

Revista Colombiana de Estadística

ISSN: 0120-1751

revcoles_fcbog@unal.edu.co

Universidad Nacional de Colombia
Colombia

Nieto, Ana B.; Galindo, M. Purificación; Leiva, Víctor; Vicente-Galindo, Purificación
A Methodology for Biplots Based on Bootstrapping with R
Revista Colombiana de Estadística, vol. 37, núm. 2, diciembre, 2014, pp. 367-397
Universidad Nacional de Colombia
Bogotá, Colombia

Available in: <http://www.redalyc.org/articulo.oa?id=89933712008>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System
Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal
Non-profit academic project, developed under the open access initiative

A Methodology for Biplots Based on Bootstrapping with R

Una metodología para biplots basada en bootstrapping con R

ANA B. NIETO^{1,a}, M. PURIFICACIÓN GALINDO^{1,b}, VÍCTOR LEIVA^{2,3,c},
PURIFICACIÓN VICENTE-GALINDO^{1,d}

¹DEPARTAMENTO DE ESTADÍSTICA, UNIVERSIDAD DE SALAMANCA, ESPAÑA

²FACULTAD DE INGENIERÍA Y CIENCIAS, UNIVERSIDAD ADOLFO IBÁÑEZ, CHILE

³INSTITUTO DE ESTADÍSTICA, UNIVERSIDAD DE VALPARAÍSO, CHILE

Abstract

A biplot is a graphical representation of two-mode multivariate data based on markers for rows and columns often provided in a two-dimensional space. These markers define parameters that help to interpret goodness of fit, quality of the representation and variability and relationships between variables. However, such parameters are estimated as point values by the biplot, thus no information on the accuracy of the corresponding estimators is obtained. We propose a graphical methodology, that may be considered as an inferential version of a biplot, based on bootstrap confidence intervals for the mentioned parameters. We implement our methodology in an R package and validate it with simulated and real-world data.

Key words: Bootstrap Confidence Interval, Graphical Methods, Multivariate Data, Quantiles, Software.

Resumen

Un biplot es una representación gráfica de datos multivariantes de dos vías basada en marcadores para filas y columnas proporcionada usualmente en un espacio bidimensional. Estos marcadores definen parámetros que ayudan a interpretar bondad de ajuste, calidad de representación y variabilidad y relaciones entre variables. Sin embargo, tales parámetros son estimados puntualmente en el biplot, sin proporcionar información acerca de la precisión de los estimadores. Se propone una metodología gráfica, que puede

^aAssociate Professor. E-mail: ananieto@usal.es

^bProfessor. E-mail: pgalindo@usal.es

^cProfessor. E-mail: victor.leiva@yahoo.com

^dProfessor. E-mail: purivg@usal.es

ser considerada como una versión inferencial de un biplot, basada en intervalos de confianza bootstrap para los parámetros mencionados. La metodología es implementada en un paquete R y validada con datos simulados y reales.

Palabras clave: cuantiles, datos multivariantes, intervalos de confianza bootstrap, métodos gráficos, software.

1. Introduction and Bibliographical Review

Analyses of high dimension data matrices for individuals and variables can be performed using multivariate techniques, which reduce this dimensionality projecting the data onto an optimal subspace, conserving the patterns of similarity between individuals and of covariation between variables. Differences among these techniques depend on the type of variables and metrics used into the respective subspaces. The biplot methods (biplots in short) proposed by Gabriel (1971) are part of such techniques, but biplots did not diffuse at the same speed as other multivariate techniques, due to the absence of software. Biplots are a graphical display in the context of principal component analysis (PCA in short, and PC for principal components), jointly representing a multivariate data matrix by markers for its rows (individuals) and columns (variables), permitting interrelations between them to be captured visually in a low-dimensional space. Biplots allow us to make description and also modeling and diagnostics (Bradú & Gabriel 1978) and are a powerful data visualization tool that can be considered as a multivariate version of the scatterplot, because biplots are usually performed in the two-dimensional space. This is the classical biplot of Gabriel, which has two parts. First, it approximates the data matrix by a singular value decomposition (SVD). Then, this matrix is factorized to obtain a low dimension Euclidean map through row and column markers represented by points/vectors, similarly to the case of the factorial correspondence analysis (CA). However, in biplots, the interpretation is based on the geometric properties of the scalar product between the rows and columns, allowing an approximation of the data matrix elements to be obtained.

Gower & Harding (1988), Gower (1992) and Gower & Hand (1996) provided a different focus of the classical biplot, ordering the individuals by scaling and then superimposing the variables so that a joint graphical interpretation is possible, as usual in biplots. The two most used biplots are known as GH and JK. Galindo (1986) proved that, with a suitable choice of the markers, it is possible to represent the rows and columns simultaneously on the same Euclidean space with an optimal quality, which is called the HJ biplot. Its coordinates for columns are the column markers in the GH biplot and the coordinates for rows are the row markers in the JK biplot. HJ biplot of Galindo has been applied in several fields. Orfao, González, San-Miguel, Ríos, Caballero, Sanz, Calmuntia, Galindo & López-Borrasca (1988) applied this biplot to histopathology; Rivas-Gonzalo, Gutiérrez, Polanco, Hebrero, Vicente-Villardón, Galindo & Santos-Buelga (1993) to enology; Mendes, Fernández-Gómez, Galindo, Morgado, Maranhão, Azeiteiro & Bacelar-Nicolau (2009) to limnology; Vilorio, Gil, Durango & García (2012) to biotechnology; Díaz-Faes, González-Albo, Galindo & Bordons (2013) to bibliome-

try; García-Sánchez, Frías-Aceituno & Rodríguez-Domínguez (2013) to sociology; and Gallego-Álvarez, Galindo & Rodríguez-Rosa (2014) to sustainability. Another biplot is known as GGE which displays the genotype main effect (G) and the genotype by environment interaction (GE) in two-way (two-mode) data. The GGE biplot originates from data graphical analysis of multi-environment variety trials. Technical details of the GH, HJ and JK biplots are provided in Section 2 and of the GGE biplot in Frutos, Galindo & Leiva (2014).

Ter-Braak (1986) used biplots fitted with linear models in the context of direct gradient analysis, which allows a set of species to be ordered according to its relationship with a set of environmental variables. Gauch (1988) employed the biplots for validating and selecting models when the interaction between genotype and environment is studied. Ter-Braak (1990) and Ter-Braak & Looman (1994) took advantage of the relationship between biplot and regression methods to introduce the biplot of the regression coefficient matrix and to propose a biplot based on reduced rank regression. Cárdenas & Galindo (2003) investigated the inferential aspects of biplots using the generalized bilinear models, extending their fitting with external information for variables related to the exponential family.

Vairinhos (2003) showed that the biplots are an ideal basis for the development of a data mining system, because most of the data analysis techniques can be expressed as particular cases of biplots. Amaro, Vicente-Villardón & Galindo (2004) studied the properties of MANOVA biplots within the context of the multivariate general linear model, developing methods for their interpretation. Hernández (2005) studied the performance of biplots in the presence of outliers and Ramírez, Vásquez, Camardiel, Pérez & Galindo (2005) proposed biplots to detect multicollinearity. As an alternative to the multiple CA for the case of presence/absence variables associated with the binomial distribution, Vicente-Villardón, Galindo & Blázquez (2006) considered the prediction biplots and applied it to biplots fitted by generalized linear regression, proposing logistic biplots, later extended by Demey, Vicente-Villardón, Galindo & Zambrano (2008).

Bradú & Gabriel (1974) and Bradú & Gabriel (1978) studied the fitting of bilinear models in two-way tables, analyzing collinearity between rows and columns on the biplots. Gabriel & Zamir (1979) also worked on the fitting of these bilinear models, but they proposed iterative techniques to obtain approximations at low rank using weighted least squares. Denis (1991), Falguerolles (1995), Choulakian (1996) and Gabriel (1998) used biplots to study interactions in two-and-three-way tables. Gabriel (1998) developed diagnostics in models based on contingency tables. Sepúlveda, Vicente-Villardón & Galindo (2008) used biplots as a diagnostic tool of local dependence in latent class models.

Methods for three-way data analysis have shown to be variants of the PCA of the two-way supermatrix, being the two most common ones: (i) TUCKER3 (Tucker 1966) and (ii) STATIS (L'Hermier des Plantes 1976). In (i), the data are summarized by three-mode components, and for their entities (individuals, sampling sites, etc.), component loadings are yielded. In (ii), data are compared on several occasions (time instants) by a PCA linked into column vectors (variables), belonging to different occasions. Based on TUCKER3, Carlier & Kroonenberg

(1996) generalized the SVD to a three-mode table proposing interactive and joint biplots to capture the information from the data. The difference between these two biplots is how the initial data matrix is treated, because in the interactive biplot two modes are combined, whereas the joint biplot is conditioned to one of the modes. Martín-Rodríguez, Galindo & Vicente-Villardón (2002) proposed meta-biplots following the meta-PC and procrustes methods, allowing biplots to be compared for studying individuals with variables, alternatively to the interactive and joint biplots.

Vallejo-Arboleda, Vicente-Villardón & Galindo (2006) and Vallejo-Arboleda, Vicente-Villardón, Galindo, Fernández, Fernández & Bécáres (2008) proposed the canonical STATIS, a biplot for multi-table data. Frequently, multivariate data taken over multiple occasions are found to produce a multi-table experiment. Neither the separate analysis of each occasion, using MANOVA or canonical analysis, nor the joint analysis using STATIS for multiple tables, are adequate to capture the real structure of the data matrices. This is because the former one accounts for group structure, but for not time evolution, whereas the last one confuses between and within group variabilities. Canonical STATIS permits a data group structure to be accounted, as well as time evolution on various occasions, by obtaining common or stable canonical variables across multiple occasions or data sets.

We focus on the classical biplot of Gabriel (1971); see Cárdenas, Galindo & Vicente-Villardón (2007) for a review and the books by Gower & Hand (1996), Greenacre (2010) and Gower, Gardner-Lubbe & Le-Roux (2011) for more details.

The bootstrap method was proposed by Efron (1979, 1987, 1993) and is used for facilitating calculations from statistical inference, which need the modern computer power since they are intensive. Bootstrapping corresponds to a resampling method useful for estimating the standard error (SE) of an estimator and then bootstrap confidence intervals (CIs) can be constructed. Because it is difficult to obtain closed expressions for sampling distributions of statistics associated with biplots (or with the SVD components), bootstrapping seems to be sound and adequate for approximating these distributions. Marcenko & Pastur (1967), Wachter (1978), Stewart (1980), McKay (1981), Edelman (1988), Lambert, Wildt & Durand (1990), Milan & Whittaker (1995), Díaz-García, Leiva & Galea (2002), Díaz-García, Galea & Leiva (2003), Díaz-García & Leiva (2003), Caro-Lopera, Leiva & Balakrishnan (2012) and Sánchez, Leiva, Caro-Lopera & Cysneiros (2015) discussed sampling distributions of SVDs and other decompositions and random matrices. Chatterjee (1984), Daudin, Duby & Trécourt (1988), Holmes (1989) and Linting, Meulman, Groenen & Van der Kooij (2007) combined bootstrapping with several multivariate techniques to provide more accurate results. Meulman (1982), Greenacre (1984), Gifi (1990), Timmerman, Kiers, Smilde & Stouten (2009), Kiers (2004) and Van Ginkel (2011) used bootstrapping in the context of multi-mode data.

