

**BIPLOT CON
INFORMACION EXTERNA
BASADO EN
MODELOS BILINEALES
GENERALIZADOS**

**Olesia C. Cárdenas C.
M^a Purificación Galindo V.**

Universidad Central de Venezuela
Consejo de Desarrollo Científico y Humanístico

Caracas, 2001

BIPLOT CON INFORMACION EXTERNA BASADO EN MODELOS BILINEALES GENERALIZADOS

Olesia C. CÁRDENAS C.

*Profesor Asociado de la Escuela de Economía
Universidad Central de Venezuela*

M^a Purificación GALINDO V.

*Profesor Titular del Departamento de Estadística
Universidad de Salamanca*

Universidad Central de Venezuela
Consejo de Desarrollo Científico y Humanístico

Caracas, 2001

A nuestras familias

Agradecimientos

En primer lugar deseamos agradecer a la Universidad Central de Venezuela y a la Universidad de Salamanca la realización de este trabajo de investigación, que fue posible gracias al Convenio Específico de Cooperación: “Programa de Investigación y Doctorado en Estadística Multivariante” firmado en Abril de 1999, el cual se deriva del Convenio Marco N° 133-90 celebrado en 1990 entre ambas Universidades y es producto de una Carta de Intención suscrita en Julio de 1997.

Agradecemos al Departamento de Estadística de la Universidad de Salamanca, a través del Doctorado en Estadística Multivariante Aplicada, la posibilidad de desarrollar esta investigación en su seno, continuando con una de sus líneas de investigación.

Muy especialmente agradecemos al Dr. José Luis Vicente Villardón, Profesor Titular del Departamento de Estadística de la Universidad de Salamanca, por sus valiosas sugerencias y por la realización del Programa de Computación utilizado en la aplicación práctica.

Finalmente nuestro agradecimiento al Consejo de Desarrollo Científico y Humanístico de la Universidad Central de Venezuela por su apoyo institucional y financiero para la publicación de esta obra.

INDICE

PRESENTACION **i**

CAPITULO I – LOS METODOS BIPLOTS **1**

1.-INTRODUCCION	2
2.- FUNDAMENTACION TEORICA DE LOS BIPLOT CLÁSICOS	8
3.-TIPOS DE BIPLLOT	13
3.1.- BIPLOTS CLASICOS	13
3.1.1.- CMP BIPLLOT	14
3.1.2.- RMP BIPLLOT	16
3.1.3.- SQRT BIPLLOT	18
3.2.- HJ BIPLLOT	18
3.3.- BIPLOTS GENERALIZADOS	20
3.4.- BIPLOTS DE PREDICCION E INTERPOLACIÓN	23
4.- PROPIEDADES DE LOS MARCADORES	26
4.1.- MARCADORES COLUMNA	26
4.1.1.- MARCADORES TIPO B = VD	26
4.1.2.- MARCADORES TIPO B = V	30
4.2.- MARCADORES FILAS	33
4.2.1.- MARCADORES TIPO A = UD	33
4.2.2.- MARCADORES TIPO A = U	36
5.- INTERPRETACION DE LAS REPRESENTACIONES BIPLLOT	39
5.1.- INTERPRETACION GEOMETRICA	39
5.2.- BONDAD DE LA APROXIMACION BIPLLOT	46
5.3.- INDICADORES COMPLEMENTARIOS EN LA INTERPRETACION	49

CAPITULO II - FORMAS ALTERNATIVAS DE APROXIMACIONES BILOT	51
1.- INTRODUCCIÓN	52
2.- APROXIMACIONES BILOT A TRAVES DE LA DESCOMPOSICION EN VALORES SINGULARES DE UNA MATRIZ	55
3.- APROXIMACIONES BILOT A TRAVES DE MINIMOS CUADRADOS ALTERNADOS	59
4.- MODELOS LINEALES GENERALIZADOS	64
4.1.- INTRODUCCIÓN	64
4.2.- FAMILIA EXPONENCIAL	65
4.3.- DEFINICION DE MODELOS LINEALES GENERALIZADOS	67
4.4.- FUNCIONES LINK	70
4.5.- ESTIMACIÓN	72
4.6.- BONDAD DE AJUSTE	78
5.- APROXIMACIONES BILOT A TRAVES DE MODELOS BILINEALES	81
5.1.- INTRODUCCIÓN	81
5.2.- AJUSTE DE MODELOS BILINEALES CLÁSICOS	83
5.3.- AJUSTE DE MODELOS BILINEALES GENERALIZADOS	87

**CAPITULO III.- BIPLOTS DE REGRESION
AJUSTADOS A TRAVES DE MODELOS
BILINEALES GENERALIZADOS** **95**

1.- INTRODUCCIÓN	96
2.- BIPLOTS DE REGRESION LINEAL	98
2.1.- AJUSTE DE BIPLOTS DE REGRESION LINEAL	98
2.2.- GEOMETRIA DE LOS BIPLOTS DE REGRESION LINEAL	106
3.- BIPLOTS DE REGRESION NO LINEAL	113
3.1.- AJUSTE DE BIPLOTS DE REGRESION NO LINEAL	113
3.2.- GEOMETRIA DE LOS BIPLOTS DE REGRESION NO LINEAL	115

**CAPITULO IV – BIPLOTS DE REGRESIÓN
CON INFORMACION EXTERNA** **123**

1.- INTRODUCCIÓN	124
2.- BIPLOTS DE REGRESION CON INFORMACION EXTERNA	132
3.- REPRESENTACIONES DE LOS BIPLOTS DE REGRESIÓN CON INFORMACION EXTERNA	140
3.1.- REPRESENTACION BILOT DE LOS COEFICIENTES DE REGRESION AD'	141
3.2.- REPRESENTACION BILOT DE LOS COEFICIENTES DE REGRESION BC'	144

3.3.- REPRESENTACION BILOT DE LOS COEFICIENTES DE REGRESION AB´	146
3.4.- REPRESENTACION BILOT DE LOS COEFICIENTES DE REGRESION BA´	147
4.- INTERPRETACION DE BILOTS DE REGRESIÓN CON INFORMACION EXTERNA	148
4.1.- BILOT DE LOS COEFICIENTES DE REGRESION AD´	151
4.2.- BILOT DE LOS COEFICIENTES DE REGRESION BC´	160
4.3.- BILOT DE LOS COEFICIENTES DE REGRESION AB´	163
4.4.- BILOT DE LOS COEFICIENTES DE REGRESION BA´	165
5.- PROPIEDADES DE LOS MARCADORES EN LAS REPRESENTACIONES DE LOS BILOTS DE REGRESION CON INFORMACION EXTERNA	169
CAPITULO V.- ESTIMACION MAXIMO VEROSÍMIL EN BILOT DE REGRESIÓN	175
<hr/>	
1.-INTRODUCCION	176
2.- ESTIMACION MAXIMO VEROSIMIL CONJUNTA EN BILOTS DE REGRESIÓN	180
2.1.- MODELOS SIN TERMINO INDEPENDIENTE	182
2.1.1.- BILOTS DE REGRESION LINEAL	190
2.1.2.- BILOTS DE REGRESION NO LINEAL	195
2.2.- MODELOS CON TERMINO INDEPENDIENTE	200

3.- ESTIMACION MAXIMO VEROSIMIL SEPARADA EN BIPLOTS DE REGRESIÓN	208
3.1.- MODELOS SIN TERMINO INDEPENDIENTE	208
3.2.- MODELOS CON TERMINO INDEPENDIENTE	212
4.- ESTIMACION EN BIPLOTS DE REGRESION CON INFORMACION EXTERNA	214
5.- PROPIEDADES DE LOS ESTIMADORES MAXIMO VEROSIMILES EN LOS BIPLOTS DE REGRESIÓN	216
6.-ESTIMACION PARA DATOS MIXTOS	218
7.- ESTIMACION MAXIMO VEROSIMIL CONJUNTA EN BIPLOTS DE REGRESIÓN: UNA APLICACION	219
CONCLUSIONES	236
<hr/>	
BIBLIOGRAFÍA	240
<hr/>	
APENDICES	252
<hr/>	
APENDICE I – NOTACION UTILIZADA	253
APENDICE II – DEFINICIONES Y RELACIONES UTILIZADAS	255

PRESENTACION

Entre los métodos de análisis multivariante, los Biplot Clásicos introducidos por GABRIEL en 1971, constituyen una herramienta de gran utilidad para explorar y comprender la estructura de grandes masas de datos, ya que tienen la característica especial de proporcionar la reproducción gráfica aproximada de los mismos, permitiendo captar las interrelaciones entre conjuntos de individuos y variables, además de las relaciones entre los elementos de cada uno de esos conjuntos.

En la última década, los aportes en el desarrollo de los métodos Biplot han sido considerables, contribuyendo en ello el Departamento de Estadística de la Universidad de Salamanca, a través de una de sus líneas de investigación en el Doctorado en Estadística Multivariante Aplicada.

