VNiVERSiDAD
D SALAMANCA

HJ – BIPLOT AUMENTADO

MARTA MARCOS HIDALGO

DIRECTOR:

DR. D. JOSÉ LUIS VICENTE VILLARDÓN

Universidad de Salamanca

Departamento de Estadística

UNIVERSIDAD DE SALAMANCA
Departamento de Estadística

HJ – BIPLOT AUMENTADO

Trabajo de investigación correspondiente al Programa de Doctorado en Estadística Multivariante Aplicada; presentado por:

Marta Marcos Hidalgo

Salamanca, 2011

Vº Bº

Director

ÍNDICE

INTRODUCCIÓN

CAPÍTULO 1: FUNDAMENTOS TEÓRICOS DE LOS MÉTODOS BIPLOT

1.1 Introducción

1.2 Definición

1.3 Fundamentación Teórica

1.4 Biplots Clásicos

CAPÍTULO 2: BIPLOTS CLÁSICOS

2.1 Introducción

2.2 GH – Biplot

2.3 JK – Biplot

2.4 SQRT – Biplot

2.5 Bondad de Ajuste

CAPÍTULO 3: HJ – BIPLOT

3.1 Introducción

3.2 Definición

3.3 Bondad de ajuste

3.4 HJ – Biplot: Aplicaciones en distintos Campos

CAPÍTULO 4: HJ – BIPLOT AUMENTADO

4.1 Introducción

4.2 Propiedades de los marcadores del HJ - Biplot

4.3 HJ – Biplot: Representación clásica

4.4 Representación del HJ – BIPLOT AUMENTADO

4.5 Algoritmo HJ – BIPLOT Aumentado

CAPÍTULO 5: IMPLEMENTACIÓN CON MATLAB

5.1 Introducción

5.2 Código fuente del HJ – Biplot

5.3 Implementación del código de la Representación Alternativa.

BIBLIOGRAFÍA

INTRODUCCIÓN

Las técnicas de representación simultánea de una matriz de datos $X_{n\times p}$, siendo n el número de individuos y p el de variables; son de gran interés en Análisis Multivariante. Las más utilizadas son, Análisis Factorial de Correspondencias y el método Biplot.

El Análisis Factorial de Correspondencias, considera dos nubes, $N_I \; y \; N_J$ de los perfiles fila y columna, con masa iguales a los respectivos totales marginales, en los espacios multidimensionales $R^p \; y \; R^n$, ambos espacios estructurados por la métrica ji-cuadrado, la cual permite la determinación unívoca de sucesivos ejes principales de inercia.

El método Biplot propuesto por Gabriel, consiste en interpretar los elementos de la matriz de datos $X_{n\times p}$, como productos internos de vectores, los cuales se utilizan para representar las filas y las columnas de la matriz de datos.

En este trabajo se hace una revisión de las distintas técnicas clásicas de la representación Biplot, esto es:

- GH – Biplot
- JK – Biplot
- HJ – Biplot

Biplot permite representar en un subespacio de baja dimensión, las relaciones entre variables e individuos en el espacio multivariante de partida; proyectando dicho espacio multivariante en un subespacio de dimensión dos o tres.

Pero, ¿esta proyección en el subespacio de baja dimensión refleja la verdadera relación entre variables e individuos del espacio de partida?

En este trabajo se propone la construcción de un Biplot, partiendo del HJ – Biplot, por ser una representación simultánea en sentido estricto, con alta calidad de representación tanto para filas como para columnas, en el que poder juzgar si la representación en baja dimensión es una buena representación, y por tanto interpretable.

Al nuevo gráfico Biplot que obtendremos, lo denominaremos:

HJ – Biplot Aumentado

En el que podremos decidir si es una buena representación Biplot y qué variables presentan una mejor calidad de representación simplemente observando el gráfico.

CAPÍTULO 1

FUNDAMENTOS TEÓRICOS DE LOS MÉTODOS BIPLOT

1.1 INTRODUCCIÓN

Un BIPLOT (GABRIEL, 1971) es una representación gráfica de datos multivariantes. De la misma manera que un diagrama de dispersión muestra la distribución conjunta de dos variables, un BIPLOT representa tres o más variables (GABRIEL y ODOROFF, 1990).

El BIPLOT aproxima la distribución de una muestra multivariante en un espacio de dimensión reducida, normalmente de dimensión dos, y superpone sobre la misma representaciones de las variables sobre las que se mide la muestra (GOWER 1996). Las representaciones de las variables son normalmente vectores, y coinciden con las direcciones en las que mejor se muestra el cambio individual de cada variable.

1.2 DEFINICIÓN

Consideramos la matriz de datos $X_{n\times p}$, siendo n el número de individuos y p el de variables. Suponemos $p < n$, y $rg(X) = r \leq p$.

Sin pérdida de generalidad, podemos suponer centradas las columnas de la matriz X.

Un Biplot para una matriz de datos X es una representación gráfica mediante marcadores $g_1, \dots, g_n$ para las filas de X y $h_1, \dots, h_p$ para las columnas de X, de forma que el producto interno $g_i' h_j$ aproxime de la mejor manera posible al elemento x_{ij} de la matriz de partida, con $i = 1, \dots, n$ y $j = 1, \dots, p$.

Si consideramos las matrices $G\ y\ H$, formadas por los marcadores fila g_i y los marcadores columna h_j, respectivamente; entonces podemos expresar lo anterior:

$$X \cong GH'$$

1.3 FUNDAMENTACIÓN TEÓRICA

El método Biplot está basado en la descomposición en valores y vectores singulares de la matriz X, esto es:

$$X = U_{n\times r}\Sigma_r\, V'_{r\times p}$$

Siendo:

$U_{n\times r}$, matriz cuyas columnas contienen los vectores propios de XX'

$V_{r\times p}$, matriz cuyas columnas contienen los vectores propios de $X'X$

$\Sigma_r = diag(\lambda_1, \dots, \lambda_r)$, con $\lambda_1 \geq \cdots \geq \lambda_r \geq 0$ valores propios de X

Las matrices anteriores verifican:

$$U'U = V'V = I_r$$

con I_r matriz identidad de orden r.

Por lo tanto la mejor aproximación de X en rango q, que denotamos por $X_{(q)}$, es:

$$X_{(q)\,n\times p} = U_{(q)\,n\times q}\Sigma_{(q)q}\, V'_{(q)\,q\times p} = \sum_{k=1}^{q} \lambda_k\, u_k\, v'_k$$

Siendo,

$U_{(q)}\ y\ V_{(q)}$, matrices construidas con las q primeras columnas de $U\ y\ V$, respectivamente.

$\Sigma_{(q)}$, matriz diagonal que contiene a los q mayores valores propios distintos de cero de la matriz X.

Para garantizar la unicidad en la representación, se realiza una factorización tal que:

$$X = (U\Sigma^{\gamma})\left(V\Sigma^{1-\gamma}\right)' = AB' \qquad (1)$$

Por lo tanto,

$$X = AB' = U\Sigma V'$$

La selección de las distintas factorizaciones nos llevará a los distintos tipos de marcadores y por tanto a distintos Biplots.