The main objective of our work is to introduce a new methodology for biplots based on bootstrapping. We implement it in a graphical user interface (GUI) package developed in the statistical software R (www.r-project.org), named

biplotbootGUI. R is an integrated suite of software facilities for data manipulation, calculation and graphical display; see R-Team (2013).

The paper is organized as follows. In Section 2, we provide the technical background of this work. In Section 3, we propose a biplot methodology with bootstrapping and the state-of-art of the software developed for biplots. In addition, in this section, we detail the features of the **biplotbootGUI** package. In Section 4, we perform the numerical application of the proposed computational implementation by using simulated and real-world data sets. Finally, in Section 5, we sketch some discussions, conclusions and future works.

2. Background and Technical Preliminaries

In this section, we provide some technical preliminaries useful for facilitating the understanding of the results proposed in this paper.

2.1. Biplot Representations

Any $I \times J$ two-way data matrix \mathbf{X} can be expressed as the product of two matrices: \mathbf{A} with I rows and S columns and \mathbf{B} with S rows and J columns. If S is equal to two, then each row in \mathbf{A} and each column in \mathbf{B} have two values defining a point in a two-dimensional plot. When both of I rows of \mathbf{A} and J columns of \mathbf{B} are displayed in a single graphical representation, this is called a biplot. Thus, a biplot is a graph of a matrix $\mathbf{X}_{I \times J}$ with row and column markers $\mathbf{a}_1, \dots, \mathbf{a}_I$ and $\mathbf{b}_1, \dots, \mathbf{b}_J$, respectively, chosen in such a way that the inner product $\mathbf{a}_i^\top \mathbf{b}_j$ is the element x_{ij} of \mathbf{X} . The rows and columns of this marker matrix are the coordinate points in an Euclidean space related to the same orthogonal axes. A property of a biplot is that each of the $I \times J$ values can be recovered by viewing its $I + J$ points, which is a display of a matrix of rank equal to two (rank-two). Decomposition of a matrix \mathbf{X} into its component \mathbf{A} and \mathbf{B} is called a SVD, obtaining as result S PCs. A two-way data matrix rarely has rank-two, so that approximating \mathbf{X} by a rank-two matrix means that only the first two PCs are used for representing it. If these explain an important proportion of the total variability of \mathbf{X} , then it is sufficiently approximated by a rank-two matrix and can be displayed in a biplot.

Let \mathbf{X} be a data matrix composed by I individuals measured on J variables. The SVD of \mathbf{X} is defined as $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top$, where \mathbf{U} is a matrix whose column vectors are orthonormal and correspond to the eigenvectors of $\mathbf{X}\mathbf{X}^\top$, \mathbf{V} is a matrix whose column vectors are also orthonormal and correspond to the eigenvectors of $\mathbf{X}^\top\mathbf{X}$, and $\mathbf{\Lambda}$ is a diagonal matrix containing the singular values arranged in decreasing order. An element of \mathbf{X} may be written generically as $x_{ij} = \sum_{s=1}^{\min(I,J)} \lambda_s u_{is} v_{js}$. The first S elements of \mathbf{u}_s and of \mathbf{v}_s combined with the singular values λ_s in different ways are used as the coordinates for a graphical display of the data. The most common types of biplots are shown in Figure 1.

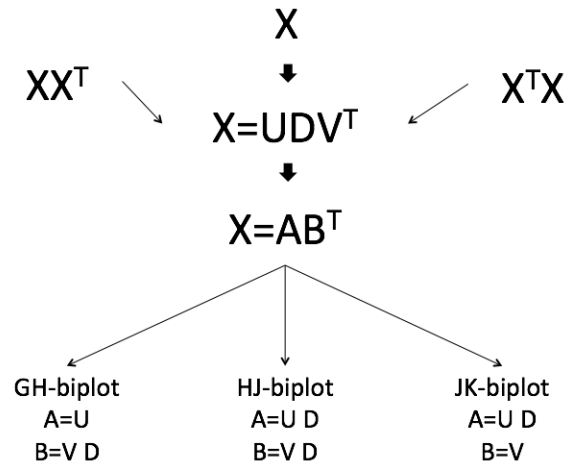


FIGURE 1: Types of biplots.

In a biplot, the column markers \mathbf{b}_j are shown as arrows and the row markers \mathbf{a}_i^T as points; see Figure 1. The biplot representation makes the projection of the row markers onto the column markers easier. The relationships between individuals and variables are studied through the projection of the points representing individuals onto the vectors representing variables, that is, $x_{ij} \approx \mathbf{a}_i^T \mathbf{b}_j$ implies $x_{ij} \approx \|\text{proj } \mathbf{a}_i / \mathbf{b}_j\| \text{sign } \mathbf{b}_j \|\mathbf{b}_j\|$, where $\|\text{proj } \mathbf{a}_i / \mathbf{b}_j\|$ is the length of the segment from the origin to the point \mathbf{a}_i (length of the projection from \mathbf{a}_i to \mathbf{b}_j), $\text{sign } \mathbf{b}_j$ is the sign of \mathbf{b}_j and $\|\mathbf{b}_j\|$ is the module of \mathbf{b}_j (length of the segment from the origin to \mathbf{b}_j). This means that x_{ij} is approximately the module of the projection of \mathbf{a}_i onto \mathbf{b}_j multiplied by the length of \mathbf{b}_j , with its corresponding sign. The direction of the vector \mathbf{b}_j shows the increasing direction of the corresponding variable values. The projections of the points \mathbf{a}_i onto a column vector approximate the j th column of \mathbf{X} and provide an ordination of the individuals respect to the corresponding variable. Once a way of representation is defined, it can be interpreted. Thus:

- The distance between points are dissimilarities between the corresponding individuals, specially if they are well represented. Individuals that are far away from each other have a larger Euclidean distance (ED) and vice versa. In Figure 2, the largest ED is observed between individuals \mathbf{a}_1 and \mathbf{a}_8 and the smallest ED is obtained between \mathbf{a}_5 and \mathbf{a}_6 .
- In the JK biplot, the line length approximates the variance of the variable. Hence, the longest line is the largest variance. From Figure 2, the variable \mathbf{b}_3 has the largest variance among the variables, while the variable \mathbf{b}_2 has the smallest variance. The cosine of the angle between the vectors approximates the correlation between the variables they represent. Thus, as this angle goes to 90 (or 270) degrees, the corresponding correlation decreases. An angle of 0 or 180 degrees reflects a correlation of 1 or -1 , respectively. The

biplot in Figure 2 shows a strong relationship between the variables b_4 and b_5 , and a weak relationship between the variables b_2 and b_3 , and between b_1 and b_3 . The correlation between the variables b_3 and b_6 is negative. The variables with the same direction involve multicollinearity, such as observed in Figure 2 for variables b_1 and b_2 . Also, biplots show multivariate outliers that can be used to detect clusters, such as the group formed by individuals a_1 , a_2 , a_7 and a_9 .

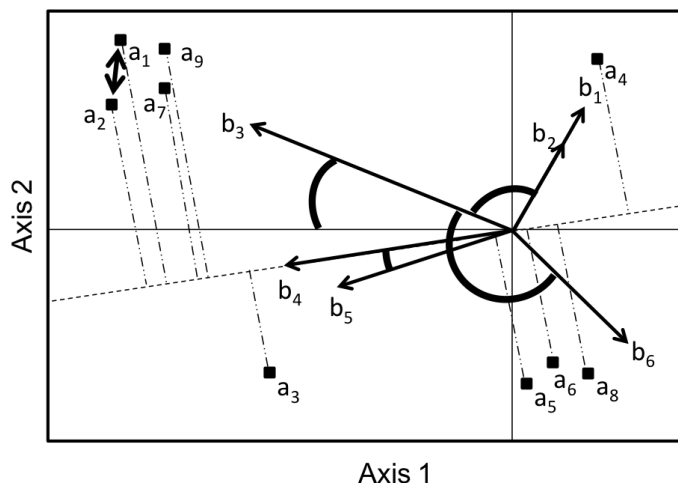


FIGURE 2: Biplot representation of a matrix with 6 variables and 9 individuals.

- The relationships between individuals and variables can be interpreted in terms of scalar products, that is, through the projections of the points onto the arrows. It permits us to know what variables differentiate among groups of individuals. If the projection falls on the origin, the value of the observation is approximately the average of the respective variable. Thus, as the projection of an individual goes increasing onto the direction of a vector representing a variable, this individual goes moving away from the average of that variable, whereas the opposite occurs when the projection goes increasing onto reverse direction. Therefore, in Figure 2, individual a_2 stands out with the largest value of the variable b_4 , followed by a_1 , a_7 and a_9 .
- In the HJ biplot, the search for the variables that differentiate individuals is made by the interpretation of the factorial axes, that is, the new variables that are a linear combination of the original variables and the relationships of new variables with the observed variables.
- The measure of the relationship between the axes of biplots and each of the observed variables is called relative contribution (RC) of the factor to the element, which represents the variability proportion of each variable explained by the factor. This measure is interpreted such as the coefficient of determination in regression. In fact, if the data are centered, this is

the coefficient of determination in the regression of each variable on the corresponding axis. The RC permits us to know what variables are more related to each axis (Axis 1 and Axis 2) and, therefore, allow us to know the variables involved in the order of the individuals on the projections in each axis. Because the axes are built to be independent, the RCs of each axis to each of the variables are independent and then it is possible to calculate the RC of a plane adding the RCs of the axes that form the plane.

Properties of the markers in the JK biplot. In this biplot, we use the metric $\mathbf{B}^\top \mathbf{B} = \mathbf{I}$, such that:

- The scalar products of the individuals of \mathbf{X} with the identity metric are the scalar products of row markers included in \mathbf{A} for the full space $\mathbf{X}\mathbf{X}^\top = \mathbf{A}\mathbf{A}^\top$.
- The ED between two individuals of \mathbf{X} and the ED between row markers in the full space are the same, that is, $(\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_j) = (\mathbf{a}_i - \mathbf{a}_j)^\top (\mathbf{a}_i - \mathbf{a}_j)$.
- The row markers and the individual coordinates are equal in the PC space, that is, if $\mathbf{\Psi}$ is a matrix containing the individual coordinates in the PC space, then $\mathbf{\Psi} = (\mathbf{U}\mathbf{D}\mathbf{V}^\top)\mathbf{V} = \mathbf{U}\mathbf{D} = \mathbf{A}$.
- The column coordinates of \mathbf{X} are the projection of the original axes onto the PC space, that is, the projection of each row marker onto column markers is an approximation of individual values on corresponding variables.
- The quality of representation of the rows is better than the columns.

Properties of the markers in the GH biplot. In this biplot, we use the metric $\mathbf{A}^\top \mathbf{A} = \mathbf{I}$, such that:

- The scalar products of the columns of \mathbf{X} are the scalar products of the column markers $\mathbf{X}^\top \mathbf{X} = \mathbf{B}\mathbf{B}^\top$.
- If \mathbf{X} has been centered by columns, the squared length of the vectors representing column markers approximate the covariance of the corresponding variables and as consequence the three following properties arise:
 - The squared length of the column vector approximates the variance of the corresponding variable, whereas the length of the vector approximates the standard deviation (SD) of these variables, that is, $\|\mathbf{b}_j\| = \|\mathbf{x}_j\| = \sqrt{\text{Var}(\mathbf{x}_j)}$.
 - The cosine of the angle formed by two column markers approximates the correlation between the corresponding variables, that is, $\cos(\mathbf{b}_i, \mathbf{b}_j) = \text{Cor}(\mathbf{x}_i, \mathbf{x}_j)$.
 - The ED between two variables is the ED between the corresponding column markers, that is, $d^2(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_j) = \|\mathbf{x}_i\|^2 + \|\mathbf{x}_j\|^2 - 2(\mathbf{x}_i^\top \mathbf{x}_j) = \|\mathbf{b}_i\|^2 + \|\mathbf{b}_j\|^2 - 2(\mathbf{b}_i^\top \mathbf{b}_j) = d^2(\mathbf{b}_i, \mathbf{b}_j)$.

