Analogue of univariate means plots

For this final section, we consider only cases in which the samples form a 1-way layout, i.e. there is a single (*a priori* defined) factor whose levels divide the data into groups, the samples in a group being considered replicates of that factor level. This factor could, of course, be from a combined factor for a higher-way crossed layout, e.g. all combinations of sites and times with (real) replicates for each combination. For a univariate response, such as a diversity index, the first step would be to test if there were significant differences among the groups and, if so, this is followed by a *means plot*, showing the respective means for each group with some measure of how reliably that mean has been determined, usually a confidence interval surrounding the mean. This is the type of output generated by all statistical packages (and by the **Plots>Means Plot** routine, seen in Section 15 for the taxonomic distinctness diversity index). It is uncommon, for example, to see a plot of the actual replicate values juxtaposed for each group in the univariate case. A glance at Fig. 18.1 of CiMC will soon tell you why! By comparison with a means plot, displaying the replicates will often make it difficult to see that pattern among the means clearly. Yet, in the multivariate context, there seems to be a (false) perception that the correct ordinations to plot have always to be ones that contain a point for each sample, in spite of the high stress solutions that often result – with that stress often due to random sampling variability within each group rather than to an inherently high-d structure in the relationship among groups. *n*MDS is surprisingly successful in capturing underlying group structure from replicate data, partly because it is able to non-linearly stretch and squeeze the dissimilarities into the low-d MDS distance scale, compressing the scale for smaller dissimilarities, which are often those from within groups. However, the logic from univariate practice (and it is sound logic) is that for *a priori* defined structures, we should run a testing procedure (ANOSIM, RELATE, PERMANOVA, ...) to justify interpreting the data and follow this with an ordination means plot, to display and interpret the among-group patterns. On several occasions in this manual and in CiMC, we have followed through this logic to good effect. We have also seen that there are a number of ways of producing a means plot, i.e. different ways of defining a *measure of central tendency* (to use the accepted jargon of statistics). This is particularly so in the multivariate context, where we have already referred to relationships among means calculated from: averaging the raw data, the transformed data matrix, the similarities, the rank similarities, and (in PERMANOVA+) generating centroids in the high-d PCO space. Chapter 18 of CiMC discusses the pros and cons of some of the different methods, but the practical reality (fortunately) is that the relationship among means is often very similar whichever method is used. The possible exception to this is averaging raw data in cases where a strong transformation is appropriate – the means can then get hijacked by species with large outlying counts in a single replicate, something which the usual process of transformation before computation of similarities is able to resolve. Ordination plots of means, based on averaging a data matrix with **Tools>Average** (rather than averaging biological-type similarities – also by **Tools>Average**, but on the resemblance matrix) are therefore preferably better carried out on the transformed form of that data sheet. (This is especially true if that averaging is over the levels of other factors.)

Revision #8
Created 10 November 2024 23:44:35 by Arden
Updated 26 February 2025 03:58:31 by Abby Miller