

LOS MÉTODOS BILOT: EVOLUCIÓN Y APLICACIONES

Olesia Cárdenas*
POSTGRADO EN ESTADÍSTICA, FACES-UCV

Galindo, P.
Vicente-Villardón, J. L.
UNIVERSIDAD DE SALAMANCA, ESPAÑA.

Resumen:

Entre los métodos de análisis multivariante de datos, los Biplot han sido fuente de continuas contribuciones a la ciencia desde su origen en 1971. Su objetivo general es la aproximación de una matriz de datos, de tal forma que permita su descripción o modelización a través de mapas geométricos contruidos como proyecciones de nubes de puntos filas y columnas sobre subespacios de ajuste óptimo. Las múltiples contribuciones realizadas en las últimas décadas, nos han motivado a la realización del presente ensayo, donde se recogen en forma bastante completa las diferentes tendencias y se destacan las principales contribuciones. Esperamos que esta revisión pueda resultar de utilidad a investigadores interesados en el tema, ya que les permite seleccionar el método Biplot más adecuado para investigaciones posteriores, así como también conocer las líneas de investigación actuales que le sirvan como base para otras contribuciones en cualquier campo de las ciencias, naturales y sociales.

Palabras Claves: Análisis multivariante, Biplot, contribuciones.

1- INTRODUCCIÓN

El análisis estadístico de grandes matrices de datos constituidas por individuos u observaciones muestrales en las filas y variables en las columnas, se puede realizar utilizando técnicas de análisis multivariante, las cuales se basan en la reducción de la dimensionalidad del problema a través de la proyección de los datos originales sobre un subespacio de ajuste óptimo, de tal forma que se conserven los patrones fundamentales respecto a la variación conjunta de los individuos y las variables. Las diferencias principales entre las distintas técnicas dependen de las matrices y tipos de variables que utilizan y la forma de sus representaciones.

Entre los métodos de análisis multivariante los investigadores han dado un gran impulso a las aplicaciones de los Biplot en diversos campos de la ciencia,

* olesiac@gmail.com

dada la alta confiabilidad en sus resultados. En sus inicios ésta técnica no se difundió con la misma velocidad que otras como, Componentes Principales y Correspondencias, debido a la no incorporación a paquetes computacionales de amplio uso en estadística, lo que a través de los años se superó programando con software de análisis matemático. Las múltiples contribuciones en las últimas décadas nos motivó a la realización de este ensayo, en el cual se presentan los fundamentos teóricos, las tendencias actuales y los principales aportes, tal que permita a investigadores interesados en el tema la realización de nuevas contribuciones teóricas y aplicaciones prácticas.

2- LOS DATOS

La aplicación de los métodos Biplot requiere considerar la finalidad del análisis y la matriz de datos a utilizar.

Cuando la finalidad es descriptiva, se consideran los datos organizados en *matrices rectangulares* Y de orden $(n \times p)$, n individuos según p variables (cuantitativas o categóricas), que conforman las i filas ($i = 1, \dots, n$) y las j columnas ($j = 1, \dots, p$), respectivamente. El elemento genérico y_{ij} corresponde a la medición en el individuo i -ésimo de la variable j -ésima. En algunos casos estas matrices pueden referirse a datos longitudinales, o sea, en filas t momentos distintos, o en columnas la misma variable en t momentos.

Cuando la finalidad es analizar la interacción o independencia de dos variables categóricas, los datos se presentan en *tablas de contingencia de 2 vías*, o matriz X de orden $(I \times J)$, constituida por las frecuencias absolutas de n individuos sobre los que se han medido dos variables cada una de ellas con diferentes categorías ($i = 1, \dots, I; j = 1, \dots, J$), por lo que cada celda en la tabla contiene la frecuencia cruzada de las 2 vías. El elemento genérico x_{ij} corresponde al número de individuos que poseen simultáneamente las categorías i y j de ambas variables, siendo el total de individuos $n = \sum_i \sum_j x_{ij}$. Esta definición se puede generalizar para el caso de 3 vías, por lo que el número de vías se refiere a la dimensión geométrica de la tabla en cuestión.

Cuando la finalidad es analizar la evolución de los individuos o la estructura de grupos de variables, los datos se encuentran en *tablas múltiples* constituidas por matrices de 3 modos, que no son exactamente de 3 vías ya que los modos no están totalmente cruzados. El número de modos se refiere a los conjuntos considerados, pudiendo tratarse de 2 modos donde uno de ellos se refiere a varios conjuntos, por ejemplo, un único conjunto de individuos sobre los que se miden varios conjuntos de variables, o varios conjuntos de individuos sobre los

que se miden las mismas variables. Cuando se trata de 3 modos referidos a 3 conjuntos diferentes estas tablas coinciden con las de 3 vías.

3- LOS MÉTODOS BILOT

El origen de los Biplot se remonta a la década de los 70, cuando Gabriel en 1971 los introduce con el objetivo principal de describir aproximadamente una *matriz rectangular* utilizando una representación gráfica en baja dimensión, que permita visualizar las interrelaciones entre individuos y variables, además de las relaciones entre ambos conjuntos.

La concepción inicial es ampliada al análisis de *tablas de dos vías* cuando Bradu y Gabriel (1978) demuestran que, la colinealidad entre vectores filas y columnas en la tabla aproximada sobre el gráfico Biplot, puede dar indicios para el ajuste de ciertos modelos bilineales.

En 1981 Gabriel analiza sus ventajas y relación con otros métodos multivariantes. Posteriormente Cox y Gabriel en 1982, hacen una comparación con el Análisis Exploratorio de Datos (Tukey, 1977), demostrando que los Biplot son más intuitivos para describir un conjunto de datos. Greenacre en 1993, analiza las condiciones bajo las cuales el Análisis de Correspondencias puede ser entendido como un Biplot y en 1995a, Gabriel los usa como alternativa al Análisis de Correspondencias Múltiples.

3.1. Biplot clásicos

La aproximación Biplot de una matriz, realizada sin hacer supuestos sobre distribuciones probabilísticas subyacentes, proporciona gráficamente una ventana sobre la estructura geométrica de los datos, mostrando la variabilidad del conjunto de individuos y variables. El prefijo *bi* se refiere a la representación simultánea de filas y columnas de la matriz.

Teóricamente en un Biplot se aproxima una matriz rectangular \mathbf{Y} de orden $(n \times p)$ y rango r , por otra de rango q ($q < r$), a través de su Descomposición en Valores Singulares (DVS), o sea, $\mathbf{Y} \cong \mathbf{U}\Sigma\mathbf{V}'$, donde \mathbf{U} y \mathbf{V} son matrices de vectores singulares ortonormales tal que $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}$ (siendo \mathbf{I} la matriz identidad) y Σ es una matriz diagonal que contiene los α_k mayores valores singulares. Para garantizar la unicidad en la representación se realiza una factorización tal que, $\mathbf{Y} \cong (\mathbf{U}\Sigma^s)(\Sigma^{1-s}\mathbf{V}') = \mathbf{A}\mathbf{B}'$, siendo \mathbf{A} y \mathbf{B} las matrices que contienen las coordenadas de los $(n+p)$ vectores o marcadores filas \mathbf{a}_i y columnas \mathbf{b}_j a utilizar

simultáneamente sobre el gráfico ($i = 1, \dots, n; j = 1, \dots, p$). Dependiendo de los valores dados al escalar s en esa factorización se pueden obtener los Biplot Clásicos, el GH ($s = 0$), el JK ($s = 1$) y el SQRT ($s = \frac{1}{2}$), cada uno de los cuales se utiliza en la práctica dependiendo de los intereses de la investigación, el GH cuando se hace énfasis en las variables, el JK cuando se enfatiza en los individuos y el SQRT cuando se hace énfasis en la matriz en general.