Sin embargo, la interpretación Biplot siempre se realiza a partir de los productos escalares, independientemente de la factorización elegida.

1.4 BIPLOTS CLÁSICOS

La selección de distintas factorizaciones conduce a distintos tipos de marcadores y por lo tanto a los distintos Biplots. Sin embargo, la interpretación de los mismos siempre se realiza a partir de los productos escalares, independientemente de la factorización elegida.

En el caso de la métrica identidad, se pueden elegir los marcadores de acuerdo a los distintos valores de γ en la descomposición (1),

$$A = U\Sigma^{\gamma} \qquad B = V\Sigma^{1-\gamma}$$

En la que según el valor de γ ($\gamma = 0, 1, 1/2$), tenemos los Biplots Clásicos, Gabriel (1971, 1981a):

GH – Biplot ($\gamma = 0$):

$$A = U \qquad B = V\Sigma$$

JK – Biplot ($\gamma = 1$):

$$A = U\Sigma \qquad B = V$$

SQRT – Biplot ($\gamma = 1/2$):

$$A = U\Sigma^{1/2} \qquad B = V\Sigma^{1/2}$$

CAPÍTULO 2

BIPLOTS CLÁSICOS

2.1 INTRODUCCIÓN

En este capítulo estudiaremos los Biplots Clásicos que hemos mencionado en el capítulo anterior.

2.2 GH – BIPLOT

Como hemos visto en el capítulo anterior, en la descomposición (1) haciendo $\gamma = 0$, tenemos la factorización:

$$A = U \qquad\qquad B = V\Sigma$$

Donde se verifica lo siguiente:

$$X'X = (AB')'(AB') = BA'AB' = BB'$$

ya que $A'A = U'U = I$.

Por lo tanto, este tipo de Biplot preserva la métrica euclídea usual entre las columnas pero no entre las filas, obteniéndose por tanto alta calidad de representación para las columnas, pero no así para las filas.

El nombre de este Biplot (GH-Biplot) que presentamos se debe a que Gabriel adoptó la notación G para los marcadores fila y H para los marcadores columna.

$$G = U \qquad\qquad H = V\Sigma$$

Greenacre en 1984, introduce una nueva terminología para el GH - Biplot, denominándolo **CMP - Biplot** (**Column Metric Preserving**) en referencia al hecho de preservación de la métrica para las columnas.

2.3 JK – BIPLOT

Si en la descomposición (1) consideramos $\gamma = 1$, obtenemos la siguiente factorización:

$$A = U\Sigma \qquad B = V$$

Verificándose,

$$XX' = (AB')(AB')' = AB'BA' = AA'$$

ya que $B'B = V'V = I$.

Por lo que queda probado que este tipo de Biplot preserva la métrica euclídea usual entre las filas, pero no entre las columnas, obteniéndose alta calidad de representación para las filas.

Gabriel, a este tipo de Biplot, lo denominó JK-Biplot, ya que utilizó J para denotar la matriz de marcadores fila y K para la matriz de marcadores columna.

$$J = U\Sigma \qquad K = V$$

Greenacre en 1984, introduce una nueva terminología para el JK – Biplot, denominándolo **RMP - Biplot** (**Row Metric Preserving**), en referencia al hecho de que este Biplot preserva de la métrica para las filas.

2.4 SQRT – BIPLOT

Considerando en la descomposición (1) $\gamma = 1/2$, obtenemos la siguiente factorización, que asigna papeles simétricos tanto a filas como a columnas:

$$A = U\Sigma^{1/2} \qquad B = V\Sigma^{1/2}$$

En este caso, la factorización no es única, puesto que:

$$A'A = \left(U\Sigma^{1/2}\right)'\left(U\Sigma^{1/2}\right) = \Sigma^{1/2}U'U\Sigma^{1/2} = \Sigma \neq I$$

Análogamente, para B se tiene que $B'B \neq I$.

Por la falta de unicidad en la factorización, este método sólo es utilizado en aquellos casos en los que el objetivo del análisis consiste fundamentalmente en la aproximación de los elementos de la matriz X.

2.5 BONDAD DE AJUSTE

Dada una matriz X de rango r $(r > 2)$, la bondad de ajuste en el plano viene dada por la siguiente expresión,

$$\rho_{(2)} = \frac{\lambda_1 + \lambda_2}{\sum_{t=1}^{r} \lambda_t}$$

Siendo $\lambda_t, t = 1, \dots, r$ los auto-valores de $X'X$.

Pero no nos interesa sólo conocer la bondad de ajuste para la matriz de datos, sino que sobre todo nos interesa conocer cuál es la bondad de ajuste para las filas (individuos) y para las columnas (variables).

En la siguiente tabla se muestran las distintas calidades de representación:

Representación Simultánea	**BONDAD DE AJUSTE**	
	FILAS (Individuos)	**COLUMNAS (Variables)**
GH – Biplot	$\frac{2}{r}$	$\frac{\lambda_1^2 + \lambda_2^2}{\sum_{t=1}^{r} \lambda_t^2}$
HJ – Biplot	$\frac{\lambda_1^2 + \lambda_2^2}{\sum_{t=1}^{r} \lambda_t^2}$	$\frac{2}{r}$

Por tanto, aunque Gabriel señala expresamente que el nombre de método BIPLOT se debe a que la técnica permite la representación conjunta de todas las líneas de la matriz, esta representación **no es una representación simultánea en sentido estricto** puesto que la bondad de ajuste no es la misma para las filas que para las columnas.

CAPÍTULO 3

HJ BIPLOT

3.1 INTRODUCCIÓN

El HJ-Biplot, propuesto por Galindo en 1985, es una representación gráfica multivariante de marcadores fila y columna, elegidos de forma que se puedan superponer en el mismo sistema de referencia con máxima calidad de representación.

El objetivo principal de este Biplot, es el de conseguir una interpretación simultánea de las relaciones entre individuos y variables, filas y columnas respectivamente, de la matriz de datos X.

