UNIVERSITAT DE BARCELONA

ARQUB R Scripts

Jaume Buxeda i Garrigós

December 28, 2023

arqub
Material Culture and Archaeometry UB

GRACPE
Research Group on Archaeology of Complex Societies and Processes of Social Change

Material Culture and Archaeometry UB (ARQUB)
Research Group on Archaeology of Complex Societies and Processes of Social Change (GRACPE)
Facultat de Geografia i Història, Departament d'Història i Arqueologia,
Universitat de Barcelona, C/ de Montalegre, 6, 08001, Barcelona, jbuxeda@ub.edu

Introduction

These R scripts are in-house code written to perform the usual analysis on archaeometric compositional data according to the procedure used in the ARQUB (GRACPE) research group at the Universitat de Barcelona.

These scripts produced Figures 3, 5 and 7 A in the paper by Ots, M. J., Buxeda i Garrigós, J., Madrid i Fernández, M., Cahiza, P. A., Small-scale pottery production and distribution in the southern confines of the Inca empire. An archaeometric insight to define the provincial style, submitted to *Archaeological and Anthropological Science* (December 2023).

These scripts need the installation of the following R packages: **compositions** (van den Boogaart *et al.* 2022), **devEMF** (Johnson 2020), **lattice** (Sarkar 2008), **latticeExtra** (Sarkar and Andrews 2019), **plotrix** (Lemon 2009) and **MASS** (Venables and Ripley 2002).

The author of these scripts is Jaume Buxeda i Garrigós.

The comments in the scripts are in Catalan. The scripts are in progress and are provided here without warranty of any kind, express or implied. In no event shall the author or copyright holder be liable for any claim, damages or other liability. If you find any problem or bug, please report it to the author, and he will try to fix it.

inicialCurt

Description

Calculates the variation matrix, the total variation and the information entropy and produces several graphs to summarise all this information.

Usage

```
xvar <- inicialCurt(x, talls = c(0.3, 0.5, 0.9), nom = "Dades", nonum = 0, idioma = 1)
```

Arguments

`x`	composition or dataset of compositions
`talls`	vertical dotted lines expressing different $tv/\tau_{.j}$ values in the compositional evenness graph. Components to the left of the dotted line are below those values. If only the value 0 is given, no dotted lines will be displayed
`nom`	gives the default title of "Data" (according to the selected language) for the compositional evenness graph. This default title can be changed to something relevant to the case study. If there is a title, it will automatically provide the number of cases (n). "NA" will exclude the title and the number of cases
`nonum`	indicates the columns with non-numeric variables that will be excluded. By default, it is 0, indicating that all variables should be used
`idioma`	by default: 1. Indicates the language of the compositional evenness graph and variation matrix: 1, Catalan; 2, Spanish; 3, English; 4, French

Details

Using the determined concentrations or the raw data, this script calculates the variation matrix as explained in the paper by Buxeda (1999), and it performs all possible bivariate graphics with the values $\tau_{.j}$ on the x-axis and the values τ_{ji} on the y-axis ($j = 1, \ldots, j-1, j+1, \ldots, S$) and their correlations. The script also performs the compositional evenness graph. The latter is automatically saved in emf and pdf formats in the current directory with the name "uniformitat" (uniformitat.emf and uniformitat.pdf). The script calculates the total variation (tv) as a measure of compositional variability in the dataset and the information entropy (H_2) (Shannon 1948) or Shannon index as a measure of evenness in contributing to the compositional variability for all retained components (Buxeda and Madrid 2017). It also calculates the percentage of the information entropy over the maximum attainable. The compositional evenness graph is inspired by the rank/abundance graph used in biodiversity studies (see, for example, Magurran 2004).

This script uses the script *evariationmatrix2* to calculate the variation matrix, the script *entropia02* to calculate the information entropy, and the script *etiqueteselements* to produce the components labels in the compositional evenness graph.