Aún cuando, los métodos Biplot se han venido utilizando generalmente solo en forma descriptiva, sin hacer supuestos sobre modelos subyacentes ni distribuciones poblacionales, se puede interpretar el Biplot de una matriz de datos \mathbf{Y} de orden $(n \times p)$, n individuos por p variables, como un Modelo Bilineal Multiplicativo (GOLLOB, 1968). Esa posibilidad, introduce un aspecto diferencial importante con respecto a los Biplots Clásicos, ya que se pueden utilizar también para describir aspectos resaltantes en tablas de dos vías, en las que se mide sólo una variable que se agrupa de acuerdo a dos factores de clasificación.

En ese contexto, específicamente cuando se utilizan los Biplot, para describir la interacción entre filas y columnas en Modelos Bilineales Aditivos y Multiplicativos, se pueden encontrar muchos aportes, entre otros: DENIS (1991), FALGUEROLLES (1995), VAN EEUWIJK (1995) y CHOULAKIAN (1996).

Paralelamente, pero utilizando las características geométricas de los Biplot en Modelos Bilineales Multiplicativos, para detectar posibles modelos a ajustar a la matriz de datos, se encuentran entre otros, los trabajos de: BRADU & GABRIEL (1978), GABRIEL, GALINDO y VICENTE-VILLARDON (1998).

Para la aproximación de los Biplot, los autores antes citados, se fundamentan en la generalización de los métodos de estimación utilizados en los Modelos Bilineales a los Modelos Bilineales Generalizados, considerándolos como extensiones de los Modelos Lineales Generalizados (NELDER & WEDDERBURN, 1972), cuando la variable respuesta tiene cualquiera de las distribuciones de la familia Exponencial (Normal, Binomial, Poisson, Multinomial, etc).

Aunque son muchos los trabajos en ese sentido, la mayoría de los métodos propuestos de estimación en los Modelos Bilineales Generalizados, se fundamentan en generalizaciones heurísticas. Resalta también el hecho, de que las representaciones Biplot son utilizadas finalmente para describir la interacción en tablas de dos vías, olvidando el propósito original que es la descripción de una matriz de individuos por variables.

Por su parte, GOWER (1992), GOWER & HARDING (1988), y GOWER & HAND (1996), aunque utilizan los Biplot para la descripción de una matriz (individuos por variables), dan otro enfoque diferente al de GABRIEL (1971).

Estos últimos autores, obtienen primero una ordenación de los individuos utilizando métodos de escalamiento, para luego superponer en los Biplots las variables, de forma tal que sea posible una interpretación conjunta. Este enfoque se puede relacionar, con la forma factorial clásica de la escuela francesa de análisis de datos y con los métodos de ordenación de la escuela biométrica.

Otro aspecto importante de estos últimos autores, es que describen la geometría de los Biplot, en términos de proyecciones de subespacios, en contraposición a la geometría a posteriori utilizada en la diagnosis de modelos a través de Biplots. En este contexto, VICENTE-VILLARDON & GALINDO (1998), describen la geometría de los Biplots de Regresión para el caso lineal y el caso logístico, cuando las coordenadas de los individuos se han obtenido a priori, mediante cualquier método de escalamiento.

En otro contexto muy diferente, la forma factorial para variables con distribuciones de la familia exponencial, es comparable a la obtención de variables latentes continuas en las Ciencias Sociales, tal como sucede por ejemplo, en la Teoría de Respuesta al Item (ver BAKER, 1992). Resalta en este ámbito, el hecho de que no se haya desarrollado una interpretación Biplot, considerando la comparación entre esas dos corrientes de investigación.

En esa misma línea, puede ser de utilidad práctica, la interpretación de la ordenación obtenida a través de un factor, utilizando para ello variables externas que expliquen el comportamiento del factor latente. En este sentido, TER BRAAK en 1986, propone el Análisis Canónico de Correspondencias, BLAZQUEZ en 1998, generaliza la idea para cualquier tipo de matriz, y GABRIEL en 1998, propone un procedimiento de estimación para Modelos Bilineales Generalizados, utilizando información externa, pero para tablas de dos vías.

La finalidad de este trabajo de investigación, está enmarcada dentro de las diferentes líneas de investigación de los autores citados, o sea, tiene el propósito de describir una matriz de datos Y (individuos x variables), para variables con distribuciones pertenecientes a la familia exponencial, aproximando los Biplots a través de Modelos Bilineales Generalizados Multiplicativos, proponiendo formalmente un método de estimación similar al utilizado para la obtención de variables latentes en Ciencias Sociales, analizando la geometría para la interpretación de los resultados, y contemplando además la posibilidad de inclusión de información externa.

De allí que, los objetivos en esta investigación son los siguientes:

1. Construir un marco teórico general, que permita el ajuste de un Biplot a matrices de datos (individuos por variables), en las que las variables tengan distribuciones de la familia exponencial, utilizando la metodología de los Modelos Bilineales Generalizados y analizando las suposiciones necesarias en este contexto.

2. Estudiar la geometría para la interpretación de los Biplot basados en Modelos Bilineales Generalizados, en el caso de matrices de individuos por variables.
3. Analizar la posibilidad de inclusión de Información Externa en los Biplot, profundizando en la interpretación de los mismos.
4. Generalizar el ajuste de los Biplot con Información Externa, para variables con distribuciones de la familia exponencial.
5. Proponer formalmente métodos de estimación para los Biplot ajustados a través de Modelos Bilineales Generalizados.
6. Desarrollar un programa de computación para aplicar los métodos de estimación propuestos.

Para cumplir con esos objetivos el trabajo se ha estructurado de la siguiente forma:

Los dos primeros capítulos conforman el marco teórico de la investigación. En el capítulo I, se presentan todo lo concerniente a los métodos Biplots, fundamentación teórica, tipos, propiedades de las representaciones gráficas e interpretación de los resultados.

En el capítulo II, se presentan las distintas formas para aproximar los Biplots, desde el punto de vista clásico a través de la Descomposición en Valores Singulares de una matriz de datos, y también desde el punto de vista del ajuste a través de Modelos Bilineales.

En el capítulo III, se formaliza una propuesta para el ajuste de Biplots a través de Modelos Bilineales Generalizados, aplicando los últimos desarrollos en cuanto al ajuste de Modelos Bilineales. También se analiza la geometría de estos Biplots como base para la interpretación gráfica de los mismos.

En el capítulo IV, se presenta una propuesta para el ajuste de Biplots a través de Modelos Bilineales Generalizados, incluyendo Información Externa. Se realiza una especificación general y se analizan detalladamente las distintas representaciones Biplot que se pueden obtener.

En el capítulo V, se propone un método alternativo para la estimación de los parámetros en los Modelos Bilineales Generalizados utilizados en el ajuste de los Biplots. Se estiman los parámetros generalizando el método de máxima verosimilitud, se calculan los errores de estimación, se analizan las ventajas y desventajas del método propuesto, y finalmente se realiza una aplicación práctica.

CAPITULO I

LOS METODOS BIPLLOT

LOS METODOS BIPLLOT

1.- INTRODUCCION

Los métodos Biplot han sido fuente de continuas contribuciones a la ciencia desde su origen en 1971 hasta la fecha.

Los distintos tipos de Biplots, sus usos y aplicaciones en otros contextos, han producido una gran diversidad de investigaciones, las cuales se pueden agrupar en los siguientes dos grupos de acuerdo a la finalidad de sus aplicaciones:

- Con la finalidad de describir matrices de datos multivariantes.
- Con la finalidad de diagnosis de modelos.

Grupo 1: Representación Biplot de una matriz de datos multivariantes, con “Finalidad Descriptiva”.

Los aportes en el desarrollo de los métodos Biplot comienzan con GABRIEL (1971).

Los métodos Biplot Clásicos introducidos por GABRIEL en 1971, tienen como objetivo principal proporcionar la representación gráfica aproximada en

baja dimensión de una matriz de datos rectangular (n individuos \times p variables), con una calidad en la representación que permita captar visualmente gracias a sus propiedades geométricas, las interrelaciones entre conjuntos de individuos y variables, además de las relaciones entre los elementos de cada uno de esos conjuntos.

Se aproxima la matriz de datos a través de la Descomposición en Valores Singulares, para luego realizar una factorización Biplot en marcadores fila y columna que permita la representación gráfica, cuya interpretación se basa en las propiedades geométricas del producto escalar entre vectores fila (denominados marcadores fila) y vectores columna (denominados marcadores columna), de tal forma que dicho producto reproduzca aproximadamente cada elemento en la matriz de datos.

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

GABRIEL propone dos tipos: el GH-Biplot y el JK-Biplot. Con el GH-Biplot se consigue una alta calidad en la representación de las variables y muy pobre para los individuos. Con el JK-Biplot sucede lo contrario. De allí que la representación no es simultánea para ninguno de los dos tipos de Biplot, ya que no se consigue la misma bondad de ajuste para las filas y las columnas de la matriz de datos.

Entre algunas de las aplicaciones de los Biplots Clásicos: en el campo de la medicina STRAUSS & GABRIEL (1979), GABRIEL & ODOROFF (1990); y en el campo de la meteorología TSIANCO & GABRIEL (1981, 1984).

Como una alternativa a los Biplot Clásicos de GABRIEL (1971, 1981a), GALINDO (1985, 1986) demuestra que es posible con una conveniente selección de los marcadores, que ambos conjuntos se puedan representar simultáneamente sobre un mismo sistema de coordenadas cartesianas, con la misma calidad en la representación tanto para las filas como para las columnas.