3.2 DEFINICIÓN

El HJ-Biplot es una representación gráfica multivariante de las líneas de una matriz $X_{n\times p}$ mediante los marcadores $j_1, \dots, j_n$ para sus filas y $h_1, \dots, h_p$ para sus columnas, elegidos de forma que ambos marcadores puedan ser superpuestos en un mismo sistema de referencia con máxima calidad de representación. (Galindo, 1985; Galindo y Cuadras, 1986)

En la descomposición en valores y vectores singulares de la matriz X, que siguiendo la notación de capítulos anteriores es:

$$X = U_{n\times r} \Sigma_r V'_{r\times p}$$

Se elige la factorización:

$$A = U\Sigma \qquad B = V\Sigma$$

Obtenemos lo que Galindo denominó HJ-Biplot, ya que siguiendo la nomenclatura de Gabriel, adoptó la siguiente notación para las matrices de marcadores fila y columna:

$$J = U\Sigma \qquad H = V\Sigma$$

La elección de tales marcadores permite representar las coordenadas de las filas y de las columnas con referencia a unos mismos ejes cartesianos, es decir los espacios de representación de las filas y de las columnas pueden ser superpuestos para obtener una representación conjunta, ya que los auto-valores para los análisis de las dos nubes permite identificar los ejes factoriales correspondientes al mismo valor propio y representar sobre el mismo gráfico ambas nubes. Es decir:

"Los marcadores para las columnas de una matriz cualquiera según un GH – Biplot y los marcadores para las filas según un JK – Biplot pueden ser representados sobre un mismo sistema de referencia, que será el sistema de los ejes factoriales, cuyo origen coincide con el punto de equilibrio de las nubes"

En efecto,

Teniendo en cuenta las relaciones que ligan a $U\,y\,V$ podemos escribir:

$$U\Sigma = XV\Sigma^{-1}\Sigma = XV$$

De donde deducimos que $U\Sigma$ coincide la proyección de los n puntos que representan a las filas sobre el espacio de mejor ajuste a esa nube de puntos en el sentido de los mínimos cuadrados. Por lo que los marcadores para la filas de un JK – Biplot coinciden con las coordenadas de las filas respecto a los ejes factoriales.

De manera análoga,

$$\Sigma V = \Sigma\Sigma^{-1}X'U = X'U$$

Por lo que ΣV coincide con la proyección de los puntos que representan a las variables sobre el espacio de máxima inercia, es decir los marcadores para las columnas de un GH – Biplot coinciden con las coordenadas de las columnas respecto a los ejes factoriales.

Mediante las relaciones anteriores y haciendo uso del álgebra matricial es sencillo comprobar que los espacios de representación de las filas y de las columnas pueden ser superpuestos para obtener una representación conjunta. A la que denominaremos, según Galindo (1996), Representación Simultánea HJ o HJ – Biplot.

Debido al hecho de que el HJ-Biplot realiza una representación simultánea tanto de filas como de columnas, también recibe el nombre de **RCMP – Biplot** (**Row Column Metric Preserving**).

3.3 BONDAD DE AJUSTE

Según la elección de marcadores del HJ – Biplot, marcadores para las columnas según un GH – Biplot y los marcadores para las filas según un JK – Biplot, se consigue la **máxima calidad de representación tanto para filas como para columnas**, y como ya demostramos en el Capítulo 2, será en ambos casos:

$$\frac{\lambda_1^2 + \lambda_2^2}{\sum_{t=1}^{r} \lambda_t^2}$$

En la representación HJ – Biplot, por las propiedades mencionadas anteriormente, es posible representar las relaciones existentes entre variables e individuos proyectando en un subespacio de dimensión dos o tres.

Estas proyecciones no siempre consiguen capturar los patrones implícitos en los datos y son necesarios varios índices para conocer la bondad de ajuste y decidir qué variables son las que verdaderamente contribuyen a la clasificación exhibida en el gráfico factorial. Sin embargo, es posible interpretar la proximidad entre los individuos en términos de similitud entre ellos, y la proximidad entre las variables en términos de covariación.

Igualmente, podemos interpretar la distancia de un individuo a una variable en el sentido de que una variable próxima a un individuo indica que esa variable ha tomado un valor alto en ese individuo.

3.4 APLICACIONES DE LOS MÉTODOS BIPLOT EN DISTINTOS CAMPOS

Los métodos Biplot, entre los métodos de análisis multivariante de datos, han sido fuente de continuas contribuciones a la ciencia desde su origen en 1971. En particular el HJ –Biplot lo ha sido desde que fue propuesto por Galindo en 1985.

En diferentes campos de las ciencias, tanto naturales como sociales, se han hecho múltiples contribuciones gracias al uso de los métodos Biplot para conseguir la aproximación de una matriz de datos y poder obtener su descripción a través de mapas geométricos en subespacios de baja dimensión con ajuste óptimo.

Los campos de aplicación son diversos, desde la medicina hasta la economía, pasando por campos tan dispares como la enología o la minería de datos.

Entre los muchos artículos en los distintos campos podemos citar los siguientes:

"El HJ – Biplot como base para el análisis de la evolución del índice de producción industrial en la Comunidad Europea en los últimos años", S. Vicente Tavera, M.P. Galindo Villardón (1992).

En el que mediante el uso de distintas técnicas multivariantes entre las que está el HJ – Biplot, se estudia la evolución del mencionado índice en los países de la Comunidad Europea haciendo uso del Índice de Producción Industrial (IPI), ya que éste puede expresar la evolución física del Producto Interior Bruto referido a la Industria.

"Biplot analysis applied to enological parameters in the geographical classification of young red wines", J.C. Rivas-Gonzalo et al. (1993).

Se realiza un estudio de 45 vinos tintos jóvenes de la D.O. Ribera de Duero y D.O. Toro y se obtienen mediciones de 15 variables (grado alcohólico, acidez, pH,…). Se emplea el HJ – Biplot para analizar si con los datos obtenidos es posible clasificar a partir de los valores de dichas variables el vino según su denominación de origen.

"El HJ – Biplot como base para la búsqueda de Clusters en función de la distribución de parados según profesiones en la Comunidad de Castilla y León", S. Vicente Tavera, G. Ramirez, P. Galindo (1994).

En 1992, Vicente-Tavera utiliza el HJ – Biplot para realizar una clasificación ascendente jerárquica, conformando los conglomerados con la información obtenida del HJ y utilizando el criterio de la inercia descrito por Benzecri (1973).

En este artículo se pone de manifiesto que la información resultante de la representación geométrica de una matriz de datos multivariante puede ser tomada como base para la obtención de una clasificación ascendente cuyas clases, si bien pueden no verificar las propiedades teóricas óptimas de compacidad y/o separabilidad, proporcionan clases fácilmente interpretables, para las cuales se pueden conocer las variables responsables de esa configuración. HJ –Biplot es la técnica de representación simultánea en la que se basa dicho estudio.

"Alternativa al análisis canónico de correspondencias basada en los métodos Biblot", Mª José Fernández Gómez, M.P. Galindo, I. Barrera, J.L. Vicente-Villardón, A. Martin. (1996)

En este trabajo, se propone el HJ-Biplot como alternativa al Análisis Canónico de Correspondencias (CCA) (ter BRAAK, 1986) en estudios de comunidades ecológicas. Concluyéndose que el HJ-Biplot proporciona mejores resultados que el CCA ya que obtiene la misma calidad de representación para filas y columnas. Se realiza una aplicación a datos reales, y se determinan los óptimos de las especies respecto de las variables analizadas y el ranking de las mismas en relación al gradiente para dichas variables.

"Análisis HJ – Biplot de la evolución de la productividad agraria de la Comunidad de Castilla y León a lo largo del quinquenio 1991-1995", A. Dorado, S. Vicente, A. Blázquez y J. Martín (1999).

En el que se analizan las estructuras de similaridad y los patrones de covariación obtenidos a partir de la representación HJ-Biplot de las matrices de datos correspondientes a cada año del análisis, de 1991 a 1995 en las provincias de Castilla y León y se comparan con los obtenidos en el análisis global del quinquenio para describir la evolución de los patrones.