Las propiedades de los marcadores fila y columna en la representación dependen de la factorización elegida, la cual depende de la métrica introducida en el espacio de las filas o en el espacio de las columnas, de donde:

a) En el *GH Biplot* se aproxima la matriz \mathbf{Y} bajo la restricción $\mathbf{U}'\mathbf{U} = \mathbf{I}$, siendo en este caso resaltantes los marcadores columnas \mathbf{b}_j , ya que preservan la métrica entre las columnas y tienen una calidad de representación óptima. Un GH es recomendable para la aproximación de las varianzas en la matriz $(\mathbf{Y}'\mathbf{Y})$ a través de la matriz $(\mathbf{B}\mathbf{B}')$, pero no lo es para las distancias euclídeas en $(\mathbf{Y}\mathbf{Y}')$ ya que a través de $(\mathbf{A}\mathbf{A}')$ lo que se reproducen son las distancias de Mahalanobis contenidas en $[\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}']$.

b) En el *JK Biplot* se aproxima la matriz \mathbf{Y} bajo la restricción $\mathbf{V}'\mathbf{V} = \mathbf{I}$, siendo los marcadores filas \mathbf{a}_i los resaltantes ya que preservan la métrica entre las filas, demostrándose que tienen una calidad de representación óptima. Es por ello que el JK es recomendable para la aproximación de las distancias entre individuos en $(\mathbf{Y}\mathbf{Y}')$ a través de $(\mathbf{A}\mathbf{A}')$, pero no lo es para las varianzas ya que a través de la matriz $(\mathbf{B}\mathbf{B}')$ lo que se aproxima son varianzas ponderadas contenidas en $[\mathbf{Y}'(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{Y}]$.

c) En el *SQRT Biplot* se dan papeles simétricos tanto a las filas como a las columnas, por lo que la aproximación no es única ya que se verifica que $\mathbf{A}'\mathbf{A} \neq \mathbf{I}$ y $\mathbf{B}'\mathbf{B} \neq \mathbf{I}$, razón por la cual este tipo de Biplot solo se usa cuando el objetivo del análisis consiste fundamentalmente en la aproximación de los elementos y_{ij} de la matriz \mathbf{Y} , como en el caso de la diagnosis de modelos en tablas de contingencia.

La representación Biplot sobre un plano permite, considerando las propiedades geométricas del producto interno entre marcadores fila y columna, reconstruir los datos originales: $y_{ij} \cong \mathbf{a}'_i \mathbf{b}_j = \sum_{k=1,q} \alpha_k \mathbf{u}_k \mathbf{v}'_k$. En el gráfico, las filas \mathbf{a}_i de \mathbf{A}

pueden verse como puntos, y las columnas \mathbf{b}_j de \mathbf{B} como vectores cuyo sentido representa la dirección en la que aumentan los valores de la variable correspondiente. Es por ello que su interpretación es muy sencilla como puede verse en la figura 1: las proyecciones ortogonales de los marcadores filas sobre los marcadores columnas aproximan el orden de los individuos y sus similitudes, el producto escalar entre dos marcadores columna aproxima la covarianza entre las

variables correspondientes, la longitud de los marcadores columna aproxima la desviación estándar de las variables y el coseno del ángulo entre dos marcadores columna aproxima la correlación entre las variables correspondientes.

En un Biplot es necesario considerar la bondad de la aproximación o calidad global de representación, ya que generalmente el rango de la matriz \mathbf{Y} es superior a dos, por lo que no se puede obtener su representación exacta sobre un plano. Es posible determinar un indicador global de la calidad de la aproximación (CA_{global}), y también las calidades de representación para cada fila y columna

($CR_{\text{fila/col}}$), de la siguiente forma: $CA_{\text{global}} = \frac{\lambda_1 + \lambda_2}{\sum_{i=1}^r \lambda_i}$, $CR_{\text{col}} = \frac{\mathbf{b}_{jk}^2}{\sum_{k=1}^q \mathbf{b}_{jk}^2}$, $CR_{\text{fila}} = \frac{2}{r}$,

siendo los λ_i los autovalores de $\mathbf{Y}'\mathbf{Y}$.

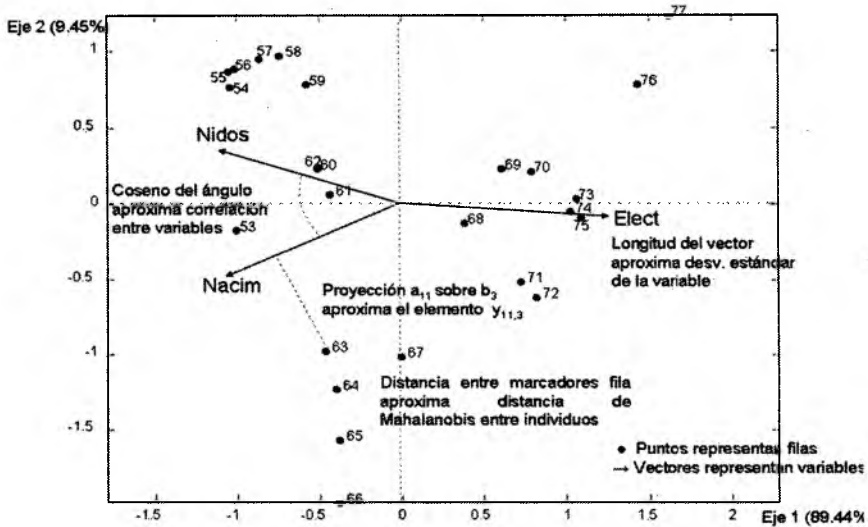
Esas fórmulas en particular garantizan alta calidad para las columnas y baja para las filas, luego corresponden a un GH Biplot, lo que se cumple en forma inversa para un JK Biplot, ya que en ese caso son las filas las que tienen alta calidad de representación.

Respecto a la medición de la bondad de ajuste individuos/variables en los distintos tipos de Biplot, Gabriel (2002) utiliza una función que depende de la razón entre los dos primeros autovalores y la factorización Biplot elegida. Realiza la comparación con el Análisis de Correspondencias (AC) y demuestra que dicha función es indicadora de buenos ajustes en el SQRT, HJ y AC, y solo en el JK y el GH cuando la razón entre los dos primeros autovalores no se aleja mucho de 0.95.

Entre algunas de las aplicaciones de los Biplots Clásicos tenemos: en el campo de la medicina Strauss y Gabriel (1979); en el campo de la meteorología Gabriel (1972) y Tsianco y Gabriel (1984).

Para ilustrar la técnica, consideramos los datos del ejemplo de Gabriel y Odoroff (1990), organizados en una matriz \mathbf{Y} , de orden (25x3), en filas 25 años desde 1953 hasta 1977, y en columnas 3 variables, número de nidos de cigüeñas en Dinamarca (Nidos), tasa de nacimientos de personas x1000 (Nacim) y consumo de electricidad per capita 1000 kwh (Elect). Se transforma la matriz inicial estandarizando por columnas por tratarse de variables con diferentes unidades de medida y ajustamos un GH Biplot. Los resultados se muestran en la figura 1.

Figura 1. GH Biplot (Gabriel y Odoroff, 1990)



La bondad en el ajuste (CA_{global}) para el primer plano factorial (ejes 1 y 2) es del 98.88%. Ninguna de las variables está mal representada, ya que las calidades de representación (CR_{col}) acumuladas para los dos primeros ejes son para las variables Nidos, Nacim y Elect, 989%, 997% y 979% respectivamente. Respecto a las calidades de las representaciones de las filas (CR_{filas}) se obtienen altas calidades para todos los años.

Se observa en esa figura la agrupación de los años dependiendo de sus relaciones con las diferentes variables, así tenemos que en la década de los 50 hay un alto número de nidos de cigüeñas, alta tasa de nacimientos y bajo consumo de electricidad. En la década de los años 60 hay un alto tasa de nacimientos y nidos de cigüeñas y un consumo medio de electricidad. En los 70 hay un alto consumo de electricidad y baja tasa de nacimientos y nidos de cigüeñas. Con respecto a las variables, se observa que nidos y nacimientos tienen (dado el ángulo entre sus vectores) una alta correlación positiva, presentando ambas variables una correlación negativa con el consumo de electricidad.

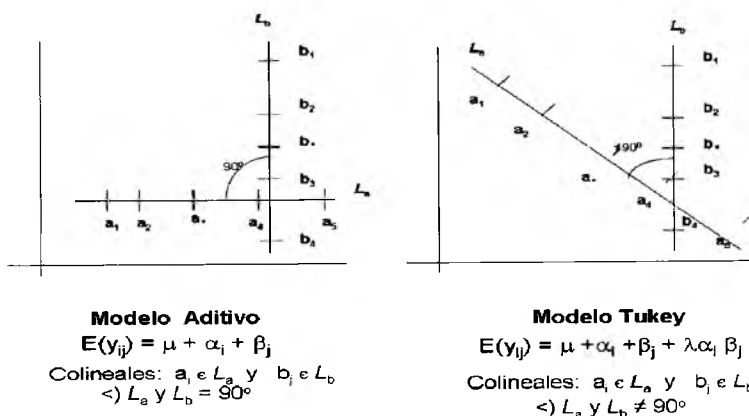
3.2. Biplot para diagnosis de modelos

Esta corriente de análisis Biplot se fundamenta en las investigaciones de Gollob (1968) y Mandel (1969). Gollob, enfoca el producto escalar entre dos vectores como una expresión que constituye una forma bilineal, mientras que

Mandel aproxima la interacción en los modelos bilineales¹ a través de términos multiplicativos tal que se puedan representar en un Biplot. Partiendo de esas premisas Bradu y Gabriel (1978) demuestran que, enfocando una aproximación Biplot como un ajuste realizado a través de un modelo bilineal de tipo multiplicativo ($y_{ij} = a_i' b_j$) se puede utilizar en la diagnosis de modelos. Posteriormente en 1979, Gabriel y Zamir, estudian la interacción en los modelos bilineales a través de una aproximación a bajo rango.