GALINDO denomina a ese tipo de Biplot HJ-Biplot, en el cual se eligen los marcadores en la misma forma que GABRIEL, pero permitiendo además una representación gráfica en la misma forma que el Análisis de Correspondencias (BENZECRI, 1973).

Entre algunas aplicaciones del HJ-Biplot tenemos: en el campo de la medicina PEDRAZ y GALINDO (1986), GALINDO y cols. (1986, 1987); en el campo de la economía VICENTE y GALINDO (1993); en el campo de la biología PEREZ MELLADO & GALINDO (1986), SANTOS y cols. (1991), RIVAS y cols. (1993).

VICENTE-TAVERA en 1992, utiliza el HJ-Biplot para realizar una Clasificación Ascendente Jerárquica. Conformar los conglomerados en un Análisis de Clusters con la información obtenida del HJ. El criterio que utiliza

se corresponde al criterio de la inercia descrito por BENZECRI (1973). Una aplicación de esta metodología en el campo de la tecnología ambiental puede encontrarse en GARCIA-GALEGON, TRUJILLANO & VICENTE-TAVERA (1995).

Como usualmente los métodos Biplot utilizan distancias pitagóricas, y dado que en la práctica muchas situaciones necesitan métricas más generales, que permitan considerar la importancia de los diferentes individuos o las relaciones entre las variables, VICENTE-VILLARDON en 1992 define los Biplot Generalizados los cuales permiten la inclusión de diferentes métricas para filas y columnas, siguiendo las ideas de GABRIEL (1971) y GREENACRE (1984).

Se pueden enfocar como casos particulares de los Biplot Generalizados la mayoría de las técnicas de Análisis de Datos Multivariantes, entre ellas: el Análisis de Componentes Principales, el Análisis de Correlaciones Canónicas, el Análisis Factorial de Correspondencias, el Análisis Canónico de Poblaciones, tal como demuestra VICENTE-VILLARDON, 1992.

GOWER & HARDING (1988) y GOWER (1992), definen los Biplot no Lineales (denominados también Biplot Generalizados), obteniendo en el ajuste trayectorias no lineales para representar a las variables.

VASQUEZ en 1995, aporta como contribución a los métodos Biplot, la formalización algebraica de los Biplots Generalizados (con diferentes

métricas), enfocados como Modelos Bilineales. Este enfoque resalta la importancia que diferentes tipos de centrado y/o estandarización de los datos tiene en la distancia euclídea entre filas y entre columnas de la matriz de datos. Una aplicación de ésta metodología en el campo de la medicina puede verse en VASQUEZ y cols. (1999).

GOWER & HAND (1996), definen los Biplot de Interpolación y Predicción, basados en puntos de vista diferentes a los Biplots Clásicos de GABRIEL (1971). Los Biplot de Predicción pueden ser utilizados en la interpretación de las ordenaciones obtenidas mediante Multidimensional Scaling, en el sentido de KRUSKAL & WISH (1978).

Grupo 2: Representación Biplot de una matriz de datos multivariantes, para “*Diagnosis de Modelos*”.

Los aportes en este sentido, se fundamentan en las investigaciones de GOLLOB (1968) y MANDEL (1969).

GOLLOB en 1968, enfoca el producto escalar entre dos vectores como una expresión que constituye una “forma bilineal”, obtenida al efectuar el producto entre el vector de la fila “i” y el vector de la columna “j” de la matriz de datos, tal que admita una representación gráfica a través de la proyección ortogonal del vector fila sobre el vector columna y viceversa.

MANDEL en 1969, aproxima la interacción en los modelos, a través de términos bilineales, tal que se puedan representar en un Biplot.

Posteriormente, BRADU & GABRIEL (1974, 1978), demuestran que la colinealidad entre vectores fila y vectores columna sobre el gráfico Biplot, puede dar indicios para el ajuste de ciertos Modelos Bilineales en tablas de dos vías.

GABRIEL & ZAMIR en 1979, estudian la interacción en los Modelos Bilineales a través de una aproximación a bajo rango.

Entre otros autores que utilizan los gráficos Biplot para explicar las interacciones en tablas de dos y tres vías, tenemos los siguientes: DENIS (1991), FALGUEROLLES (1995), VAN EEUWIJK (1995), CHOULAKIAN (1996), DIAZ-LENO (1995), BLAZQUEZ (1998), GABRIEL, GALINDO y VICENTE-VILLARDON (1998).

En este mismo grupo pero en otro contexto, utilizando los Biplots para describir una cierta matriz de datos, pero considerando la distribución probabilística que genera a los datos, tenemos: VICENTE-VILLARDON & GALINDO, (1998); VICENTE-VILLARDON, GALINDO & CARDENAS, (2000).

La adecuación de todas las contribuciones en este grupo, para ser utilizadas con “la finalidad de describir una matriz de datos utilizando información externa, y aproximando la matriz a través del ajuste de modelos bilineales generalizados”, es la finalidad de éste trabajo de investigación.

A continuación, en este capítulo, se expondrán los fundamentos básicos de los métodos Biplot, para después en el capítulo II presentar en forma resumida, todo lo concerniente a las distintas formas de aproximar una matriz para obtener una representación Biplot.

2.- FUNDAMENTACION TEORICA DE LOS BILOT CLASICOS

El objetivo principal de los Biplot Clásicos (GABRIEL, 1971) es proporcionar la reproducción gráfica aproximada de los datos, con una calidad en la representación que permita, gracias a sus propiedades geométricas, captar visualmente las interrelaciones entre conjuntos de individuos y variables, además de las relaciones entre los elementos de cada uno de esos conjuntos.

La notación a utilizar en esta investigación es la que usualmente se utiliza en la terminología estadístico-matemática, o sea: las matrices se denotan en letras mayúsculas en negrita, los vectores en letras minúsculas en negrita, los vectores de parámetros en letras griegas, las matrices de parámetros en letras mayúsculas en negrita, los sub-índices denotan los elementos fila y/o columna en las matrices (los sub-índices “i” y “s” se usan para las filas, mientras que “j” y “k se usan para las columnas).

Para facilidad del lector, en cada capítulo y/o apartado se irá introduciendo la notación específica a utilizar, presentándose además un resumen de toda la notación utilizada, en el Apéndice I.

La fundamentación teórica de los Biplot Clásicos, se basa en la aproximación de una cierta matriz de datos \mathbf{Y} de orden $(n \times p)$ y de rango r , por una de bajo rango $(q < r)$ a través de la Descomposición en Valores Singulares, para luego hacer una factorización en matrices de marcadores filas \mathbf{A} de orden $(n \times q)$ y de marcadores columnas \mathbf{B} de orden $(q \times p)$ tal que:

$$\mathbf{Y}_{(r)} \cong \mathbf{Y}_{(q)} = \mathbf{U}_{(q)} \mathbf{D}_{(q)} \mathbf{V}'_{(q)} = \sum_{k=1}^q \alpha_k u_k v'_k$$

$$\mathbf{Y}_{(q)} = \mathbf{A}_{(q)} \mathbf{B}'_{(q)} \quad (1.1)$$

Donde:

$\mathbf{Y}_{(r)}$ es la matriz de datos de orden $(n \times p)$ y rango r

$\mathbf{Y}_{(q)}$ es la aproximación de la matriz \mathbf{Y} de rango q ($q < r$)

$\mathbf{U}_{(q)}$ es una matriz de orden $(n \times q)$, cuyos q vectores columna son los vectores singulares por la izquierda de $\mathbf{Y}\mathbf{Y}'$, los cuales son ortonormales, por lo que

$$\mathbf{U}'\mathbf{U} = \mathbf{I}$$

$\mathbf{V}_{(q)}$ es una matriz orden $(p \times q)$ cuyos vectores columna son los vectores singulares por la derecha de $\mathbf{Y}'\mathbf{Y}$, ortonormales, luego $\mathbf{V}'\mathbf{V} = \mathbf{I}$

$\mathbf{D}_{(q)}$ es una matriz diagonal que contiene los q mayores valores singulares α_k asociados a los λ_k valores propios de la matriz $\mathbf{Y}'\mathbf{Y}$, ($\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_q$)

$\mathbf{A}_{(q)}$ es la aproximación de rango q de la matriz \mathbf{A} $(n \times q)$ de marcadores fila

$\mathbf{B}_{(q)}$ es la aproximación de rango q de la matriz \mathbf{B} ($q \times p$) de marcadores columna.

Si el rango de la matriz \mathbf{Y} es $q = 2$, entonces el Biplot representa en forma exacta a la matriz sobre el plano. Cuando el rango de \mathbf{Y} es $q > 2$, entonces la factorización Biplot garantiza la representación gráfica aproximada de la matriz, ya que cada y_{ij} puede reconstruirse de la siguiente forma:

$$y_{ij (q)} = \mathbf{a}_i'_{(q)} \mathbf{b}_j_{(q)} \quad (i = 1, 2, \dots, n; \quad j = 1, 2, \dots, p) \quad (1.2)$$

"**expresión que constituye una forma bilineal**" (GOLLOB, 1968) obtenida al efectuar el producto interno de la fila i de $\mathbf{A}_{(q)}$ y la columna j de $\mathbf{B}'_{(q)}$, admitiendo una representación gráfica a través de la proyección ortogonal de $\mathbf{a}_{i(q)}$ sobre $\mathbf{b}_{j(q)}$ y viceversa.