“El análisis HJ – Biplot como herramienta de estudio de un ecosistema acuático”. Galindo, M.P., Fernández-Gómez, M.J., Ávila, C., del Río, A., Fernández-Aláez, M. (1999).

> *En el que se pone de manifiesto la utilidad del análisis HJ – Biplot como herramienta de inspección de matrices de datos ecológicos. Aplicándose al estudio de la sucesión fitoplanctónica en el embalse de Riaño (León), poniendo de manifiesto mediante la interpretación del gráfico factorial, la sucesión tanto temporal como vertical en la columna de agua de los diferentes grupos taxonómicos de algas encontrados en este embalse.*

“Desarrollo de un sistema para Minería de Datos basado en los Métodos Biplot.”. Vairinhos V.M (2003).

> *Se demuestra que es posible y deseable usar las representaciones Biplot como núcleo de un sistema de minería de datos; ya que cuando la matriz de datos es muy extensa, es necesario que el ordenador colabore activamente con el analista sugiriendo interpretaciones para los resultados. El resultado de esa investigación ha sido una formulación matemática del problema de interpretación como un problema de aproximación de conjuntos de*

individuos/observaciones (representando el significado de grandes familias de resultados) por los significados elementales o atómicos asociados a las variables observadas y sus valores. El problema así formulado puede ser representado por un grafo de intersección, lo que permite desarrollar algoritmos, que generan sugestiones de interpretación para resultados expresados por conjuntos de individuos.

"Estilos de Personalidad y Desadaptación Social durante la Adolescencia", Paula Alarcón, Eugenia Vinet, Sonia Calvo (2005).

Estudio cuyo objetivo es el de identificar las características de personalidad en adolescentes reincidentes en conductas antisociales. Para ello se administró el Inventario Clínico para Adolescentes de Millon MACI (Millon, 1993) a una muestra de 86 adolescentes varones chilenos infractores de Ley. Los datos obtenidos fueron analizados mediante la combinación de un análisis de conglomerados y un análisis multivariado HJ-Biplot, resultado en la detección cinco agrupaciones de perfiles de personalidad. A estas agrupaciones se les aplicó un análisis HJ – Biplot,

obteniéndose buena calidad de representación de las variables involucradas en el análisis a través de la longitud de sus vectores. Las variables cuyos vectores son más largos se corresponden con mayor varianza inter-sujetos y por tanto describen mejor al grupo con el cual están asociadas al plano. Las variables con un menor ángulo entre sí son las que están más correlacionadas entre sí.

A partir de este análisis se obtienen las variables eje responsables de las agrupaciones, dando origen a perfiles de personalidad en base a escalas de patrones, preocupaciones y síndromes clínicos.

"Study of the evolution of air pollution in Salamanca (Spain) along a five-year period (1994‑1998) using HJ – Biplot simultaneous representation analysis", J.M. González Cabrera et al. (2006).

La información recogida en dos estaciones monitorizadas en la ciudad de Salamanca de distintas variables referentes a la contaminación atmosférica fue analizada y representada mediante HJ – Biplot con el fin de estudiar las posibles relaciones entre las distintas variables y poder profundizar en la evolución temporal de la contaminación en dicha ciudad.

"Suicidio Juvenil: Sociología de una realidad social", Mª Elena Pinto Rodríguez, Fernando Álvarez – Uría Rico (2006).

> *En el que mediante la revisión y análisis realizado a los datos consultados en el Instituto Nacional de Estadística (I.N.E.), en torno a las estadísticas de suicidio en España, ocurridos en el último medio siglo (1951-2000), haciendo una inferencia muy particular en el suicidio juvenil, se hace uso de las representaciones Biplot y en particular la representación HJ – Biplot para mostrar los distintos resultados obtenidos. Destacando las que representan el análisis en cuanto a las principales profesiones de los suicidas, el medio utilizado en los suicidios, la causa de los mismos y los grupos de edades.*

"Caracterización Multivariante de los perfiles de las mujeres en situación laboral irregular: el caso de Salamanca", M.P. Galindo Villardón, P. Vicente Galindo, C. Patino, J.L. Vicente (2007).

> *En el que se presenta un estudio cuyo objetivo principal es el de identificar los perfiles socioeconómicos de las mujeres salmantinas en situación de irregularidad laboral, en comparación con las mujeres que desarrollan su actividad en un marco laboral regular. HJ – Biplot es la herramienta*

utilizada para detectar los municipios con características socioeconómicas similares, siendo esto la base del estudio de empleo irregular en esta provincia.

"The study of bacterioplankton dynamics in the Berlengas Archipelago (West coast of Portugal) by applying the HJ – Biplot method", Susana Mendes (2009).

En el que se utiliza el HJ – Biplot para el estudio de la relación entre bacterioplankton y el impacto ambiental en el Archipiélago Berlengas (Costa Occidental de Portugal) que fue estudiada entre Febrero de 2006 y febrero de 2007 con dos estaciones de muestreo: Berlenga y Canal de Panamá.

"Inferencia Ecológica para la Caracterización de Abstencionistas: El caso de Portugal", Eugénia Castela, M. P. Galindo Villardón (2010).

En éste se estudio se aplican los métodos de Inferencia Ecológica (IE), métodos estadísticos que permiten estimar las celdas de una tabla de contingencia, cuando solo son conocidos los respectivos totales marginales. Y donde a partir del modelo de King (1997) y considerando las elecciones legislativas ocurridas en Portugal en 2002 y 2005, trata de encontrar los coeficientes de estabilidad (los

ciudadanos mantienen la misma actitud ante el voto en las dos elecciones, es decir, cuando votan o se abstienen, en las dos elecciones, consecutivamente) e inestabilidad electoral (los ciudadanos votan en una de ellas y se abstienen en la otra, independientemente del orden) para todos y cada uno de los municipios de Portugal.

En el caso de Portugal, el método de King no ha producido buenas estimaciones.

A partir del método HJ-Biplot se persigue obtener la homogeneidad espacial, en términos de tendencias políticas mayoritarias en las dos elecciones del estudio, proponiendo una "reorganización" territorial partidista/abstencionista.

La "reorganización" territorial, proporciona 6 grupos de provincias a los cuales se aplica el modelo de King para encontrar el porcentaje de electores que han votado y se han abstenido en las dos elecciones, así como el porcentaje de electores fluctuantes, es decir, los electoresque han votado en una de las elecciones y se abstenido en la otra.

CAPÍTULO 4

HJ – BIPLOT AUMENTADO

4.1 INTRODUCCIÓN

Como ya se ha mencionado en el Capítulo 3, el HJ – BIPLOT, permite representar las relaciones existentes entre variables e individuos proyectando en un subespacio de dimensión dos o tres. Estas proyecciones no siempre consiguen capturar los patrones implícitos en los datos y son necesarios varios índices para conocer la bondad de ajuste y decidir qué variables son las que verdaderamente contribuyen a la clasificación exhibida en el gráfico factorial.

