2.4 Generalisation to dissimilarities

Of course, in many applications that we will encounter (especially in the case of community data), the Euclidean distance may not be the most appropriate measure for the analysis. What we require is a test of homogeneity of dispersions that will allow any resemblance measure to be used. There are (at least) two potential problems to overcome in order to proceed. First, there is the issue that the centroids in the space of the dissimilarity (or similarity) measure chosen are generally *not* the same as the vector of arithmetic averages taken for each of the original variables (which is what we would normally calculate as being in the "centre" of the cloud in Euclidean space). Therefore, it is not appropriate to calculate (say) the Bray-Curtis dissimilarity between an observation and its group centroid when the centroid has simply been calculated as an arithmetic average. What is needed, instead, is the "centre" of the cloud of points for each group in *Bray-Curtis* (rather than Euclidean) space (or in whatever space has been defined by the dissimilarity measure of choice for the particular analysis) .

The solution to this is to place the points into a Euclidean space in such a way so as to preserve all of the original inter-point dissimilarities. This is achieved through the use of a method called principal coordinates analysis (PCO, see chapter 3), described by Torgerson (1958) and Gower (1966) . In essence, PCO generates a new set of variables (principal coordinate axes) in Euclidean space from the dissimilarity matrix (Legendre & Legendre (1998)). Usually, PCO is used for ordination (see chapter 3), and in this case only the first two or three PCO axes are drawn and examined. However, the full analysis actually generates a larger number of PCO axes (usually up to N – 1, where N is the total number of samples). All of the PCO axes taken together can be used to re-create the full set of inter-point resemblances, but in Euclidean space. More particularly (and here is what interests us for the moment), if we calculate the *Euclidean* distance between, say, sample 1 and sample 2, using all of the PCO axes, then this will be equal to the *dissimilarity* calculated between those same two sample points using the original variables[52]. Thus, any calculation that would be appropriate in a Euclidean context (such as calculating centroids as arithmetic averages), can be achieved in a non-Euclidean (dissimilarity) space by performing the operation on PCO axes from the resemblance matrix.

This means that we can proceed with the following steps in order to obtain the correct values for z in more general (non-Euclidean distance) cases: (i) calculate inter-point dissimilarities, using the resemblance measure of choice; (ii) do a PCO of this dissimilarity matrix; (iii) calculate centroids (arithmetic averages) of groups using the *full set* of PCO axes; and (iv) calculate the Euclidean distance from each point to its group centroid using the PCO axes. These then correspond to the dissimilarity from each point to its group centroid *in the space of the chosen resemblance measure*.

The PERMDISP routine indeed follows these four steps (see Anderson (2006) for more details) in order to calculate appropriate z values for the test of homogeneity of dispersions on the basis of any dissimilarity measure[53].

[52] The distances obtained from the PCO axes must be calculated by first keeping the axes that correspond to the negative and the positive eigenvalues separate. These two parts are then brought together to calculate the final distance by taking the square root of the difference in two terms: the positive sum of squares and the absolute value of the negative sum of squares. This will result in a real (non-imaginary) value provided the positive portion exceeds the negative portion. See Anderson (2006) and chapter 3 of this manual for more details.

[53] The use of PCO axes to calculate centroids and the use of permutation methods to obtain *P*-values are two important ways that the method implemented by PERMDISP differs from the method proposed by Underwood & Chapman (1998) .

Revision #7
Created 9 August 2022 14:28:22 by Arden
Updated 3 December 2024 22:35:07 by Abby Miller