En resumen, el Biplot de la matriz \mathbf{Y} se define como la representación gráfica sobre un plano de marcadores filas $\mathbf{a}_i = (a_{i1}, a_{i2})'$ y marcadores columna $\mathbf{b}_j = (b_{j1}, b_{j2})'$, ambos vectores de dos elementos cada uno, tal que para cada i y cada j entonces:

$$y_{ij} = \mathbf{a}_i' \mathbf{b}_j = a_{i1} b_{j1} + a_{i2} b_{j2}$$

Dicha representación sobre el plano proporciona una ventana sobre la estructura geométrica de los datos, sin realizar ningún supuesto sobre distribuciones probabilísticas subyacentes.

En la figura 1.1, se muestra la representación Biplot para una matriz \mathbf{Y} de orden (11×4) :

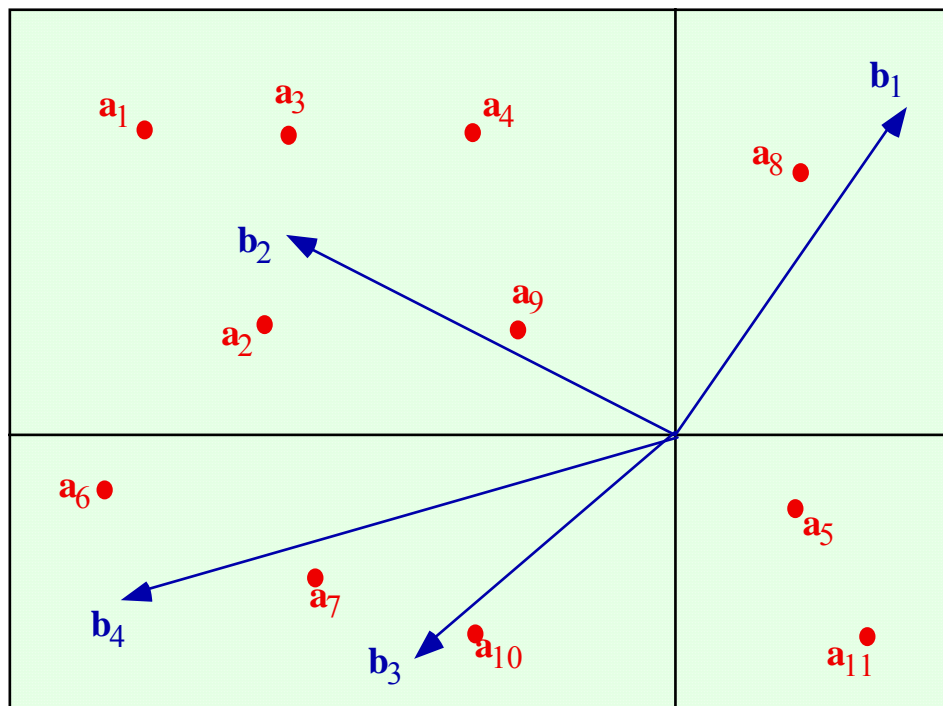


Figura 1.1.- Representación Biplot de una matriz \mathbf{Y} de orden (11×4)

Así pues, un Biplot es una representación gráfica, que muestra la distribución conjunta de varias variables, en la misma forma que un diagrama de dispersión muestra la distribución conjunta de dos variables (GABRIEL & ODOROFF, 1990).

De la aproximación Biplot de los elementos de la matriz \mathbf{Y} y aplicando la definición de producto escalar tenemos que:

$$y_{ij} = \mathbf{a}'_i \mathbf{b}_j = \|\text{Proy}(\mathbf{a}_i / \mathbf{b}_j)\|(\text{signo})\|\mathbf{b}_j\|$$

Siendo:

$\|\text{Proy}(\mathbf{a}_i / \mathbf{b}_j)\|$ = longitud de la proyección = longitud del segmento que vá desde el origen de la representación hasta el punto \mathbf{a}_i .

$\|\mathbf{b}_j\|$ = módulo de \mathbf{b}_j = longitud del segmento que une al origen con el extremo del vector \mathbf{b}_j .

O sea, y_{ij} es aproximadamente el módulo de la proyección de \mathbf{a}_i sobre \mathbf{b}_j multiplicado por la longitud de \mathbf{b}_j , con el signo correspondiente.

Así tenemos, por ejemplo en la figura 1.1 que:

$$y_{43} = \mathbf{a}'_4 \mathbf{b}_3 = \|\text{Proy}(\mathbf{a}_4 / \mathbf{b}_3)\|\|\mathbf{b}_3\|$$

3- TIPOS DE BILOT

La selección de distintas factorizaciones conduce a distintos tipos de Biplots, pero para que el Biplot sea único es necesario introducir una métrica en el espacio p-dimensional de las filas o en espacio n-dimensional de las columnas.

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

$$\mathbf{Y}_{(q)} = \mathbf{A}_{(q)} \mathbf{B}'_{(q)} = \mathbf{U}_{(q)} \mathbf{D}_{(q)}^\gamma (\mathbf{V}_{(q)} \mathbf{D}_{(q)}^{1-\gamma})' \quad (1.3)$$

De allí surgen los Biplots Clásicos de GABRIEL (1971) y el HJ-Biplot de GALINDO (1985, 1986).

En el caso de métricas distintas a la identidad se obtienen los Biplots Generalizados (VICENTE-VILLARDON, 1992).

En el caso de trayectorias no lineales para representar a las variables tenemos los Biplots no Lineales (GOWER & HARDING, 1988).

3.1.- BILOTS CLASICOS

Partiendo de la expresión (1.3) y dependiendo del valor seleccionado para γ ($\gamma = 0, 1, \frac{1}{2}$) tenemos los Biplots Clásicos de GABRIEL (1971, 1981a): el GH-Biplot, el JK-Biplot y el SQRT-Biplot.

GREENACRE en 1984 introduce una nueva terminología para el GH y el JK-Biplot, denominándolos respectivamente CMP (**C**olumn **M**etric **P**reserving) y RMP (**R**ow **M**etric **P**reserving), aludiendo al hecho que con el primer tipo se preserva la métrica para las columnas y con el segundo tipo la métrica para las filas.

Esta terminología ha sido aceptada por GABRIEL, razón por la cual en las referencias más recientes se opta por la misma.

Algunos de esos tipos de Biplot son muy utilizados en la aproximación a bajo rango de los elementos de una matriz de datos \mathbf{Y} , mientras que otros en la diagnosis de modelos.

3.1.1.- CMP – BILOT (GH – BILOT)

Para simplificar un poco la notación en adelante nos referiremos a \mathbf{Y} como la aproximación Biplot $\mathbf{Y}_{(q)}$, siendo \mathbf{A} y \mathbf{B} las aproximaciones $\mathbf{A}_{(q)}$ y $\mathbf{B}_{(q)}$ respectivamente.

Imponiendo la métrica $\mathbf{A}'\mathbf{A} = \mathbf{I}$ en el espacio p-dimensional de las filas de la matriz $\mathbf{Y}_{(n \times p)}$ para que la representación Biplot sea única (excepto por rotaciones) y eligiendo los marcadores de acuerdo a la siguiente expresión:

$$\mathbf{Y} = \mathbf{A} \mathbf{B}' = \mathbf{U} \mathbf{D}^\gamma \mathbf{D}^{1-\gamma} \mathbf{V}'$$

Cuando $\gamma = 0$ se trata del GH- Biplot o CMP- Biplot (Column Metric Preserving), luego tendremos lo siguiente:

$$\mathbf{Y} = \mathbf{A} \mathbf{B}' = (\mathbf{U} \mathbf{D}^\gamma) (\mathbf{D}^{1-\gamma} \mathbf{V}') = (\mathbf{U}) (\mathbf{D} \mathbf{V}') \quad (1.4)$$

Luego la factorización en el CMP-Biplot corresponde a la elección de los marcadores tal que:

$$\mathbf{A} = \mathbf{U} \qquad \mathbf{B} = \mathbf{V} \mathbf{D}$$

Sin embargo, haciendo honor al nombre de este tipo de Biplot (GH-Biplot) se adopta la siguiente notación para las matrices de marcadores fila ($\mathbf{A} = \mathbf{G}$) y marcadores columna ($\mathbf{B} = \mathbf{H}$), o sea:

$$\mathbf{G} = \mathbf{U} \qquad \mathbf{H} = \mathbf{V} \mathbf{D} \quad (1.5)$$

Las propiedades de estos marcadores son objeto del apartado 3 de este mismo capítulo. Esta factorización es muy útil desde el punto de vista estadístico, ya que se demuestra que:

$$\mathbf{Y}' \mathbf{Y} = \mathbf{H} \mathbf{H}'$$

De allí el nombre de este tipo de Biplot, el cual preserva la métrica euclídea usual entre las columnas pero no entre las filas, obteniéndose por lo tanto alta calidad de representación para las columnas.