En la mayor parte de las aplicaciones prácticas no es posible saber a priori si un experimento debe considerar un término de interacción o no, por lo que los hallazgos de Bradu y Gabriel son de mucha utilidad, ya que, con una simple inspección de la posición geométrica de los marcadores fila y columna (a_i, b_j) se puede diagnosticar el modelo bilineal que mejor describe los datos. Así tenemos que a una matriz X de 2 vías se le puede ajustar un modelo aditivo, si los a_i son colineales, los b_j también lo son, y las dos líneas que unen cada tipo de marcadores son ortogonales, en caso contrario existe interacción pudiéndose ajustar modelos con interacción como por ejemplo los de Tukey (ver figura 2).

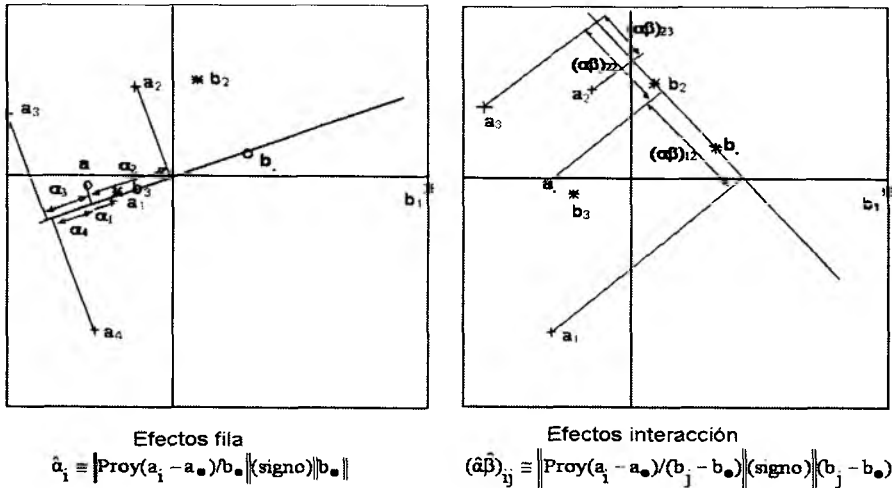
Figura 2. Colinealidad entre marcadores fila y columna en gráficos Biplot



¹ Los modelos bilineales utilizados para explicar la respuesta x_{ij} de una variable aleatoria X clasificada con respecto a dos factores en una tabla de contingencia de orden $(I \times J)$, son por definición modelos con término multiplicativo en la interacción y efectos principales aditivos. Su especificación es: $E(x_{ij}) = \mu + \alpha_i + \beta_j + \text{interacción}$, siendo $E(x_{ij})$ el valor esperado de la respuesta normalmente distribuida, μ el efecto global común, α_i el efecto del factor fila, β_j el efecto del factor columna, el efecto de interacción entre filas-columnas o debida a factores externos.

Diagnosticado el modelo bilineal subyacente a los datos, se pueden aproximar los efectos fila e interacción, tal como se muestra en la figura 3 (pudiéndose seguir para los efectos columna un procedimiento similar).

Figura 3. Aproximación de los efectos fila e interacción sobre el gráfico Biplot



Una aplicación con este enfoque la realiza Gauch (1988) para la validación y selección de modelos para estudiar la interacción genotipo-ambiente en estudios agronómicos. Gower (1990) generaliza el uso de Biplot para la diagnosis de modelos de 3 vías, en los que explica la interacción de segundo orden con un solo eje factorial. Entre otros autores que utilizan los gráficos Biplot para explicar las interacciones en tablas de dos y tres vías, tenemos: Denis (1991), Falgueroles (1995), Van Eeuwijk (1995), Choulakian (1996), Díaz-Leno (1995), Blázquez (1998), Gabriel *et al.* (1998).

4- CONTRIBUCIONES A LOS MÉTODOS BIPLLOT

Las desventajas de los métodos Biplot Clásicos frente a otros métodos como Componentes Principales y Correspondencias, han abonado el camino a los investigadores para realizar aportes en diferentes campos de aplicación. Aunque en su origen fueron dos las corrientes de investigación, la descriptiva y la de diagnosis de modelos, en la actualidad el interés parece centrarse en la descripción de tablas rectangulares y múltiples, así como en la aproximación de la matriz considerando la distribución probabilística subyacente, razón por la cual en esta revisión resaltamos importantes contribuciones siguiendo esas tres corrientes.

4.1- Biplot para descripción de matrices rectangulares

Con respecto a los Biplot Clásicos son varias las preguntas que nos podemos hacer, entre ellas: ¿es posible obtener igual calidad en la representación para filas y columnas de la matriz? ¿es posible considerar métricas diferentes a la euclídea en los espacios de las filas y columnas? ¿se pueden enfocar otros métodos de análisis multivariante como Biplot? ¿se pueden utilizar en minería de datos? ¿se pueden utilizar en el análisis de multicolinealidad entre variables? Respondiendo a esas interrogantes, presentaremos a continuación las ideas generales de algunas contribuciones, las cuales deberán ser profundizadas por los interesados utilizando la bibliografía citada en cada caso.

1) HJ Biplot: Galindo (1986) propone esta alternativa para obtener simultáneamente altas calidades de representación para filas y columnas, seleccionando la factorización $\mathbf{Y} = (\mathbf{U}\Sigma)(\Sigma\mathbf{V}') = \mathbf{A}\mathbf{B}'$. Aunque la matriz de marcadores fila $\mathbf{A} = \mathbf{U}\Sigma$ tiene las buenas propiedades del JK, y la de marcadores columna $\mathbf{B} = \mathbf{V}\Sigma$ las del GH, este tipo de Biplot presenta la desventaja que no permite reproducir los datos originales (ya que $\mathbf{Y} = \mathbf{U}\Sigma^2\mathbf{V}'$), razón por la cual algunos investigadores señalan que el HJ no es estrictamente un Biplot, sin embargo es muy utilizado dadas sus buenas propiedades, altas calidades de representación para filas y columnas, y la posibilidad de interpretar las posiciones de las filas, de las columnas y las relaciones fila-columna como en el Análisis de Correspondencias (Galindo y Cuadras, 1986). Entre sus aplicaciones podemos citar las siguientes: en el campo de la medicina Pedraz y Galindo (1986), Galindo *et al.* (1987); en el campo de la economía Vicente y Galindo (1993); en el campo de la biología Pérez Mellado y Galindo (1986), Santos *et al.* (1991), Rivas *et al.* (1993); en el campo de la tecnología ambiental García-Galegón *et al.* (1995). 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).

2) Biplot Generalizados: Vicente-Villardón en 1992 propone este tipo de Biplot como una alternativa que permite considerar la importancia de los diferentes individuos y variables. Introduce métricas definidas positivas Ω y Φ en el espacio de las filas y columnas respectivamente tal que $\mathbf{U}'\Omega\mathbf{U} = \mathbf{I}$ y $\mathbf{V}'\Phi\mathbf{V} = \mathbf{I}$, siendo necesario utilizar en la aproximación de la matriz de datos \mathbf{Y} la DVS Generalizada (DVSG) (véase Greenacre, 1984), o sea, $\mathbf{X} = [\Omega^{1/2} \mathbf{Y} \Phi^{1/2}] = \mathbf{P}\Sigma\mathbf{Q}'$, siendo $\mathbf{U} = \Omega^{-1/2}\mathbf{P}$ y $\mathbf{V} = \Phi^{-1/2}\mathbf{Q}$ matrices cuyas columnas son los vectores singulares generalizados por la izquierda y la derecha respectivamente, y Σ es una matriz que tiene en su diagonal los valores singulares generalizados. Es posible obtener los Biplot Clásicos como casos particulares del HJ-Biplot Generalizado, lo que se demuestra utilizando un enfoque similar al de Greenacre, quien considerando diferentes métricas entiende como una familia los análisis de, componen-

tes principales, correspondencias, correlaciones canónicas y canónico de poblaciones. En 1995, Vásquez realiza la formalización algebraica de los Biplot Generalizados, deduciendo sus propiedades basándose en la relación entre su aproximación y la estimación por mínimos cuadrados generalizados.

