Quantitative similarity measures

In addition to Bray-Curtis S_{17}, and its zero-adjusted modification, PRIMER 7 also calculates:

$$ S_{15} = 100 \frac{1}{p} \sum_i \left[1 - \frac{ \left| y_{i1} - y_{i2} \right| }{ R_i} \right] \text{, where } R_i=\max_j \left\{ y_{ij} \right\} - \min_j \left\{ y_{ij} \right\} \text{ \hspace{1mm} Gower's coefficient,}$$

where standardisation is by the range R_i of values for the *i*th species over all samples (effectively by the maximum since the minimum will usually be zero), and thus shares withχ^2 distance the (generally undesirable) property that adding further samples can change existing similarities;

$ S_{18} = 100 \frac{ \sum_i \min \left\{ y_{i1}, y_{i2} \right\} }{ 2 / \left[\left(1 / \sum_i y_{i1} \right) + \left(1 / \sum_i y_{i2} \right) \right] } \text{\hspace{12mm} Kulczynski similarity,} $

which can be seen from the second form of S_{17} to be related to Bray-Curtis, replacing the arithmetic mean of the sample totals in the denominator of S_{17} with a harmonic mean;

$ S_{19} = 100 \frac{1}{p_{12}} \sum_i \left[1 - \frac{ \left| y_{i1} - y_{i2} \right| }{ R_i} \right] \text{ \hspace{13mm} Gower (excluding double zeros), } $

which is S_{15} with the fixed total number of species in the matrix (p) being replaced by p_{12}, the number of non-jointly absent species in the two samples being compared – an important difference;

$ S^{Can} = 100 \left(1 - \frac{1}{p_{12}} \sum_i \frac{ \left| y_{i1} - y_{i2} \right| }{\left(y_{i1} + y_{i2} \right) } \right) \text{ \hspace{10mm} Canberra similarity,} $

in the form used by Stephenson W, Williams WT, Cook SD 1972, *Ecol Monogr* 42: 387-415, not numbered by L&L but of more use for species data than its distance form (Canberra metric) D_{10}, because of the division by the variable species numbers p_{12} (i.e. excluding double zeroes);

$ S^{M-H} = 100 \left(1 - D_{1}^{\prime 2}/ \left[\sum_i y^{\prime 2}_ {i1} + \sum_i y^{\prime 2}_ {i2} \right] \right) \text{ \hspace{10mm} Morisita-Horn similarity, } $

where $^\prime$ denotes that y's are sample-standardised before D_1 and the denominator are calculated; and

$ S^{Och} = 100 \frac{ \sum_i \min \left\{ y_{i1}, y_{i2} \right\} }{ \sqrt{ \sum_i y_{i1} \sum_i y_{i2} } } \text{ \hspace{30mm} quantitative Ochiai similarity, } $

not defined by Ochiai as such, but it reduces to Ochiai's coefficient (S_{14}) when applied to P/A data. Clarke *et al* 2006 (see above for reference) construct this coefficient – which is an intermediate form between Bray-Curtis and Kulczynski, because it replaces the denominator with a geometric rather than arithmetic or harmonic mean – to illustrate that measures with reasonable properties are not difficult to invent, explaining the plethora of coefficients available in the literature!

Revision #40
Created 6 June 2024 23:10:55 by Arden
Updated 15 January 2025 01:57:35 by Abby Miller