

# UNA ALTERNATIVA DE REPRESENTACIÓN SIMULTÁNEA: HJ-BIPLLOT

M.ª PURIFICACIÓN GALINDO VILLARDÓN

UNIVERSIDAD DE SALAMANCA

*En el presente trabajo se hace una revisión del método BIPLLOT propuesto por GABRIEL y se propone una nueva forma de representación simultánea para matrices de datos que denominamos HJ-BIPLLOT, en la cual, las coordenadas para las columnas coinciden con los marcadores para las columnas en el  $GH^T$ -biplot y las coordenadas para las filas coinciden con los marcadores para las filas en el  $JK^T$ -biplot de GABRIEL. Estas coordenadas pueden ser representadas en un mismo sistema de referencia: El sistema de los ejes factoriales.*

*Se demuestra que el HJ-biplot consigue la misma bondad de ajuste para filas y para columnas, siendo ésta de un orden muy superior al usual.*

*Se demuestra también, que el HJ-biplot, para matrices de datos positivos, da lugar a las mejores representaciones  $\beta$ -baricéntricas.*

Keywords: REDUCED RANK APPROXIMATION, BIPLLOT, BIMODEL, HJ-BIPLLOT.

## 1. INTRODUCCION.

En ciertas aplicaciones del Análisis Multivariante tiene interés representar gráficamente matrices de datos, por ejemplo, matrices  $X_{n \times p}$  correspondientes a datos tomados a partir de  $p$  variables y  $n$  individuos.

Las representaciones que comprenden el análisis simultáneo de individuos y variables son particularmente interesantes, ya que --tan importante como conocer la configuración de los individuos es conocer que variables son las responsables de esa configuración.

La técnica de representación simultánea más utilizada es el análisis factorial de correspondencias /2/ y /3/. En este análisis se consideran las nubes  $N_I, N_J$  de los perfiles fila y columna, con masas iguales a los respectivos totales marginales, en los espacios multidimensionales  $R^D$  y  $R^n$ ; ambos espacios estructurados por la métrica ji-cuadrado, la cual permite la determinación unívoca de sucesivos ejes principales de --inerencia.

Otra forma mucho menos conocida de represen

tar de manera conjunta las filas y las columnas de una matriz cualquiera, es el método BIPLLOT propuesto por GABRIEL /6/, que consiste en interpretar los elementos de la matriz  $X$  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 propone una alternativa de representación simultánea basada en el método Biplot que presenta ventajas con respecto a las anteriores.

## 2. REPRESENTACION BIPLLOT PARA UNA MATRIZ CUALQUIERA.

Dada una matriz cualquiera  $X_{n \times p}$  de rango  $r \leq$   $p$  puede ser factorizada en la forma

$$X = GH^T \quad (2.1)$$

con  $G_{n \times r}$  y  $H^T_{r \times p}$  ambas necesariamente de rango  $r$  /20/.

La factorización asigna los vectores  $g_1 \dots g_n$ , filas de la matriz  $G$ , uno a cada una de las

- M.ª. Purificación Galindo Villardón - Universidad de Salamanca - Facultad de Biología - U.D. de Bioestadística (Depto. de Ecología) - Salamanca.

- Article rebut el Desembre de 1985.

filas de la matriz  $X$  y los vectores  $h_1 \dots h_p$ , filas de la matriz  $H$ , uno a cada una de sus columnas; todos ellos de rango  $r$ . La representación de  $g_1 \dots g_n$  y  $h_1 \dots h_p$  en el subespacio  $E_r$ , es llamada  $GH^T$ -bigeometría de  $X$ .

La factorización de la matriz  $X$  no es única ya que si  $X = GH^T$  es una factorización, para cualquier matriz  $R$  no singular,  $X = GR^T (HR^{-1})^T$  es otra factorización. Tampoco es única por tanto, la correspondiente bigeometría.

La factorización (2.1) es equivalente al producto interno

$$x_{ij} = g_i^T h_j \quad (2.2)$$

donde  $x_{ij}$  es el elemento de la  $i$ -ésima fila y de la  $j$ -ésima columna de  $X$ . La relación (2.2) proporciona una  $GH^T$ -bigeometría de la matriz  $X$ , de forma que la matriz de partida viene representada por medio de  $n \times p$  vectores en el subespacio  $E_r$ .

Si  $r=p$  (XI-bigeometría), las filas vienen representadas por los vectores  $g_i$  y las columnas como vectores de longitud unidad con respecto al sistema de referencia ortogonal de  $E_p$ . Los vectores  $h_j$  preservan la configuración de las columnas.

Se verifica también, que cada elemento de la matriz se puede escribir como producto interno de los vectores  $g_i, h_j$ ; aquí las interdistancias entre las filas coinciden con las interdistancias entre los vectores  $g$ .

$$\|X_i - X_j\| = \|g_i - g_j\|$$

Si  $r=n$ , la IX-bigeometría correspondiente -- permite representar la configuración de la matriz de varianzas-covarianzas de las columnas de la matriz  $X$ .