3) Biplot de Gower: Con un enfoque diferente, Gower propone varios tipos de Biplot basándose en la obtención de los marcadores columna \mathbf{b}_j a partir de la regresión multivariante, $E(\mathbf{Y}) = \mathbf{AB}'$, en la que se considera la obtención a priori de la matriz \mathbf{A} a través de métodos de escalamiento multidimensional. Gower y Harding (1988) y Gower (1992) proponen los *Biplot no Lineales*, utilizados para obtener en el ajuste trayectorias no lineales para representar a las variables que luego proyectan sobre representaciones obtenidas a través de coordenadas principales. En 1996, Gower y Hand definen los *Biplot de Interpolación y Predicción*. Con los de interpolación es posible superponer nuevos individuos proyectándolos sobre el subespacio de la representación, mientras que con los de predicción es posible inferir valores de las variables originales dado un punto sobre la representación en dimensión reducida. Se demuestra que la interpretación de los Biplot Clásicos en términos de producto escalar está relacionada con los Biplot de Predicción. En el 2003, Gower unifica la geometría de sus Biplot.

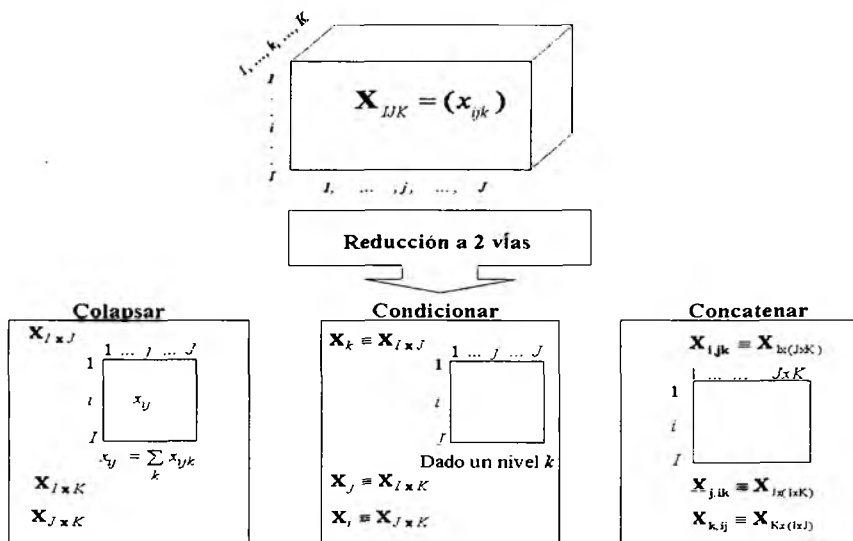
4) Biplot para minería de datos: Vairinhos incursiona en el 2003 en el campo de la minería de datos, proponiendo los Biplot como base idónea para el descubrimiento de patrones de clasificación en grandes conjuntos de datos. Propone un procedimiento basado en la formulación matemática del problema que permite aproximar conjuntos de individuos u observaciones a través de su representación por grafos de intersección. También construye un software en ambiente window para su aplicación práctica.

5) Biplot para detectar multicolinealidad: Considerando la posibilidad que brinda un Biplot en la visualización de relaciones entre variables, Ramírez *et al.* (2005) proponen un método que se basa en la descomposición espectral de la inversa de la matriz de correlaciones, demostrando que el coeficiente de inflación de varianza de una variable específica, es igual al producto interno de los correspondientes marcadores columna en el Biplot, y que la correlación parcial entre dos variables es igual, excepto por el signo, al coseno del ángulo entre los marcadores columna correspondientes.

4.2- Biplot para descripción de tablas múltiples

El análisis descriptivo de la interacción en tablas múltiples (de 3 modos) requiere su transformación a tablas de 2 vías para su aproximación Biplot a través de la DVS. Las transformaciones posibles se presentan en la figura 4.

Figura 4. Formas de reducción de una tabla de 3 vías a 2 vías



Los métodos desarrollados para su análisis pueden aplicarse también a tablas de 3 vías vistas como una sucesión de tablas de 2 vías, y los agrupamos según las siguientes corrientes:

- Integración de matrices para la obtención de una configuración consenso: Análisis Canónico de Poblaciones (Hotelling, 1936), Análisis Canónico Generalizado (Carroll, 1968; Kettenring, 1971), Análisis de Componentes Principales Doble (Bouroche y Dussaix, 1975), Statis y Statis Dual (Escofier, 1973; L'Hermier des Plantes, 1976), Análisis Factorial Múltiple (Escofier y Pagès, 1984).
- Comparación de matrices para el análisis de sus estructuras: Análisis Procrustes (Gower, 1975), Meta-Componentes Principales (Krzanowski, 1979).
- Determinación de componentes latentes para explicar cada celda en la matriz: Modelos de Tucker (1966), Métodos Tuckals (Kroonenberg y de Leeuw, 1980), Modelo de Escalamiento de 3 modos (Kroonenberg, 1983), Métodos Candecomp-Parafac e Indscal-Idioscal (Carroll y Chang, 1970, 1972; Harshman, 1970).

Cuando interesa en el análisis el descubrimiento de relaciones entre los elementos de los modos, los métodos Biplot adquieren gran importancia, por lo que presentamos a continuación una recopilación de avances que se fundamentan en los métodos de la escuela francesa para el análisis de tablas de 3 modos, el

el Statis y el Análisis Factorial Múltiple, y los métodos de la escuela anglosajona para el análisis de tablas de 3 vías, el Análisis Procrustes, Meta-Componentes Principales y los métodos Tucker.

1) Biplot Interactivo y Conjunto: Basándose en el modelo de Tucker y el método Tucker3, Carlier y Kroonenberg en 1996 consideran la generalización de la DVS a una tabla de 3 modos \mathbf{X} de orden $(I \times J \times K)$, tal que permita descomponer los residuales de interacción triple en tres matrices de marcadores, o sea: $\mathbf{X} = \mathbf{A}(\mathbf{C}' \otimes \mathbf{B}')$, siendo \mathbf{A} , \mathbf{B} y \mathbf{C} las matrices de marcadores para cada uno de los modos, \mathbf{G} una matriz diagonal de 3 vías que contiene las relaciones entre las componentes de todos los modos (generalización de la matriz Σ en la DVS de dos vías), y \otimes denota el producto Kronecker entre matrices ($\mathbf{C} \otimes \mathbf{B} = [c_{ij} \mathbf{B}]$). Para estimar las matrices \mathbf{A} , \mathbf{B} , \mathbf{C} y \mathbf{G} construyen, concatenando dos de los modos, las

siguientes matrices de dos vías: $\mathbf{X}_{i,jk} = \mathbf{A} \mathbf{G}_{i,jk} (\mathbf{B} \otimes \mathbf{C}')$, $\mathbf{X}_{j,ki} = \mathbf{B} \mathbf{G}_{j,ki} (\mathbf{C} \otimes \mathbf{A}')$ y $\mathbf{X}_{k,ij} = \mathbf{C} \mathbf{G}_{k,ij} (\mathbf{B} \otimes \mathbf{A}')$. La solución a bajo rango la obtienen a partir de un algoritmo iterativo con valores iniciales para \mathbf{A} , \mathbf{B} y \mathbf{C} . La diferencia entre los dos Biplot que proponen está en la forma como se trata la matriz inicial de datos, ya que en el *Biplot Interactivo* se combinan dos modos, mientras que en el *Biplot Conjunto* se condiciona a uno de los modos, por lo que el primer tipo es aconsejable cuando uno de los modos que interactúan está ordenado (por ejemplo modo tiempo), y el segundo tipo cuando el número de categorías de los diferentes modos es elevado. Siguiendo este enfoque, Varela (2002) introduce información externa para explicar los residuales de interacción triple.

2) Meta Biplot: Martín-Rodríguez *et al.* (2002) proponen los Meta Biplot siguiendo el enfoque de las Meta-Componentes Principales y del Análisis Procrustes. Esta contribución permite la comparación de varios Biplot, aplicados al estudio de varios individuos con varias variables a lo largo de varios momentos, varias situaciones experimentales, etc. Se parte de varias matrices de datos correspondientes a varios conjuntos de individuos (o al mismo conjunto pero en situaciones distintas), y se busca un subespacio común realizando la comparación de los ángulos que forman los subespacios respectivos a través de rotaciones hasta que coincidan lo máximo posible.