- The coordinates in \mathbf{B} are the importance of the variables on the principal axes.
- The Mahalanobis distance between two individuals can be approximated by the ED between row markers, that is, by $(\mathbf{x}_i - \mathbf{x}_j)^\top \hat{\Sigma}^{-1}(\mathbf{x}_i - \mathbf{x}_j) = (\mathbf{a}_i - \mathbf{a}_j)^\top (\mathbf{a}_i - \mathbf{a}_j)$, where $\hat{\Sigma}$ is an estimate of the corresponding variance-covariance matrix.
- If \mathbf{X} is centered by columns, the row marker coordinates are the individual coordinates in the PC space and then \mathbf{A} contains the scores on the standardized PCs.
- The scalar products of the row markers are the scalar products of the rows of \mathbf{X} with the metric $(\mathbf{X}^\top \mathbf{X})^{-1}$ in the column space, that is, $\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top = \mathbf{A}\mathbf{A}^\top$.
- The quality of representation of columns is better than that for the rows.

Properties of the markers in the HJ biplot. In this biplot, the properties of row markers are the same as in the JK biplot, whereas the column markers are the same as in the GH biplot. The rules for interpreting the HJ biplot are a combination of the rules used in classical biplots, CA, factor analysis (FA) and multidimensional scaling. Specifically, we have that:

- The distances between row markers are interpreted as inverse similarities, in such a way that closer individuals are more similar, which allows the clusters of individuals with similar profiles to be identified.
- The lengths of the column vectors approximate the SD of the variables.
- The cosines of the angles between the column vectors approximate the correlations between variables. Hence, small acute angles are associated with highly positive correlated variables; obtuse angles near to the straight angle are associated with highly negative correlated variables; and right angles are associated with non-correlated variables; analogously the cosines of the angles between the variable markers and the PCs approximate the correlations between them, whereas for standardized data they approximate the factor loadings in FA.
- The location in the plot of the orthogonal projections of the row markers onto a column marker allows us to approximate the ranking of the row elements in that column. Thus, as the projection of a point (individual) away from the center of gravity (average coordinate point), the value that this individual takes on the variable is farther from its mean.
- Row and column markers can be shown in the same Cartesian system with optimal quality of representation. In the CA context, Greenacre (1984) and Lebart, Morineau & Piron (1995) proved that the clouds of row and column points have the same eigenvalues and barycentric relationships between them

exist. The relationships proposed by Galindo (1986) are similar, that is, the relations between the eigenvectors \mathbf{U} and \mathbf{V} are $\mathbf{U} = \mathbf{XVD}^{-1}$ and $\mathbf{V} = \mathbf{X}^{\top}\mathbf{UD}^{-1}$. Hence, the markers can be written as

$$\mathbf{A} = \mathbf{VD} = \mathbf{X}^{\top}\mathbf{UD}^{-1}\mathbf{D} = \mathbf{X}^{\top}\mathbf{U} = \mathbf{X}^{\top}\mathbf{XVD}^{-1} = \mathbf{X}^{\top}\mathbf{BD}^{-1} \quad \text{and}$$

$$\mathbf{B} = \mathbf{UD} = \mathbf{XVD}^{-1}\mathbf{D} = \mathbf{XV} = \mathbf{XX}^{\top}\mathbf{UD}^{-1} = \mathbf{XAD}^{-1}$$

Therefore, the row coordinates are weighted means of the columns where the weights are the values of \mathbf{X} and the same applies for columns.

2.2. Goodness of Fit

To assess goodness of fit in S dimensions, we need to know the variability proportion of \mathbf{X} explained by the approximated matrix $\tilde{\mathbf{X}}$, that is, the proportion of total variability $= \|\mathbf{X}\|^2 = \sum_{i=1}^I \sum_{j=1}^J x_{ij}^2$. Because of the least-square properties of the SVD and the orthogonormality of \mathbf{U} and \mathbf{V} , this total variability can be split into an explained variability and a residual variability expressed in terms of the squared singular values as $\sum_{s=1}^{\tilde{S}} \lambda_s^2 = \sum_{s=1}^S \lambda_s^2 + \sum_{s=S+1}^{\tilde{S}} \lambda_s^2$, where \tilde{S} is the rank of \mathbf{X} . This expression shows that the sum of the first S squared singular values divided by the total sum of squared singular values is a way to assess the amount of total variability explained by the first S vectors. If the explained total variability is large, it means that the graph represented by the first S singular vectors has a good representation of the initial matrix. If only a small proportion of such a variability is explained by the first singular vectors, the rest of variability can be explained by vectors of higher dimensions. If the data are centered by columns, individuals located near the origin may have measures close to the variable means, or their variability is explained by higher dimensions. In the same way, variables located near the origin may have small variability or may be not well represented in these dimensions. The estimates of row and column markers for each biplot and their quality of representation are shown in Table 1.

TABLE 1: Markers and their quality of representation.

| | Rows | | Columns | |
|-----------|---------------|-----------------------------------------------------------------|---------------|-----------------------------------------------------------------|
| | Coordinate | Quality | Coordinate | Quality |
| GH biplot | \mathbf{U} | S/\tilde{S} | \mathbf{VD} | $\sum_{s=1}^S \lambda_s^4 / \sum_{s=1}^{\tilde{S}} \lambda_s^4$ |
| JK biplot | \mathbf{UD} | $\sum_{s=1}^S \lambda_s^4 / \sum_{s=1}^{\tilde{S}} \lambda_s^4$ | \mathbf{V} | S/\tilde{S} |
| HJ biplot | \mathbf{UD} | $\sum_{s=1}^S \lambda_s^4 / \sum_{s=1}^{\tilde{S}} \lambda_s^4$ | \mathbf{VD} | $\sum_{s=1}^S \lambda_s^4 / \sum_{s=1}^{\tilde{S}} \lambda_s^4$ |

2.3. Contributions

The quality of representation detailed in Subsection 2.2 is a way to globally measure the fit of an approximation. However, it is also possible to individually measure its fit related to units and variables, which is important to interpret the results from the biplot. These measures are based on the concepts of RC or

absolute contribution (AC) proposed in Galindo (1986) and Jambu (1991). The total inertia is the sum of the eigenvalues of a matrix, that is, the trace of the matrix, used as a measure of the total variability in a data matrix. It is directly related to the physical concept of inertia, which is the tendency of an object in motion to stay in motion, and the tendency of an object at rest to stay at rest. Note that the total variability of the individual cloud is equal to the total variability of the variable cloud, given by $\text{trace}(\mathbf{X}\mathbf{X}^\top) = \text{trace}(\mathbf{X}^\top\mathbf{X}) = \sum_{s=1}^S \lambda_s^2$, where $\sum_{s=1}^S \lambda_s^2 = \sum_{j=1}^J d^2(\mathbf{b}_j, \mathbf{0}) = \sum_{s=1}^S \sum_{j=1}^J b_{js}^2 = \sum_{i=1}^I d^2(\mathbf{a}_i, \mathbf{0}) = \sum_{s=1}^S \sum_{i=1}^I a_{is}^2$. The ACs of the individual i and of the variable j to the variability of the axis s are $\text{AC}_{is} = a_{is}^2$ and $\text{AC}_{js} = b_{js}^2$, respectively. The total inertia of the factor s taking into account the ACs of the individual i and of the variable j are $\sum_{i=1}^I a_{is}^2 = \lambda_s^2$ and $\sum_{j=1}^J b_{js}^2 = \lambda_s^2$, respectively. The RCs of the elements i and j to the factor s are $\text{RC}_{is} = \text{AC}_{is}/\lambda_s$ and $\text{RC}_{js} = \text{AC}_{js}/\lambda_s$, respectively, whereas the RCs of the factor s to the elements i and j are $\text{RC}_{si} = a_{is}^2/d^2(\mathbf{a}_i, \mathbf{0}) = \cos^2(\mathbf{a}_i)$ and $\text{RC}_{sj} = a_{js}^2/d^2(\mathbf{b}_j, \mathbf{0}) = \cos^2(\mathbf{b}_j)$, respectively. The RC of the element to the factor measures how this factor can be explained by such an individual or variable.

3. A Biplot Methodology with Bootstrapping

In this section, we provide some aspects related to bootstrapping, propose a biplot methodology based on bootstrapping, discuss the state-of-art of the software developed for biplots and detail the features of the `biplotbootGUI` package.

3.1. Bootstrapping

Statistical theory attempts to answer three basic questions. (i) How should the data be collected? (ii) How should the collected data be analyzed and summarized? (iii) How accurate is this data summary? The third question constitutes part of the process known as statistical inference. Bootstrapping can help to answer this question when a sampling distribution is not available. Suppose a random sample $\mathbf{X} = (X_1, \dots, X_n)^\top$ is obtained from a population with unknown distribution. Let $\mathbf{x} = (x_1, \dots, x_n)^\top$ be an observation of \mathbf{X} , from which we can obtain the estimate $\hat{\theta} = s(\mathbf{x})$ of a parameter of interest θ , corresponding to an observed value of the estimator $\hat{\theta} = s(\mathbf{X})$ for which we want to know its accuracy. A bootstrap sample $\mathbf{x}^* = (x_1^*, \dots, x_n^*)^\top$ is defined to be a sample of size n with replacement from the observed sample \mathbf{x} . A bootstrap replication of $\hat{\theta}$ results from applying the same function $s(\cdot)$ to B bootstrap samples. To calculate the accuracy of the estimator $\hat{\theta}$, the bootstrap estimate of the corresponding SE, $\text{SE}[s(\mathbf{X})]$ say, can be used. Its bias can be empirically calculated from $\text{B}[s(\mathbf{X})] = s(\mathbf{x}^*) - \theta$. Algorithm 1 summarizes the bootstrap method to calculate the mentioned SE, which is often used for constructing a CI for a parameter.

