Bootstrap regions

These averages can then be used to generate a *bootstrap region* for each of the *g* groups – at its simplest by displaying the full set of *b*$\times$*g* averages in a 2-d (or 3-d) ordination. Here is the first obvious approximation therefore, namely that a 2- or 3-d ordination is not necessarily a perfect representation of the *b*$\times$*g* samples, since they are from a higher-d variable space. But this is an issue we are well used to dealing with – we interpret 2-d ordinations cautiously if they have high stress, and look at the 3-d plots (or even subsets of higher axes, though this is rarely necessary in this case) to check whether the 2-d plot has over-simplified some aspect of the groups' structure. In fact, the stress values for a 2-d plot are often quite acceptably low, even though these are typically ordinations on a very large number of samples of bootstrap averages (the recommendation is *b*=100+ bootstraps per group, if you can run this in a viable time, i.e. an ordination on 500+ points if you have *g*=5 groups). This is because the inherent structure of the plot may be just that of the relationships among the *g* group means, and such means plots are usually low-dimensional. At least this will be the case if the original number of replicates per group is not small, so that the regions are fairly tight (and PRIMER will issue a warning if you run **Analyse>Bootstrap Averages** with groups which are definitely too small – less than 5 replicates, though many more are preferable).

The **Bootstrap Averages** routine is able to take this a stage further and, for the 2-d ordination, will construct smooth envelopes for the bootstrap average points which have a nominal 95% coverage (or 80% or 50%). As stressed above, this is not a formal 95% confidence interval, since several sources of uncertainty (such as the approximation to the 'true' dimensionality) are not catered for, but a subtle and rather complex correction is made for the well-known underestimation (of order 1-n^{-1} on both axes, where n is the number of original replicates in a group) in variance estimates from bootstrap means. The nominal 95% coverage comes from approximating the shape of the observed bootstrap average regions in 2-d by back-transformed bivariate normals from individual location-shifted power transformations, fitted to the rotated major and minor axes for each group separately (essentially the algorithm used in Section 15 for $\Delta^{\scriptscriptstyle +} / \Lambda^{\scriptscriptstyle +}$ 'ellipse' plots). This is another approximation therefore, and will not be able to fit non-convex (e.g. banana-shaped) clusters of points very convincingly – but it does incorporate the variance bias correction, so it is generally seen that the smooth envelopes contain more than 95% of the bootstrap average points.

Revision #4
Created 11 November 2024 01:07:10 by Arden
Updated 26 February 2025 04:21:48 by Abby Miller