3) Biplot Múltiples: Baccalá (2004) basándose en la comparación de las nubes de variables a través de una matriz compromiso, realiza esta propuesta para el caso de datos de tablas múltiples como una alternativa a los Biplot Interactivos y Conjuntos. Pueden utilizarse: a) partiendo de una matriz obtenida concatenando las columnas de todas las tablas, como una alternativa al método Statis Dual para el caso de varios conjuntos de individuos sobre los que se observa un mismo conjunto de variables, y b) partiendo de una matriz obtenida yuxtaponiendo las columnas de las tablas, como una alternativa al Análisis Fac-

torial Múltiple para el caso de un único conjunto de individuos sobre los que se observan distintos conjuntos de variables. El procedimiento contempla una primera fase donde se realiza un Análisis de Componentes Principales de cada una de las subtablas separadas, y una segunda fase que se realiza sobre las matrices obtenidas concatenando o yuxtaponiendo tablas. Finalmente se realizan los Biplot Múltiples, un JK Biplot para aproximar los valores de la matriz inicial y un HJ Biplot para aproximar las relaciones individuos/variables.

4) Biplot Canónicos: propuestos por Vallejo-Arboleda *et al.* (2006) para analizar datos de tres vías cuando los individuos presentan estructura de grupo (replicados en el tiempo). Realizan generalizaciones de los métodos Statistal que permitan: a) identificar conglomerados y variables responsables de la separación a través de la obtención de marcadores para las medias compromiso y para las variables de todas las ocasiones, b) identificar la trayectoria de las medias y las variables compromiso a través de la obtención de marcadores para las variables compromiso y las medias de los grupos de todas las ocasiones.

4.3- Biplot considerando la Distribución Probabilística Subyacente

Considerando el hecho de que una aproximación mínimo cuadrática también se puede hallar a través de la DVS (Householder y Young, 1938), Gabriel en 1978 demuestra que los ajustes por mínimos cuadrados alternados de las matrices de marcadores fila **A** y columna **B** en la factorización Biplot conservan las propiedades inherentes a cada tipo de Biplot. De allí que esas matrices pueden ser consideradas equivalentes a las matrices de parámetros en el modelo bilineal multiplicativo $E(\mathbf{Y}) = \mathbf{AB}'$, cuando la distribución de las p variables contenidas en la matriz **Y** tienen distribución normal. La estimación de las matrices **A** y **B** siguen la filosofía de la Regresión con Rango Reducido (Davies y Tso, 1982), ya que la matriz producto de parámetros (\mathbf{AB}') es de rango incompleto por lo que se puede factorizar como en los Biplot, por lo que estimadas ambas matrices se pueden considerar a posteriori las restricciones de ortonormalidad realizando la DVS de la solución final obtenida, para luego recalcular en forma definitiva dichas matrices, o sea: $\hat{\mathbf{Y}} = \hat{\mathbf{A}}\hat{\mathbf{B}}' = \mathbf{U}\Sigma^s\Sigma^{1-s}\mathbf{V}' = \mathbf{AB}'$ y representar cualquier tipo de Biplot.

Este enfoque brinda la posibilidad de mejorar la bondad del ajuste en los Biplot, fundamentalmente en el caso de variables con respuesta no lineal, lo que ha permitido a distintos investigadores proponer alternativas para aproximar un Biplot a través del ajuste de modelos bilineales, clásicos y generalizados, como veremos a continuación:

1) Biplot a través de Regresión Bilineal Generalizada: Gabriel en 1998, propone la aproximación Biplot a través del ajuste de modelos bilineales generalizados, contemplando la posibilidad de que las variables en la matriz de datos \mathbf{Y} tengan cualquier distribución de la familia exponencial (normal, binomial, poisson, etc.). Utiliza el modelo, $g[E(\mathbf{Y})] = \mathbf{X}\Theta\mathbf{Z}'$, donde g es una función que enlaza la respuesta (en el caso de ser no lineal) con un predictor lineal, $E(\mathbf{Y})$ es el valor esperado de \mathbf{Y} , \mathbf{X} y \mathbf{Z} son matrices dadas a priori, Θ es la matriz de parámetros (desconocida excepto por su rango r) que permite reescribir el modelo como bilineal, o sea, $g[E(\mathbf{Y})] = \mathbf{X}(\mathbf{A}\mathbf{B}')\mathbf{Z}'$. Para la estimación de la matriz Θ , reespecifica el modelo de tal forma de poder estimar todas las columnas o filas simultáneamente, para lo que transforma las matrices \mathbf{Y} , \mathbf{A} y \mathbf{B} en vectores (concatenando sus columnas)² y luego utiliza las propiedades del producto Kronecker y del producto de matrices concatenadas³. En la estimación utiliza un procedimiento iterativo de regresiones generalizadas alternadas, como una generalización del método *criss-cross* para datos gaussianos (Gabriel y Zamir, 1979).

2) Biplot Logísticos: Como una alternativa confiable al Análisis de Correspondencias Múltiple para el caso de variables presencia/ausencia con distribución binomial, Vicente-Villardón *et al.* (2006) consideran el enfoque de los Biplot de Predicción y lo aplican sobre los Biplot ajustados a través de Regresión Bilineal Generalizada. Utilizan para la aproximación Biplot el modelo con término constante, $g[E(\mathbf{Y})] = \text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \mathbf{1}\mathbf{b}_0 + \mathbf{A}\mathbf{B}'$, con función enlace g logit,

siendo π la matriz de probabilidades esperadas, $(\mathbf{1}\mathbf{b}_0)$ el término que contiene la constante, \mathbf{A} y \mathbf{B} las matrices de marcadores fila y columna. La inclusión de la constante obedece al tipo de datos por lo que no tiene sentido centrarlos, y además permite calcular la probabilidad π_{j0} en el punto (0,0). Aplican un procedimiento iterativo para realizar las estimaciones máximo verosímiles de \mathbf{A} y \mathbf{B} en forma alternada, suponiendo fija en cada iteración una de esas matrices. Del análisis geométrico de sus representaciones demuestran que aún cuando el modelo contempla una respuesta no lineal, el Biplot resultante es lineal en escala logit, por lo que la representación final es igual a la de los Biplot Clásicos pudiendo interpretarse en forma similar, pero considerando además la geometría de la proyección sobre subespacios de mejor ajuste en el sentido de los mínimos cuadrados.

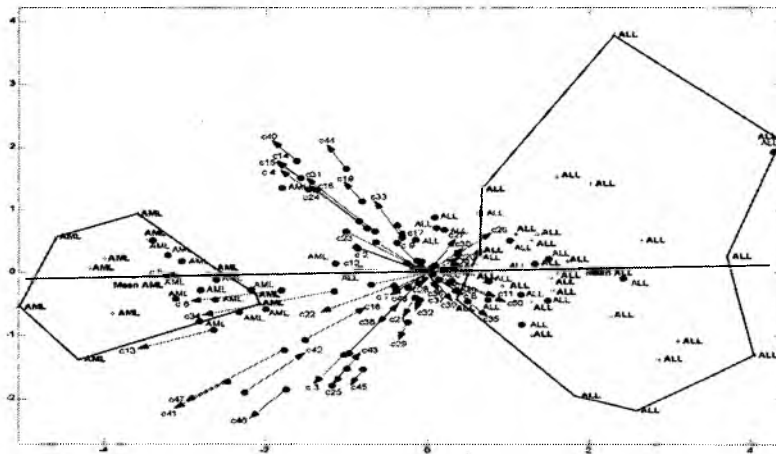
Con fines ilustrativos, se presenta un ejemplo tomado de Cárdenas *et al.* (2006), en el que se puede observar la capacidad predictiva de estos Biplot, al

² La concatenación de las columnas de una matriz \mathbf{Y} de orden $(n \times p)$, en el vector columna y de orden $(n \times 1)$ es igual a: $\mathbf{y}' = (y_1, \dots, y_j, \dots, y_p)$ siendo $y_j = (y_{j1}, \dots, y_{jn})$ para $j = 1, \dots, p$.

³ El producto \mathbf{p} de matrices concatenadas es igual a: $\mathbf{p} = (\mathbf{A}\mathbf{B}')\mathbf{C} = [\mathbf{C}' \otimes \mathbf{A}]\mathbf{b}'$

aplicarlo en la diferenciación de dos tipos de leucemia, Acute Lymphoblastic Leukemia (ALL) y Acute Myeloid Leukemia (AML) utilizando 50 genes (ADN), 22 para ALL y 28 para AML. Se utiliza una matriz Y de datos presencia-ausencia, de orden (38×50) , 27 individuos para ALL y 11 para AML. Para la validación de los resultados se proyectan sobre el gráfico 34 muestras adicionales, 20 ALL y 14 AML, como se muestra en la figura 5.

