Zero-adjusted Bray-Curtis

A simple modification to the Bray-Curtis coefficient adjusts its behaviour as samples become vanishingly sparse. Standard Bray-Curtis is undefined for two samples containing no species, and can fluctuate wildly for near-blank samples – two samples containing just a single individual can fluctuate between 100% similarity if the individuals are from the same species, to 0% if they are not. The zero-adjusted Bray-Curtis coefficient (Clarke KR, Somerfield PJ, Chapman MG 2006, *J Exp Mar Biol Ecol* 330:55-80; also CiMC, Chapter 16) damps down this behaviour – analogously to the addition of the constant 1 in the log(1+x) transformation (to cater for x=0) – by adding +2 to the denominator of the ratio in S_{17}. A simple way of viewing this is as adding a 'dummy species' to the matrix, taking the value 1 for all samples. This forces two samples with no content to be 100% similar (they share the dummy species) and two samples with a single real individual now have some similarity, whether that species is shared (100%) or not (50%). It is clear that once there are a modest number of individuals, in either sample, then the adjustment makes no difference. It can only come into force when the assemblage is virtually denuded, and should only be applied if it makes biological sense to regard two blank samples as 100% similar, because both are denuded *from the same environmental cause*. If blank samples can be present in very different treatments/ sites etc., because of small sample sizes and highly clustered spatial distributions of organisms, it is unwise to use the zero-adjustment – instead, remove the blanks and use standard Bray-Curtis.

The adjustment is made by taking: (✓Add dummy variable)>(Value:1) in the Resemblance dialog. The constant 1 is appropriate to integer counts, being the lowest non-zero value attainable. This is true whether the data sheet has previously been transformed or not (the constant remains 1 under any power transform). For data on biomass, % area cover etc., the value could sensibly be chosen similarly as the lowest non-zero entry likely to be recorded (again the analogy with the log(c+x) transform is appropriate). 'Adding a dummy variable' can be carried out with other resemblance measures, but will only be effective for those coefficients which, like Bray-Curtis, treat joint absences of species as uninformative (e.g. Kulczynski, Czekanowski mean character difference, Canberra etc.). It is not given as an option for data type Environmental (it makes no sense then).

Revision #5
Created 31 May 2024 02:47:00 by Arden
Updated 15 January 2025 01:14:31 by Abby Miller