Si la matriz \mathbf{Y} está centrada entonces la matriz simétrica $(\mathbf{Y}'\mathbf{Y})$ de orden $(p \times p)$ es proporcional a la matriz de varianzas y covarianzas, y el producto escalar $(\mathbf{h}'_j \mathbf{h}_k)$ representa aproximadamente las covarianzas.

Introduciendo un factor de escala, entonces la matriz $[(1/n-1)\mathbf{Y}'\mathbf{Y}]$ coincide exactamente con la matriz de covarianzas, y si se designan los marcadores como:

$$\mathbf{G} = \sqrt{n-1} \mathbf{U} \qquad \mathbf{H} = (1/\sqrt{n-1})(\mathbf{V} \mathbf{D}) \qquad (1.6)$$

Encontrándonos en ese caso en presencia del Biplot de Componentes Principales, donde los productos internos entre marcadores columna \mathbf{h}_j reproducen la estructura de las covarianzas entre variables.

3.1.2.- RMP – BILOT (JK-BILOT)

En este tipo de Biplot se impone la métrica $\mathbf{B}'\mathbf{B} = \mathbf{I}$ en el espacio n dimensional de las filas de la matriz $\mathbf{Y}_{(n \times p)}$, por lo que la representación es única (excepto por rotaciones).

Eligiendo los marcadores de acuerdo a la siguiente expresión:

$$\mathbf{Y} = \mathbf{A} \mathbf{B}' = \mathbf{U} \mathbf{D}^\gamma \mathbf{D}^{1-\gamma} \mathbf{V}'$$

Cuando $\gamma = 1$ se trata del JK-Biplot o RMP- Biplot (Row Metric Preserving), luego tendremos lo siguiente:

$$\mathbf{Y} = \mathbf{A} \mathbf{B}' = (\mathbf{U} \mathbf{D}^\gamma) (\mathbf{D}^{1-\gamma} \mathbf{V}') = (\mathbf{U} \mathbf{D}) (\mathbf{V}') \quad (1.7)$$

La factorización elegida corresponde a la elección de los marcadores tal que:

$$\mathbf{A} = \mathbf{U} \mathbf{D} \quad \mathbf{B} = \mathbf{V}$$

Haciendo honor al nombre de este tipo de Biplot (JK-Biplot), se adopta la siguiente notación para las matrices de marcadores fila ($\mathbf{A} = \mathbf{J}$) y marcadores columna ($\mathbf{B} = \mathbf{K}$), así tenemos que:

$$\mathbf{J} = \mathbf{U} \mathbf{D} \quad \mathbf{K} = \mathbf{V} \quad (1.8)$$

Las propiedades de estos marcadores son objeto del apartado 3 de este mismo capítulo.

Teniendo en cuenta la factorización elegida y la métrica impuesta se demuestra que:

$$\mathbf{Y} \mathbf{Y}' = \mathbf{J} \mathbf{J}'$$

De allí el nombre de este tipo de Biplot (RMP-Biplot), el cual preserva la métrica euclídea usual entre las filas pero no entre las columnas, obteniéndose alta calidad de representación para las filas.

3.1.3.- SQRT – BILOT (BILOT SIMETRICO)

Si en la expresión: $\mathbf{Y} = \mathbf{A} \mathbf{B}' = \mathbf{U} \mathbf{D}^\gamma \mathbf{D}^{1-\gamma} \mathbf{V}'$

Se elige la factorización tal que $\gamma = 1/2$, o sea:

$$\mathbf{Y} = \mathbf{A} \mathbf{B}' = (\mathbf{U} \mathbf{D}^{1/2})(\mathbf{D}^{1/2} \mathbf{V}') \quad (1.9)$$

Se trata entonces del SQRT- Biplot o Biplot-Simétrico ya que se dan papeles simétricos tanto a las filas como a las columnas.

La factorización en la expresión (1.9) corresponde a:

$$\mathbf{A} = \mathbf{U} \mathbf{D}^{1/2} \quad \mathbf{B} = \mathbf{V} \mathbf{D}^{1/2} \quad (1.10)$$

En este caso dicha factorización 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 de la matriz \mathbf{Y} , como en el caso de la diagnosis de modelos en tablas de contingencia donde el papel de filas y columnas es simétrico.

3.2.- HJ – BILOT (RCMP-Biplot)

Un HJ-Biplot (GALINDO, 1985, 1986) es una representación gráfica multivariante de marcadores fila y columna, elegidos de tal forma que puedan

superponerse en el mismo sistema de referencia con máxima calidad de representación.

En este tipo de Biplot se suponen los datos en la matriz \mathbf{Y} , centrados por filas y columnas, de donde la métrica introducida en el espacio de las filas es equivalente a la inversa de la matriz de covarianzas entre variables, mientras que en el espacio de las columnas la métrica introducida es equivalente a la inversa de la matriz de dispersión entre individuos.

Si la elección de los marcadores en la Descomposición en Valores Singulares de \mathbf{Y} referida a la relación (1.3) corresponde a:

$$\mathbf{Y} = \mathbf{A} \mathbf{B}' = (\mathbf{U} \mathbf{D})(\mathbf{D} \mathbf{V}') \quad (1.11)$$

En este caso la factorización elegida corresponde a:

$$\mathbf{A} = \mathbf{U} \mathbf{D} \quad \mathbf{B} = \mathbf{V} \mathbf{D} \quad (1.12)$$

Haciendo honor a su nombre HJ-Biplot o RCMP-Biplot (Row Column Metric Preserving) se adopta la siguiente notación para las matrices de marcadores fila y columna en la factorización Biplot elegida:

$$\mathbf{J} = \mathbf{U} \mathbf{D} \quad \mathbf{H} = \mathbf{V} \mathbf{D}$$

Las propiedades de estos marcadores son objeto del apartado 3 de este mismo capítulo.

En este tipo de Biplot se obtiene la misma calidad de representación tanto para las filas como para las columnas (GALINDO y CUADRAS, 1986), luego en este sentido la representación es simétrica permitiendo interpretar las posiciones de las filas, de las columnas y las relaciones fila-columna a través de los factores, como en el caso del Análisis Factorial de Correspondencias (BENZECRI, 1973; GREENACRE, 1984) aplicado a cualquier tipo de datos (no solo frecuencias).

También se pueden interpretar los ejes factoriales y su relación con las variables y con los individuos, a través de las contribuciones relativas del factor al elemento y del elemento al factor (GALINDO y CUADRAS, 1986).

El RCMP-Biplot es muy utilizado en la interpretación simultánea de relaciones entre filas (individuos) y columnas (variables), no siendo su objetivo principal la aproximación de los elementos de la matriz de datos, como es el caso de los Biplots Clásicos.

3.3.- BIPLOTS GENERALIZADOS

Existen además los denominados Biplots Generalizados (VICENTE-VILLARDON, 1992), definidos al introducir una métrica cualquiera (distintas de las asociada a la matriz identidad o métrica euclídea usual) en el espacio de las filas y las columnas.

Al igual que en el caso de cualquiera de los Biplots antes descritos, se aproxima la matriz de datos \mathbf{Y} pero a través de la Descomposición en Valores Singulares Generalizada, de manera tal que se pueda introducir cualquier métrica en el espacio de las filas y las columnas, buscando una matriz $\mathbf{Y}_{(q)}$ de rango q que sea lo mas aproximada a \mathbf{Y} en el sentido de los mínimos cuadrados generalizados.

Se extienden las restricciones de identificación, para que las matrices \mathbf{U} y \mathbf{V} sean ortonormales con métricas definidas positivas $\mathbf{\Omega}$ y $\mathbf{\Phi}$, o sea:

$$\mathbf{Y} = \mathbf{U} \mathbf{D} \mathbf{V}' \quad \text{siendo} \quad \mathbf{U}' \mathbf{\Omega} \mathbf{U} = \mathbf{I} \quad \mathbf{V}' \mathbf{\Phi} \mathbf{V} = \mathbf{I}$$

La aproximación tradicional de una matriz a bajo rango, a través de la Descomposición en Valores Singulares Ordinaria, no es posible si se introducen métricas en el espacio de las filas y de las columnas asociadas a matrices definidas positivas, sin embargo en la “Descomposición en Valores Singulares Generalizada”, es posible utilizar la Descomposición en Valores Singulares Ordinaria de la siguiente matriz \mathbf{X} (véase GREENACRE, 1984):

$$\mathbf{X} = [\mathbf{\Omega}^{-1/2} \mathbf{Y} \mathbf{\Phi}^{-1/2}] = \mathbf{P} \mathbf{D} \mathbf{Q}'$$

Tomando $\mathbf{U} = \mathbf{\Omega}^{-1/2} \mathbf{P}$ y $\mathbf{V} = \mathbf{\Phi}^{-1/2} \mathbf{Q}$

Las columnas de las matrices \mathbf{U} y \mathbf{V} son los “vectores singulares generalizados” por la izquierda y la derecha respectivamente, y a los valores de la matriz diagonal \mathbf{D} se les denominan “valores singulares generalizados”.