Algorithm 1 Bootstrapping

-
- 1: Select B bootstrap samples $\mathbf{x}_1^*, \dots, \mathbf{x}_B^*$ each consisting of n data drawn with replacement from \mathbf{x} .
 - 2: Calculate the estimate $\hat{\theta}_b^* = s_b(\mathbf{x}^*)$ from the b th sample corresponding to a bootstrap replication of $\hat{\theta}$ for $b = 1, \dots, B$.
 - 3: Estimate the SE of $\hat{\theta} = s(\mathbf{X})$ with the sample SD of the B bootstrap replications, that is, by $\hat{\text{SE}}[s(\mathbf{X})] = ((1/B) \sum_{b=1}^B (s_b(\mathbf{x}^*) - s(\mathbf{x}^*))^2)^{1/2}$, where $s(\mathbf{x}^*) = \sum_{b=1}^B s_b(\mathbf{x}^*)/B$.
-

Normal and t distributions-based CIs. Assume the estimator $\hat{\theta}$ is normally distributed (at least approximately) with unknown expectation θ and SE known given by $(\text{Var}[\hat{\theta}])^{1/2} = \text{SE}[\hat{\theta}]$, that is, $\hat{\theta} \sim N(\theta, \text{Var}[\hat{\theta}])$. Then, $Z = (\hat{\theta} - \theta)/\text{SE}[\hat{\theta}] \sim N(0, 1)$. Note that $P(|Z| \leq z_{1-\alpha/2}) = 1 - \alpha$ is equivalent to

$$P(\theta \in [\hat{\theta} - z_{1-\alpha/2}\text{SE}[\hat{\theta}], \hat{\theta} + z_{1-\alpha/2}\text{SE}[\hat{\theta}]] = 1 - \alpha$$

Denote $\hat{\theta}_L = \hat{\theta} - z_{1-\alpha/2}\text{SE}[\hat{\theta}]$ and $\hat{\theta}_U = \hat{\theta} + z_{1-\alpha/2}\text{SE}[\hat{\theta}]$. Hence, the random interval $[\hat{\theta}_L, \hat{\theta}_U]$ has probability $1 - \alpha$ of containing the true value of θ . Thus, a $100 \times (1 - \alpha)\%$ CI for θ is $[\hat{\theta} \pm z_{1-\alpha/2}\hat{\text{SE}}[s(\mathbf{X})]]$. These results are meaningful for large enough sample sizes, for example, for $n \geq 25$. However, if we have small samples ($n < 25$), these results still can be correct (Bickel & Krieger 1989), but inappropriate for $n \leq 5$ (Chernick 1999). In addition, if $\text{SE}[\hat{\theta}]$ is unknown, we can estimate it with the expression given in Step 3 of Algorithm 1, $\hat{\text{SE}}[s(\mathbf{X})]$ say, but in this case $Z = (\hat{\theta} - \theta)/\hat{\text{SE}}[s(\mathbf{X})]$ still follows, in an approximate way, for large enough sample sizes, a standard normal distribution. Otherwise (smaller size samples), we have $Z = (\hat{\theta} - \theta)/\hat{\text{SE}}[s(\mathbf{X})] \sim t(n - 1)$, that is, now Z is Student- t with $n - 1$ degrees of freedom distributed, but we need additionally the normality assumption for the population X . Thus, in this case, an $100 \times (1 - \alpha)\%$ CI for θ with small sample sizes is $[\hat{\theta} \pm t_{1-\alpha/2}(n - 1)\hat{\text{SE}}[s(\mathbf{X})]]$, where $t_{1-\alpha/2}(n - 1)$ denotes the $(1 - \alpha/2) \times 100$ th quantile of the $t(n - 1)$ distribution.

Bootstrap normal and t distributions-based CIs. The normal and t distributions do not adjust the CI for θ to account for skewness and/or other aspects that can result when $\hat{\theta}$ is not the sample mean. The bootstrap normal and t CIs are procedures that adjust these aspects. Thus, by using the bootstrap method, we can obtain accurate CIs without having to make the normality assumption. This procedure approaches the population distribution directly from the data and builds CIs in the same way that we have explained in the cases of normal and t distributions. Algorithm 2 summarizes this procedure.

Bootstrap quantile-based CI. An alternative way to the bootstrap t distribution-based method (boot- t) for constructing bootstrap CIs is the quantile method (boot- q). The boot- t and boot- q methods are based on a simple structure. However, several data analyses involve more complex structures such as analysis of variance, regression models or time series. Boot- t and boot- q methods used for a more complex parameter than the mean were recently proposed by Leiva,

Algorithm 2 Bootstrap normal and t CIs

-
- 1: Follow Steps 1-3 of Algorithm 1 and obtain $\hat{\text{SE}}[s(\mathbf{X})]$.
 - 2: Calculate the value $z_b^* = (\hat{\theta}_b^* - \hat{\theta})/\hat{\text{SE}}_b^*$ from the b th sample corresponding to a bootstrap replication of $z = (\hat{\theta} - \theta)/\hat{\text{SE}}[\hat{\theta}]$, where $\hat{\theta}_b^*$ and $\hat{\text{SE}}_b^*$ are the estimates of θ and of $\text{SE}[\hat{\theta}]$ for the b th bootstrap sample, \mathbf{x}_b^* say, with $b = 1, \dots, B$.
 - 3: Determine the $(1 - \alpha/2) \times 100$ th quantile of z_b^* as follows:
 - 3.1 If $n \geq 25$, use the value $\hat{z}_{1-\alpha/2}$ such that $\#\{z_b^* \leq \hat{z}_{1-\alpha/2}\}/B = \alpha/2$;
 - 3.2 If $n < 25$, use $\hat{t}_{1-\alpha/2}(n-1)$ such that $\#\{z_b^* \leq \hat{t}_{1-\alpha/2}(n-1)\}/B = \alpha/2$.
 - 4: Compute the bootstrap CI for θ as follows:
 - 4.1 If $n \geq 25$, $[\hat{\theta} \pm \hat{z}_{1-\alpha/2} \hat{\text{SE}}[s(\mathbf{X})]]$;
 - 4.2 If $n < 25$, $[\hat{\theta} \pm \hat{t}_{1-\alpha/2}(n-1) \hat{\text{SE}}[s(\mathbf{X})]]$. If $B\alpha/2$ is not an integer, assume $\alpha/2 \leq 0.5$ and compute k as the largest integer less or equal than $(B+1)\alpha/2$ and define the $(1 - \alpha/2) \times 100$ th quantile by the $(B+1-k)$ th largest value of z_b^* .
-

Marchant, Saulo, Aslam & Rojas (2014) and can be adapted to data structures still more complex, as occurs with biplots; see Subsection 3.2. Algorithm 3 summarizes the boot-q method.

Algorithm 3 Bootstrap quantile CIs

-
- 1: Follow Steps 1 and 2 of Algorithm 1 obtaining the bootstrap replications $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$.
 - 2: Order $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ obtained in Step 1 of Algorithm 3 as $\hat{\theta}_{(1)}^* < \dots < \hat{\theta}_{(B)}^*$.
 - 3: Determine the $(B\alpha/2) \times 100$ th and $B(1 - \alpha/2) \times 100$ th quantiles of the distribution of $\hat{\theta}^*$, denoted by $\hat{\theta}_{B\alpha/2}$ and $\hat{\theta}_{B(1-\alpha/2)}$, respectively.
 - 4: Construct the boot-q CI as $[\hat{\theta}_{B\alpha/2}, \hat{\theta}_{B(1-\alpha/2)}]$.
-

3.2. Biplots Based on Bootstrapping

We adapt Algorithms 2 and 3 to measure the accuracy of the following biplot parameters: (B1) goodness of fit; (B2) quality of the approximation for columns; (B3) eigenvalues; (B4) angles between variables; (B5) angles between variables and axes; (B6) RC to the total variability of the j th column element; (B7) RC of the column element j to the q th factor; and (B8) RC of the q th factor to the j th column element. Adaptation of Algorithms 2 and 3 is given in Algorithm 4.

Algorithm 4 Adaptation of Algorithms 2 and 3

-
- 1: Follow Steps 1 and 2 of Algorithm 1 obtaining the bootstrap replications $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$.
 - 2: Calculate the empirical mean, SE and bias of the estimator $\hat{\theta}$ with the bootstrap samples by using the expressions $\hat{\text{E}}[s(\mathbf{X})] = \sum_{b=1}^B s_b(\mathbf{x}^*)/B$, $\hat{\text{SE}}[s(\mathbf{X})] = ((1/B) \sum_{b=1}^B (s_b(\mathbf{x}^*) - \hat{\text{E}}[s(\mathbf{X})])^2)^{1/2}$ and $\hat{\text{B}}[s(\mathbf{X})] = s(\mathbf{x}^*) - \hat{\text{E}}[s(\mathbf{X})]$, respectively.
 - 3: Establish boot-t CIs for the parameters (B1)-(B8) with Step 4 of Algorithm 2.
 - 4: Determine boot-q CIs for the parameters (B1)-(B8) with Step 4 of Algorithm 3.
-

3.3. Software for Biplots

Macros for biplots have been implemented in main commercial and non-commercial statistical software packages. Currently, most commercial statistical software packages include a procedure or macro for generating biplots; see details in Frutos et al. (2014). Specifically, the **GGEbiplot** software, dedicated to the GGE biplot (www.ggebiplot.com), can also generate the classical biplot. The **GGEbiplot** program is a commercial software and is widely used by agronomists, crop scientists and geneticists; see Yan & Kang (2003) and Frutos et al. (2014) and references therein. Vicente-Villardón (2010) implemented in the commercial software **MatLab** (www.mathworks.com/products/matlab) a program to perform biplots called **multbiplot**. It contains classical, HJ and logistic biplots, among other biplots, as well as simple and multiple CA for contingency tables.

Most of the software available for biplots is developed for specific applications, or as part inside general purpose packages. Consequently, they are not very flexible and produce static pictures that limit the interpretation of their results. Tables 2 and 3 contain the main packages in R, which have implemented biplot decompositions and/or representations. In these tables, the name of the package, the approach on which it is based, that is, Gabriel (1971), Galindo (1986) or Gower (1992), the main references, the creation date and last update of the corresponding package are presented and its main contents and functionalities are discussed.

In Table 4, we provide a review of the R packages mentioning the word “biplot”, although it refers to the joint representation of coordinates calculated with other methods instead of using the biplot decomposition.

3.4. The BiplotbootGUI Package

Because all of the packages (commercial and non-commercial) discussed in Subsection 3.3 are not suitable for constructing bootstrap CIs for biplot parameters (B1) through (B8), we developed a new package in the R language that combines the biplots described by Gabriel (1971) and Galindo (1986) and the bootstrap method to display results of these biplots and their statistical accuracy measures.

As mentioned, a GUI is a type of user interface which allows practitioners to interact with electronic devices such as computers. It is characterized by the use of icons and visual indicators, as opposed to text-based interfaces, typed command labels or text navigation, to fully represent the information and actions available to the user. The actions are often linked through direct manipulation of the graphical elements. Below, we discuss the features of a GUI in R language of the methodology for biplots based on bootstrapping proposed in the article and implemented in the **biplotbootGUI** package.

