

# Correspondence analysis

## Description

A simple function for correspondence analysis, following the algorithm described in Section 9.4 of Legendre and Legendre (1998).

## Usage

```
CA(Y, use.svd=TRUE, color.sites="black", color.sp="red")  
# Write total inertia, eigenvalues, relative eigenvalues, cumulative rel. eigenvalues to R window  
# name_of_output_object or print(name_of_output_object, ...)  
# Either two biplots, scaling type 1 and type 2, or a single scaling type 3 biplot, are plotted in a  
# window  
biplot(name_of_output_object, scaling=12, cex=2, ...)
```

## Arguments

Y	Data matrix
use.svd	TRUE: the decomposition is done by <a href="#">svd</a> (default). FALSE: the decomposition is done by <a href="#">eigen</a> . The signs of the coefficients along any one axis may differ between the two methods.
color	Color of the site and species symbols and labels in the biplots. Defaults: color.sites="black", color.sp="red".
scaling	scaling=12 to obtain scaling 1 and 2 biplots in a single window, scaling=3 to obtain the scaling 3 biplot alone
cex	A numerical value giving the amount by which plotting text and symbols should be magnified relative to the default.
...	Other parameters passed to the <code>print</code> or <code>biplot</code> functions.

## Details

Correspondence analysis (CA) of a table of frequencies producing scaling 1 and scaling 2 biplots. The variables must be frequency-like; they must all be expressed in the same physioical dimensions. Negative values are not allowed in CA.

Scaling type 1 biplot: uses F for sites, V for species (notation as in Legendre and Legendre 1998, Section 9.4). The sites are at the centroids (barycentres) of the species. This projection preserves the chi-square distance among the sites

Scaling type 2 biplot: uses Vhat for sites, Fhat for species. The species are at the centroids (barycentres) of the sites. This projection preserves the chi-square distance among the species.

Scaling type 3 biplot: use this scaling only when analyzing a contingency table, where the rows and columns are equivalent in nature. In this hybrid scaling, the positions of the rows are as in scaling 1 whereas the positions of the columns are as in scaling 2. This is not the same as scaling 3 in Canoco.

Algorithmic notes – The data matrix is transformed into matrix Qbar of the contributions to chi-square, following equations 9.31 and 9.32 of Legendre and Legendre (1998). Then the matrix  $t(Qbar) \%*\% Qbar$  is decomposed by [svd](#). Users can choose to decompose it by [eigen](#) instead.

## Value

Function CA returns a list containing the following results and matrices:

total.inertia      Total inertia in matrix Qbar.  
 eigevalues        CA eigenvalues.  
 rel.eigen         Relative eigenvalues.  
 cum.rel.eigen    Cumulative sum of the relative eigenvalues.  
 U, Uhat, F, Fhat, V, Vhat: matrices required to produce the biplots.  
 site.names, sp.names, color: additional information needed to produce the biplots.

## References

Aart, P. J. M. (van der) and N. Smeenk-Enserink. 1975. Correlations between distributions of hunting spiders (Lycosidae, Ctenidae) and environmental characteristics in a dune area. *Neth. J. Zool.* 25: 1-45.

Legendre, P. and Legendre, L. 1998. *Numerical Ecology*. 2nd English ed. Elsevier, Amsterdam.

## Author

Pierre Legendre, Departement de Sciences Biologiques, Universite de Montreal.

## Examples

```
# Example: data from Table 9.11 of Legendre and Legendre (1998)
table = matrix(c(10,10,15,10,15,5,20,10,5),3,3)
rownames(table) = c("Site1", "Site2", "Site3")
colnames(table) = c("Sp.1", "Sp.2", "Sp.3")
res = CA(table)
res                                     # Print out the summary results
biplot(res)                           # Produce two biplots in the graphics window
summary(res)                          # Print the structure of the output object
res$U                                 # Print out the matrix of eigenvectors of  $t(Qbar) \%*\% Qbar$ 
res$Uhat                             # Print out the matrix of eigenvectors of  $Qbar \%*\% t(Qbar)$ 
```

```
# Example: the spider data of Aart and Smeenk-Enserink (1975), available in library mvpart.
# The spider data frame has 28 rows and 18 columns. The first 12 columns are abundances of
different species of spiders and the next 6 are environmental data.
library(mvpart)
data(spider)           # Note : this file does not contain site names
res = CA(spider[,1:12], color.sites="blue")
res                    # Print out the summary results
biplot(res)            # Produce two biplots in the graphics window
summary(res)           # Print the structure of the output object
res$F                  # Print out the matrix of site scores for the scaling 1 biplot
```