

Chapter 4 Multidimensional quantitative data

Melissa

December 4, 2015

Multidimensional Variables

Simple analysis of descriptors is not enough because it doesn't take into account the covariance among descriptors. *Remember, objects are set a priori and descriptors "describe" each object.* Let's assume that in this case the objects are sites and the descriptors are species. As the number of descriptors increases, the number of dimensions of the random variable increases. Therefore more axes are necessary to construct the space in which the objects are plotted.

This chapter focuses on the *dependence* among descriptors.

To sum up:

1. The p descriptors in ecological data matrices are the dimensions of a random variable "descriptors". As the number of species increases, so do the dimensions of the sites.
2. The p descriptors (species) are not independent of one another. That's why we can't use unidimensional analysis.

Variance

Variance is a measure of the dispersal of a random variable y around its mean. ie. how much does a variable deviate from its mean.

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_j)^2 \quad (1)$$

Covariance

Is the extension to two descriptors of variance. It measures the joint dispersion of two random variables y_i and y_k around their means.

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_j)(y_{ik} - \bar{y}_k) \quad (2)$$

When the covariance is positive it means that both descriptors have a positive relationship. A negative covariance means that the descriptors have a negative relationship.

Dispersion matrix S

Contains the variances and covariances of the p descriptors. Therefore S is an association matrix. All eigenvalues of S are positive or null. Ideally, the matrix S should be estimated from a number of observations n larger than the number of descriptors p . When $n \leq p$ then the matrix has null eigenvalues but usually the first few are not affected.