Esta aproximación de la matriz \mathbf{Y} a través de la Descomposición en Valores Singulares Generalizados, se puede generalizar a todas las representaciones Biplot ya descritas: CMP-Biplot, RMP-Biplot, SQRT-Biplot y HJ-Biplot.

La elección de métricas en el espacio de las filas y/o columnas de la matriz de datos, unida a transformaciones adecuadas de los mismos, permiten definir las siguientes técnicas del análisis de datos como soluciones particulares del HJ-Biplot Generalizado: el Análisis de Componentes Principales, en Análisis de Correspondencias, el Análisis de Correlación Canónica y el Análisis Canónico de Poblaciones (ver VICENTE-VILLARDON, 1992).

La formalización algebraica de los Biplots Generalizados, enfocados como modelos bilineales, forma parte de la contribución de VASQUEZ (1995) a los métodos Biplot.

Ese enfoque resalta la importancia que diferentes tipos de centrado y/o estandarización de los datos tiene en la distancia euclídea entre filas y entre columnas, ya que la distancia de un punto fila al origen está influenciada por su nivel promedio y también por la dispersión que presentan los datos en esa fila.

Por su parte la distancia al cuadrado entre variables se encuentra influenciada por la varianza de la diferencia entre las dos variables, y las diferencias entre los promedios de las variables.

VASQUEZ, propone indicadores contruídos de la desagregación de la distancia en sumandos, obteniendo luego a través de aproximaciones GH-Biplot las desemejanzas entre variables, mientras que a través de aproximaciones JK-Biplot obtiene las desemejanzas entre individuos.

GOWER y HARDING (1988) y GOWER (1992), definen los Biplots no Lineales, denominados también Biplots Generalizados (desde otro punto de vista diferente a VICENTE-VILLARDON, 1992). En estos Biplot se obtienen trayectorias no lineales para representar a las variables.

Esas trayectorias para distintos valores de cada una de las variables (manteniendo constantes las restantes), se proyectan sobre representaciones obtenidas a través de Coordenadas Principales.

Para esos autores la diferencia entre Biplots no Lineales y Biplots Generalizados, está en el hecho de que cuando se proyecta la trayectoria para una de las variables, el resto de las variables se considera constante. En los Biplots no Lineales les dan el valor cero al resto de las variables, mientras que en los Biplots Generalizados les asignan el valor promedio de distintas combinaciones de las demás variables.

3.4.- BIPLOTS DE PREDICCIÓN E INTERPOLACION

GOWER & HAND (1996) definen los Biplots de Interpolación y los Biplots de Predicción. Se trata de versiones diferentes a los Biplots Clásicos de GABRIEL (1971, 1981a) basadas en puntos de vista diferentes.

Mientras que GABRIEL considera fundamentalmente los Biplots para analizar relaciones entre variables, GOWER & HAND utilizan los Biplot fundamentalmente para analizar similitud entre individuos, introduciendo información sobre las variables a través de distintos procedimientos.

En el análisis de la similitud entre individuos, GOWER & HAND ordenan los individuos aplicando distintas versiones del análisis de proximidades, como por ejemplo: Multidimensional Scaling (MDS), Análisis de Componentes Principales (análisis de proximidades métrico aplicado a la matriz de distancias euclídeas entre los individuos) y Análisis de Correlaciones Múltiples.

Los ***Biplots de Predicción*** permiten la predicción de la matriz de datos ($\mathbf{Y} = \mathbf{AB}'$) a través de la Descomposición en Valores Singulares (en forma similar a los Biplot Clásicos de GABRIEL, (1981).

Una forma alternativa, se produce cuando se desea predecir los valores de \mathbf{Y} dados los de \mathbf{A} a través de regresión lineal, luego $\mathbf{Y} = \mathbf{AB}'$. En ese caso la aproximación se hace a través de regresiones alternadas (véase para mayor detalle el apartado 3 en el capítulo II), o sea:

Considerando fijas las coordenadas filas de \mathbf{A} entonces se estiman las coordenadas para las columnas a través de regresión lineal para obtener:

$$\mathbf{B}' = (\mathbf{A}'\mathbf{A})^{-1} \mathbf{A}'\mathbf{Y} \quad (1.13)$$

Las coordenadas dadas a priori de la matriz \mathbf{A} , se obtienen por ejemplo a través de algún método del Análisis de Proximidades.

Recíprocamente, si se fijan las coordenadas columna de \mathbf{B} entonces las coordenadas en \mathbf{A} podrán estimarse por regresión lineal como:

$$\mathbf{A}' = (\mathbf{B}'\mathbf{B})^{-1} \mathbf{B}'\mathbf{Y}' \quad (1.14)$$

Alternando ambas regresiones finalmente el producto de $(\mathbf{A}\mathbf{B}')$ converge a la Descomposición en Valores Singulares de \mathbf{Y} (ver BLAZQUEZ, 1998).

Utilizando esas ecuaciones en forma separada, se construyen los Biplots de Predicción basados en regresiones. Colocando las p filas de cada columna de \mathbf{B} , sobre el espacio de \mathbf{A} , nos da los puntos unidad para interpolación sobre los ejes Biplot.

Por otra parte, los *Biplots de Interpolación* permiten superponer un nuevo individuo u objeto (fila) de \mathbf{Y} (\mathbf{y}_{n+1}), si se conocen sus distancias o similitudes con los n individuos originales, proyectándolas en un subespacio q -dimensional de baja dimensión.

Resumiendo, en los Biplots de Interpolación es posible superponer un nuevo individuo sobre el gráfico Biplot, mientras que en los Biplot de Predicción se aproximan valores de las variables originales dado un punto de la representación. En general, los ejes Biplot para Interpolación y Predicción no coinciden y no tienen porque tener las mismas escalas.

4.- PROPIEDADES DE LOS MARCADORES

De acuerdo al tipo de Biplot, y dependiendo de la métrica introducida en el espacio de las filas o de las columnas, los marcadores tienen distintas propiedades las cuales son muy útiles a la hora de la interpretación de una representación Biplot.

4.1.- MARCADORES COLUMNA

4.1.1.-MARCADORES TIPO $B = VD$ (Métrica $A'A = I$)

1) El producto escalar de las columnas de Y coincide con el producto escalar de los marcadores columna, o sea:

$$Y'Y = BB' \quad \text{de donde} \quad y_j' y_k \approx b_j' b_k$$

Sea $Y = AB' = U(VD)'$ entonces $Y'Y = BA'AB' = BB'$

También se puede demostrar considerando el hecho de que cada columna de Y se puede aproximar a través de:

$$y_j = A b_j \quad (\forall j= 1,2,\dots,p)$$

Sustituyendo tendremos que:

$$\mathbf{y}_j' \mathbf{y}_k = (\mathbf{A} \mathbf{b}_j)' (\mathbf{A} \mathbf{b}_k) = \mathbf{b}_j' \mathbf{A}' \mathbf{A} \mathbf{b}_k = \mathbf{b}_j' \mathbf{b}_k$$

2) En el caso de los Biplots de Componentes Principales (cuando los datos en \mathbf{Y} han sido centrados por columnas), la longitud al cuadrado de los vectores que representan a los marcadores columna, aproximan la covarianza entre las variables correspondientes:

$$\mathbf{b}_j' \mathbf{b}_k = \text{Cov}(\mathbf{y}_j, \mathbf{y}_k)$$

En este caso, la matriz de varianzas y covarianzas \mathbf{S} es igual a:

$$\mathbf{S} = [1/(n-1)] \mathbf{Y}' \mathbf{Y}$$

De la primera propiedad entonces: $\mathbf{S} = [1/(n-1)] \mathbf{B}' \mathbf{B}$

y ya que: $\mathbf{B} = (1/\sqrt{n-1})(\mathbf{V} \mathbf{D})$

entonces $\mathbf{b}_j' \mathbf{b}_k = \mathbf{S}_{jk}$

Como una consecuencia de esta segunda propiedad, entonces:

2a) La longitud de los marcadores columna, aproximan la desviación estándar de las variables:

$$\|\mathbf{y}_j\| = \|\mathbf{b}_j\| = \sqrt{\text{Var}(\mathbf{y}_j)}$$

2b) El coseno del ángulo entre dos marcadores columna aproxima la correlación entre variables correspondientes:

$$\text{Cos}(\mathbf{b}_j, \mathbf{b}_k) \approx (\mathbf{b}_j' \mathbf{b}_k) / \|\mathbf{b}_j\| \|\mathbf{b}_k\| = \text{Cos}(\mathbf{y}_j, \mathbf{y}_k) = (\mathbf{y}_j' \mathbf{y}_k) / \|\mathbf{y}_j\| \|\mathbf{y}_k\|$$

De donde:

$$\text{Cos}(\mathbf{b}_j, \mathbf{b}_k) = \text{Corr}(\mathbf{y}_j, \mathbf{y}_k)$$

Luego: $\text{Cos}(\mathbf{b}_j, \mathbf{b}_k) = \mathbf{S}_{jk} / \mathbf{S}_j \mathbf{S}_k = \text{Corr}(\mathbf{y}_j, \mathbf{y}_k) = r_{jk}$

2c) La distancia euclídea entre dos vectores columna coincide con la distancia entre los marcadores columna respectivos:

$$\begin{aligned} d^2(\mathbf{y}_j, \mathbf{y}_k) &= \sum_{i=1}^n (y_{ij} - y_{ik})^2 = (\mathbf{y}_j - \mathbf{y}_k)'(\mathbf{y}_j - \mathbf{y}_k) = \\ &= \|\mathbf{y}_j\|^2 + \|\mathbf{y}_k\|^2 - 2(\mathbf{y}_j' \mathbf{y}_k) = \|\mathbf{b}_j\|^2 + \|\mathbf{b}_k\|^2 - 2(\mathbf{b}_j' \mathbf{b}_k) = d^2(\mathbf{b}_j, \mathbf{b}_k) \end{aligned}$$

3) Las coordenadas de la matriz de marcadores columna \mathbf{B} ($\mathbf{B} = \mathbf{VD}$) en el Biplot de Componentes Principales (PEARSON, 1901; HOTTELING, 1936), son equivalentes a la importancia de las variables a lo largo de los k-ésimos ejes principales, o sea:

$$\varphi = \mathbf{Y}'\mathbf{U} = \mathbf{VD} = \mathbf{B}$$

De allí que, las columnas de la matriz \mathbf{B} definen un nuevo individuo, combinación lineal de los n individuos originales, o sea, son equivalentes a la importancia de las variables a lo largo de los ejes principales.

