CLUSTER analyses

The Basic MVA wizards then run a cluster analysis, again on the respective resemblance matrices. This component routine is **Analyse>Cluster>CLUSTER**>(Cluster mode•Group average), without taking the (✓SIMPROF test) option since the latter is the appropriate test (rather than ANOSIM) when an *a priori* group structure is not defined. That is, if we had chosen to ignore the structure of sites within 5 creeks and simply treated the 27 samples as just 27 Fal estuary locations, the primary thrust of the analysis would not have been the ANOSIM tests and MDS display (see below) of those creek groups. Instead, it would have been a more exploratory analysis of whether the sites fell into clusters of similar communities (or environmental variables) at all – and, if so, which sites constituted those groups. The SIMPROF test is then important in deciding which sub-clusters in the hierarchical group-average cluster analysis (UPGMA) we are entitled to interpret as distinguishable groups, statistically – and, if we did not tick the ✓ANOSIM (1-way) box in the **Basic multivariate analysis** wizard, it would instead run a series of SIMPROF tests on the nodes of the cluster analysis dendrogram (Section 6) to determine this. As it is, the clustering in this Fal example is secondary and *Graph2* simply displays the dendrogram of the 27 sites, without SIMPROF tests. However, it is interesting to note that the dendrogram does largely divide the 27 samples into the 5 creeks, with an exception or two, which is consistent with the clear distinction among creeks seen in ANOSIM. You might like to accentuate this point by **Graph>Sample Labels & Symbols**>(Symbols✓Plot)> (✓By factor Creek) and look also at **Graph>Special** options, e.g. re-orienting the dendrogram.

PRIMER has other clustering tools (Section 6): a hierarchical binary divisive cluster analysis in unconstrained, **Analyse>Cluster>UNCTREE**, or constrained form, **>LINKTREE** (in which only divisions which have an 'explanation' in terms of a threshold on an environmental variable, say, are permitted). Both these methods share a common structure, consistent with the non-parametric treatment of resemblance matrices (which applies to tests such as ANOSIM, RELATE, BEST and ordinations such as non-metric MDS etc), namely each group is successively sub-divided so as to maximise the ANOSIM R statistic (PRIMER's key measure of group separation in multivariate space) between the two groups formed. A further non-hierarchical clustering method is available in the **Analyse>Cluster>kRCLUSTER** routine, a generalisation of classical *k-means* clustering to any resemblance matrix but again using only ranks. SIMPROF tests can be applied to all methods.

Revision #7
Created 12 March 2024 22:34:17 by Arden
Updated 7 January 2025 23:14:45 by Abby Miller