Value

A list of three with the following content:

`MVC`	the variation matrix as explained in the paper by Buxeda (1999)
`Probabilitat`	the probability used in the calculation of the information entropy and its percentage over the maximum attainable
`Entropia`	the entropy used in the calculation of the information entropy and its percentage over the maximum attainable

Author

For the scripts *inicialCurt*, *evariationmatrix2*, *entropia02*, and *etiqueteselements*:

Jaume Buxeda i Garrigós https://github.com/jbuxeda/ARQUB

Examples

```
Inkavar <- inicialCurt(Inka, nom = "CWA Provincial Inca Pottery", nonum = c(1:3),
                       idioma = 3)
Inkavar <- inicialCurt(Inka, nom = "CWA Provincial Inca Pottery", idioma = 3)
```

The latter will produce the Figure 3 of the paper. The first example assumes that the first three variables will not be used.

ACPacomp

Description

It performs a principal component analysis via the singular value decomposition of the double-centred clr-transformed data. It can produce biplots by groups and display their probabilistic ellipsoids.

Usage

```
xACPacomp <- ACPacomp(x, nom = "NA", grup = 0, llegenda = grup, ggrup = grup,
                      gllegenda = ggrup, titol1 = 0, titol2 = 0, idioma = 1,
                      paleta = 1, paleta0 = NA, color = "chocolate", cps = 3,
                      simbol = 0, simbol0 = NA, midall = 0.8, elli = T, pelli = 0.95,
                      nelli = 2, noelli = 0)
```

Arguments

`x`	composition or dataset of compositions
`nom`	is the title to use in the biplots. By default ("NA"), there is no title
`grup`	by default, it is 0 and then no groups are employed. The biplots only display the labels of the individuals. If there are groups, 0 must be changed for the column number where the group is expressed. Groups must be indicated by integer numbers, starting at 1 and continuing until the last one without empty numbers. This column must be a factor. If groups are indicated, the biplots are made twice, first with the labels of the individuals and then with symbols and colours indicating the group
`llegenda`	if there are groups, it indicates the column number with the labels of the groups. By default, the column number is the `grup` column number. If another column is used, it must be defined as character. There is no need to put the group label on all individuals. The script will only use the label of the first individual of each group in `x`. These labels will be used to produce the legend of the groups in the biplots
`ggrup`	if there are groups, we can use different symbols in the biplots. By default, `ggrup` is `grup`; therefore, the symbols may change for every group. However, we can have a different column for symbols to express a second factor other than group membership. Then, `ggrup` must be a column number where we express, by integer numbers starting at 1 and without empty numbers, this second factor
`gllegenda`	if there is a second factor, `gllegenda` can be used to indicate the column number with the labels for this second factor. By default, the column number is the `ggrup` column number. If another column is used, it must be defined as character. There is no need to put the group label on all individuals. The script will only use the label of the first individual of each level of this second factor in `x`. These labels will be used to produce the legend of the second factor in the biplots. In that case, the legend in the biplots will be doubled by group and by this second factor
`titol1`	by default (0), the legend of the biplot has no title. If the user wants a title for the legend, then the user can put the title here
`titol2`	if a second factor is used, the user can add a second title for the doubled legend here that will be used as the title of this second factor. By default (0), there is no title for this second factor

`idioma` by default: 1. Indicates the language used in labelling the biplots: 1, Catalan; 2, Spanish; 3, English; 4, French

`paleta` if there are groups, `paleta` indicates the colour palette to be used for the symbols. By default, the palette to be used is 1, a predefined palette by the ARQUB unit. The 15 colours in this palette are: gray75, grey9, cyan, red, goldenrod1, dodgerblue, darkgoldenrod4, chartreuse1, darkgreen, indianred1, blue, darkmagenta, maroon1, aquamarine, and lightpink (colours from R Color Chart: https://rstudio-pubs-static.s3.amazonaws.com/3486_79191ad32cf74955b4502b8530aad627.html). Palette 2 is another predefined palette by the ARQUB unit that contains 7 colours in a grey scale for a black and white palette: white, grey90, grey70, grey50, grey30, grey10, black (colours from R Color Chart). Palettes 3 to 7 are R predefined palettes: 3, rainbow; 4, heat.colors; 5, terrain.colors; 6, topo.colors; 7, cm.colors. The user can use a free palette, providing a value of 0

`paleta0` when the user chooses to use a free palette (`paleta` = 0), a vector of colour names must be provided here. By default, the palette in use is ARQUB, and then no custom palette is used (`paleta0` = NA)

