Environmental data

For environmental-type data, such as the Fal environment sheet, it is often appropriate to transform individual variables selectively, rather than all in the same way, since they may be of very disparate types. Here, the main objective is to avoid strong skewness in the distribution over samples, since large outliers will dominate both computation of (normalised) Euclidean distances and the Principal Component Analysis (**Analyse>PCA**), which is often the multivariate analyses chosen for abiotic data. The degree of skewness, or presence of outliers, is visually assessed using **Plots>Histogram Plot** or **Plots>Draftsman Plot** on active sheet Fal environment (you may wish to increase symbol size on the draftsman plot – do this by **Graph>Sample Labels & Symbols** and Size: 150, say). If there is strong right-skewness, those variables might need a log transform by highlighting them and taking **Pre-treatment>Transform (individual)**>(Expression: log(V+1)), Section 4. Alternatively take the rank transform, **Tools>Rank Variables**, which certainly gets rid of outliers! Although there is skewness here, there are no strong outliers and, for this demo, omit any transformation. So, run **Wizards>Basic multivariate analysis** on Fal environment and take all the defaults, examining the different choices made for this environmental-type matrix (e.g. normalising variables onto a common dimensionless scale; Euclidean distance resemblance; PCA ordination, see Section 12).

Revision #3
Created 12 March 2024 22:34:16 by Arden
Updated 7 January 2025 23:01:26 by Abby Miller