Figura 5. Biplot Logístico (Cárdenas, et al., 2006)



Nota: Las marcas en los ejes biplot (•) y (→) identifican los percentiles 50 y 75 respectivamente. Los marcadores fila AML y ALL, de la muestra original se representan con (+), y los de la muestra de validación se representan con (•).

Se observa claramente la diferenciación entre los dos tipos de leucemia. La probabilidad de presencia de cada gen c_i , se puede aproximar proyectando los marcadores fila sobre cada eje biplot, por lo que con el fin de facilitar la predicción, se introducen marcas para los percentiles 50 y 75. Se obtiene un buen ajuste para todos los genes, siendo el porcentaje de bien clasificados superior al 75%, por lo que la predicción de presencia-ausencia de cada gen es bastante aproximada.

Dada la geometría de los Biplots Logísticos los genes con baja calidad de representación, son los que coinciden en el gráfico con los ejes más largos, tal como sucede por ejemplo con c_6 , c_{47} y c_{31} . En forma inversa los ejes más cortos indican los genes con mayor poder de discriminación (los mejor representados) como son c_4 , c_5 , c_8 , c_{10} , etc. Dado el ángulo entre vectores asociados a los marcadores columna, se observan correlaciones altas positivas entre los siguientes grupos de genes, para AML (c_5 , c_6 , c_{13} , c_{22} , c_{34}), (c_{25} , c_{29} , c_{32} , c_{41} , c_{45} , c_{46} , c_{47}), (c_4 , c_{14} , c_{16} , c_{19} , c_{24} , c_{40} , c_{49}), y para ALL (c_1 , c_8 , c_{11} ,

c35, c36, c39, c49, c50), (c10, c20, c26, c27). Este hecho pareciera conformar subgrupos dentro de cada tipo de leucemia, lo cual por supuesto debe ser analizado por expertos en la materia (para mayores detalles de los datos véase Golub *et al.* 1999).

3) Biplot de Regresión: Como una generalización de los Biplot Logísticos, para el caso de variables con cualquier distribución de la familia exponencial, Cárdenas y Galindo (2004), proponen este tipo de Biplot considerando la inclusión de información externa tal que permitan explicar patrones de variación de los individuos y las variables. Utilizan un modelo bilineal generalizado con término constante, $g[E(Y)] = X(1b'_o + AB')Z'$, donde la $E(Y)$ se enlaza con predictores lineales a través de funciones g , como la logit, probit, logarítmica, etc., el término $(1b'_o)$ permite la inclusión del término independiente, X y Z son matrices de variables externas. Para la estimación de las matrices de parámetros, generalizan el método de Regresiones Bilineales Segmentadas (Gabriel, 1998) y utilizan un procedimiento iterativo que realiza alternadamente esas regresiones. En este caso, es posible realizar diferentes representaciones, y obtener gráficamente indicadores de las relaciones entre las matrices Y , X y Z . La interpretación se hace tal como en los Biplot Logísticos, pero dependiendo de la función g utilizada en el ajuste. Se demuestra que resultan una alternativa confiable a la Regresión con Componentes Principales, cuando el análisis no contempla dependencia funcional entre las variables (Cárdenas *et al.*, 2006).

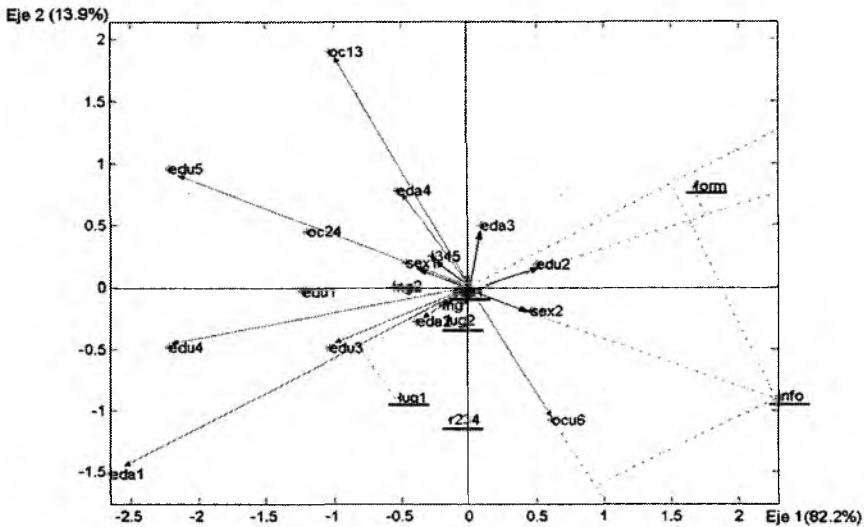
A continuación, se ejemplifica el método con datos de la realidad venezolana, correspondientes a la Encuesta de Hogares por Muestreo, Área Metropolitana de Caracas (AMC), segundo semestre del año 2000. Se analizan las características de la fuerza de trabajo, de acuerdo a la dicotomía sector formal e informal, lugar de nacimiento y tiempo de residencia en el AMC (Cárdenas, *et al.*, 2005). Los datos corresponden a 3.329 personas en la muestra, sobre los cuales se realizó un análisis de conglomerados. Se ajusta un Biplot de Regresión Logístico con Información Externa, a uno de los conglomerados obtenidos, constituido por 565 personas, 70% de sexo femenino, y 74% del sector informal. Se construyen dos matrices presencia-ausencia, la matriz Y de datos a ajustar, de orden (565×17) , y una matriz X variables externas, de orden (565×6) .

En la matriz Y , las filas corresponden a las 565 personas, y las columnas a 17 categorías-variables: edad (eda1: *menos 24 años*; eda2: *25 a 44*; eda3: *45 a 64*; eda4: *más de 65*), sexo (sex1: *masculino*; sex2: *femenino*), nivel educativo (edu1: *inferior a primaria*; edu2: *primaria*; edu3: *media*; edu4: *técnica*; edu5: *universitario*), ingresos (ing1: *menos 80 mil bolívares*; ing2: *entre 80 y 143 mil*; ing345: *entre 144 y 532 mil*), ocupación (oc13: *empleados*; oc24: *obreros*; oc6: *trabajador por cuenta propia*). En la matriz X , las filas corresponden a 565 personas, y las columnas a 6 categorías-variables: sector en el que trabaja (form:

formal, info: informal), lugar de nacimiento (lug1: venezolanos; lug2: extranjeros), tiempo de residencia en el AMC (r1: siempre; r234: menos de 9 años).

Se transforman las matrices iniciales centrando por filas y columnas, y se ajusta el Biplot. Los resultados se muestran en la figura 6.

Figura 6. Biplot de Regresión Logístico con Información Externa (Cárdenas, et al., 2005)



A diferencia de los ejemplos en las figuras 1 y 5, el ajuste considerando información externa, conlleva a la no representación (en forma directa) de los 565 marcadores filas. Estos marcadores, al ser determinados a través de combinaciones lineales de las categorías en la matriz X , se reducen a tantos como columnas tenga esa matriz, o sea 6 marcadores filas (subrayados en azul en la figura 6). Se interpretan las proyecciones ortogonales de los marcadores fila, sobre los marcadores columnas, como correlaciones entre las categorías de las variables en X e Y , lo que a su vez permite una ordenación de los individuos caracterizados según las categorías en X . En el gráfico obtenido, se interpretan las proyecciones de los las categorías en X y sus correlaciones con las de la matriz Y , así como también los ángulos entre categorías en Y .

La bondad en el ajuste, para el primer plano factorial (ejes 1 y 2), es del 96.1%. Se obtiene un buen ajuste para todas las categorías con un alto porcentaje de bien clasificados. Dada la geometría de los Biplot Logísticos, las categorías con baja calidad de representación, son las que coinciden en el gráfico con

los ejes más largos, tal como sucede por ejemplo, con oc13, eda1 y edu5. En forma inversa los ejes más cortos, indican las categorías mejor representadas, como son, sex2, eda3, edu2, etc. De la figura 6, se obtienen los siguientes resultados, los cuales deben ser analizados por expertos en la materia:

De las proyecciones ortogonales de los marcadores fila sobre los marcadores columna, se observan correlaciones positivas, entre: 1) *Sector informal* (info) con sex2, oc6, edu2, eda2, ing1; 2) *Sector formal* (form) con sex2, edu2, eda3; 3) *Venezolanos* (lug1) con eda1, edu3; 4) *Extranjeros* (lug2) con sex1, eda2, ing1; 5) *Siempre residenciados en el AMC* (res1) con ing2; 6) *Residencia menor a 9 años en el AMC* (res234) con oc6. También se observan correlaciones negativas, entre: 1) *Sector informal* (info) con sex1, oc13, oc24, eda1, eda4, edu5, ing345; 2) *Sector formal* (form) con eda1, eda2, ing1, edu3, edu4; 3) lug1 y res234 con eda3.

