

Section: [Numerical classification](#)

K-means (non-hierarchical classification)

Theory [R functions](#) [Examples](#)

K-means is a non-hierarchical agglomerative clustering algorithm, based on Euclidean distances among samples and using an iterative algorithm to find the solution. It minimizes the total error sum of squares (TESS), the same objective function as in the case of Ward's algorithm. The number of clusters (k) is defined by the user. Other than Euclidean distance can be used, but they need to be converted into metric distances and submitted to PCoA. For example, in the case of Bray-Curtis distance, which is not metric, one may calculate square-rooted Bray-Curtis distances (which are metric), submit them to PCoA, and then use all PCoA axes as the input matrix in K-means method instead of the raw data. The K-means algorithm, similarly to other iterative methods (like NMDS) can get trapped in local minima, and it may be useful to repeat the analysis many times and choose the solution with the lowest overall TESS.

The method can run in two modes, unsupervised or supervised. In the unsupervised mode, it searches for optimal clustering of samples into a predefined number of clusters; in the supervised mode, the user supplies the k centroids (e.g. typical samples) and the method searches for an optimal solution how to cluster the samples in the dataset around these centroids.

From:

<https://www.davidzeleny.net/anadat-r/> - **Analysis of community ecology data in R**

Permanent link:

<https://www.davidzeleny.net/anadat-r/doku.php/en:non-hier>

Last update: **2021/03/03 19:56**