4) La calidad de representación de las columnas de la matriz \mathbf{Y} es óptima, siendo mayor incluso que la calidad de representación global de \mathbf{Y} .

De la Descomposición en Valores Singulares de la matriz \mathbf{Y} la calidad de representación global se calcula como el cociente entre la suma de cuadrados de $\mathbf{Y}_{(q)}$ y la suma de cuadrados de los elementos de \mathbf{Y} .

Como la suma de cuadrados de \mathbf{Y} es igual a la traza de $(\mathbf{Y}\mathbf{Y}')$, y dado que:

$$tr(\mathbf{Y}\mathbf{Y}') = \sum_{k=1}^r \lambda_k = \sum_{k=1}^r \alpha_k^2$$

Entonces la calidad de representación global será igual a:

$$CRG = \left(\frac{\sum_{k=1}^q \alpha_k^2}{\sum_{k=1}^r \alpha_k^2} \right) \times 100$$

Donde α_k , son los q mayores valores singulares asociados a los λ_k valores propios de la matriz $\mathbf{Y}'\mathbf{Y}$ ($\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_q$).

En el caso de la matriz de varianzas y covarianzas, donde $\mathbf{S} = [1/(n-1)] \mathbf{Y}'\mathbf{Y}$. Entonces sustituyendo $\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}'$, se demuestra que: $\mathbf{S} = \mathbf{V}\mathbf{D}^2\mathbf{V}'$

De donde la calidad de representación de las columnas será:

$$CRC = \left(\frac{\sum_{k=1}^q \alpha_k^4}{\sum_{k=1}^r \alpha_k^4} \right) \times 100$$

4.1.2.-MARCADORES TIPO $\mathbf{B} = \mathbf{V}$ (Métrica $\mathbf{B}'\mathbf{B} = \mathbf{I}$)

1) Las coordenadas para las columnas ($\mathbf{B} = \mathbf{V}$) coinciden con las proyecciones de los ejes originales (ejes canónicos) en el espacio de las componentes principales.

El vector canónico define la dirección del j-ésimo eje canónico en el espacio original de las filas.

En efecto
$$\varphi = \mathbf{I} \mathbf{V} = \mathbf{B}$$

Luego
$$\mathbf{e}'_j \mathbf{V} = (v_{j1}, v_{j2}, \dots, v_{jq}) = \mathbf{b}'_j$$

Teniendo en cuenta que las coordenadas de los vectores que forman la base canónica forman la matriz identidad \mathbf{I}_p y la proyección de las mismas sobre el espacio de las componentes principales es:

$$\mathbf{I}_p \mathbf{V} = \mathbf{V} = \mathbf{B}$$

Entonces, las coordenadas para las columnas marcan la unidad para las escalas de predicción.

Como cada una de las variables originales están asociadas con un eje en el espacio p -dimensional, las proyecciones de los vectores directores correspondientes se han denominado “Ejes Biplot”, cada uno de los cuales identifica a la variable asociada (GOWER & HARDING, 1988).

La proyección de cada uno de los marcadores fila sobre estos ejes, es una aproximación de los valores que toman los individuos en las variables correspondientes.

2) El producto interno entre marcadores columna aproxima el producto interno entre columnas de \mathbf{Y} con la métrica $(\mathbf{Y}\mathbf{Y}')$, o sea:

$$\mathbf{y}'_j (\mathbf{Y}\mathbf{Y}') \mathbf{y}_k = \mathbf{b}'_j \mathbf{b}_k$$

Sabemos que, en una aproximación Biplot la columna j -ésima de la matriz \mathbf{Y} se puede aproximar mediante:

$$\mathbf{y}_j = \mathbf{A} \mathbf{b}_j \quad (\forall j= 1,2,\dots,p)$$

Sustituyendo tendremos que:

$$\mathbf{y}'_j \mathbf{y}_k = (\mathbf{A} \mathbf{b}_j)' (\mathbf{A} \mathbf{b}_k) = \mathbf{b}'_j \mathbf{A}' \mathbf{A} \mathbf{b}_k$$

Multiplicando por $(\mathbf{Y}\mathbf{Y}')^{-1}$ en ambos miembros y sustituyendo $\mathbf{Y} = \mathbf{A}\mathbf{B}'$, tendremos que:

$$\mathbf{y}_j'(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{y}_k = \mathbf{b}'_j \mathbf{A}' (\mathbf{Y}\mathbf{Y}')^{-1} \mathbf{A} \mathbf{b}_k = \mathbf{b}'_j \mathbf{A}' (\mathbf{A}\mathbf{A}')^{-1} \mathbf{A} \mathbf{b}_k = \mathbf{b}'_j \mathbf{b}_k$$

3) La similitud entre columnas se aproxima utilizando la inversa de la matriz de dispersión entre individuos, no siendo posible interpretar los ángulos en términos de correlación, ya que:

$$(\mathbf{y}_j - \mathbf{y}_k)' (\mathbf{Y}\mathbf{Y}')^{-1} (\mathbf{y}_j - \mathbf{y}_k) \approx (\mathbf{b}_j - \mathbf{b}_k)' (\mathbf{b}_j - \mathbf{b}_k)$$

La demostración se hace sustituyendo la aproximación de cada columna de la matriz por su expresión: $\mathbf{y}_j = \mathbf{A} \mathbf{b}_j$

4) La calidad de representación para las columnas es muy baja.

Para las columnas de \mathbf{Y} la suma de cuadrados de los elementos de la matriz de proyección $[\mathbf{Y}(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{Y}']$ sobre el espacio de las filas de \mathbf{Y} (GRAYBILL, 1983), es igual al rango r de \mathbf{Y} .

Al aproximar en dimensión q tenemos que:

$$\mathbf{Y}(\mathbf{Y}\mathbf{Y}')^{-1} \mathbf{Y}' \cong \mathbf{B} \mathbf{B}'$$

La suma de cuadrados de esa aproximación es igual a q , luego la Calidad de Representación para las columnas, con la métrica $(\mathbf{Y}\mathbf{Y}')^{-1}$ es:

$$\text{CRC} = (q / r) \times 100$$

Por otra parte, por la primera propiedad las coordenadas de las columnas aproximan las correlaciones entre variables originales y ejes, de allí la mala calidad en la representación de las columnas.

4.2.- MARCADORES FILA

4.2.1.- MARCADORES TIPO $A = UD$ (Métrica $B'B = I$)

1) El producto escalar, con la métrica identidad, de las filas de la matriz \mathbf{Y} , coincide con el producto escalar entre marcadores filas:

$$\mathbf{Y}\mathbf{Y}' = \mathbf{A}\mathbf{A}'$$

Sustituyendo la expresión de la matriz $\mathbf{Y} = \mathbf{A}\mathbf{B}'$ y considerando la métrica impuesta ($\mathbf{B}'\mathbf{B} = \mathbf{I}$) tenemos que:

$$\mathbf{Y}\mathbf{Y}' = (\mathbf{A}\mathbf{B}')(\mathbf{B}\mathbf{A}') = \mathbf{A}\mathbf{I}\mathbf{A}' = \mathbf{A}\mathbf{A}'$$

Lo que también se puede demostrar ya que sabemos que las filas de \mathbf{Y} se pueden aproximar a través de: $\mathbf{y}_i = \mathbf{B}\mathbf{a}_i$ ($\forall i = 1, 2, \dots, n$)

Por lo tanto, sustituyendo tendremos que:

$$\mathbf{y}_j' \mathbf{y}_k = (\mathbf{B} \mathbf{a}_j)' (\mathbf{B} \mathbf{a}_k) = \mathbf{a}_j' \mathbf{B}' \mathbf{B} \mathbf{a}_k$$