TABLE 2: Biplots in R

| Package Method | References | Content | Date Update |
|---------------------------------------------------|---------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| calibrate Gower | (Graffelman 2013) | It draws calibrated scales with tick marks on non-orthogonal variable vectors in biplots. | 21-01-06 20-03-12 |
| BiplotGUI Gower | (La Grange, Le-Roux & Gardner-Lubbe 2009, La Grange, Le-Roux, Rousseeuw, Ruts & Tukey 2013) | It provides a GUI to construct and interact with biplots and displays the variables as calibrated axes. Then, it is not possible to interpret the variable lengths. It allows us to change the title, show labels and points or hide them, change the type, color and size of lines and font, the color and orientation of labels and tick marks, draw convex-hulls and alpha bags. It also performs non-linear and MDS biplots and allows us to choose the distance and way to calculate the coordinates. It shows the variable correlations and provides interactive 3D graphs. | 13-08-08 19-03-13 |
| bpca Gabriel, Galindo | (Faria & Demetrio 2012) | It shows biplots in 2D-and-3D and provides variable lengths, angles between variables, correlations, coordinates to individuals and variables, eigenvalues, eigenvectors and quality of representation. It displays a graph with the correlations and their approximations and the 3D graph is interactive. | 17-08-08 21-02-12 |
| GGEbiplotGUI Gabriel Galindo Yang | (Yan, Hunt, Sheng & Szlavnic 2000, Yan & Kang 2003, Frutos & Galindo 2013) | It is a GUI to construct and interact with GGE biplots. It provides eigenvalues, % of variability explained by each of them, coordinates of individuals and variables, contributions of factors to elements. Also, this GUI allows us to change the background color, genotype labels, environments labels and title, font, graph title, in addition to showing genotypes and environments, as well as to hide title, axes and symbols. Furthermore, with this GUI it is possible to move the labels by the mouse button and change the color and text of labels. | 29-08-11 22-06-13 |
| multibiplotGUI Gabriel, Galindo | (Nieto, Baccalá, Vicente-Galindo & Galindo 2012) | It provides a GUI to construct and interact with multibiplots. It allows us to obtain the quality of representation, contributions, goodness of fit, eigenvalues and possibility of selecting the number of axes. It shows 2D-and-3D graphs (2D graph moves or removes labels, changes color, size and symbol of the points and selects the axes shown in the graph; 3D graph rotates and makes zoom). | 29-10-12 |

TABLE 3: (continued) biplots in R.

| Package Method | References | Content | Date Update |
|--------------------------------------------------|---------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| nominallogisticbiplot Gabriel, Galindo | (Hernández & Vicente-Villardón 2013a) | It produces a matrix analysis of polytomous items using nominal logistic biplots, extending the binary logistic biplot to polytomous nominal data. | 17-09-13 |
| biplot{stats} Gabriel | (R-Team 2013) | It is part of the basics of R and produces a biplot from the output of princomp or prcomp . | 25-09-13 |
| ordinallogisticbiplot Gabriel, Galindo | (Hernández & Vicente-Villardón 2013b) | It produces a matrix analysis of polytomous items using ordinal logistic biplots, extending the binary logistic biplot to polytomous ordinal data. | 30-10-13 26-11-13 |
| dynbiplotGUI Gabriel, Galindo | (Egido 2014) | It is a GUI to solve dynamic, classic and HJ biplots and tries with 2-and-3 way data matrices. | 04-11-13 08-01-14 |

TABLE 4: R packages which mention biplots.

| Package | References | Content | Dates |
|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| vegan | (Oksanen, Blanchet, Kindt, Legendre, Minchin, O'Hara, Simpson, Solyomos, Stevens & Wagner 2013) | It provides tools for describing community ecology. This package has the basic functions of diversity, community ordination and dissimilarity analyses. In addition, it shows biplots from results of redundancy, canonical correlation and canonical correspondence analyses, which can be used for other types of data as well. | 06-09-01 19-03-13 |
| ade4 | (Chessel, Dufour & Thioulouse 2004, Dray & Dufour 2007, Dray, Dufour & Chessel 2007, Chessel, Dufour, Dray, Jombart, Lobry, Ollier & Thioulouse 2013) | It is characterized by the implementation of graphical and statistical functions, availability of numerical data and writing of technical and thematic documentation. It includes bibliographic references and has functions to show biplots from results of the implemented analysis. | 10-12-02 11-04-13 |
| ade4TkGUI | (Thioulouse & Dray 2007, Thioulouse & Dray 2012) | It is a Tcl/Tk GUI for some basic functions of the ade4 package. | 29-09-06 13-11-12 |
| ca | (Nenadic & Greenacre 2007, Greenacre & Nenadic 2012) | It computes and visualizes simple, multiple and joint CA and shows biplots from the results of the previous analysis. | 28-07-07 12-06-12 |
| caGUI | (Markos 2012) | It is a Tcl/Tk GUI for functions of the ca package | 04-10-09 29-10-12 |
| ThreeWay | (Del Ferraro, Kiers & Giordani 2013) | It allows us to do component analysis for 3-way data arrays by means of Candecomp/Parafac, Tucker1, Tucker2 and Tucker3 models and shows joint biplots from Tucker3 models. | 29-10-12 11-06-13 |

After the R software has been downloaded from `cran.r-project.org` and installed, the user must download and install the `biplotbootGUI` package and its dependencies, which are the `rgl`, `tcltk`, `tcltk2`, `tkrplot` and `vegan` packages; see Adler & Murdoch (2012), Grosjean (2012), Tierney (2012) and Oksanen et al. (2013). Then, to load the `biplotbootGUI` package into the R software, the command `library(biplotbootGUI)` must be entered at the R prompt. Once all these instructions have been followed, the data must be loaded. Hence, one starts the GUI by entering the command `biplotboot(data)` in the R console, where data to be analyzed must be in a data frame; see details and examples in Section 4 of applications. Once the GUI has been initialized, a window entitled “bootstrap on classical biplots” emerges; see Figure 3. This window allows us to enter the number of replications and the confidence level to calculate the CIs. Also, it is possible to choose the parameters to be considered by the user.

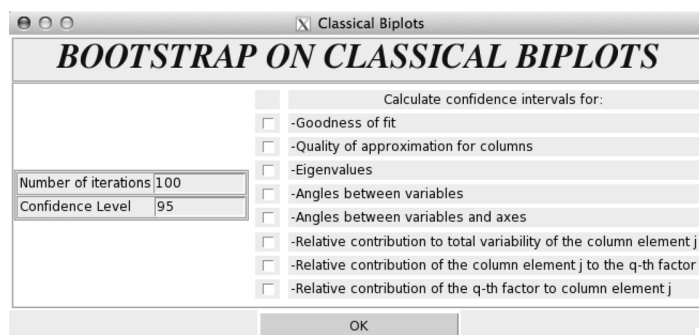


FIGURE 3: Main window.

After entering and selecting the parameters, one must click on the OK button and a window titled “Options” appears; see Figure 4. In this window the following options are available:

- Select the type of biplot to be executed (HJ, GH or JK).
- Select the transformation to be performed on the data considering:
 - Subtract the global mean.
 - Center by columns.
 - Standardize by columns.
 - Center by rows.
 - Standardize by rows.
 - Raw data.
- Change the color, size, label and symbol representing individuals in the graph.
- Change the color, size and label representing variables in the graph.
- Show the axes in the graph.

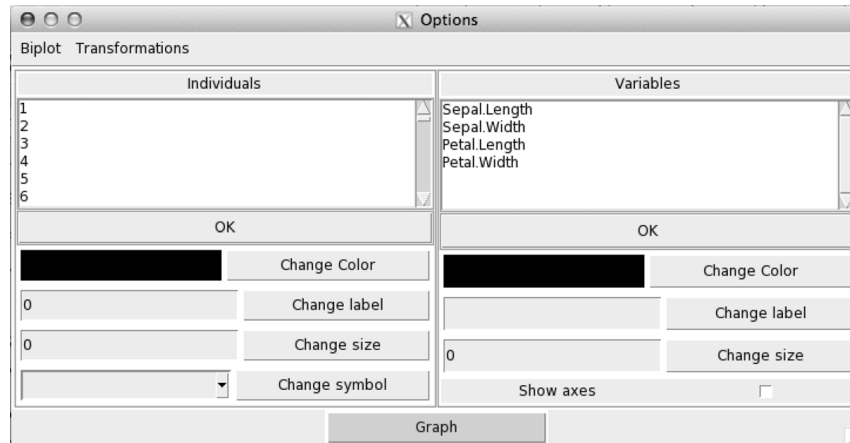


FIGURE 4: Window of options.

Given that not all the data are well represented by the first two axes, a window after clicking the button “graph” emerges with the option to choose the number of axes to be retained, according to the variability explained by each axis. After choosing the number of axes to be retained and clicking the button “choose”, a window showing the resulting graph in 2D appears; see Figure 5. This window displays the labels for the two axes indicating the percentage of variability explained by each of them (72.96 by axis 1 and 22.85% by axis 2). The user can select the axes to be displayed in the graph. At the top of the window, two menus with options to save the graph and show the biplot in 3D are displayed, whereas three text boxes where the user can change the axes displayed in the graph are at the bottom. Also, the user can move or remove the label of a specific element by clicking the left-mouse button and change the graphical displays of such an element by clicking the right-mouse button. This window contains two dropdown menus. In the first one, options to copy, save the graph in different file formats (PDF, postscript, BMP, PNG, JPG/JPEG) or exit are available, whereas the second one provides a 3D-graph made by the `rgl` package; see Adler & Murdoch (2012).

The user can rotate or make zoom in this graph by clicking the left-mouse or right-mouse button. Together with this window, a graph showing the coordinates for variables computed for all of the replications is shown. The GUI provides two text files. In the first one, the parameters of the biplot analysis (see B1-B8) are saved, whereas in the second one, tables with the values for the mean, SE, bias and lower and upper limits of the bootstrap CIs are provided. These two text files are automatically saved together with all the graphs containing the histogram and quantile versus quantile (QQ) plot of the estimates calculated by bootstrapping of the selected parameters in the first window. In the histogram, the solid line represents the estimate of the biplot parameter obtained from bootstrapping, whereas the dashed line is its value obtained from the biplot. In the x-axis of the QQ plot are the theoretical quantiles and in the y-axis the empirical quantiles.

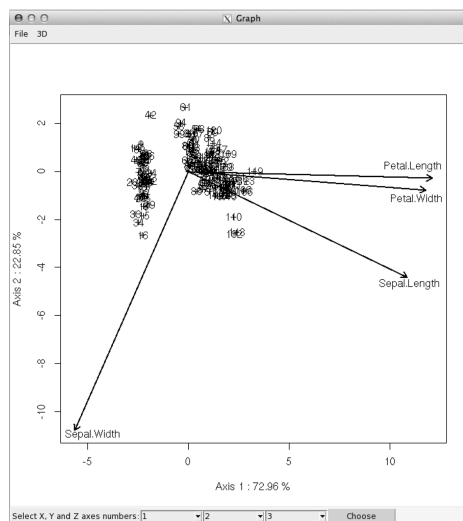


FIGURE 5: Window with a biplot representation in two dimensions.

4. Numerical Applications

In this section, we evaluate the performance and potentiality of our methodology by means of the `biplotbootGUI` package using both simulated and real-world data.

4.1. Simulated Data

To evaluate the performance of the `biplotbootGUI` package, an HJ biplot with the transformation “centering by columns” has been performed. We simulate data of 100 individuals on 5 variables (V_1, \dots, V_5) normally distributed, generated to have correlations $\text{Cor}(V_1, V_2) = 0.50$, $\text{Cor}(V_2, V_3) = 0.80$ and $\text{Cor}(V_4, V_5) = 0.90$. The number of bootstrap replications is 1,000 and the confidence level 95%. The time involved in a bootstrap replication is usually small. For example, the time spent in the calculations of a $1,000 \times 5$ matrix is less than four minutes for 1,000 replications. First, we explain the main results of the classical biplot. In Table 5, we observe the variability explained by each axis (Axis 1, Axis 2 and Axis 3). Note that the first eigenvalue explains more than 50% and the first three axes explain more than 94.27% of the total variability. Table 6 shows the RCs of the factor to the column elements in the first three axes. Note that all the variables are well represented by the first two axes, except the variable V_1 , which is in the third axis. The biplot representation using the first two axes (Axis 1 and Axis 2) is shown in Figure 7(left). The covariation structure shows a very high correlation between the variables V_4 and V_5 , and V_2 and V_3 , represented by acute angles. Variables V_2 and V_3 have a high correlation with V_1 , however they present no correlation with V_4 and V_5 , since they are almost orthogonal; see Table 7. Second, we explain

the results of applying the bootstrap method. Goodness of fit and eigenvalues are explained next. Figure 8 shows the histogram and QQ plot representing the values of the quality of approximation of 1,000 bootstrap replications.