En este trabajo proponemos una modificación del HJ – BIPLOT que llamaremos HJ – Biplot Aumentado, siguiendo la terminología de Bartkowiak y Szuztalewiez, con la que es posible conocer esta información directamente desde el gráfico.

4.2 PROPIEDADES DE LOS MARCADORES DEL HJ – BIPLOT

1) La matriz HH' reproduce, salvo una constante igual a $\frac{1}{n-1}$ la matriz de covarianzas S de la matriz de datos, esto es $S = \frac{1}{n-1} X'X$.

En efecto,

Sea $X = U\Sigma V'$ la descomposición en valores y vectores singulares de la matriz X.

Entonces, tenemos lo siguiente:

$$S = \frac{1}{n-1} X'X \Rightarrow S = \frac{1}{n-1} (U\Sigma V')'(U\Sigma V') \Rightarrow$$

$$S = \frac{1}{n-1} V\Sigma U'U\Sigma V' = \frac{1}{n-1} V\Sigma\Sigma V \Rightarrow S = \frac{1}{n-1} (\Sigma V)'(\Sigma V)$$

Por lo tanto,

$$S = \frac{1}{n-1} HH'$$

c.q.d.

De esta propiedad, deducimos que el elemento S_{ij} (fila i-ésima, columna j-ésima) de la matriz S, puede ser calculado de la siguiente manera:

$$S_{ij} = \frac{1}{n-1} h_i' h_j$$

De donde, teniendo en cuenta la definición de producto escalar, obtenemos:

$$S_{ij} = \frac{1}{n-1} \|h_i\| \|h_j\| \cos(\alpha_{ij}), \text{ siendo } \alpha_{ij} = (\widehat{h_i, h_j})$$

En el caso particular, en el que $i = j$ tenemos lo siguiente:

$$S_{jj} = S_j^2 = \frac{1}{n-1} \|h_j\|^2, \ j = 1, \dots, p$$

Lo que significa que el cuadrado de la longitud del vector $h_j, \ j = 1, \dots, p$, es proporcional a la varianza de la variable $j - ésima$, donde el coeficiente de proporcionalidad es $n - 1$ para todo $j = 1, \dots, p$.

2) Si consideramos la descomposición HJ-Biplot sobre la matriz $\bar{\boldsymbol{X}} = (\bar{\boldsymbol{x}}_{\boldsymbol{ij}})$ ***donde*** $\bar{\boldsymbol{x}}_{\boldsymbol{ij}} = \frac{\boldsymbol{x_{ij} - x_{\bullet j}}}{\boldsymbol{S_j\sqrt{n-1}}}$ ***, siendo*** $\boldsymbol{x_{\bullet j}} = \frac{\boldsymbol{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x_{ij}}$ ***con*** $\boldsymbol{j = 1, \dots, p}$***; tenemos que*** $\boldsymbol{HH' = R}$***, siendo R la matriz de correlaciones de las variables consideradas.***

En efecto,

Sea $Z = \bar{X}'\bar{X}$, entonces:

$$z_{jk} = \frac{\sum_{i=1}^{n}(x_{ij} - \bar{x}_{\bullet j})(x_{ik} - \bar{x}_{\bullet k})}{S_j S_k (n-1)} = r_{jk}$$

Luego efectivamente coincide con el coeficiente de correlación entre la variable $j - ésima$ y la variable $k - ésima$. Por lo tanto,

$$R = \bar{X}'\bar{X}$$

Teniendo en cuenta la descomposición considerada, HJ – Biplot sobre $\bar{X}$, se verifica:

$$R = \bar{X}'\bar{X} = (U\Sigma V')'(U\Sigma V') = V\Sigma U'U\Sigma V' = V\Sigma\Sigma V' = (V\Sigma)(V\Sigma)'$$

$$\Rightarrow R = H'H$$

c.q.d.

De esta segunda propiedad deducimos dos importantes propiedades de los marcadores para las columnas, h_j , $j = 1, \dots, p$.

Si $j = k$, tenemos que $r_{jj} = h_j' h_j = \left\| h_j \right\|^2 = 1$, por lo que los vectores h_j, son de módulo unidad, para todo $j = 1, \dots, p$.

Si $j \neq k$, tenemos que $r_{jk} = h_j' h_k = \left\| h_j \right\| \| h_k \| \cos(\alpha_{jk})$, siendo $\alpha_{jk} = (\widehat{h_j,\ h_k})$. Lo que significa que el coseno del ángulo formado por los vectores $h_j \ y \ h_k \ \ con \, j \neq k$, coincide con el coeficiente de correlación r_{jk}.

4.3 HJ – BIPLOT: REPRESENTACIÓN CLÁSICA

En la representación Biplot, utilizamos la información que aparece en la descomposición de la matriz X.

En realidad sólo se hace uso de la información contenida en las dos primeras columnas de la matriz H y en las dos primeras filas de la matriz J, es decir, utilizamos la siguiente aproximación:

$$X_{n\times p}^{(2)} = \begin{pmatrix} j_{11} & j_{12} \\ j_{21} & j_{22} \\ \cdots & \\ j_{n1} & j_{n2} \end{pmatrix} \begin{pmatrix} h_{11} & h_{21} & \ldots & h_{p1} \\ h_{12} & h_{22} & \ldots & h_{p2} \end{pmatrix}$$

$$\Rightarrow \quad X_{n\times p}^{(2)} = \begin{pmatrix} j_1^{(2)} \\ j_2^{(2)} \\ \vdots \\ j_n^{(2)} \end{pmatrix} \begin{pmatrix} h_1^{(2)} & h_2^{(2)} & \ldots & h_p^{(2)} \end{pmatrix}$$

Siendo $j_k^{(2)} = (j_{k1} \quad j_{k2}),\ k = 1, \ldots, n$ la aproximación en dimensión dos de los marcadores fila de la matriz X; por lo que contienen las coordenadas de los individuos en el espacio de dimensión reducida y que representaremos mediante puntos en dicho espacio.

Mientras que $h_w^{(2)} = (h_{w1} \quad h_{w2}),\ w = 1, \ldots, p$ la aproximación en dimensión dos de los marcadores columna de la matriz X, y que por tanto contienen las coordenadas de las variables en el espacio de dimensión dos y que en la representación Biplot Clásica se representan mediante vectores.

De esta manera, y como hemos visto en el Capítulo anterior, los marcadores fila y los marcadores columna pueden ser representados sobre un mismo sistema de referencia, que será el sistema de los ejes factoriales, cuyo origen coincide con el punto de equilibrio de las nubes, con máxima calidad de representación para ambos marcadores. Es decir es lo que se denomina una **representación simultánea en sentido estricto**; tal representación es lo que conocemos como representación HJ – Biplot Clásica.

A partir de dicha representación en un subespacio de baja dimensión, interpretamos las relaciones entre puntos – individuos y vectores – variables.

Pero, ¿esta proyección en el subespacio de baja dimensión refleja la verdadera relación entre variables e individuos del espacio de partida?

4.4 REPRESENTACIÓN DEL HJ – BIPLOT AUMENTADO

En la representación clásica, a partir de la representación Biplot no es posible valorar si la calidad de representación es buena y por tanto interpretable, sino que se necesita del uso de otra serie de índices para tal valoración.