Las columnas de  $X$  vienen representadas por los vectores  $h$  y las filas como vectores de longitud unidad a lo largo de los ejes ortogonales que forman el sistema de referencia en  $E_n$ .

Supongamos la matriz  $X_{n \times p}$  descompuesta en valores singulares /5/:

$$X = U \Sigma V^T \quad (2.3)$$

donde  $U$  es la matriz cuyas columnas son ortogonales y vectores propios de  $XX^T$ ;  $\Sigma$  es la matriz diagonal de valores singulares de  $X$  y  $V$  es una matriz ortogonal cuyas columnas son vectores de  $X^T X$ . Como sólo las  $r$  primeras filas de  $V$  son no nulas, la matriz  $X$  puede ser representada en  $E_r$  por  $GH^T$ -bigeometría con

$$G = U \quad \text{y} \quad H^T = \Sigma V^T$$

lo cual equivale a proyectar sobre las  $r$  primeras componentes principales para el espacio de las columnas.

Los vectores  $h$  tienen la misma configuración que las columnas de  $X$ ; o sea

$$h_j^T h_j = X_j^T X_j,$$

donde  $X_j$  denota la  $j$ -ésima columna de la matriz  $X$ .

Los vectores  $g_i$  son las proyecciones sobre los ejes principales.

$$\text{Como además } GG^T = X(X^T X)^{-1} X^T \quad (2.4)$$

los vectores  $g$  representan la configuración de las filas de  $X$  en forma estandar.

La configuración de las filas no se modifica si la XI-bigeometría es rotada sobre cualquier otro conjunto de ejes ortogonales.

Consideremos pues, la JK<sup>T</sup>-bigeometría en la cual  $J=U\Sigma$  y  $K^T=V^T$ . Los productos internos de las filas vendrán reproducidos por los productos internos de los marcadores  $j$ ;

$$X_i^T X_j = j_i^T j_j,$$

donde  $X_i$  es la fila  $i$ -ésima de la matriz  $X$ .

$$\text{Como } KK^T = X(XX^T)^{-1} X \quad (2.5)$$

las filas de  $X$  vienen representadas en forma standard, lo cual equivale a tomar en el espacio de las filas la métrica asociada a la matriz identidad.

Las  $GH^T$ -bigeometría y la  $JK^T$ -bigeometría, difieren sólo por las escalas recíprocas a lo largo de los ejes principales ya que:

$$J_{i\alpha} \lambda_{\alpha}^{-1} = g_{i\alpha}$$

$$K_{j\alpha} \lambda_{\alpha} = h_{j\alpha}$$

La relación, pues, entre ambos es obvia.

Si la matriz X es de rango 2 la bigeometría se reduce a un BIPILOT (representación conjunta de filas y columnas de X en el plano) y si es de rango 3 a un BIMODEL /8/.

### 3. BIPILOT APROXIMADO PARA UNA MATRIZ CUALQUIERA.

Cuando r es mayor que dos (o tres), la bigeometría puede ser proyectada sobre un plano (espacio) pero en ese caso obtendremos un BIPILOT (BIMODEL) aproximado para la matriz X /6/, /8/, /9/. Si conseguimos aproximar la matriz X de rango r por otra de rango dos  $X_{(2)}$  y esa aproximación es satisfactoria, el BIPILOT correspondiente a la matriz  $X_{(2)}$  puede resultar útil para representar a X. En este caso los productos internos de los marcadores correspondientes a las filas y columnas son sólo valores aproximados a los valores  $x_{ij}$  de X.

#### 3.1 APROXIMACIONES A BAJO RANGO.

GABRIEL /6/ utiliza el conocido método para aproximaciones a bajo rango debido a HOUSEHOLDER y YOUNG /16/ que consiste en minimizar la suma de cuadrados de las desviaciones entre los elementos de X y los de la matriz aproximada de rango menor. GOLUB y REINSCH /14/ parten de la descomposición en valores singulares de ECKART y YOUNG /5/, según la cual dada una matriz  $X_{n \times p}$  real, se puede descomponer en la forma (2.3).

GABRIEL y ZAMIR /10/ proponen un método iterativo de ajuste mínimo cuadrático a bajo rango para una elección cualquiera de pesos.

Dada una matriz X ajustarla por una de más - bajo rango q equivale a ajustarla por un producto  $GH^T$  donde G y H son  $n \times q$  y  $q \times p$  respectivamente.

Si a cada elemento  $x_{ij}$  de la matriz X le suponemos provisto de un peso  $w_{ij}$ , el criterio de minimización consiste en hacer mínima la

expresión:

$$\psi = \sum_{i=1}^n \sum_{j=1}^p w_{ij} (x_{ij} - g_i^T h_j)^2 \quad (3.1.1.)$$

donde  $g_i$  y  $h_j$  denotan las filas de G y H respectivamente.