TABLE 5: Eigenvalues and variability % explained by each of them with simulated data.

| No. | Eigenvalue | Variability | Accumulated variability |
|-----|------------|-------------|-------------------------|
| 1 | 16.06 | 53.47 | 53.47 |
| 2 | 12.15 | 30.59 | 84.06 |
| 3 | 7.01 | 10.21 | 94.27 |

TABLE 6: RCs of the factors to the column elements for simulated data.

| Variable | Axis 1 | Axis 2 | Axis 3 |
|----------|--------|--------|--------|
| V_1 | 226.48 | 16.48 | 737.49 |
| V_2 | 282.43 | 118.64 | 201.18 |
| V_3 | 281.58 | 89.64 | 46.09 |
| V_4 | 112.91 | 421.49 | 9.12 |
| V_5 | 96.60 | 353.75 | 6.12 |

TABLE 7: Angles between variables for simulated data.

| Variable | V_1 | V_2 | V_3 | V_4 | V_5 |
|----------|-------|-------|-------|-------|-------|
| V_1 | 0.00 | 14.59 | 11.58 | 67.15 | 66.89 |
| V_2 | 14.59 | 0.00 | 3.00 | 81.73 | 81.48 |
| V_3 | 11.58 | 3.00 | 0.00 | 78.73 | 78.47 |
| V_4 | 67.15 | 81.73 | 78.73 | 0.00 | 0.26 |
| V_5 | 66.89 | 81.48 | 78.47 | 0.26 | 0.00 |

We denote by “lower-t” and “upper-t” the lower and upper limits of the CIs based on the boot-t method, respectively, whereas these limits are denoted by “lower-q” and “upper-q” for the boot-q method. Table 11 provides the observed values for the mean, SE, bias and these limits. Notice that the observed value and its approximation are very close. These same results for eigenvalues are provided in Table 8. Figure 6 shows the histogram and QQ plot for the first eigenvalue (a similar behavior is observed for the other four eigenvalues, whose plots are omitted here, but are available under request for interested users). Note that the observed and estimated values practically do not differ and a similar conclusion is reached for the CIs. Each of the five eigenvalues resulting from the SVD of the simulated data shows the values calculated by 1,000 bootstrap replications.

TABLE 8: Results for the eigenvalues with simulated data.

| No. | Eigenvalue | Mean | SE | Bias | lower-t | upper-t | lower-q | upper-q |
|-----|------------|-------|------|-------|---------|---------|---------|---------|
| 1 | 16.06 | 16.09 | 1.13 | 0.03 | 13.85 | 18.33 | 13.87 | 18.24 |
| 2 | 12.15 | 11.92 | 0.78 | -0.22 | 10.37 | 13.47 | 10.27 | 13.33 |
| 3 | 7.01 | 6.9 | 0.63 | -0.12 | 5.64 | 8.15 | 5.74 | 8.13 |
| 4 | 4.51 | 4.39 | 0.31 | -0.12 | 3.78 | 5.00 | 3.82 | 5.05 |
| 5 | 2.69 | 2.61 | 0.17 | -0.08 | 2.28 | 2.94 | 2.28 | 2.93 |

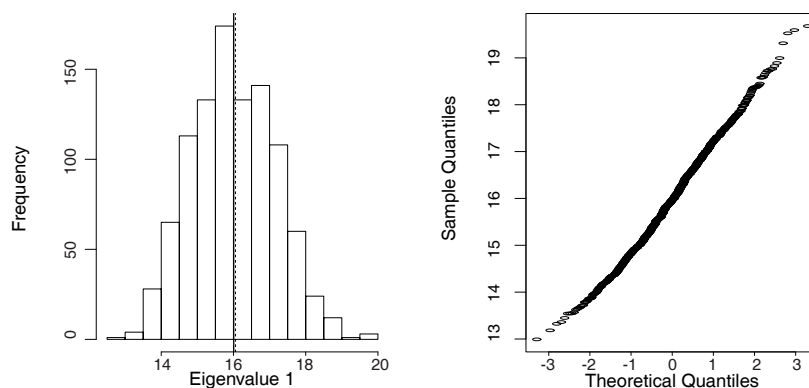


FIGURE 6: Histogram (left) and QQ plot (right) for the first eigenvalue of the simulated data SVD.

4.2. Real-World Data

To illustrate the potentiality of the `biplotbootGUI` package, we use real-world data collected by Anderson (1935) and contained in the R software, which can be loaded once the user installs it. The data set corresponds to the measurements in cm of the variables: sepal length (Y_1) and width (Y_2) and petal length (Y_3) and width (Y_4), for 50 flowers from each of three species of iris. The species are iris setosa, versicolor and virginica. An HJ biplot with the transformation “standardize by columns” is performed. Once again the number of replications entered is 1,000 and the confidence level 95%. First, we show the main results of the HJ biplot. Table 9 presents the percentage (%) of variability explained by each axis, from where the first eigenvalue explains more than 70% and the first three axes explain almost the 100% of the total variability.

TABLE 9: Eigenvalues and variability % explained by each of them for iris data.

| No. | Eigenvalue | Variability | Accumulated variability |
|-----|------------|-------------|-------------------------|
| 1 | 20.85 | 72.96 | 72.96 |
| 2 | 11.67 | 22.85 | 95.81 |
| 3 | 4.68 | 3.67 | 99.48 |

Table 10 provides the RCs of the factor to the column elements in the first three axes. Notice that all the variables are well represented by the first axis, except the variable Y_4 , which is well represented by the second axis. The biplot representation using the first two axes is shown in Figure 7(right). The covariation structure shows a very high correlation between Y_3 and Y_4 represented by an acute angle. Both variables have a high correlation with the variable Y_1 . However, there is no relation with Y_2 due to a right angle is obtained. Table 10 also explains the angles between variables in the plane representing the first two axes. Figure 8 shows the histogram and QQ plot representing the values of the quality of approximation of the 1,000 bootstrap replications.

TABLE 10: RCs of the factors to the columns and angles between variables for iris data.

| Variable | Axis 1 | Axis 2 | Axis 3 | Y_1 | Y_2 | Y_3 | Y_4 |
|----------|--------|--------|--------|-------|--------|--------|--------|
| Y_1 | 793.52 | 130.38 | 76.09 | 0.00 | 95.47 | 20.71 | 18.27 |
| Y_2 | 211.80 | 779.43 | 8.77 | 95.47 | 0.00 | 116.18 | 113.74 |
| Y_3 | 996.44 | 0.56 | 3.00 | 20.71 | 116.18 | 0.00 | 2.44 |
| Y_4 | 936.50 | 4.12 | 59.38 | 18.27 | 113.74 | 2.44 | 0.00 |

Table 11 provides the observed values for the mean, SE, bias, lower-t, upper-t, lower-q and upper-q for simulated and real-world (iris) data. Note that there is no difference between the observed value and its approximation, whereas the endpoints of both intervals are similar. Table 12 provides the RCs to the total variability of the variables based on 1,000 bootstrap replications. Note that there are no differences between observed values and their estimates, whereas the width of the CIs is small suggesting a high accuracy of our methodology. Figure 8 shows the histogram and QQ plot for the RCs to total variability of the variable Y_1 . A similar behavior is observed for the other three variables, whose plots are omitted here, but are available under request for interested users.

TABLE 11: Results of the approximation quality for the indicated data set.

| Data set | Value | Mean | SE | Bias | lower-t | upper-t | lower-q | upper-q |
|-----------|-------|-------|------|------|---------|---------|---------|---------|
| Simulated | 94.27 | 94.46 | 0.75 | 0.19 | 92.98 | 95.95 | 92.9 | 95.80 |
| Iris | 99.48 | 99.49 | 0.08 | 0.01 | 99.34 | 99.64 | 99.33 | 99.62 |

TABLE 12: Results of the contributions to the total variability for iris data.

| Variable | Value | Mean | SE | Bias | lower-t | upper-t | lower-q | upper-q |
|----------|--------|--------|------|-------|---------|---------|---------|---------|
| Y_1 | 250.95 | 250.93 | 0.16 | -0.01 | 250.62 | 251.24 | 250.65 | 251.27 |
| Y_2 | 251.22 | 251.20 | 0.17 | -0.02 | 250.86 | 251.55 | 250.90 | 251.57 |
| Y_3 | 247.96 | 247.99 | 0.31 | 0.03 | 247.39 | 248.59 | 247.36 | 248.53 |
| Y_4 | 249.87 | 249.87 | 0.13 | 0.00 | 249.63 | 250.12 | 249.63 | 250.10 |

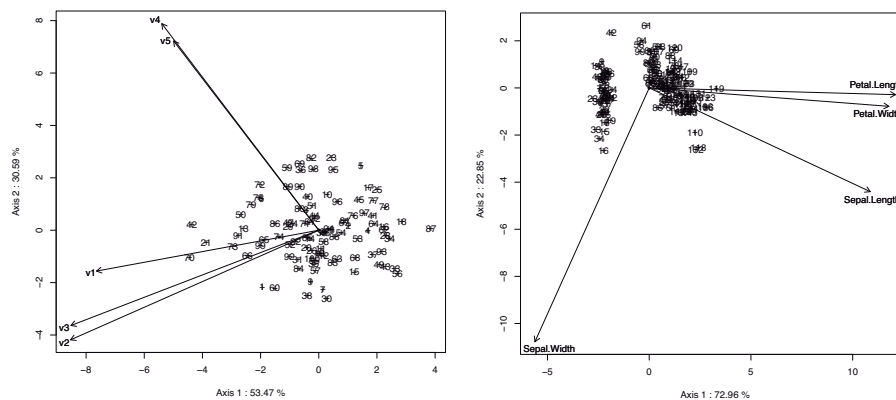


FIGURE 7: Biplots of simulated (left) and iris (right) data sets.

