in which a number of sites where characterized by the abundances of five species of cockroach.

<div style="border:1px solid navy">

Format of Ludwig.csv data file

	C.spp	C.cucullatus	C.delicatulus	..
BCI	0	14	28	..
LC	0	4	0	..
FORT	0	1	1	..
BOQ	0	1	0	..
MIR	0	4	1	..
CORG	1	0	0	..

</div>

Open the Ludwig data file.

Q4-1. Examine the whole data set. Are there any patterns

Q4-2. Calculate the **Bray-Curtis(Czekanowski) dissimilarity coefficients** amongst the sites using all cockroach species.

    a. Which sites are most similar?

    _____

Q4-3. Perform an **MDS** with 2 dimensions on the dissimilarity matrix.

    a. What was the final stress value? _____ ☐

    b. What does this stress value suggest about the success of the MDS?

    c. The Sheppard diagram (plot) represents the relationship between the original distances (y-axis) and the new MDS distances (x-axis). How would you describe the shape of this curve, and base on this is metric or non-metric MDS more appropriate?

    _____

    d. The final ordination plot summarizes the relationship between the sites. What would you conclude from this?

Welcome to the end of Worksheet 6