Average and Sum on data matrices; Average on resemblance matrices

Tools>Average and **Sum** operate in the same way on data sheets. For example, when (Samples• Averages for factor: area) & (Variables•No averaging) is selected, they average (or sum) across all samples with the same level of the specified factor, separately for each variable (species), here creating a derived data sheet of averaged (or totalled) communities for each area, which can be input into the same multivariate analysis options as the original matrix. Averages are taken for the specified factor (not across it), e.g. if the above set of 277 locations (identified by a factor *site*) had been sampled at several *times*, then **Tools>Average** for factor *site* gives time-averaged site means. All factors in the original matrix are taken across to the new sheet, and factors such as *area* would still be well-defined over the 277 sites. (However, if the averaging had been for *times*, across all the sites, then the *area* column in the Factor sheet would consist only of Undefined! entries, since the averaging has mixed different areas). If the number of sites in each of the 9 areas is balanced then **Average** and **Sum** leads to the same ordination because the sheets differ only by a constant factor – most resemblance measures are unaltered by an overall scale change. If replication is unbalanced, however, then it is unwise to use **Sum**, because the outcome (using Bray-Curtis at least) would be sensitive to the different total abundances from the differing group sizes – **Average** is preferable.

A less common option is, for example, **Tools>Sum**>(Samples•No summing) & (Variables•Sums for indicator: class#) which would retain all 277 samples but total the matrix over the species to give just two new variables, class 1 (Chondrichthyes) and 2 (Osteichthyes). Pooling abundances to higher taxonomic levels is quite a common requirement but this is more naturally achieved with the **Tools>Aggregate** routine, discussed below. It is possible to **Sum** (or **Average**, though that is very unlikely) on both the axes, e.g. (Samples•Sums for factor: area) & (Variables•Sums for indicator: class#) would give a 2$\times$9 sheet of totals of each of the 2 classes in each of the 9 areas.

The main difference between **Tools>Average** or **Sum** and the **Analyse>Summary Stats** routine, new in PRIMER 7, is that the former computes means or sums within groups of samples (and/or variables) whereas **Summary Stats** will calculate these (and several other) summary statistics only over the full set of samples or variables (and in succession, not both at once, if both are required).

In PRIMER 7, the averaging facility extends to resemblance matrices: **Tools>Average**>(Factor/ indicator for groups: area) takes the average, e.g. for area 1 and 2, of all resemblances between pairs of samples, the first in area 1 and the second in area 2. It does this for all pairs of areas, thus giving a (9 $\times$ 9) triangular matrix of area resemblances. (These are the values at the head of each SIMPER table, defining dissimilarities between pairs of groups, which are then broken down into species contributions, Section 10 – but they are now more conveniently held in resemblance form). As the dialog implies, the averaging could also take place on variable resemblances, if groups are defined over those – perhaps coherent species groups from Type 3 SIMPROF tests

(Section 10).

Revision #4
Created 21 August 2024 21:16:24 by Arden
Updated 5 February 2025 05:02:01 by Abby Miller