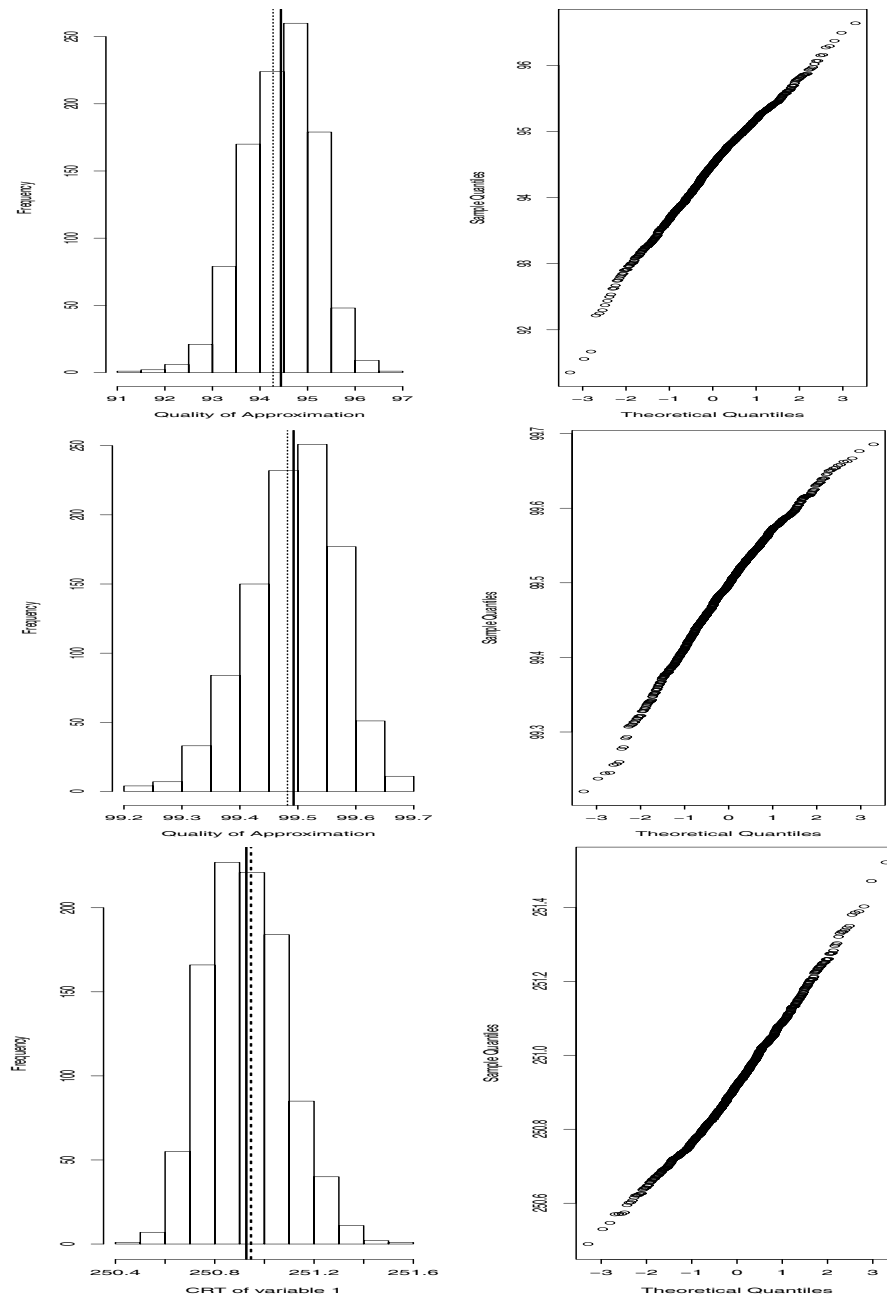


FIGURE 8: Histograms (left) and QQ plots (right) for quality of approximation with simulated (1st panel) and iris (2nd panel) data sets and RCs to total variability of the first variable presented with iris data (3rd panel).

5. Discussion and Conclusions

Factorial analysis techniques only provide to researchers point estimates for their results. In this work, we have proposed a methodology that combines bootstrap and biplots methods to calculate confidence intervals for the results from biplots in order to provide measures of their accuracy. This idea has been applied in several multivariate techniques that incorporate a singular value decomposition. Despite there are some packages in the R software to perform biplots, such as detailed in this paper, these packages only provide estimated results as point values and no information about their accuracy is available. For such a reason, we have developed a new package in this software to implement our methodology.

Specifically, in this paper, we have proposed a graphical methodology based on confidence intervals for the main parameters of biplots based in bootstrapping. These parameters help to interpret the contribution from the elements and axes of the biplot and correspond to goodness of fit, quality of the representation, and variability and relationships among variables. The proposed methodology may be considered as an inferential version of classical biplots and has been implemented in the new `biplotbootGUI` R package. We have detailed the features of this package and validated our methodology with numerical applications based on simulated and real-world data. The numerical results have shown the good performance and potentiality of our methodology, as well as the simple and easy manner to work with the `biplotbootGUI` package. As a supplement to our work, we have also provided a review on the key theoretical contributions and the computational implementations for biplot methods, covering the period from 1971 to the present.

Other ways to calculate measures of accuracy, such as jackknife, Markov chain Monte Carlo and permutation methods, are proposed in the literature, as well as ways to calculate confidence intervals other than the intervals proposed in this paper. In future works, some of these methods may be considered by us to provide different measures of accuracy for the results obtained by biplots methods.

Acknowledgement

The authors wish to thank the Editors of the Special Issue on “Current Topics in Statistical Graphics”, headed by Dr. Fernando Marmolejo-Ramos, the Editor-in-Chief of the journal, Dr. Leonardo Trujillo, and two anonymous referees for their constructive comments on an earlier version of this manuscript which resulted in this improved version. Our research was partially supported by the Chilean Council for Scientific and Technological Research under the project grant FONDECYT 1120879.

[Recibido: mayo de 2014 — Aceptado: octubre de 2014]

References

- Adler, D. & Murdoch, D. (2012), *The rgl R package version 0.92.894: 3D visualization device system (open GL)*, R project.
*cran.r-project.org/package=rgl
- Amaro, I., Vicente-Villardón, J. & Galindo, M. (2004), 'MANOVA biplot for treatment arrays with two factors based on multivariate general linear models', *Interciencia* **29**, 26–32.
- Anderson, E. (1935), 'The irises of the Gaspé peninsula', *Bulletin of the American Iris Society* **59**, 2–5.
- Bickel, P. & Krieger, A. (1989), 'Confidence bands for a distribution function using the bootstrap', *Journal of the American Statistical Association* **84**, 95–100.
- Bradu, D. & Gabriel, K. (1974), 'Simultaneous statistical inference on interactions in two-way analysis of variance', *Journal of the American Statistical Association* **29**, 428–436.
- Bradu, D. & Gabriel, K. (1978), 'The biplot as a diagnostic tool for models of two-way tables', *Technometrics* **20**, 47–68.
- Cárdenas, O. & Galindo, M. P. (2003), *Biplot with External Information based on Generalized Bilinear Models*, Council of Scientific and Humanistic Development of the Central University of Venezuela, Caracas.
- Cárdenas, O., Galindo, M. & Vicente-Villardón, J. (2007), 'Biplot methods: Evolution and applications', *Revista Venezolana de Análisis de Coyuntura* **13**, 279–303.
- Carlier, A. & Kroonenberg, P. (1996), 'Decompositions and biplots in three-way correspondence analysis', *Psychometrika* **61**, 355–373.
- Caro-Lopera, F., Leiva, V. & Balakrishnan, N. (2012), 'Connection between the Hadamard and matrix products with an application to a matrix-variate Birnbaum-Saunders distribution', *Journal of Multivariate Analysis* **104**, 126–139.
- Chatterjee, S. (1984), 'Variance estimation in factor analysis: An application of the bootstrap', *British Journal of Mathematical and Statistical Psychology* **37**, 252–262.
- Chernick, M. (1999), *Bootstrap Methods: A Practitioner's Guide*, Wiley & Sons, New York, US.
- Chessel, D., Dufour, A., Dray, S., Jombart, T., Lobry, J., Ollier, S. & Thioulouse, J. (2013), *The ADE4 R package version 1.5-2: Analysis of ecological data: Exploratory and Euclidean methods in environmental sciences*, R project.
*cran.r-project.org/package=ade4

- Chessel, D., Dufour, A. & Thioulouse, J. (2004), 'The ADE4 R package-I: One-table methods', *R Journal* **4**, 5–10.
- Choulakian, V. (1996), 'Generalized bilinear models', *Psychometrika* **61**, 271–283.
- Daudin, J., Duby, C. & Trécourt, P. (1988), 'Stability of principal components studied by the bootstrap method', *Statistics* **19**, 241–258.
- Del Ferraro, M., Kiers, H. & Giordani, P. (2013), *The ThreeWay R package version 1.1.1: Three-way component analysis*, R project.
*cran.r-project.org/package=ThreeWay
- Demey, J., Vicente-Villardón, J., Galindo, M. & Zambrano, A. (2008), 'Identifying molecular markers associated with classifications of genotypes by external logistic biplot', *Bioinformatics* **24**, 28–32.
- Denis, J. (1991), 'Ajustements de modelles lineaires et bilineaires sous contraintes lineaires avec donnes manquantes', *Statistique Applique* **39**, 5–24.
- Díaz-Faes, A., González-Albo, B., Galindo, M. & Bordons, M. (2013), 'HJ-biplot as tool of matrix inspection for bibliometrical data', *Revista Española de Documentación Científica* **36**, 1–16.
- Díaz-García, J., Galea, M. & Leiva, V. (2003), 'Influence diagnostics for multivariate elliptic regression linear models', *Communications in Statistics: Theory and Methods* **32**, 625–641.
- Díaz-García, J. & Leiva, V. (2003), 'Doubly non-central t and F distribution obtained under singular and non-singular elliptic distributions', *Communications in Statistics: Theory and Methods* **32**, 11–32.
- Díaz-García, J., Leiva, V. & Galea, M. (2002), 'Singular elliptic distribution: Density and applications', *Communications in Statistics: Theory and Methods* **31**, 665–681.
- Dray, S. & Dufour, A. (2007), 'The ADE4 package: Implementing the duality diagram for ecologists', *Journal of Statistical Software* **22**, 1–20.
- Dray, S., Dufour, A. & Chessel, D. (2007), 'The ADE4 package-II: Two-table and K-table methods', *R Journal* **7**, 47–52.
- Edelman, A. (1988), 'Eigenvalues and condition numbers of random matrices', *SIAM Journal on Matrix Analysis and Applications* **9**, 543–560.
- Efron, B. (1979), 'Bootstrap methods: Another look at the jackknife', *The Annals of Statistics* **7**, 1–26.
- Efron, B. (1987), 'Better bootstrap confidence intervals', *Journal of the American Statistical Association* **82**, 171–185.
- Efron, B. (1993), *An Introduction into the Bootstrap*, Chapman and Hall, New York, US.

- Egido, J. (2014), *The dynBiplotGUI R package version 1.0.1: full interactive GUI for dynamic biplot*, R project.
*cran.r-project.org/web/packages/dynBiplotGUI
- Falguerolles, A. (1995), *Generalized Bilinear Models and Generalized Biplots: Some Examples*, Publications du Laboratoire de Statistique et Probabilités. Université Paul Sabatier, Toulouse.
- Faria, J. & Demetrio, C. (2012), *The bpca R package version 1.0-10: Biplot of multivariate data based on principal component analysis*, R project.
*cran.r-project.org/package=bpca
- Frutos, E. & Galindo, M. (2013), *The GGEbiplotGUI R package version 1.0-6: interactive GGE biplots in R*, R project.
*cran.r-project.org/package=GGEbiplotGUI
- Frutos, E., Galindo, M. & Leiva, V. (2014), 'An interactive biplot implementation in R for modeling genotype-by-environment interaction', *Stochastic Environmental Research and Risk Assessment* **28**, 1629–1641.
- Gabriel, K. (1971), 'The biplot graphic display of matrices with application to principal component analysis', *Biometrika* **58**, 453–467.
- Gabriel, K., G. M. . V.-V. J. (1998), Use of biplots to diagnose independence models in three-way contingency tables, in J. Blasius & M. Grenacre, eds, 'Visualization of Categorical Data', Academic Press, London, UK, pp. 391–404.
- Gabriel, K. & Zamir, S. (1979), 'Lower rank approximation of matrices by least squares with any choice of weights', *Technometrics* **21**, 489–498.
- Galindo, M. (1986), 'An alternative for simultaneous representation: HJ-biplot', *Questão* **10**, 12–23.
- Gallego-Álvarez, I., Galindo, M. & Rodríguez-Rosa, M. (2014), 'Analysis of the sustainable society index worldwide: A study from the biplot perspective', *Social Indicators Research* **120**, 29–65.
- García-Sánchez, I., Frías-Aceituno, J. & Rodríguez-Domínguez, L. (2013), 'Determinants of corporate social disclosure in Spanish local governments', *Journal of Cleaner Production* **39**, 60–72.
- Gauch, H. (1988), 'Model selection and validation for yield trials with interaction', *Biometrics* **44**, 705–715.
- Gifi, A. (1990), *Nonlinear Multivariate Analysis*, Wiley, Chichester, UK.
- Gower, J. (1992), 'Generalized biplots', *Biometrika* **79**, 475–493.
- Gower, J., Gardner-Lubbe, S. & Le-Roux, N. (2011), *Understanding Biplots*, Wiley, New York, US.