En la alternativa de representación que proponemos en este trabajo sí que es posible, a partir simplemente del gráfico, deducir si la calidad de representación es alta, y por tanto las interpretaciones que de ella se deduzcan sean buenas.

Para cada marcador columna, h_w, $w = 1, \dots, p$, que representa la variable $w - ésima$, nos encontramos con la siguiente situación:

Longitud del vector en la representación tradicional:

$$\left\| h_w^{(2)} \right\| = \left(\sum_{i=1}^{2} h_{iw}^2 \right)^{1/2}$$

Longitud real del vector:

$$\left\| h_w^{(r)} \right\| = \left(\sum_{i=1}^{r} h_{iw}^2 \right)^{1/2}$$

La representación tradicional hace uso de los vectores $h_w^{(2)}$, $w = 1, \dots, p$, y por tanto muestra la longitud de los mismos sin tener en cuenta la longitud real de éstos, cuando es posible representar ambas longitudes.

Teniendo en cuenta lo anterior, proponemos como alternativa a la representación tradicional, la siguiente:

Representar primeramente el HJ – Biplot tradicional, es decir, incluyendo los vectores $\boldsymbol{h_w^{(2)}, w = 1, \dots, p}$; y posteriormente prolongar cada uno de los vectores, mediante trazo discontinuo, hasta conseguir su longitud real, esto es $\boldsymbol{h_w^{(r)}, w = 1, \dots, p}$.

De esta manera obtendremos la representación alternativa, que siguiendo la terminología utilizada por Anna Bartkowiak, denominaremos

HJ – Biplot Aumentada

En esta representación, simplemente observando el gráfico, podemos deducir si es o no una buena representación Biplot, e incluso cuáles son las variables que presentan alta calidad de representación y por tanto interpretables las relaciones o deducciones que de ellas se deriven.

4.5 ALGORITMO HJ – BIPLOT AUMENTADO

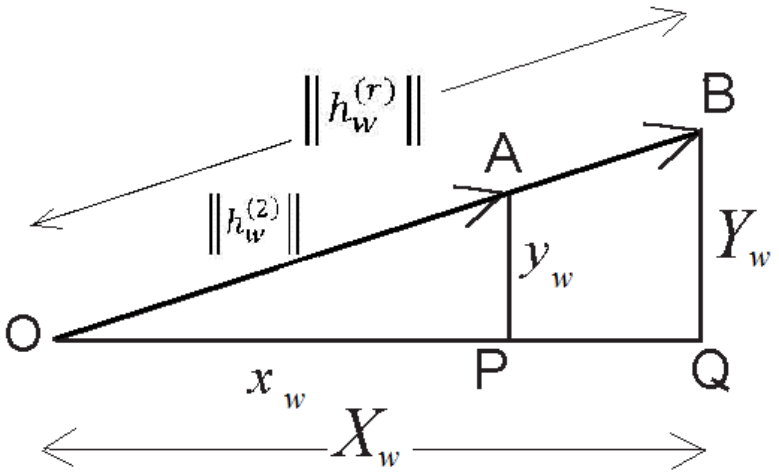
Sean $(x_w, y_w) = (h_{w1} \quad h_{w2})$, las coordenadas del vector $h_w^{(2)}$ que representa la variable $w - \acute{e}sima$, con $w = 1, \dots, p$, en la representación clásica.

Entonces, si denotamos por (X_w, Y_w) a las coordenadas del vector que representa la variable $w - \acute{e}sima$ en la representación alternativa, éstas se obtienen como sigue:

$$X_w = x_w \frac{\left\| h_w^{(r)} \right\|}{\left\| h_w^{(2)} \right\|}$$

$$Y_w = y_w \frac{\left\| h_w^{(r)} \right\|}{\left\| h_w^{(2)} \right\|}$$

En efecto, para su comprobación basta tener en cuenta la semejanza entre los triángulos $\widehat{OQB}$ y $\widehat{OPA}$ de la siguiente figura:



De donde se deduce que:

$$\frac{x_w}{\|h_w^{(2)}\|} = \frac{X_w}{\|h_w^{(r)}\|}$$

$$\frac{y_w}{\|h_w^{(2)}\|} = \frac{Y_w}{\|h_w^{(r)}\|}$$

c.q.d.

CAPÍTULO 5

IMPLEMENTACIÓN CON MATLAB

5.1 INTRODUCCIÓN

El objetivo principal de este capítulo es el de implementar el código fuente diseñado con MATLAB y que mostraremos en el siguiente apartado con el código necesario para conseguir la representación alternativa.

En líneas generales, nos basaremos en el hecho de que la representación HJ – Biplot alternativa, muestra primeramente los vectores $h_w^{(2)}, w = 1, \dots, p$; y posteriormente prolonga cada uno de éstos, mediante trazo discontinuo, hasta conseguir su longitud real, es decir $\left\| h_w^{(r)} \right\|, w = 1, \dots, p$.

De manera esquemática nuestro objetivo a la hora de representar pretende obtener la figura 2, a partir de la figura 1; es decir representar los vectores con la dirección de los obtenidos en dimensión reducida, $h_w^{(2)}, w = 1, \dots, p$ y con la longitud real de los vectores originales.

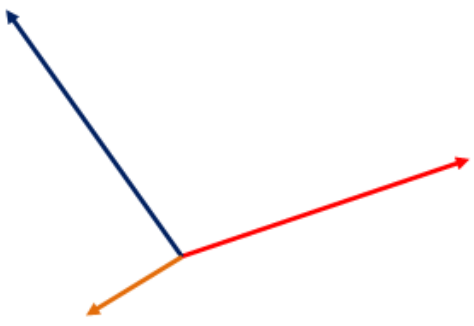


Figura 1

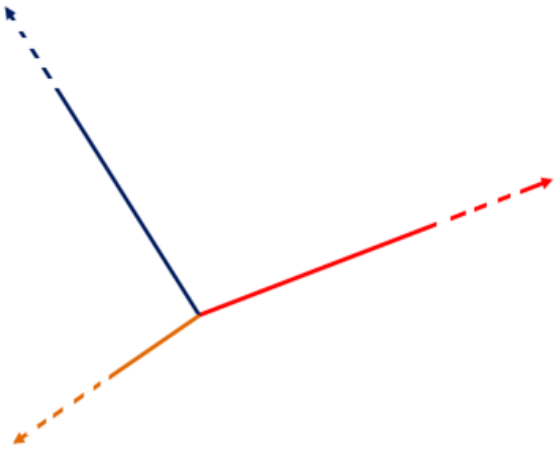


Figura 2

5.2 MATLAB: IMPLEMENTACIÓN DE LAS REPRESENTACIONES HJ – BIPLOT TRADICONAL Y ALTERNATIVA

A continuación se muestra en distintos bloques el código fuente de la Representación alternativa.

El código está implementado para poder realizar la elección entre la representación HJ – Biplot clásica y la HJ – Biplot Aumentada.