`color` the colour of the bars in the bar chart with the per cent of variance explained by the first principal components. By default, it is chocolate, but it must be changed by the user giving the desired name

`cps` the number of the first principal components that will be displayed in all possible pairwise biplots. By default, there are 3

`simbol` in the colored biplots by group, indicates the symbol used to display the individuals. By default (`simbol` = 0), it is the point shape (pch) 21 (filled circle). A value different from 0 obliges the use of `simbol0`

`simbol0` if the user wants to use a different symbol or several symbols for groups or for a second factor, the user can provide a vector of point shapes with values between 21 and 25 (both included). The length of the vector must equal the number of groups in `grup` if only groups are displayed, or the number of levels in `ggrup` if a second factor is desired

`midall` refers to the size of the elements of the legend when present. By default, it is 0.8

`elli` is a logic argument that, when using groups (column `grup`), displays the probabilistic ellipsoids of the groups in the biplots with symbols, which is the option by default (`elli` = TRUE). To not display any ellipsoid, `elli` must be turned to FALSE

`pelli` is a value or a vector of values indicating the probability of the ellipsoids being displayed. By default, the probability ellipsoid for 0.95 will be displayed

`nelli` establishes the maximum number of individuals in a group not to display the probabilistic ellipsoids. By default, only groups with more than 2 individuals will be displayed

`noelli` is a vector with the number of groups (column `grup`) whose probabilistic ellipsoids will not be displayed regardless of the group's number of individuals. By default (`noelli` = 0), all groups will be used

Details

Using the **compositions** library (van den Boogaart *et al.* 2022), this script performs a principal component analysis in the Aitchison geometry (*i.e.*, clr-transform) of the simplex (van den Boogaart and Tolosana-Delgado 2013). The original library uses an acomp-dataset, *i.e.*, the composition or dataset of compositions transformed to a vector of class "acomp" representing one closed composition or a matrix of class "acomp" representing multiple closed compositions each in one row. This script uses the `x` composition or dataset of compositions (the determined concentrations or raw data) and transforms this into a class "acomp". Then, the R "princomp" command performs a singular value decomposition of the double-centred clr-transformed dataset.

The script performs the singular valued decomposition and displays a bar chart with up to the first 10 principal components' per cent variance explained. The plot is automatically saved in emf and pdf formats in the current directory named "BarplotVar" (BarplotVar.emf and BarplotVar.pdf). Then, for the number of desired components (`cps`), it produces all pairwise biplots of form and covariance using the labels of the individuals. These biplots are automatically saved in emf and pdf formats in the current directory named "EtBiplotForm12" (labelled form biplot for principal components 1 and 2) and "EtBiplotCov12" (labelled covariance biplot for principal components 1 and 2) (EtBiplotForm12.emf, EtBiplotForm12.pdf, EtBiplotCov12.emf, and EtBiplotCov12.pdf). In case groups are indicated (in column `grup`), it also produces all pairwise biplots of form and covariance using the desired palettes indicating group colour. The user can choose the desired symbols for all individuals or every group (using `simbol` and `simbol0`). Moreover, in those biplots, a second column (`ggrup`) might be used to choose symbols not for groups but for a second factor. Groups are differentiated by colour, and symbols indicate differences by the second factor. In addition, those plots can display probabilistic ellipsoids (using `elli`) expressing the desired probability (using `pelli`) for groups over a given number of individuals (using `nelli`) and excluding undesired groups (using `noelli`). In those colored biplots, a legend is automatically inserted using the legend text in column `llegenda`, for groups, and `gllegenda`, for a second factor. The legend might have a title for the groups (`titol1`) or two titles if a second factor is used (`titol2`). When producing each of these biplots, the program stops, and the cursor changes to a cross inside the plot when asking for the uppermost left corner of the legend box. Then, the user must locate this point with the mouse's left button and press it. The script moves to the next biplot. These biplots are automatically saved in emf and pdf formats in the current directory with the name "CoBiplotForm12" (colored form biplot for principal components 1 and 2) and "CoBiplotCov12" (colored covariance biplot for principal components 1 and 2) (CoBiplotForm12.emf, CoBiplotForm12.pdf, CoBiplotCov12.emf, and CoBiplotCov12.pdf).

