Variability weighting

Pre-treatment>Variability Weighting is a new option in PRIMER 7, which bears similarities to the idea of Dispersion Weighting. This was introduced by Hallett CS, Valesini FJ, Clarke KR 2012, *Ecol Indicat* 19: 240-252 in a context where the variables were 'health indices' of fish communities and is exemplified here in a comparable case of a 'biomarker' suite, measured on individual fish, from locations with putatively differing contaminant impacts. Such indices typically behave more like environmental-type variables, with differing measurement scales and without the presence/ absence structure of community matrices, so that transformation, normalisation and then a distance resemblance measure (e.g. Euclidean) would be appropriate (see earlier this section). The downside of normalising is, however, that all variables are essentially given equal weight in that calculation – but how else can one sum variable contributions over different units other than to shrink or stretch their scales to a common 'spread' (SD of 1)? (The location shift involved in normalising is actually irrelevant as far as distance measures such as Euclidean or Manhattan are concerned, because they are a function only of differences between sample values for each variable, so a subtracted constant disappears – the key thing is only the scale change to each variable). One possible answer is to scale each variable to a common spread (e.g. SD of 1) of its replicates within groups, not the full set of values across the groups (where the groups are the combinations of site, time etc.). As with Dispersion Weighting, the idea is that some indices may be inherently less reliable than others, with erratic values for genuinely independent replicate observations within groups, so that it is desirable to give more weight to variables with lower (average) replicate variability. The variables now, though, are no longer 'quantities' – indeed after some transformations (e.g. log) they may take negative values, the mean may even be zero and dividing by the Index of Dispersion (ratio of variance to mean) will make no sense. Instead, the **Variability Weighting** dialog offers a range of possible rescalings of replicates, by: •Pooled SD (as would be calculated from 1-way ANOVA, by square-rooting the residual variance estimate, the logically best option for normally distributed variables with common replicate variance across groups); •Averaged SD (a simple mean of SDs computed separately for each group); •Averaged range (mean of the separate ranges – if used with Manhattan distance this is a more subtle version of the Gower coefficient, see Section 5); and •Averaged IQ range (mean of the inter-quartile ranges for each group, potentially a more relevant spread measure than SD for non-normally distributed – but continuous – replicate observations). As with Dispersion Weighting, all samples for each variable are simply divided through by their own averaged replicate 'spread', a new sheet formed and the divisors given in a Results window.

Revision #3
Created 22 May 2024 01:05:54 by Arden
Updated 15 January 2025 00:42:50 by Abby Miller