El caso en que el criterio de minimización consistía en hacer mínima la norma euclídea de la matriz de residuales  $\|X - GH^T\|^2$  /7/ puede considerarse como un caso particular de lo expuesto, en el cual  $w_{ij} = 1 \quad \forall i, j$ .

Para este caso particular GABRIEL /7/ señala la solución propuesta por GOOD /15/ según la cual las columnas g y h, de las matrices G y H se obtienen en el paso r-ésimo teniendo en cuenta los valores que han tomado en pasos anteriores, y conocida, además, la aproximación a X en el paso r-1, es decir:

$$X^{(r-1)} = X \sum_{\ell=1}^{r-1} g_{\ell} h_{\ell}^T \quad (3.1.2)$$

Las columnas de G y H vienen determinadas por las ecuaciones:

$$\sum_i g_{ir}^2 h_{jr} = \sum_i g_{ir} x_{ij}^{(r-1)} \quad (3.1.3)$$

$$\sum_j h_{jr}^2 g_{ir} = \sum_j h_{jr} x_{ij}^{(r-1)} \quad (3.1.4)$$

Partiendo de un valor  $g_r$  inicial, el proceso continua hasta conseguir la convergencia.

El proceso se reduce, pues, a una regresión de las columnas de  $X^{(r-1)}$  sobre las  $g_r$  para obtener los  $h_r$  como coeficientes, y a una regresión de las filas de  $X^{(r-1)}$  sobre las  $h_r$  para obtener las  $g_r$  como coeficientes.

Estos ajustes, paso a paso, son posibles por que las g y las h son ortogonales.

En el caso en que se consideren pesos  $w_{ij}$  la generalización del método es inmediata. Las ecuaciones (3.1.3) y (3.1.4) se escribirían así:

$$\left( \sum_i w_{ij} g_{ir}^2 \right) h_{jr} = \sum_i w_{ij} g_{ir} x_{ij}^{(r-1)} \quad (3.1.5)$$

$$\left( \sum_j w_{ij} h_{jr}^2 \right) g_{ir} = \sum_j w_{ij} h_{jr} x_{ij}^{(r-1)} \quad (3.1.6)$$

Los autores señalan dos posibles problemas. A no ser que los pesos sean iguales, las sucesivas  $g_i, h_j$  no son necesariamente ortogonales, en cuyo caso el método no proporcionaría el mejor ajuste.

También es importante tener en cuenta que el proceso iterativo que nos permite resolver las ecuaciones de regresión cuyos coeficientes proporcionan los valores de las sucesivas  $g_i, h_j$  puede converger a un mínimo local el cual no sea el deseado para el mejor ajuste, problema que puede solucionarse con una elección adecuada de la inicialización de la regresión.

#### 4. BONDAD DE AJUSTE PARA LAS LINEAS DE LA MATRIZ X EN LA REPRESENTACION GH<sup>T</sup> Y JK<sup>T</sup>-BI-PLOT.

Dada una matriz X de rango r mayor que dos, la bondad de ajuste en el plano viene dada por.

$$\rho_{(2)} = \frac{\lambda_1 + \lambda_2}{r \sum_{i=1} \lambda_i} \quad (4.1)$$

donde los  $\lambda_i$  denotan como es usual, los autovalores de  $X^T X$ .

Sin embargo, no nos interesa conocer sólo la bondad de ajuste para la matriz de datos sino que además y sobre todo, nos interesa conocer cual es la bondad de ajuste para las filas (individuos, poblaciones) y para las columnas (variables). Una bondad de ajuste aceptable para la matriz X puede ser debida, por ejemplo, a que las filas están muy bien representadas y las columnas muy mal o quizás, a que tanto las filas como las columnas tengan la misma bondad de ajuste. Es evidente que en problemas reales las dos situaciones hipotéticas anteriores no nos permitirían una misma interpretación de resultados.

Analizaremos, pues, cual es la bondad de ajuste para las distintas líneas de la matriz al hacer una representación GH<sup>T</sup> y JK<sup>T</sup>-BI-PLOT.

Si tenemos en cuenta que la matriz de varianzas-covarianzas es simétrica y que por tanto

en la descomposición singular los vectores U y V son iguales y los valores singulares coinciden con los autovalores, podemos asegurar que el biplot de una matriz de covarianzas puede ser reducido al gráfico de un conjunto de vectores que son precisamente la matriz H para la representación GH<sup>T</sup>-Biplot de X.

Como la bondad de ajuste para la matriz de covarianzas es:

$$\rho_{(2)} = \frac{\lambda_1^2 + \lambda_2^2}{r \sum_{i=1} \lambda_i^2} \quad (4.2)$$

esa será también la bondad de ajuste para las columnas de la matriz X. Sin embargo, para las filas la bondad de ajuste es:

$$\rho_{(2)} = 2/r \quad (4.3)$$

Un razonamiento análogo pero ahora sobre la matriz de dispersión de los individuos (poblaciones) nos permite asegurar que la bondad de ajuste para las filas de X en un JK<sup>T</sup>-biplot es del orden (4.2) y para las columnas del orden (4.3), /6/.