Value

An object of type c("princomp.acomp","princomp") whose content is explained in the manual of the package **compositions** (van den Boogaart *et al.* 2022).

Author

Jaume Buxeda i Garrigós https://github.com/jbuxeda/ARQUB

Examples

```
inka2ACPacomp01 <- ACPacomp(Inka21a7[ , -c(3:9)], nom = "Inka ?  Viluco pottery",
                      grup = 1, llegenda = 2, idioma = 3, paleta = 0,
                      paleta0 = c("navy", "dodgerblue", "thistle1", "chartreuse",
                      "red", "chocolate1", "yellow"))
inka2ACPacomp01 <- ACPacomp(Inka21a7[ , -c(3:9)], nom = "Inka ?  Viluco pottery",
                      grup = 1, llegenda = 2, idioma = 3, paleta = 0,
                      paleta0 = c("navy", "dodgerblue", "thistle1", "chartreuse",
                      "red", "chocolate1", "yellow"), simbol = 1,
                      simbol0 = c(21, 22, 23, 24, 25, 21, 25), midall = 0.8)
inka2ACPacomp01 <- ACPacomp(Inka21a7, grup = 1, llegenda = 2, ggrup = 3,
                      gllegenda = 4, nom = "Inka ?  Viluco pottery", idioma = 3,
                      paleta = 0, paleta0 = c("navy", "dodgerblue", "thistle1",
```

```
"chartreuse", "red", "chocolate1", "yellow"), simbol = 0,
simbol0 = c(21, 24, 25), midall = 0.8, nelli = 3, noelli = 7)
```

The latter will produce the form and covariance biplots in Figure 4 of the paper. Inka21a7 contains the determined concentrations of the retained components. The first column is a factor indicating group membership, and the second column indicates the label for each group in the legend. The third column contains a second factor with the type of material. The labels for this factor in the legend are in the fourth column. The doubled legend has no titles. A probabilistic ellipsoid at 0.95 is displayed for groups with more than 3 individuals, excluding group 7, which corresponds to the ungrouped individuals. The first two examples assume that the columns 3 to 9 will not be used.

LDAclr

Description

It performs a principal component analysis via the singular value decomposition of the double-centred clr-transformed data. It can produce biplots by groups and display their probabilistic ellipsoids.

Usage

References

[1] Buxeda i Garrigós, J. (1999) Alteration and contamination of archaeological ceramics: the perturbation problem, *Journal of Archaeological Science* 26: 295–313. https://doi.org/10.1006/jasc.1998.0390

[2] Buxeda i Garrigós, J., Madrid i Fernández, M. (2017) Designing Rigorous Research: Integrating Science and Archaeology. In Hunt, A. (ed.) *The Oxford Handbook of Archaeological Ceramic Analysis*: 19–47. Oxford University Press, Oxford. https://doi.org/10.1093/oxfordhb/9780199681532.013.3

[3] Johnson, P. (2020) *devEMF: EMF Graphics Output Device*. R package version 4.0-2. https://CRAN.R-project.org/package=devEMF

[4] Lemon, J. (2006) Plotrix: a package in the red light district of R, *R-News*, 6(4): 8–12.

[5] Magurran, A. E. (2004) *Measuring Biological Diversity*, Blackwell Science Ltd., Oxford.

[6] Sarkar, D. (2008) *Lattice: Multivariate Data Visualization with R*. Springer, New York.

[7] Sarkar, D., Andrews, F. (2019 *latticeExtra: Extra Graphical Utilities Based on Lattice*. R package version 0.6-29. https://CRAN.R-project.org/package=latticeExtra

[8] Shannon, C. E. (1948) A Mathematical Theory of Communication, *The Bell System Technical Journal* 27: 379–423, 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[9] van den Boogaart, K. G., Tolosana-Delgado, R. (2013) *Analysing Compositional Data with R*, Springer-Verlag, Berlin Heidelberg.

[10] van den Boogaart, K. G., Tolosana-Delgado, R., Bren, M. (2022) *compositions: Compositional Data Analysis*, R package version 2.0-4. https://CRAN.R-project.org/package=compositions

[11] Venables, W. N., Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York.