Dispersion weighting of species

When variables are on different measurement scales, there is little viable alternative to normalising each variable (as above) thus equalising, in effect, their contributions to the multivariate analysis. When variables are (ostensibly) on the same scale, e.g. species abundances, then their respective contributions to commonly-used similarity coefficients, such as Bray-Curtis, will differ, based on the relative magnitude of counts (or transformed counts). Larger abundances are always given more weight (unless 'transformed out' to purely presence/absence). This may not always be desirable, however. For example, some numerically dominant species may give highly erratic counts over replicate samples within a site (or time or condition), perhaps due to an innately high degree of spatial clumping of individuals (individuals of that species arrive in the sample in clusters). This is likely to add 'noise' rather than 'signal' to the multivariate analysis, and downweighting of such species is called for, in relation to other species which are not spatially clustered, but have the lower variance associated with Poisson counts (the individuals arrive in the sample independently of each other). The weighting is achieved by the **Pre-treatment>Dispersion weighting** procedure, (Clarke KR, Chapman MG, Somerfield PJ, Needham HR, 2006, *Mar Ecol Prog Ser* 320: 11-27), covered in detail in Chapter 9 of the CiMC manual.

The differential downweighting is carried out by dividing the counts for each species by their index of dispersion $\overline{D}$ (variance to mean ratio, a 'clumping' measure), calculated from replicates within a group (site/time/treatment etc.), and then averaged across groups. The weighting is valid under rather general conditions, not unrealistic, but the original derivation did require: a) data to be real species counts, not densities standardised to some unit volume or substrate area; b) independent replicates within each of a set of sample groups, so that there is a basis for assessing within-group variance structure; and c) those replicates to be of a uniform size (strictly 'quantitative sampling'). Downweighting is only applied where a species shows significant evidence of clumping, this being tested by an exact permutation test, valid for the very small counts that are typical of many species. The resulting dispersion-weighted matrix has a common (Poisson-like) variance structure across species but unchanged relative responses of species in different groups. This is an important point: there is no attempt here to place greater emphasis on those species which best show up a given group structure (e.g. best separate control from polluted conditions). Such 'constrained' methods run the risk of circular arguments: selecting out only those species that tell you the answer you wanted in the first place! All that dispersion weighting does is divide through each row of the matrix (species) by a constant, so that a different balance of species contributions will be obtained by the subsequent analysis. These weights are calculated solely using information from replicates within each group, not across groups, so a consistent species (low variance-to-mean ratio within groups) will be given a high relative weight even if it shows no difference at all between groups.

If dispersion weighing of a count matrix is contemplated, this pre-treatment step must be carried out before any transformation. It may still make sense then to transform the dispersion weighted data sheet: a species which has large mean abundance at some sites, and is found in very consistent numbers in all replicates from those sites, will still tend to dominate the similarities. Transforming now has the strict objective of balancing contributions of consistent abundant species

with equally consistent but less numerous species. Previously, it was really used for this purpose and to reduce the impact of large but erratic counts of some species – but the latter can now be catered for by dispersion weighting. Whilst this will eliminate the need for transformation in some cases, it will still be required in others (Clarke KR, Tweedley JR, Valesini FJ 2014, *J Mar Biol Ass UK* 94: 1-16), to down-weight large counts which are also consistent. (The example there is of counts of small-bodied fish species, and demonstrates the usefulness of shade plots – seen earlier in this section – in determining whether/what transform may be needed after dispersion weighting.)

Chapter 9 of CiMC also discusses generalising the dispersion weighting concept to data which are not strict counts, but are density, area cover or biomass, etc. For 'quantity' data of this type, on a common measurement scale, it can still make sense to apply dispersion weighting, e.g. colonial species in large patches can have high variability for their mean area, over replicate quadrats (as measured by grid intersections, perhaps), and thus less inherent reliability than individual small-bodied, motile species with the same mean area cover. However, a dispersion index of 1 no longer has meaning (values depend on the measurement units) and permutation testing of $\overline{D} = 1$ thus also makes no sense. The PRIMER 7 dialog for **Pre-treatment>Dispersion Weighting** now gives a tick box not to perform this test, and division of entries by $\overline{D}$ then takes place whatever its value.

Revision #4
Created 21 May 2024 23:59:15 by Arden
Updated 15 January 2025 00:26:21 by Abby Miller