Por tanto, aunque GABRIEL señala expresamente que el nombre de método BI-PLOT 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.

#### 5. REPRESENTACION SIMULTANEA EN SENTIDO ESTRICTO: HJ-BI-PLOT

Sea  $X_{n \times p}$  la matriz de rango p correspondiente a datos tomados a partir de p variables y n individuos.

Introduzcamos en el espacio de las filas la métrica asociada a la inversa de la matriz de covarianzas de las variables y en el espacio de las columnas la métrica asociada a la inversa de la matriz de dispersión de los individuos (filas y columnas se suponen centradas).

Esto es equivalente a tomar  $U \Sigma$  como marca-

dores para las filas de la matriz  $X_{n \times p}$  y  $\Sigma V$  como marcadores para las columnas.

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

$$U\Sigma = XV \Sigma^{-1} \Sigma = XV \quad (5.1)$$

entonces  $U\Sigma$  coincide con 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 tanto, podemos asegurar que los marcadores para las filas en un  $JK^T$ -biplot coinciden con las coordenadas de las filas respecto a los ejes factoriales.

Un razonamiento análogo nos permite escribir:

$$\Sigma V = \Sigma \Sigma^{-1} X^T U = X^T U \quad (5.2)$$

luego  $\Sigma V$  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 en un  $GH^T$ -biplot coinciden con las coordenadas de las columnas respecto a los ejes factoriales.

Ambas representaciones, aparentemente, no están relacionadas, pero teniendo en cuenta que:

$$X^T U = X^T X V \Sigma^{-1} \quad (5.3)$$

si a  $X^T U$  la llamamos  $B$  y a  $XV$  la llamamos  $A$ , podemos escribir:

$$B = X^T A \Sigma^{-1} \quad (5.4)$$

$$A = X B \Sigma^{-1} \quad (5.5)$$

Resulta, pues de (5.4) que la  $h$ -ésima coordenada de la variable  $j$ -ésima se puede expresar en función de las  $h$ -ésimas coordenadas de las  $n$  filas:

$$b_{jh} = \{x_{1j} a_{1h} + \dots + x_{nj} a_{nh}\} (1/\sqrt{\lambda_h}) \quad (5.6)$$

Además, de (5.5) se deduce que la  $h$ -ésima coordenada de la población  $i$ -ésima se expresa en función de las  $h$ -ésimas coordenadas de las  $p$  variables según la relación:

$$a_{ih} = \{x_{i1} b_{1h} + \dots + x_{ip} b_{ph}\} (1/\sqrt{\lambda_h}) \quad (5.7)$$

Es decir, cada coordenada de la variable  $j$ -ésima es una media de las coordenadas de

los individuos, donde cada una de esas coordenadas está ponderada por el valor que toma la variable  $X_j$  sobre los individuos; análogamente, cada individuo ocupa el punto de equilibrio del conjunto de las variables /4/.

Si una variable toma un valor preponderante para un individuo, el punto que representa a esa variable estará próximo al punto que representa a ese individuo.

Las relaciones anteriores nos permiten 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 autovalores 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. A la representación conjunta de los marcadores  $H$  y los marcadores  $J$  la llamaremos REPRESENTACION SIMULTANEA HJ o HJ-BIPLLOT para seguir la misma terminología de GABRIEL /6/, /8/, /9/. Para más detalles ver /11/, /12/, /13/.

Los autores que hasta ahora han utilizado el método biplot consideran dos sistemas de referencia simultáneamente a fin de conseguir la impresión de una representación simultánea; es decir, sitúan las nubes de puntos dentro de un cuadrado, una de ellas referida a una escala que considera sobre dos lados del cuadrado, y la otra nube referida a otra escala que toman sobre los otros dos lados del cuadrado. Es evidente que interpretar las relaciones entre puntos de una misma nube es válido. Guardamos reservas, desde luego, en cuanto a considerar las relaciones entre las dos nubes.

## 6. CALIDAD DE REPRESENTACION PARA LAS LINEAS DE UNA MATRIZ CUALQUIERA X EN UNA REPRESENTACION HJ-BIPLLOT.

Una medida absoluta de la bondad del ajuste para la matriz  $X$  puede ser definida por:

$$\rho(2) = 1 - (\|X - X_{(2)}\|^2 / \|X\|^2) \quad (6.1)$$

Si  $\rho(2)$  es próximo a cero la aproximación es buena. No conocemos, sin embargo, cual es la

bondad de ajuste de las filas y las columnas por separado para el nuevo procedimiento propuesto. Al estudiar el  $GH^T$  y el  $JK^T$ -biplot - vimos que en un  $GH^T$  la calidad de representaci6n para las columnas era del orden;

$$\rho'_{(2)} = \frac{\lambda_1^2 + \lambda_2^2}{\sum_{i=1}^r \lambda_i^2}$$

y en un  $JK^T$ -biplot la calidad para las filas es de ese mismo orden.