– Lectura del archivo de datos, debiendo ser éste un archivo Excel 2010:

```
% Lectura de archivo EXCEL 2010
[filename, pathname]=uigetfile('*.xlsx', 'Seleccione archivo de Excel
2010');
file = fullfile(pathname,filename);
[X,labels] = xlsread(file);
```

– Elección del tipo de Biplot a representar:

```
alfa=input('Tipo de Biplot: HJ-Biplot Representación Clásica=1,
Representación Alternativa=2: ');
disp('Opciones de representación: dimensiones');
dx=input('Representar en la abcisa la dimensión: ');
dy=input('Representar en la ordenada la dimensión: ');
```

– Lectura de las etiquetas de individuos y variables, así como del tamaño de la matriz.

```
X = X(1:end,:);
[n, m] = size(X);m0=m;
labelf = labels(2:end,1);
labelc = labels(1, 2:end,1);
```

– Estandarización de la matriz de datos y descomposición en valores y vectores propios.

```
X=(X-(1/n)*ones(n)*X)./(ones(n,1)*(diag(sqrt((1/n)*
(X'*X-(1/n)*X'*ones(n)*X))))');
[U,D,V]=svd(X);
if (alfa==2)
    F=U*D;
    C=V*D';
else
    F=U*(D*D')^(alfa/2);
    C=V*(D'*D)^((1-alfa)/2);
end
```

– Eliminación de las filas y columnas nulas de las matrices F y C, es decir de las matrices teóricas A y B.

```
% Reducción de las matriz de coordenadas al mínimo
  F0=F;
  C0=C(:,1:m);
  D0=D(1:m,1:m);
  d=max(dx, dy);
  F=F(:, 1:d);
  C=C(:, 1:d);
```

– En los siguientes bloques del código se etiquetan tanto variables – columnas, como individuos – filas y dimensión. Calculándose además, las contribuciones y la calidad de representación global.

```
%*************************************************************************
**
%                          CALCULO DE CONTRIBUCIONES
%*************************************************************************
**
file=[pathname, filename];
warning off MATLAB:xlswrite:AddSheet;
r=min(m, n);

% Etiquetas para las variables (columna)
sobran=size(labelc, 2)-m;
etic{1}=' ';
j=1;
for i=1+sobran:m+sobran,
    j=j+1;
    etic(j)=labelc(i);
end

% Etiquetas para los elementos (fila)
if (length(labelf)==0),
    etif=cell(1,n+1);
    etif{1}=' ';
    for i=1:n,
        etif{i+1}=['i', num2str(i)];
    end

elseif (length(labelf)~=0),
    etif{1}=' ';
    for i=1:n,
        etif(i+1)=labelf(i);
    end
end

% Etiquetas para las dimensiones
etid=cell(1,m+1);
etid{1}=' ';
for i=1:m,
    etid{i+1}=['Dim ', num2str(i)];
end
```

```
% Calidad de representación global
CRG=cell(r+1,2);
CRG{1,1}='Eigenvalores';
CRG{1,2}='Inercia (%)';
CRG{1,3}='Inercia acumulada (%)';
total=trace(D0.^2);
acumulado(1)=0;
for i=1:r,
    CRG{i+1,1}=(m/total)*D0(i,i)^2;
    CRG{i+1,2}=100*(D0(i,i)^2/total);
    acumulado(i+1)=acumulado(i)+CRG{i+1,2};
    CRG{i+1,3}=acumulado(i+1);
end
xlswrite(file, CRG, 'Global');
```

```
% Contribuciones del elementos al factor
CAEfF=F0.^2;
CREFF=cell(n+1, r+1);
CREFF (1,:)=etid';
CREFF (:,1)=etif;
for j=1:r,
    for i=1:n,
        CREFF{i+1,j+1}=100*CAEfF(i,j)/sum(CAEfF(:,j));
    end
end
xlswrite(file, CREFF, 'Fila al factor');

% Contribuciones de la columna al factor
CAEcF=C0.^2;
CRECF =cell(m+1, r+1);
CRECF (1,:)=etid';
CRECF (:,1)=etic';
for j=1:r,
    for i=1:m,
        CRECF{i+1,j+1}=100*CAEcF(i,j)/sum(CAEcF(:,j));
    end
end
xlswrite(file, CRECF, 'Columna al factor');

% Contribuciones del factor a fila
CAFEf=F0.^2;
CRFEF=cell(n+1, r+1);
CRFEF (1,:)=etid';
CRFEF (:,1)=etif';
for j=1:r,
    for i=1:n,
        CRFEF{i+1,j+1}=round(1000*CAFEf(i,j)/sum(CAFEf(i,:)));
    end
end
xlswrite(file, CRFEF, 'Factor a fila');

% Contribuciones del factor a columna
CAFEc=C0.^2;
CRFEC=cell(m+1, r+1);
CRFEC (1,:)=etid';
CRFEC (:,1)=etic';
for j=1:r,
    for i=1:m,
        CRFEC{i+1,j+1}=round(1000*CAFEc(i,j)/sum(CAFEc(i,:)));
    end
end
xlswrite(file, CRFEC, 'Factor a columna');
```

– En este último bloque se implementa la representación gráfica del Biplot, tanto del HJ – Biplot tradicional como de la representación alternativa. Para esta última, como se ha indicado en capítulos anteriores representamos tanto longitud real como la longitud "tradicional" del vector en el espacio 2 – dimensional. Por lo que si la elección de representación es la del Biplot tradicional, obtendremos en dicha representación únicamente los vectores $h_w^{(2)}, w = 1, \dots, p$; mientras que si la elección de representación es la alternativa, obtendremos simultáneamente tanto los vectores 2 – dimensionales anteriores con su longitud correspondiente y mediante un trazo diferenciado prolongaremos dichos vectores hasta conseguir su longitud real, es decir la de los vectores $h_w^{(r)}, w = 1, \dots, p$ respectivamente.

