11.5 Further ‘BEST’ variations

Entering variables in groups

In some contexts, it makes good sense to utilise an *a priori* group structure for the explanatory variables and enter or drop all variables within a single group simultaneously, e.g. if locations of sites expressed in latitude and longitude are two of the variables, it does not make sense to enter one into the ‘explanation’ and leave out the other. Valesini, Tweedley, Clarke *et al.* (2014) *{e}* give a more major example of an estuarine fish study, where abiotic variables potentially driving the assemblages over different spatial scales were divided into those measuring wave exposure, substrate/vegetation type, extent of marine water intrusion, and more dynamic water quality parameters – with multiple variables in each group – all within a categorical structure, e.g. of different microtidal estuaries in Western Australia. Groups were entered into the BEST Bio-Env routine as indivisible units, to determine which variable *type,* or types, best explained the fish communities (at sites aggregated by SIMPROF into homogeneous clusters of their fish communities). Both BEST and the global BEST test need thus to be run on these (aggregated) samples by searching all combinations of *groups* of explanatory variables, which involves a much smaller number of combinations – and consequently lower selection bias to allow for in the permutation test – than if all variables had been separately entered.¶

Constrained (‘two-way’) BEST analyses

A further BEST modification parallels the two-way ANOSIM test of Chapter 6 and two-way SIMPER breakdown of Chapter 7. A strong categorical factor, clearly dominating the main differences observed in community structure among samples in an ordination, may sometimes not be comfortably incorporated into a set of quantitative explanatory variables to enter into BEST, e.g. if the factor has several levels which are in no sense ordered. An example could again be found in the Valesini, Tweedley, Clarke *et al.* (2014) study in which the suite of c. 15 quantitative environmental variables are measured at a wide range of sites within each of a number of different estuaries. Rather than attempt to convert the estuary factor into a quantitative form†, or simply ignore it on the grounds (say) that the major differences noted between estuaries should be identified by one of the measurement variables, in some circumstances it may be appropriate to accept that the differing locations will have differing assemblages and *remove* this categorical estuarine factor. For each considered combination of explanatory variables (or groups of variables perhaps, in the previous section), the matching statistic ρ is calculated separately *within* each of the levels (each estuary) and its values then averaged over those levels. The variable combination giving the largest average ρ is the *constrained BEST* match, and it can be tested for departure from the null hypothesis of ‘no genuine match’ by the same style of global BEST test as previously, but with constrained permutation of sample labels only *within* each level, then recalculating the largest average ρ, etc. The 2-way crossed ANOSIM analogy is very clear.

¶ *The option to group variables, using a pre-defined indicator, is implemented in the PRIMER BEST routine and its associated test, as is the conditional BEST analysis which follows.*

† *Clearly it would usually be inappropriate to number estuaries 1, 2, 3, 4, and then treat this as a quantitative variable, since it forces estuaries 1 and 4 to be 'further apart' environmentally than 1 and 3, which may be arbitrary. Instead, the trick is usually to replace this single factor by four new binary factors. (Is the sample in estuary 1? If so score 1, otherwise 0. Is it in estuary 2? … etc). Such binary variables are quantitative and now ordered.*

Revision #8
Created 6 March 2022 10:30:24 by Arden
Updated 6 November 2024 03:11:18 by Abby Miller