En un HJ-biplot las coordenadas para las columnas coinciden con los marcadores para las columnas en un  $GH^T$ -biplot y las coordenadas para las filas con los marcadores para las filas en un  $JK^T$ -binplot, por tanto, con la t6cnica propuesta la calidad de representaci6n para las filas y columnas es la misma (4.2), es por tanto, una representaci6n simult6nea en sentido estricto; adem6s esa calidad es de orden muy superior al usual. Todo lo dicho se sintetiza en esta tabla:

TABLA 1

CUADRO RESUMEN DE LAS CARACTERISTICAS MAS IMPORTANTES DE LOS TRES BIPLOTS CONSIDERADOS

R. Simult6nea	Coordenadas filas	Coordenadas columnas	Bondad ajuste para filas	Bondad ajuste columnas
<b>GH'- biplot</b>	U	$\Sigma V$	$2/r$	$\frac{\lambda_1^2 + \lambda_2^2}{\sum_{\alpha=1}^r \lambda_{\alpha}^2}$
<b>JK'- biplot</b>	$U\Sigma$	V	$\frac{\lambda_1^2 + \lambda_2^2}{\sum_{\alpha=1}^r \lambda_{\alpha}^2}$	$2/r$
<b>HJ'- biplot</b>	$U\Sigma$	$\Sigma V$	$\frac{\lambda_1^2 + \lambda_2^2}{\sum_{\alpha=1}^r \lambda_{\alpha}^2}$	$\frac{\lambda_1^2 + \lambda_2^2}{\sum_{\alpha=1}^r \lambda_{\alpha}^2}$

Z. BUSQUEDA DE LA MEJOR REPRESENTACION SIMULTANEA DE LAS LINEAS DE UNA MATRIZ DE DATOS.

Podemos justificar tambi6n la representaci6n HJ-biplot planteando el problema como la b6squeda de una mejor representaci6n simult6nea.

Sean  $B_1 = (b_1, \dots, b_p)^T$  las coordenadas de las p variables respecto al primer eje de representaci6n. Sean an6logamente,  $A_1 = (a_1, \dots, a_n)^T$  las coordenadas de las n poblaciones.

Analicemos la siguiente relaci6n r6ciproca entre  $A_1$  y  $B_1$ .

$$a_i = \sum_j x_{ij} b_j \quad i=1, \dots, n \tag{7.1}$$

$$b_j = \sum_i x_{ij} a_i \quad j=1, \dots, p$$

Sin embargo, tales relaciones no ser6n en general posibles, por lo que introduciremos una constante  $\beta > 0$  y escribiremos:

$$a_i = \beta \sum_j x_{ij} b_j \quad b_j = \beta \sum_i x_{ij} a_i \quad i=1, \dots, n, j=1, \dots, p \tag{7.2}$$

Adem6s, para que (7.2) puedan interpretarse como una "media ponderada" (salvo el factor  $\beta$ ), impondremos las condiciones siguientes:

$$x_{ij} \geq 0 \text{ para todo } i, j; \quad \sum_{ij} x_{ij} \leq 1 \tag{7.3}$$

La primera condición se verificará en la práctica. Por ejemplo, en el caso de datos que expresen frecuencias, porcentajes, abundancias. Para la segunda condición bastará convertir  $X=(x_{ij})$  en  $\alpha X=(\alpha x_{ij})$  donde  $\alpha$  es una constante adecuada (podríamos tomar  $\alpha=(\sum_{ij} x_{ij})^{-1}$ ).

Podemos escribir entonces

$$a_i = (\beta/\alpha) \sum_j (\alpha x_{ij}) b_j \quad b_j = (\beta/\alpha) \sum_i (\alpha x_{ij}) a_i$$

es decir

$$a_i = \beta' \sum_j y_{ij} b_j \quad b_j = \beta' \sum_i y_{ij} a_i \quad (7.4)$$

siendo

$$\beta' = \beta/\alpha, \quad y_{ij} = \alpha x_{ij}$$

Vemos que este cambio de escala no afecta al cálculo de  $A_1$  y  $B_1$ .

Volviendo ahora a la notación (7.2), pero suponiendo que se verifica (7.3), tenemos que  $a_i$  es una media ponderada de  $b_1, \dots, b_p$ , salvo el factor  $\beta$ .

Luego buscaremos el factor  $\beta > 0$  lo más próximo posible a la unidad, a fin de que la relación recíproca (7.1) sea lo más aproximada posible.

