PC scores

The final table in the results window is headed *Principal Component Scores* – these can instead be sent to a new worksheet by checking (✓Scores to worksheet) in the **Analyse>PCA** dialog, which facilitates their further use in PRIMER. An example would be to compute Euclidean distances among sites in PC spaces of different dimension, which could then be input to **Analyse>RELATE** (Section 14) to give a matrix correlation with the original 11-d Euclidean distances. (This is another way of measuring the fidelity of the observed low-d ordination structure to the high-d relationships, an idea we met as *cophenetic correlation* in Section 6 on the fidelity of cluster analyses, and such matrix correlations are fundamental within PRIMER, met especially in the next two sections). The PC scores are simply the *x*, *y* (or *x*, *y*, *z* etc.) co-ordinates of the samples on the PCA plot – their values on each PC, obtained by substituting the (normalised) variable values into the above linear equations for PC1, PC2, etc. It is the ability to generate a numerical score for a fresh set of values for the same suite of variables which is one of the strengths of PCA. If values from a new site *a* are recorded as (Cu_a, Mn_a, Co_a, …) you can see where it fits on the contaminant scale by calculating:

$$ \text{PC1} = 0.378 \left\{ [(ln \mathit{Cu}_a) – 4.046]/0.924 \right\} – 0.213 \left\{ [(ln \mathit{Mn}_a) – 6.062]/0.757 \right\} + … $$

where the means and standard deviations used in the normalisations were given in the *Mean & SD* worksheet from the normalising of the original logged data set (see output on the previous page). This is the main downside to using rank variables in a PCA, which on other grounds has much going for it – it is harder to relate new sites to the PCA from the original set of samples.

Increasing the default value of (Maximum number of PCs: 5) when running **Analyse>PCA** will print more columns of PC vectors in the results window (PC6, PC7, etc.), and will allow selection of these higher PCs to be plotted in pairs or triples in the 2-d or 3-d PC configuration. However, it is rarely helpful to interpret more than the first 3 or 4 PCs, so the default computation of the first 5 is usually perfectly adequate. It is important to note that nothing changes at all in the first 5 sets of vectors if it is decided to calculate axes 6 to 10, say. Each lower-d configuration is a projection from the higher-d solution, which therefore just involves dropping out the higher axes. This is not true of MDS ordination, for which the 2-d solution is recalculated from scratch, and not just the first two dimensions of the 3-d solution.

Revision #13
Created 27 September 2024 22:19:42 by Arden
Updated 12 February 2025 00:39:48 by Abby Miller