UNIVERSITAT DE BARCELONA

ARQUB R Scripts

Jaume Buxeda i Garrigós

December 21, 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**, **devEMF**, **lattice**, **latticeExtra**, **plotrix** and **MASS**.

The author of these scripts is Jaume Buxeda i Garrigós.

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. 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). 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

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 variation matrix (MVC), and the probability (Probabilitat) and entropy (Entropia) used in the calculation of the information entropy and its percentage over the maximum attainable.

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)

This 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	it 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. 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 choose to use a free palette (paleta = 0) a vector of colour names must be provided here

Details

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

[4] 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