Sustituyendo obtenemos:

$$a_i = \beta^2 \sum_j x_{ij} (\sum_i x_{i'j} a_{i'}) = \beta^2 \sum_i (\sum_j x_{ij} x_{i'j}) a_i$$

Luego  $A_1$  es autovector de  $XX^T$  de autovalor  $1/\beta^2$

$$XX^T A_1 = A_1 (1/\beta^2)$$

Repetiendo el mismo razonamiento para las coordenadas en los demás ejes vemos que

$$XX^T A = AD \quad (7.5)$$

siendo  $D$  la matriz diagonal:

$$D = \text{diag}(1/\beta_1^2, \dots, 1/\beta_p^2)$$

Por otro lado, como  $A = XB\Sigma^{-1}$ ,  $B = X^T A \Sigma^{-1}$ , sustituyendo

$$A = XX^T A \Sigma^{-1} \Sigma^{-1} = XX^T A \Sigma^{-2}$$

es decir

$$XX^T A = A \Sigma^{-2} \quad (7.6)$$

Comparando (7.5) con (7.6) obtenemos  $D = \Sigma^2$ . Luego

$$\beta_i = \sigma_i^{-1} = 1/\Sigma^{-2} = 1/\sqrt{\lambda_i} \quad i=1, \dots, p \quad (7.7)$$

siendo  $\lambda_i$  el  $i$ -ésimo autovalor de  $XX^T$ ,  $\sigma_i$  el correspondiente valor singular de  $X$ .

Estudieemos ahora los autovalores de  $M = XX^T$ . Es fácil probar que los elementos de  $M$  y de cualquier potencia  $M^k$  verifican también la condición (7.3). Sea  $V$  un autovector de  $M$  de autovalor  $\lambda$ . Entonces

$$M^k V = \lambda^k V$$

Supongamos que  $V = (V_1, \dots, V_n)^T$  verifica

$$V_i \geq 0 \quad \sum_i |V_i| \leq 1 \quad (7.8)$$

Entonces  $M^k V = W$  también verifica (7.3).

Si fuera  $\lambda > 1$  tendríamos que para  $k$  suficientemente grande,  $\lambda^k V$  sería tal que la suma de sus componentes no verificaría (7.8). Esto es absurdo. Luego debe ser  $\lambda \leq 1$ .

Volviendo ahora a las relaciones (7.2), vemos que  $\beta_1 = 1/\sqrt{\lambda_1}$ , donde  $\lambda_1 \leq 1$  es el mayor valor propio de  $XX^T$ . Luego  $\beta_1$  es el valor más próximo a 1 para el cual se verifica la relación recíproca (7.2). Examinando (7.5) (7.6), podemos extender esta propiedad a los demás ejes, es decir, hallar  $(\beta_1, \beta_2, \dots)$  cuyos valores den la mejor representación  $\beta$ -baricéntrica en 2, 3, ... dimensiones.

Llegamos pues a la conclusión de que el HJ-biplot, para matrices de datos positivos, da lugar a las mejores representaciones  $\beta$ -baricéntricas, es decir, obtenemos la misma propiedad que se cumple en el análisis de correspondencias /17/.

## 8. EJEMPLO.

Sea  $X$  una matriz de datos positivos. Supongamos, por ejemplo, que cada elemento  $x_{ij}$  representa el número de individuos de la población  $i$ -ésima en los cuales está presente la característica  $j$ -ésima (Ver tabla 2).

TABLA 2

MATRIZ DE DATOS

Los datos forman parte de un amplio trabajo sobre la clasificación del género PODARCIS en la P. Ibérica /18/ y /19/.

POBLACIONES						POBLACIONES					
VAR	1	2	3	4	5	VAR	1	2	3	4	5
1	49	31	20	13	1	2	14	4	31	236	19
3	7	6	40	40	17	4	0	4	4	1	0
5	13	23	36	99	22	6	0	0	14	7	1
7	80	65	0	0	5	8	78	66	41	350	60
9	1	0	63	35	0	10	0	0	2	0	0
11	1	0	34	0	0	12	0	0	3	1	0
13	81	55	23	383	60	14	0	0	98	0	0
15	0	0	16	0	0	16	0	1	1	21	7
17	0	0	5	36	6	18	1	1	15	204	18
19	41	16	27	76	26	20	33	19	30	13	4
23	37	5	2	269	29	24	44	41	3	60	3
25	0	0	141	7	5	26	59	32	1	275	51
27	22	18	2	85	4	28	1	0	28	14	5
29	0	0	116	0	0						

La matriz  $X_{5 \times 29}$  viene representada en las fig. 1, 2 y 3 en las cuales los marcadores para las columnas vienen representadas por vectores y los marcadores para las filas -- por números.

La figura 1 corresponde a una representación  $JK^T$ -biplot, la figura 2 a un  $GH^T$ -biplot y la figura 3 a un  $HJ$ -biplot.

En cualquiera de los tres casos la tasa de inercia en el primer plano principal es del 94.64%.

La bondad de ajuste para las poblaciones en el  $JK^T$ -biplot es del 99%, sin embargo, la bondad de ajuste para las variables es sólo del 40%. En el  $GH^T$ -biplot la bondad de ajuste para las poblaciones es del 40% y para las variables del 99%. En el  $HJ$ -biplot la calidad de representación, tanto para filas como para columnas, es del 99%.