Dado el ángulo entre vectores asociados a los marcadores columna, se observan correlaciones positivas entre: 1) *Sexo masculino* (sex1) con ing345, eda4, oc24, edu5; 2) *Sexo femenino* (sex2) con ing1, eda2, eda4, oc13, edu2. Se observan también correlaciones negativas entre: 1) *Sexo masculino* (sex1) con edu2, oc6, oc13; 2) *Sexo femenino* (sex2) con edu5, oc24, oc13, ing345.

4) Manova Biplot: Amaro et al. (2004) generalizan los Manova Biplot de dos vías (Gabriel, 1972) en el contexto del Modelo Lineal General Multivariante. Realizan una representación ponderada en baja dimensión de la matriz de medias que permite estudiar y contrastar la separación entre grupos, la influencia de las variables en la separación, el efecto interacción, la importancia de las variables en la interacción y los efectos fila y columna. Para contrastar las hipótesis transforman el problema en univariante usando el principio de unión-intersección, formando combinaciones lineales de las variables originales tal que variando los coeficientes en ellas generan todas las posibles hipótesis, los estadísticos de contraste y las regiones de rechazo.

5) Biplot Genotipo-Ambiente: Yan en el 2001, utiliza los Biplot enfocándolos como modelos bilineales para el análisis de la interacción Genotipo-Ambiente (filas-columnas) en el campo de la Agronomía, tal que posibiliten el estudio del grado de estabilidad de los genotipos al ser probados en diferentes ambientes. Son varias las aplicaciones en ese sentido Yan et al. (2001), Yan y Hunt (2002), Crossa et al. (2002), Yan y Kan (2003). En el mismo campo, utilizando información externa Varela (2002) propone una generalización de la Regresión con Rango Reducido al caso de varios factores y considera la diagnosis de modelos para datos continuos de n vías a partir de los Biplot Interactivos.

6) Biplot Robustos: Hernández (2005), considerando que la DVS es muy susceptible a la presencia de valores atípicos, también aproxima los Biplot a través de modelos bilineales, para lo cual aborda el problema de la estimación de las matrices de marcadores considerando dos casos: a) eliminar los valores atípicos, para lo cual propone utilizar un ajuste HJ Biplot como una alternativa al método de Peña y Prieto (2001), b) no eliminar los valores atípicos y realizar estimaciones robustas utilizando el método de mínimo determinante de covarianza (Rousseeuw y Van Driessen, 1999). Utiliza en la estimación el algoritmo propuesto por Hubert y Engelen (2004) que toma como punto de partida el resultado de un Análisis de Componentes Principales Robusto.

7) Biplot basados en Modelos de Rasgos Latentes: Yáñez (2003) generaliza los Biplot Logísticos proponiendo un método para el ajuste de los Biplot basados en modelos de rasgos latentes para datos dicotómicos y politómicos. Posteriormente Sepúlveda (2004) utiliza los Biplot para detectar la violación del principio de independencia local en los modelos de clases latentes a través de la evaluación gráfica de diferentes hipótesis de independencia condicional entre distintos pares de variables manifiestas y latentes.

5- DISCUSIÓN

Dado el carácter del trabajo no es fácil formular conclusiones ya que no se obtuvo resultado alguno, más bien podría decirse, que se espera que de esta revisión, surjan nuevas alternativas a los métodos Biplot y se multipliquen sus aplicaciones prácticas en cualquier campo de las ciencias, naturales y sociales. La comprensión de los métodos aquí expuestos, requiere su estudio minucioso remitiéndose a la bibliografía citada en cada caso, dado que la presentación por nuestra parte, necesita la realización de un compendio con carácter monográfico, el cual pensamos realizar en un futuro no lejano.

Hemos observado un vacío en las contribuciones siguiendo las dos últimas corrientes en el análisis multivariante de datos, la que se interesa nuevamente por la estadística inferencial clásica, y la que considera la asignación de roles asimétricos a las variables (dependientes e independientes), por lo que constituyen áreas abiertas a la investigación, la utilización en las aproximaciones Biplot de técnicas de remuestreo para el análisis de la estabilidad en las estimaciones y determinación de distribuciones asintóticas de los estimadores, y la inclusión de relaciones de dependencia funcional entre las variables.

6- REFERENCIAS BIBLIOGRÁFICAS

- Amaro, I., Vicente-Villardón, J. L. y Galindo, P. (2004), "Manova biplot para arreglos de tratamientos con dos factores basados en modelos lineales generales multivariantes", *Interciencia*, 29 (1): 26-32.
- Baccalá, N. (2004), *Contribuciones al análisis de matrices de datos multivía: tipología de las variables*, Tesis Doctoral, Universidad de Salamanca (USAL), España.
- Benzècri, J. (1973), *L'Analyse des données*, Tournai: La taxinomie, Dunod, París.
- Blázquez, A. (1998), *Análisis biplot basado en modelos lineales generalizados*, Tesis Doctoral, USAL, España.
- Bouroche, J., Dussaix, A. (1975), "Several alternatives for three-way data analysis", *Metra*, 14: 299-319.
- Bradu, D., Gabriel, K. R. (1978), "The biplot as a diagnostic tool for models of two-way tables", *Technometrics*, 20: 47-68.
- Cárdenas, O., Galindo, P. (2004), *Biplot con información externa basado en modelos bilineales generalizados*, Ediciones del CDCH-UCV.
- Cárdenas, O., Noguera, C., Gurrea, I., Alvarez, W. (2005), *Caracterización multivariante de la oferta de trabajo en Venezuela*, LV Convención Anual de ASOVAC, Caracas.
- Cárdenas, O., Noguera, C., Galindo, P., Vicente-Villardón, J. L. (2006), "Alternativa a la regresión con componentes principales basada en biplots de regresión", *Interciencia*, 31 (3): 160-167.
- Carlier, A., Kroonenberg, P. M. (1996), "Decompositions and biplots in three-way correspondence analysis", *Psychometrika*, 61 (2): 355-373.
- Carroll, J. (1968), "A generalized of canonical correlation analysis to three or more sets of variables", *Proceedings of 76th annual convention of the American Psychological Associations*: 227-228.
- Carroll, J., Chang, J. (1970), "Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition", *Psychometrika*, 35: 283-320.
- (1972), "IDIOSCAL (Individual Differences In Orientation Scaling): A Generalization of INDSCAL allowing idiosyncratic reference systems as well as an analytic approximation to INDSCAL", *Artículo presentado en la Psychometric Society*, Marzo Princeton, NJ.

- Choulakian, V. (1996), "Generalized bilinear models", *Psychometrika* 61 (2), 271-283.
- Cox, C., Gabriel, K. (1982), "Some comparisons of biplot display and pencil-and-paper E.D.A. methods", Launer, R.L. and Siegel, A.F. (eds.), *Modern data analysis*, 45-82, Academic Press, London.
- Crossa, J., Cornelius, P., Yan, W. (2002), "Biplots of linear-bilinear models for studying crossover Genotype X Environment interaction", *Crop Science* 42, 619-633.
- Davies, P., Tso, M. (1982), "Procedures for reduced rank regression", *Applied Statist* 31: 244-255.
- Denis, J. B. (1991), "Ajustements de modèles lineaires et bilineaires sous contraintes lineaires avec données manquantes", *Statistique Appliquée XXXIX* (2): 5-24.
- Díaz-Leno, M. S. (1995), *Los métodos biplot como herramienta de diagnóstico en la modelización de datos multidimensionales*, Tesis Doctoral, USAL, España.
- Escófier, B. (1973), "Le traitement des variables vectorielles", *Biometrics* 29: 751-760.
- Escófier, B., Pagès, J. (1984), "Multiple factorial analysis: A method to compare groups of variables", *Data Analysis and Informatics* 3: 41-55.
- Falguerolles, A. (1995), *Generalized bilinear models and generalized biplots: Some Examples*, Publications du Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse.
- Gabriel, K. R. (1971), "The biplot-graphic display of matrices with application to principal component analysis", *Biometrika* 58: 453-467.
- (1972), "Analysis of meteorological data by means of canonical decomposition and Biplots", *Journal of Applied Meteorology* 11: 1071-1077.
- (1978), "Least squares approximation of matrices by additive and multiplicative models", *Journal of the Royal Statistical Society* 40B (2): 186-196.
- (1981), "Biplot display of multivariate matrices for inspection of data and diagnosis", Barnett V. (ed.), *Interpreting Multivariate Data*, 147-173, Wiley, London.
- (1995), "Biplot display of multivariate categorical data, with comments on multiple correspondence analysis", W. J. Krzanowski (ed.), *Recent Advances in Descriptive Multivariate Analysis*, 190-226, Clarendon Press, Oxford.
- (1995), "Manova biplots for two-way contingency tables", W. J. Krzanowski (ed.), *Recent Advances in descriptive Multivariate Analysis*, 227-268, Clarendon Press, Oxford.

