

10.1 Ordinations for multi-factor designs

Rationale

When considering the response of a whole set of variables (such as the abundances of species or taxa) simultaneously to a suite of several factors (e.g., arising from a multi-factor experiment or sampling design), it can be difficult to visualise salient structures and patterns in the data. One common problem is that multi-factor designs, when appropriately replicated, can yield a large total number of sampling units. A non-metric (or metric) multi-dimensional scaling (MDS) ordination done on a large number of samples can be very difficult to interpret. First, the 2D (and even 3D) stress might be quite high (> 0.20), precluding interpretability. Second, the residual variation (i.e., variation among the sampling units within each cell of the study design) is often quite large, and can mask essential patterns happening across the main factors of interest.

In univariate analyses of groups of samples (e.g., as in an ANOVA), one would commonly examine plots of means, rather than plots of raw sample values, to visualise patterns. In a similar way, for multivariate analyses, it is very useful to be able to visualise *distances among the centroids* in the space of a chosen resemblance measure. We can usefully construct ordinations from distance matrices among centroids that have been constructed from:

- levels of factors that are the main effects (*main effects plots*);
- combinations of levels of factors that are crossed with one another (*interaction plots*).

Ordination plots of distances among centroids were described by [Anderson \(2017\)](#). In PRIMER 7, it was possible to calculate distances among centroids, based on any given grouping factor (using **PERMANOVA+ > Distance Among Centroids...**). One can use this tool to generate interaction plots of interest by first creating factors that consist of combinations of levels of some chosen factors (e.g., using **Edit > Factors... > Combine...**).

In PRIMER 8, one can generate resemblance matrices and ordination plots of either (i) main-effect centroids or (ii) interaction centroids, automatically, by reference to a specific Design file. We shall outline briefly here (below) how resemblance matrices among centroids are constructed. We will then demonstrate these new practical tools and their utility for visualising and interpreting salient patterns in a multi-factor design by way of an example.

Distances among centroids

Let $\mathbf{Y}$ be a matrix of N rows (sampling units) by p columns (variables). Let $\mathbf{D} = \{d_{ij}\}$, $i = 1, \dots, N$; $j = 1, \dots, N$ be the distances or dissimilarities between every pair (i,j) of sampling units. If $\mathbf{D}$ contains Euclidean distances, then the distances

among centroids are equivalent to Euclidean distances among the arithmetic averages calculated separately for each variable. This equivalence does not hold, however, for non-Euclidean dissimilarities. Distances among centroids based on some other chosen dissimilarity measure (such as Bray-Curtis) are calculated as follows:

Step 1 - Calculate Gower's $\mathbf{G}$ matrix from $\mathbf{D}$

As in [Gower \(1966\)](#), obtain $(N \times N)$ matrix $\mathbf{G}$ by first defining matrix $\mathbf{A} = \{a_{ij}\} = \{-0.5 \cdot d_{ij}^2\}$, then centring the elements of this matrix by its row-means, $\bar{a}_{i\cdot}$, its column-means, $\bar{a}_{\cdot j}$, and its overall mean, $\bar{a}_{\cdot\cdot}$, to yield $\mathbf{G}$, i.e.,

$$g_{ij} = a_{ij} - \bar{a}_{i\cdot} - \bar{a}_{\cdot j} + \bar{a}_{\cdot\cdot}$$

Step 2 - Obtain the full set of principal coordinate (PCO) axes from matrix $\mathbf{G}$

This is done by performing an eigenvalue decomposition of matrix $\mathbf{G}$. The resulting eigenvectors are each standardised by the absolute value of their respective eigenvalue. At this step, it is important to keep track of those eigenvectors that are associated with *positive* eigenvalues, and those that are associated with *negative* eigenvalues (if any), as two separate sets.

Step 3 - Calculate centroids as averages along PCO axes

Suppose there are $\ell = 1, \dots, c$ cells (or specified groups of sampling units), and we require a centroid for each of these. The centroids are obtained as the arithmetic averages of the sampling units belonging to each cell (or group), calculated separately along each PCO axis.

Step 4 - Calculate distances among centroids

For every pair of centroids (ℓ, ℓ') , $\ell = 1, \dots, c$ and $\ell' = 1, \dots, c$, calculate Euclidean distances separately in each of two sets: one based on PCO axes corresponding to non-negative eigenvalues $(d_{\ell\ell'}^+)$ and one based on those corresponding to negative eigenvalues $(d_{\ell\ell'}^-)$, if any. Next, the $(c \times c)$ matrix of distances among centroids in the space of the chosen dissimilarity measure is then:

$$D^{\left[C \right]} = \{d_{\ell\ell'}^{\left[C \right]}\}, \text{ where } d_{\ell\ell'}^{\left[C \right]} = \sqrt{|(d_{\ell\ell'}^+)^2 - (d_{\ell\ell'}^-)^2|}$$

It is worth noting here that the distances among centroids can also be calculated directly from matrix $\mathbf{G}$, without using PCO axes. See section 5.1 in [Anderson \(2017\)](#) for details.

Revision #42

Created 6 June 2025 01:27:19 by Marti

Updated 30 November 2025 19:34:29 by Marti