- Gower, J. & Hand, D. (1996), *Biplots*, Chapman & Hall, London, UK.
- Gower, J. & Harding, S. (1988), 'Nonlinear biplots', *Biometrika* **75**, 445–455.
- Graffelman, J. (2013), *The calibrate R package version 1.7.1: Calibration of scatterplot and biplot axes*, R project.
*cran.r-project.org/package=calibrate
- Greenacre, M. J. (1984), *Theory and Application of Correspondence Analysis*, Academic Press, London.
- Greenacre, M. J. (2010), *Biplots in Practice*, Publications of BBVA Fundation, Spain.
- Greenacre, M. J. & Nenadic, O. (2012), *The ca R package version 0.53: simple, multiple and joint correspondence analysis*.
*cran.r-project.org/package=ca
- Grosjean, P. (2012), *SciViews-R: A GUI API for R*, MONS, Belgium,
www.sciviews.org/SciViews-R.
- Hernández, J. & Vicente-Villardón, J. (2013a), *The NominalLogisticBiplot R package version 0.1: Biplot representations of categorical data*, R project.
*cran.r-project.org/web/packages/NominalLogisticBiplot/index.html
- Hernández, J. & Vicente-Villardón, J. (2013b), *The OrdinalLogisticBiplot R package version 0.2: Ordinal logistic biplots*, R project.
*cran.r-project.org/web/packages/OrdinalLogisticBiplot/index.html
- Hernández, S. (2005), *Robust Biplot*, PhD Dissertation, University of Salamanca, Spain.
- Holmes, S. (1989), 'Using the bootstrap and the RV coefficient in the multivariate context', *Proceedings of the conference on Data Analysis, Learning Symbolic and Numeric Knowledge* pp. 119–131.
- Jambu, M. (1991), *Exploratory and Multivariate Data Analysis*, Academic Press, Orlando, US.
- Kiers, H. (2004), 'Bootstrap confidence intervals for three-way methods', *Journal of Chemometrics* **18**, 22–36.
- La Grange, A., Le-Roux, N. & Gardner-Lubbe, S. (2009), 'Biplotgui: Interactive biplots in R', *Journal of Statistical Software* **30**, 12–37.
- La Grange, A., Le-Roux, N., Rousseeuw, P., Ruts, I. & Tukey, J. (2013), *The biplotGUI R package version 0.0-7: Interactive biplots*, R project.
*cran.r-project.org/package=BiplotGUI
- Lambert, Z., Wildt, A. & Durand, R. (1990), 'Assessing sampling variation relative to number-of-factors criteria', *Educational and Psychological Measurement* **50**, 33–48.

- Lebart, L., Morineau, A. & Piron, M. (1995), *Statistique Exploratoire Multidimensionnelle*, Dunod, Paris, France.
- Leiva, V., Marchant, C., Saulo, H., Aslam, M. & Rojas, F. (2014), ‘Capability indices for Birnbaum-Saunders processes applied to electronic and food industries’, *Journal of Applied Statistics* **41**, 1881–1902.
- L’Hermier des Plantes, H. (1976), *Structuration Des Tableaux A Trois Indices De La Statistique: Theorie et Application d’une Méthode d’Analyse Conjointe*, Master’s thesis, Université Des Sciences et Techniques Du Languedoc, Montpellier.
- Linting, M., Meulman, J. J., Groenen, P. J. F. & Van der Kooij, A. J. (2007), ‘Stability of nonlinear principal components analysis. An empirical study using the balanced bootstrap.’, *Psychological Methods* **12**(3), 359–379.
- Marcenko, V. & Pastur, L. (1967), ‘Distributions of eigenvalues for some sets of random matrices’, *Mathematics of the USSR-Sbornik* **1**, 457–483.
- Markos, A. (2012), *The GUI ca R package version 0.1-4: a Tcl/Tk GUI for the functions*, R project.
*cran.r-project.org/package=caGUI
- Martín-Rodríguez, J., Galindo, M. & Vicente-Villardón, J. (2002), ‘Comparison and integration of subspaces from a biplot perspective’, *Journal of Statistical Planning and Inference* **102**, 411–423.
- McKay, B. D. (1981), ‘The expected eigenvalue distribution of a large regular graph’, *Linear Algebra and Applications* **40**, 203–216.
- Mendes, S., Fernández-Gómez, M., Galindo, M., Morgado, F., Maranhão, P., Azeiteiro, U. & Bacelar-Nicolau, P. (2009), ‘The study of bacterioplankton dynamics in the Berlengas archipelago (west coast of Portugal) by applying the HJ-biplot method’, *Arquipelago Life and Marine Sciences* **26**, 25–35.
- Meulman, J. J. (1982), *Homogeneity Analysis of Incomplete Data*, DSWO Press, Leiden.
- Milan, L. & Whittaker, J. (1995), ‘Application of the parametric bootstrap to models that incorporate a singular value decomposition’, *Applied Statistics* **44**, 31–49.
- Nenadic, O. & Greenacre, M. (2007), ‘Correspondence analysis in R, with two- and three-dimensional graphics: The ca package’, *Journal of Statistical Software* **20**, 1–13.
- Nieto, A., Baccalá, N., Vicente-Galindo, P. & Galindo, M. (2012), *The multi-biplotGUI R package version 0.0-1: Multibiplot analysis*, R project, cran.r-project.org/package=multibiplotGUI.

- Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, B., Simpson, G., Solymos, P., Stevens, M. & Wagner, H. (2013), *The vegan R package version 2.0-8: Community ecology*, R project.
*cran.r-project.org/package=vegan
- Orfao, A., González, M., San-Miguel, J., Ríos, A., Caballero, M., Sanz, M., Calmuntia, M., Galindo, M. & López-Borrascas, A. (1988), 'Bone marrow histopathologic patterns and immunologic phenotype in B-cell chronic lymphocytic leukaemia', *Blut* **57**, 19–23.
- R-Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria., www.R-project.org.
- Ramírez, G., Vásquez, M., Camardiel, A., Pérez, B. & Galindo, M. (2005), 'Graphical detection for the multicollinearity by the h-plot of the inverse matrix of correlations', *Revista Colombiana de Estadística* **28**, 207–219.
- Rivas-Gonzalo, J., Gutiérrez, Y., Polanco, A., Hebrero, E., Vicente-Villardón, J., Galindo, M. & Santos-Buelga, C. (1993), 'Biplot analysis applied to enological parameters in the geographical classification of young red wines', *American Journal of Enology and Viticulture* **44**, 302–308.
- Sánchez, L., Leiva, V., Caro-Lopera, F. & Cysneiros, F. (2015), *On matrix-variate Birnbaum-Saunders distributions and their estimation and application*, Brazilian Journal of Probability and Statistics.
*http://dx.doi.org/10.1214/14-BJPS247 (in press)
- Sepúlveda, R., Vicente-Villardón, J. & Galindo, M. (2008), 'The biplot as a diagnostic tool of local dependence in latent class models: A medical application', *Statistics in Medicine* **27**, 1855–1869.
- Stewart, G. (1980), 'The efficient generation of random orthogonal matrices with application to condition estimators', *SIAM Journal on Numerical Analysis* **17**, 403–409.
- Ter-Braak, C. (1986), 'Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis', *Ecology* **5**, 1167–1179.
- Ter-Braak, C. (1990), 'Interpreting canonical correlation analysis through biplot of structure and weights', *Psychometrika* **55**, 519–531.
- Ter-Braak, C. & Looman, C. (1994), 'Biplots in reduced-rank regression', *Biometrical Journal* **36**, 983–1003.
- Thioulouse, J. & Dray, S. (2007), 'Interactive multivariate data analysis in R with the ade4 and ade4TkGUI packages', *Journal of Statistical Software* **22**, 1–14.
- Thioulouse, J. & Dray, S. (2012), *The ade4TkGUI R package version 0.2-6: ade4 Tcl/Tk graphical user interface*, R project.
*cran.r-project.org/package=ade4TkGUI

- Tierney, L. (2012), *The tkrplot R package version 0.0-23: TK Rplot*, R project.
*cran.r-project.org/package=tkrplot
- Timmerman, M., Kiers, H., Smilde, A. & Stouten, J. (2009), 'Bootstrap confidence intervals in multi-level simultaneous component analysis', *British Journal of Mathematical and Statistical Psychology* **62**, 299–318.
- Tucker, L. (1966), 'Some mathematical notes on three-mode factor analysis', *Psychometrika* **31**, 279–311.
- Vairinhos, V. (2003), Development of a System for Data Mining based on Biplot Methods, PhD Dissertation, University of Salamanca, Spain.
- Vallejo-Arboleda, A., Vicente-Villardón, J. & Galindo, M. (2006), 'Canonical STATIS: Biplot analysis of multi-table group structured data based on STATIS-ACT methodology', *Computational Statistics & Data Analysis* **51**, 4193–4205.
- Vallejo-Arboleda, A., Vicente-Villardón, J., Galindo, M., Fernández, M., Fernández, C. & Bécares, E. (2008), 'Analysis of time evolution for group structured data: Canonical dual statis and doubly multivariate repeated measures model', *Revista Colombiana de Estadística* **31**, 321–340.
- Van Ginkel, J. and Kiers, H. (2011), 'Constructing bootstrap confidence intervals for principal component loadings in the presence of missing data: A multiple-imputation approach', *British Journal of Mathematical and Statistical Psychology* **64**, 498–515.
- Vicente-Villardón, J. (2010), *MULTBILOT: A Package for Multivariate Analysis using Biplots*, Matlab software.
*biplot.usal.es/ClassicalBiplot/index.html
- Vicente-Villardón, J., Galindo, M. & Blázquez, A. (2006), *Logistic Biplots*, Chapman & Hall, New York, US.
- Viloria, J., Gil, J., Durango, D. & García, C. (2012), 'Physicochemical characterization of propolis from the region of Bajo Cauca Antioqueño (Antioquia, Colombia)', *Biotechnología en el Sector Agropecuario y Agroindustrial* **10**, 77–86.
- Wachter, K. (1978), 'The strong limits of random matrix spectra for sample matrices of independent elements', *The Annals of Probability* **6**, 1–18.
- Yan, W., Hunt, L., Sheng, Q. & Szlavnick, Z. (2000), 'Cultivar evaluation and mega-environment investigation based on GGE biplot', *Crop Science* **40**, 597–605.
- Yan, W. & Kang, M. (2003), *GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists*, CRC Press, Boca Raton, US.