- (1998), "Generalised bilinear regression", *Biometrika* 85 (3): 689-700.
- (2002), "Goodness of fit of biplots and correspondence analysis", *Biometrika* 89 (2): 423-436.
- Gabriel, K. R., Zamir, S. (1979), "Lower rank approximation of matrices by least squares with any choice of weights", *Technometrics* 21: 489-498.
- Gabriel, K. R., Odoroff, C. L. (1990), "Biplots in biomedical research", *Statistics in Medicine* 9 (5): 469-485.
- Gabriel, K. R., Galindo, M. P., Vicente-Villardón, J. L. (1998), "Use of biplots to diagnose independence models in three-way contingency tables", J. Blasius and M. Greenacre (eds.), *Visualization of Categorical Data*, 391-404, Academic Press, London.
- Galindo, P. (1986), "Una alternativa de representación simultánea: HJ-Biplot", *Questiio* 10 (1): 13-23.
- Galindo, P., Cuadras, C. (1986), "Una extensión del método Biplot y su relación con otras técnicas", *Publicaciones de Bioestadística y Biomatemática* 17, Universidad de Barcelona, España.
- 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.
- 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 Commission, Dg Xii D-I.
- Gauch, H. (1988), "Model selection and validation for yield trials with interaction", *Biometrics* 44: 705-715.
- Gollob, H. (1968), "A statistical model which combines features of factor analytic and analysis of variance techniques", *Psychometrika* 33: 73-115.
- Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., Caligiuri, M., Bloomfield, C., Lander, E. (1999), "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring", *Science* 286: 531-537.
- Gower, J. (1975), "Generalized procrustes analysis", *Psychometrika* 40: 33-51.
- (1990), "Three-dimensional biplots", *Biometrika* 77 (4): 773-785.

- (1992), "Generalized biplots", *Biometrika* 79 (3): 475-493.
- (2003), "Unified biplot geometry", Ferligoj A. and Mrvar A (eds.), *Developments in Applied Statistics*, Metodološki zvezki, 19, Ljubljana: FDV.
- Gower, J., Harding S. (1988), "Nonlinear biplots", *Biometrika* 75: 445-455.
- Gower, J., Hand D. (1996), *Biplots*, Chapman and Hall, London.
- Greenacre, M. (1984), *Theory and applications of correspondence analysis*, Academic Press, London.
- (1993), "Biplots in correspondence analysis", *Journal of Applied Statistics* 20 (2).
- Harshman, R. (1970), "Foundations of the PARAFAC procedure: models and conditions for an explanatory multi-mode factor analysis", *Working Papers in Phonetics* 16: 1-84, UCLA, USA.
- Hernández, S. (2005), *Biplot robustos*, Tesis Doctoral, USAL, España.
- Hotteling, H. (1936), "Simplified calculation of principal components", *Psychometrika* 1: 27-35.
- Householder, A., Young, G. (1938), "Matrix approximation and latent roots", *Am. Math. Monthly* 45: 165-171.
- Hubert, M., Engelen, S. (2004), "Robust PCA and classification in biosciences", *Bioinformatics* 20: 1728-1736.
- Kettenring, J. (1971), "Canonical analysis of several sets of variables", *Biometrika* 58: 433-460.
- Kroonenberg, P. (1983), *Three-mode principal components analysis*, DSWO Press, Leiden.
- Kroonenberg, P., de Leeuw, J. (1980), "Principal component analysis of three-mode data by means of alternating least squares algorithms", *Psychometrika* 45: 69-97.
- Krzanowski, W. (1979), "Between-groups comparison of principal components", *Journal of the American Statistical Association* 74 (367): 703-707.
- L'Hermier des Plantes, H. (1976), *Structuration des tableaux a trois indices de la statistique*, PhD Thesis, Université des Sciences et Techniques du Languedoc, Francia.
- Mandel, J. (1969), The partitioning of interaction in analysis of variance, *J. Nat. Bur. Stand.* 73B: 309-328.

- Martín-Rodríguez, J., Galindo, P., Vicente-Villardón, J. L. (2002), "Comparison and integration of subspaces from a biplot perspective", *Journal of Statistical Planning and Inference* 102(2): 411-423.
- Pedraz, C., Galindo, P. (1986), "Study of socio-cultural factors influencing the decision to breast-feed instead of bottle-feed", *Arch. Pediat.* 36: 469-477.
- Peña, D., Prieto, F. (2001), "Multivariate outlier detection and robust covariance matrix estimation", *Technometrics* 43 (3): 286-300.
- 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.
- Ramírez, G., Vasquez, M., Camardiel, A., Pérez, B., Galindo, P. (2005), "Detección gráfica de la multicolinealidad mediante el h-plot de la inversa de la matriz de correlaciones", *Revista Colombiana de Estadística* 28 (2): 207-219.
- 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 Viticultura* 44 (3): 302-308.
- Rousseeuw, P.J., van Driessen, K. (1999), "A fast algorithm for the minimum covariance determinant estimator", *Technometrics* 41: 212-223.
- Santos, C. et. al. (1991), "Characterization of young red wines by application of HJ-Biplot Analysis to anthocyanin profiles", *Journal of Agricultural and food chemistry* 39: 1086-1090.
- Sepúlveda, R. (2004), *Contribuciones al análisis de clases latentes en presencia de dependencia local*, Tesis Doctoral, USAL, España.
- Strauss, J., Gabriel, K. R. (1979), "Do psychiatric patients fit their diagnosis, pattern of symptomatology as described with the biplot", *Journal of Nervous and Mental Disease* 167: 105-113.
- Tsianco, M., Gabriel, K. R. (1984), "Modeling temperature data: An illustration of the use of biplot and bimodels in non-linear modeling", *Journal of Climate and Applied Meteorology* 23: 787-799.
- Tucker, L. (1966), "Some mathematical notes on three-mode factor analysis", *Psychometrika* 31: 279-311.
- Tukey, J. (1977), *Exploratory data analysis*, Addison-Wesley.
- Vairinhos, V. (2003), *Desarrollo de un sistema para minería de datos basado en los métodos Biplot*, Tesis Doctoral, USAL, España.

- Van Eeuwijk, F. (1995), "Multiplicative interaction in generalized linear models", *Biometrics* 51: 1017-32.
- Vallejo-Arboleda, A., Vicente-Villardón, J. L., Galindo-Villardón, P. (2006), "Canonical STATIS: Biplot analysis of multi-table group structured data based on STATIS-ACT methodology", *Computational Statistics & Data Analysis*. Article in Press. Available online <http://dx.doi.org/10.1016/j.csda.2006.04.032>.
- Varela, M. (2002), *Los métodos Biplot como herramienta de análisis de interacción de orden superior en un modelo lineal/bilineal*, Tesis Doctoral, USAL, España.
- Vásquez, M. (1995), *Aportaciones al análisis Biplot: Un enfoque algebraico*, Tesis Doctoral, USAL, España.
- Vicente, S., Galindo, M. P. (1993), "Análisis gráfico y descripción estructural de la variabilidad de cultivos en Castilla-León", *Investigación Agraria. Economía* 8 (3): 17-329.
- 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, USAL, España.
- Vicente-Villardón, J. L. (1992), *Una alternativa a las técnicas factoriales basada en una generalización de los métodos Biplot*, Tesis Doctoral, USAL, España.
- Vicente-Villardón, J. L., Galindo, P., Blázquez, A. (2006), "Logistic Biplots", Greenacre M and Blasius J (eds.), *Multiple Correspondence Analysis and related methods*: 491-509. Chapman & Hall, London.
- Yáñez, M. (2003), *Modelos de rasgos latentes basados en métodos Biplot*, Tesis Doctoral, USAL, España.
- Yan, W. (2001), "GGE Biplot. A windows application for graphical analysis of multi-environment trial data and other types of two-way data", *Agronomy Journal* 93: 1111-1118.
- Yan, W., Cornelius, P., Crossa, J., Hunt, L. (2001), "Two types of GGE biplots for analyzing multi-environment trial data", *Crop Science* 41: 656-663.
- Yan, W., Hunt, L. (2002), "Biplot analysis of diallel data", *Crop Science* 42: 21-30.
- Yang, W., Kang, M. (2003), *GGE biplot analysis: a graphical tool for breeders, geneticists, and agronomists*, CRC Press LLC, New York.