

Standardising samples

How the data are treated, prior to computation of a resemblance matrix (e.g. similarities), can have an important influence on the final analysis, and such decisions often depend on the practical context rather than any statistical considerations. For example, standardising the samples (by total) divides each entry in the data sheet by the total abundance in that sample, across all variables (species). This would turn assemblage counts for each sample into relative percentages (what is referred to by statisticians as compositional data), all samples then adding to 100% across species. It thus removes all differences in total abundance in each sample from the multivariate comparison of samples. Sometimes this may be desirable, e.g. where the unit of sampling cannot be tightly controlled. An example is the data we have just been working with (on W Australian fish diets), analysing the prey taxa in the gut contents of fish predators: the quantity of food in the gut varies across individual fish in an uncontrollable way so is not relevant to a multivariate comparison of the prey composition, and the data should initially be sample-standardised. On the other hand, a typical marine impact study, using sediment-dwelling fauna sampled by a corer of fixed size, more strictly controls the quantity of material in each sample. It might then be important to use the fact that a potentially impacted site contains 5 times fewer individuals, in total, than a control site, so sample standardisation would be undesirable. The philosophy in PRIMER 7 is that users control all such pre-treatment decisions, combining them in an order under their choice, appropriate to the context. Each pre-treatment step results in display of a revised datasheet so the user can see its effect, before proceeding to analysis (or in some cases a further pre-treatment step).

Re-open the workspace **WA fish ws** from the directory C:\Examples v7\WA fish diets, or if not previously saved, **File>Open>Filename: WA fish diets %vol.pri**. (Note that if you had a selection in place at the time the workspace was saved, this will still be operational. You can leave this on or deselect it with **Select>All**, but it might make sense to leave samples A9, B3 and B4 excluded, because of their very low sample totals – gut fullness <<10% – and thus unreliable % composition after standardising). Take **Pre-treatment>Standardise>(Standardise•Samples) & (By•Total) & (✓Stats to worksheet)**. You will see from the resulting sheet (probably named Data3) that samples are now expressed as % composition of each prey category, the columns adding to 100.

The screenshot displays the PRIMER 7 software interface. The 'Analyse' menu is open, showing options like 'Transform(overall)...', 'Transform(individual)...', 'Standardise...', 'Cumulate Samples...', 'Weight Variables...', 'Dispersion Weighting...', 'Variability Weighting...', and 'Normalise Variables'. The 'Standardise' dialog box is open, with 'Samples' selected under 'Standardise' and 'Total' selected under 'By'. The 'Stats to worksheet' checkbox is checked. The 'OK' button is highlighted.

In the background, a data table titled 'Diets of 7 nearshore fish species from WA' is visible. It shows biomass data for various dietary categories across different samples.

	B1	B2	B5	B6	B7
Nematoda	0.44248	4.7499	0	0	0
Oligochaeta	0	0	0	0	0
Combined polychaeta	48.357	31.526	85.484	43.373	
Calanoid	0	0	0	0	0
Harpacticoid	51.201	54.897	0	0	0

Another table titled 'Other Variables' is also visible, showing totals for various samples.

	Total
B1	63.28
B2	47.58
B5	18.6
B6	16.6
B7	28.1
B8	61.14
B9	29.78
B10	33.54
E1	40
E2	78
E3	19
E4	41

Revision #1

Created 21 May 2024 22:15:22 by Arden

Updated 21 May 2024 22:22:50 by Arden