Si observamos las fig. 1 y 3 vemos como la configuración de las poblaciones es la misma en ambas representaciones, como cabía esperarse, ya que la bondad de ajuste para las poblaciones es la misma en el  $JK^T$  que en el  $HJ$ -biplot. Sin embargo, los puntos que representan a las variables aparecen en la fig. 1 en una maraña en torno al origen que hace imposible su interpretación -- conviene recordar que la bondad de ajuste era sólo del 40%.

En la fig. 3 las variables, para las cuales la bondad de ajuste es del mismo orden que la de las poblaciones (99%), se configuran de manera que la interpretación es evidente.

Algo análogo ocurre al comparar las figs. 2 y 3: En la representación  $GH^T$  biplot, la bondad de ajuste para las variables es del 99%, del mismo orden que en el  $HJ$ -biplot, pero -- las poblaciones para las cuales la calidad de representación es muy baja (40%) aparecen también en una maraña en torno al origen. En el  $HJ$ -biplot (fig. 3) variables y poblaciones están bien representadas y por tanto la interpretación de las relaciones entre poblaciones y variables es factible.

8. CONCLUSIONES.

Los marcadores para las columnas en un  $GH^T$ -biplot coinciden con las proyecciones de los puntos de la nube  $N_j$  sobre los ejes principales del espacio de los individuos; además, podemos asegurar que los marcadores para las filas en un  $JK^T$ -biplot coinciden con las --- coordenadas de los individuos referidos éstos a unas nuevas variables, combinación lineal de los de partida, con máxima varianza, incorreladas y cada una de ellas menos importante que la anterior desde el punto de vista descriptivo, o sea, representar a los individuos mediante un  $JK^T$ -biplot equivale a descomponer la variabilidad global en sus --



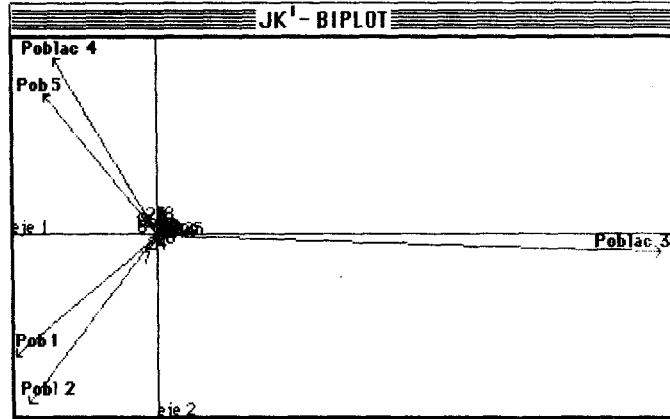


FIG 1: Representación JK¹-Biplot de la matriz  $X_{5 \times 29}$  que aparece en la tabla 1.

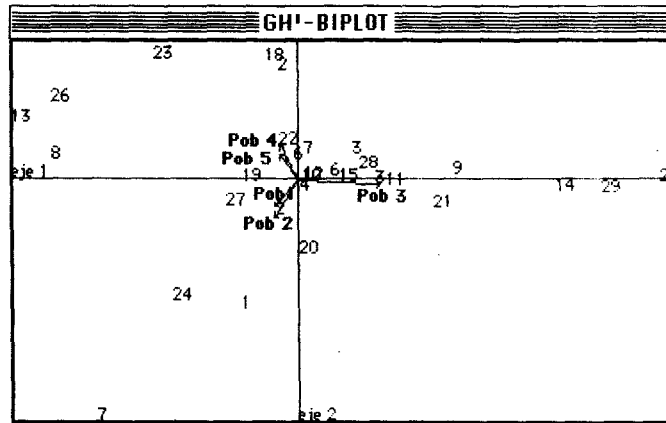


FIG 2: Representación GH¹-Biplot de la matriz  $X_{5 \times 29}$  que aparece en la tabla 1.

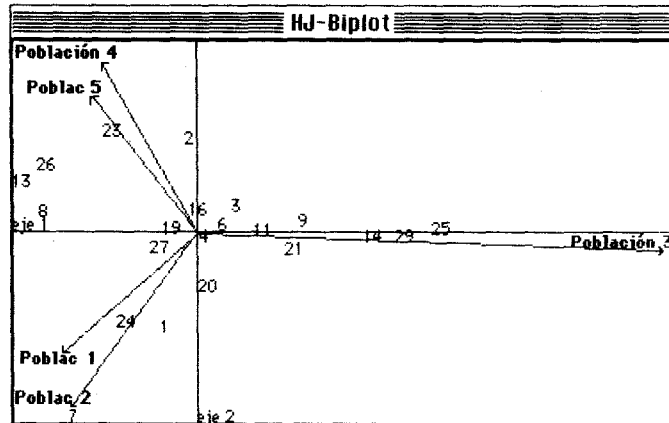


FIG 3: Representación HJ-Biplot de la matriz  $X_{5 \times 29}$  que aparece en la tabla 1.

componentes principales.

