

Euclidean distances

Euclidean distance, an appropriate measure for environmental (and other) data types, is defined as:

$$D_1 = \sqrt{\sum_i \left(y_{i1} - y_{i2} \right)^2}$$

where the y_{i1} & y_{i2} result from pre-treatment by transformation (sometimes) and subsequent normalisation (often). The outcome is a triangular distance matrix, which orders in the opposite direction to similarity: high similarity = low distance (= low dissimilarity). Note, however, that the user does not have to worry about which way round the resemblances are ordered: all routines will utilise the information given in the Resemblance type to make sensible choices.

Re-open the Ekofisk workspace **Ekofisk ws** from the \Ekofisk macrofauna directory; you should have available the transformed and normalised environmental data (**Data4** perhaps) from Section 4, on which to calculate Euclidean distance. The **Analyse>Resemblance** dialog box now gives the default as Measure•Euclidean because Data type has been defined as Environmental, so you can take the defaults here. The result is a resemblance matrix of type Distance; the History box on the **Edit>Properties** dialog shows its derivation as Euclidean distance on normalised data. Compute Manhattan distance also (see next page) and rename the sheets as **Euclid** and **Manhattan** by clicking (twice, slowly) on their default *Resem names* in the Explorer tree. Most other measures in the lists below are not suitable for normalised environmental data but are designed for positive 'quantities'.

Revision #4

Created 4 June 2024 00:17:24 by Arden

Updated 15 January 2025 01:27:04 by Abby Miller