En esta representación alternativa, simplemente observando el gráfico, podemos deducir si es o no una buena representación Biplot, e incluso cuáles son las variables que presentan alta calidad de representación y por tanto interpretables las relaciones o deducciones que de ellas se deriven.

```
%*************************************************************************
**
%                          REPRESENTACIÓN GRAFICA
%*************************************************************************
**
% Construcción de los ejes de representación
  mxf=min(F(:,dx));
  Mxf=max(F(:,dx));
  myf=min(F(:,dy));
  Myf=max(F(:,dy));
  mxc=min(C(:,dx));
  Mxc=max(C(:,dx));
  myc=min(C(:,dy));
  Myc=max(C(:,dy));
  mx=min (mxf, mxc);
  Mx=max (Mxf, Mxc);
  my=min (myf, myc);
  My=max (Myf, Myc);
  rx=0.05*(Mx-mx);
  ry=0.05*(My-my);
  mx=mx-rx;
  Mx=Mx+rx;
  my=my-ry;
  My=My+ry;

% Construcción de los títulos para los ejes de representación
  dim='Dimensión ';
  xl=int2str(dx);
  yl=int2str(dy);
  xl=[dim xl];
  yl=[dim yl];


% Representación de los marcadores fila (o individuos)
   plot (F(:,dx),F(:,dy),...
   'Color',[0.04314 0.5176 0.7804],...
   'LineStyle','none',...
   'Marker','diamond',...
   'MarkerSize',3,...
   'MarkerEdgeColor',[0 0 0],...
   'MarkerFaceColor',[0 0.749 0.749])
    axis ([mx Mx my My])
    axis ('equal')
    ls=axis;
    xlabel (xl, 'FontSize', 12)
    ylabel (yl, 'FontSize', 12)
    hold all


if (alfa==1)
%HJ-Biplot tradicional
%Representación de los marcadores columna (o variables)
a=[0 1]';
for i=1:m
    c=a*C(i,:);
    plot (c(:, dx), c(:, dy), 'color', 'r', 'marker', 'none')
end

else
```

```
%HJ-Biplot: Representación alternativa
%Representación de los marcadores columna (o variables) con longitud
real
a=[0 1]';
for i=1:m
    cr=a*(norm(C0(i,:),2)/norm(C(i,:),2))*C(i,:);
    plot (cr(:, dx), cr(:, dy), 'color', 'g', 'marker', 'none')
end
% Representación de los marcadores columna (o variables)
a=[0 1]';
for i=1:m
    c=a*C(i,:);
    plot (c(:, dx), c(:, dy), 'color', 'r', 'marker', 'none')
end
end

% Representación de los ejes seleccionadas para la gráfica
text (F(:,dx),F(:,dy),labelf);
text (C(:,dx),C(:,dy),labelc);
ls=axis;
plot ([ls(1);ls(2)],[0;0],'Color',[0.502 0.502 0.502], 'LineStyle',
':')
plot ([0;0],[ls(3);ls(4)],'Color',[0.502 0.502 0.502],'LineStyle',
':')

% Final del programa
```

BIBLIOGRAFÍA

Gabriel, K.R. (1971). “The biplot-graphic display of matrices with application to principal component analysis”. Biometrika, 58: 453-467.

Gabriel, K.R. (1981a). “Biplot display of multivariate matrices for inspection of data and diagnosis”. In V. Barnett (ed.) Interpreting Multivatiate Data. Wiley. Chichester. U.K.

Gabriel, K.R. (1995). “Biplot”. Encyclopedia of Statistical Sciences. (S. Kotz & N.L. Johnson, eds.). Wiley. New York.

Galindo, M.P. (1986). “Una alternative de representación simultánea: HJ – Biplot”. Questíio, 10(1): 13-23.

Gower, J.C. & Hand, D.J. (1996). Biplots. Chapman & Hall. London.

Bartkowiak A., Szustalewicz A.: The augmented biplot and some examples of its use. Machine Graphics & Vision, 4, no 3 – 4, 161 – 185.

Galindo, M.P., Lorente, F., Romo, A., Martín, M. (1987), “Inspección de matrices de datos multivariantes utilizando el método HJ – Biplot: Aplicación a un problema médico”, Cuadernos de Bioestadística y sus Aplicaciones Informáticas 5(1): 88-101.

Pérez-Mellado, V., Galindo, P. (1986), "Biplot graphic display of Iberian and North African populations of podarcis", Rocek Z (ed.), Studies in Herpetology, 197-200.

Rivas, J. et. al. (1993), "HJ – Biplot analysis applied to enological parameters in the geographical classification of young red wines", American Journal of Enology and Vinicultura 44 (3):302-308.

Santos, C. et. Al. (1991), "Characterization of young red wines by application of HJ – Biplot Analysis to anthocianin profiles", Journal of Agricultural and food chemistry 39: 1086-1090.

Vicente, S. Galindo, M.P. (1993), "Análisis gráfico y descripción estructural de la habilidad de cultivos en Castilla y León", Investigación Agraria. Economía 8 (3):17-329.

García-Galegón, A. C., Trujillano, R., Vicente-Tavera, S. (1995), "Influence of consolidation and hydrofugation treatments of the physical properties of Avila natural and artificially aged granite: a statistical approach", Degradation and conservation of granitic rocks in monuments, European Commision, Dg Xii D-I.

Vicente-Tavera, S. (1992), "Las técnicas de representación de datos multidimensionales en el estudio del índice de producción industrial en la C.E.E.". Tesis Doctoral. Universidad de Salamanca.

Dorado, et. al. (1999), "Análisis HJ – Biplot de la evolución de la productividad agraria de la Comunidad de Castilla y León a lo largo del quinquenio 1991-1995", A. Dorado et. al. Invest. Agr.: Prod. Prot. Veg. Vol. 14 (3), 1999.

Mª José Fernández Gómez, M.P. Galindo, I. Barrera, J.L. Vicente-Villardón, A. Martin, (1996), "Alternativa al análisis canónico de correspondencias basada en los métodos Biblot". Mediterránea. Serie de estudios biológicos. Pág.63-71.

Galindo, M.P., Fernández-Gómez, M.J., Ávila, C., del Río, A., Fernández-Aláez, M. (1999) "El análisis HJ – Biplot como herramienta de estudio de un ecosistema acuático". Statistics and Environment pp:85-94.

M.P. Galindo Villardón, P. Vicente Galindo, C. Patino, J.L. Vicente (2007), "Caracterización Multivariante de los perfiles de las mujeres en situación laboral irregular: el caso de Salamanca". Pecvnia, 4, pp. 49-79.

Susana Mendes, et. al. (2009), "The study of bacterioplankton dynamics in the Berlengas Archipelago (West coast of Portugal) by applying the HJ – Biplot method", Arquipélago. Life and Marine Sciences 26: 25-35.

S. Vicente Tavera, M.P. Galindo Villardón (1992), "El HJ – Biplot como base para el análisis de la evolución del índice de producción industrial en la Comunidad Europea en los últimos años". III Congreso de Economía Regional de Castilla y León. Comunicaciones 2: 647-661.

Eugénia Castela, M. P. Galindo Villardón, (2010), "Inferencia Ecológica para la Caracterización de Abstencionistas: El caso de Portugal". Discussion Spatial and Organizational Dynamics. Quantitative Methods Applied to Social Sciences. Discussion Papers Nº3: 6-25.

Benzècri, J. (1973), L'Analyse des données, Toumo I: La taxinomie, Dunond, París.

S. Vicente Tavera, G. Ramirez, P. Galindo (1994), "El HJ – Biplot como base para la búsqueda de Clusters en función de la distribución de parados según profesiones en la Comunidad de Castilla y León". IV

Congreso de Economía Regional de Castilla y León. Comunicaciones 2: 822-835.

Pinto Rodríguez, Mª Elena, Álvarez – Uría Rico, Fernando (2006), "Suicidio Juvenil: Sociología de una realidad social". Tesis doctoral. Universidad Complutense de Madrid.

Paula Alarcón, Eugenia Vinet, Sonia Calvo (2005), "Estilos de Personalidad y Desadaptación Social durante la Adolescencia". Psykhe, 2005, Vol. 14, Nº1, 3-16.