Lo más importante, probablemente, es que los marcadores para las columnas de una matriz cualquiera según un  $GH^T$ -biplot y los marcadores para las filas según un  $JK^T$ -biplot pueden ser representados sobre un mismo sistema de referencia (el sistema de los ejes factoriales) cuyo origen coincide con el punto de equilibrio de las nubes, lo cual hace posible interpretar la proximidad entre los individuos en términos de similitud entre ellos, y la proximidad entre variables en términos de covariación. A esta representación la hemos denominado HJ-biplot.

Además de poder comparar las posiciones relativas de distintos pares de puntos-filas con respecto al conjunto de variables y viceversa, es posible 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.

Cuanto más distantes aparezcan los puntos -- que representan a los caracteres del centro de gravedad, más variabilidad habrán presentado esos caracteres en el estudio, cuanto menor sea el ángulo que formen dos vectores que unan el centro de gravedad con los puntos que representan a las variables, más correlacionados están los caracteres. Además -- cuanto mayor sea el módulo de ese vector, mayor será la contribución del elemento a la inercia.

La calidad de representación es la misma para individuos que para variables siendo ésta de un orden muy superior al usual.

El HJ-biplot para matrices de datos positivos da lugar a las mejores representaciones  $\beta$ -baricéntricas.

## 9.- BIBLIOGRAFIA.

- /1/ V. BARNETT: "Interpreting Multivariate Data". John Wiley. (1980).
- /2/ J.P. BENZECRI et.al.: "L'Analyse des Données. Tomo II: L'Analyse des Correspondences". 4<sup>a</sup>. Ed. Dunod. (1973-1982).
- /3/ J.P. BENZECRI et.al.: "Pratique de L'Analyse des Données". Dunod (1981).
- /4/ C.M. CUADRAS: "Conceptos Metodológicos, Probabilísticos y Geométricos en Bioestadística". Publicaciones de Bioestadística y Biomatemática N<sup>o</sup> 4. Universidad de Barcelona (1983).
- /5/ K. ECKARD & G. YOUNG: "Approximation of One Matrix by Another of Lower Rank". Psychometrika, 1:211-218. (1936).
- /6/ K.R. GABRIEL: "The BIPLLOT Graphic Display of Matrices with application to principal Component Analysis". Biometrika 58:453-467. (1971).
- /7/ K.R. GABRIEL: "Least Squares Approximation of matrices by Additive and Multiplicative Models". J.R. Statist. Soc.B. 40(2):186-196. (1978).
- /8/ K.R. GABRIEL: "Biplot". Encyclopedia of Statistical Sciences (S.Ktoz and N.L. Johnson, eds.), New York: John Wiley. (1981a).
- /9/ K.R. GABRIEL: "Biplot Display of Multivariate Matrices for Inspection of Data and Diagnosis". Chapter 8 of Interpreting Multivariate Data, (V. Barnett, ed.), London: John Wiley. (1981b).
- /10/ K.R. GABRIEL & S. ZAMIR: "Lower Rank Approximation of Matrices By Least Squares with any Choice of Weights". Technometrics 21: 489-498. (1979).
- /11/ M.P. GALINDO: "Contribuciones a la representación simultánea de datos multidimensionales". Tesis Doctoral. Universidad de Salamanca (1985).
- /12/ M.P. GALINDO: "Contribuciones a la representación simultánea de datos multidimensionales" Ser. Resum. Tesis Doc. Univ. Salamanca 33 pgs. (1985).
- /13/ M.P. GALINDO y C.M. CUADRAS: "Una extensión del método Biplot y su relación con otras técnicas". Publicaciones de Bioestadística y Biomatemática de la Universidad de Barcelona. 91 pgs. (1986).

- /14/ GOLUB, G.H. & REINSCH, C: "Singular Value Descomposition and Least Squares Solution". Numer. Math. 14:403-420. (1970).
- /15/ I.J. GOOD: "Some Applications of the Singular Descomposition of a Matrix". Technometrics, 11. N<sup>o</sup> 4:823-831.(1969).
- /16/ A.S. HOUSEHOLDER & G. YOUNG: "Matriz Approximation and Latent Roots". American Mathematics Monthly, 45:165-171. (1938).
- /17/ L. LEBART, A. MORINEAU, J.P. FENELON: "Tratamiento Estadístico de Datos". Marcombo. Barcelona (1985).
- /18/ V. PEREZ MELLADO, M.P. GALINDO: "Biplot Graphic Display if Iberian and North African Populations of the Genus Podarcis. WAGLER 1830 (Sauria, Lacertidae)". Communication: Third General Meeting Societas Europa Herpetologica. Prague. (1985).
- /19/ V. PEREZ MELLADO, M.P. GALINDO: "Sistémática de Podarcis (Sauria, Lacertidae) Ibéricas y Norteafricanas mediante técnicas multidimensionales". Publicaciones de la Universidad de Salamanca 214 pags. (1985).
- /20/ C.R. RAO: "Linear Statistical Inference and its Applications" Seg. Ed. John Wiley. New-York (1965/1973).