De donde:

$$\mathbf{y}_i' \mathbf{y}_s = \mathbf{a}_i' \mathbf{a}_s$$

A partir de allí se puede demostrar fácilmente que:

$$\|\mathbf{y}_i\| = \|\mathbf{a}_i\|$$

$$\text{Cos}(\mathbf{y}_i, \mathbf{y}_s) = \text{Cos}(\mathbf{a}_i, \mathbf{a}_s)$$

$$d(\mathbf{y}_i, \mathbf{y}_s) = d(\mathbf{a}_i, \mathbf{a}_s)$$

2) La distancia euclídea entre filas de la matriz \mathbf{Y} coincide con la distancia entre marcadores fila:

$$(\mathbf{y}_i - \mathbf{y}_s)' (\mathbf{y}_i - \mathbf{y}_s) = (\mathbf{a}_i - \mathbf{a}_s)' (\mathbf{a}_i - \mathbf{a}_s)$$

Sabemos que las filas de la matriz \mathbf{Y} se pueden aproximar por: $\mathbf{y}_i = \mathbf{B} \mathbf{a}_i$, para $i = 1, 2, \dots, n$, luego sustituyendo tenemos que:

$$(\mathbf{y}_i - \mathbf{y}_s)' (\mathbf{y}_i - \mathbf{y}_s) = (\mathbf{B} \mathbf{a}_i - \mathbf{B} \mathbf{a}_s)' (\mathbf{B} \mathbf{a}_i - \mathbf{B} \mathbf{a}_s) = (\mathbf{a}_i - \mathbf{a}_s)' \mathbf{B}' \mathbf{B} (\mathbf{a}_i - \mathbf{a}_s)$$

Considerando la métrica impuesta ($\mathbf{B}'\mathbf{B} = \mathbf{I}$) entonces finalmente se demuestra que:

$$(\mathbf{y}_i - \mathbf{y}_s)' (\mathbf{y}_i - \mathbf{y}_s) = (\mathbf{a}_i - \mathbf{a}_s)' (\mathbf{a}_i - \mathbf{a}_s)$$

3) Los marcadores fila coinciden con las coordenadas de las filas en el espacio de las componentes principales:

$$\boldsymbol{\psi} = \mathbf{Y}\mathbf{V} = (\mathbf{U}\mathbf{D}\mathbf{V}')\mathbf{V} = \mathbf{U}\mathbf{D} = \mathbf{A}$$

Por lo tanto, los marcadores filas \mathbf{a}_i en la representación Biplot aproximan las similitudes entre individuos u objetos, coincidiendo con las componentes de proyección de \mathbf{y}_i sobre los q primeros ejes del subespacio q-dimensional de mejor ajuste a las filas de la matriz \mathbf{Y} .

Estas componentes, constituyen las coordenadas principales en la representación de las filas sobre el espacio reducido q-dimensional.

Así pues, la k-ésima columna de la matriz de marcadores \mathbf{A} contiene las coordenadas que permiten representar los individuos sobre el k-ésimo eje del espacio reducido ($\mathbf{A}^\alpha = \mathbf{Y}\mathbf{V}^\alpha$), por lo que el vector de coordenadas fila puede interpretarse como una variable artificial.

4) La calidad de representación para las filas es máxima.

La demostración es igual que para la propiedad 4 en el apartado 4.1.1, tan solo que se cambia el rol de las columnas por las filas.

4.2.2.- MARCADORES TIPO $A = U$ (Métrica $A'A = I$)

1) En el Biplot de Componentes Principales la distancia de Mahalanobis entre dos filas de \mathbf{Y} se aproxima mediante la distancia euclídea entre dos marcadores fila:

$$(\mathbf{y}_i - \mathbf{y}_s)' \mathbf{S}^{-1} (\mathbf{y}_i - \mathbf{y}_s) \approx (\mathbf{a}_i - \mathbf{a}_s)' (\mathbf{a}_i - \mathbf{a}_s)$$

Por definición la distancia de Mahalanobis entre dos filas es:

$$(\mathbf{y}_i - \mathbf{y}_s)' \mathbf{S}^{-1} (\mathbf{y}_i - \mathbf{y}_s)$$

Como cada fila de la matriz se puede escribir como $\mathbf{y}_i = \mathbf{B} \mathbf{a}_i$, sustituyendo tenemos:

$$\begin{aligned} (\mathbf{B}\mathbf{a}_i - \mathbf{B}\mathbf{a}_s)' \mathbf{S}^{-1} (\mathbf{B}\mathbf{a}_i - \mathbf{B}\mathbf{a}_s) &= (\mathbf{a}_i - \mathbf{a}_s)' \mathbf{B}' \mathbf{S}^{-1} \mathbf{B} (\mathbf{a}_i - \mathbf{a}_s) = \\ &= (\mathbf{a}_i - \mathbf{a}_s)' \mathbf{D}\mathbf{V}' \mathbf{S}^{-1} \mathbf{V}\mathbf{D} (\mathbf{a}_i - \mathbf{a}_s) = (\mathbf{a}_i - \mathbf{a}_s)' \mathbf{D}\mathbf{V}' (\mathbf{V}\mathbf{D}^2\mathbf{V})^{-1} \mathbf{V}\mathbf{D} (\mathbf{a}_i - \mathbf{a}_s) \end{aligned}$$

De donde finalmente se obtiene que:

$$(\mathbf{B}\mathbf{a}_i - \mathbf{B}\mathbf{a}_s)' \mathbf{S}^{-1} (\mathbf{B}\mathbf{a}_i - \mathbf{B}\mathbf{a}_s) = (\mathbf{a}_i - \mathbf{a}_s)' (\mathbf{a}_i - \mathbf{a}_s)$$

2) En el caso del Biplot de Componentes Principales ($\mathbf{A} = \sqrt{n-1} \mathbf{U}$), el i -ésimo marcador fila coincide con las coordenadas del i -ésimo individuo sobre las r primeras componentes principales estandarizadas, por lo que \mathbf{A} se puede

interpretar como el score de las componentes principales estandarizadas:

$$\mathbf{a}_i = \psi_i = \sqrt{(n-1)} \mathbf{u}_i = \mathbf{y}_i' \mathbf{V} [\sqrt{(n-1)} \mathbf{D}^{-1}]$$

3) El producto interno entre filas de \mathbf{Y} , introduciendo en el espacio de las columnas la métrica $(\mathbf{Y}'\mathbf{Y})^{-1}$, es aproximadamente igual al producto interno entre marcadores filas:

$$\mathbf{Y} (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}' = \mathbf{A} \mathbf{A}'$$

Sea $\mathbf{y}_i = \mathbf{B} \mathbf{a}_i$ y como se demuestra que $[\mathbf{B}' (\mathbf{B}\mathbf{B}')^{-1} \mathbf{B}] = \mathbf{I}$, entonces se tiene sustituyendo que:

$$\mathbf{y}_i' (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{y}_s = (\mathbf{B} \mathbf{a}_i)' (\mathbf{B}\mathbf{B}')^{-1} (\mathbf{B} \mathbf{a}_s) = \mathbf{a}_i' \mathbf{B}' (\mathbf{B}\mathbf{B}')^{-1} \mathbf{B} \mathbf{a}_s = \mathbf{a}_i' \mathbf{a}_s$$

4) La calidad de representación de las filas de la matriz \mathbf{Y} es muy baja.

La demostración (cambiando el rol de columnas por filas) se hace en forma similar a la de la propiedad 4 en el apartado 4.1.2.

En la tabla 1.1, se presenta un resumen de las propiedades de los marcadores en los Biplot.

Tipo de Marcadores y Propiedades	Métrica
<p>B=VD</p> <p>Propiedades: El producto escalar de las columnas de Y, coincide con el producto escalar de los marcadores columna, o sea: $\mathbf{y}_j' \mathbf{y}_j \approx \mathbf{b}_j' \mathbf{b}_j$. La longitud al cuadrado de los vectores que representan a los marcadores columna, aproximan la covarianza entre las variables correspondientes, de donde: $\mathbf{b}_j' \mathbf{b}_j = s_{jj} = s_j^2$ La longitud de los marcadores columna, aproxima la desviación estándar de las variables El coseno del ángulo entre dos marcadores columna, aproxima la correlación entre variables correspondientes: $\text{Cos}(\mathbf{b}_j, \mathbf{b}_k) \approx \text{Corr}(\mathbf{y}_j, \mathbf{y}_k)$ Las coordenadas de la matriz de marcadores columna B en el Biplot de Componentes Principales, son equivalentes a los pesos de las variables a lo largo de los ejes principales. Es óptima la calidad de representación de las columnas</p>	<p>A'A=I</p>
<p>B=V</p> <p>Propiedades: Las coordenadas para las columnas coinciden con las proyecciones de los ejes originales en el espacio de las componentes principales: $\mathbf{I}_p \mathbf{V} = \mathbf{V} = \mathbf{B}$ La similitud entre columnas, se aproxima utilizando la inversa de la matriz de dispersión entre individuos Es baja la calidad de representación para las columnas</p>	<p>B'B=I</p>
<p>A=UD</p> <p>Propiedades: El producto escalar, con la métrica identidad, de las filas de Y coincide con el producto escalar entre marcadores filas: $\mathbf{Y}\mathbf{Y}' = \mathbf{A}\mathbf{A}'$ La distancia euclídea entre filas de Y, coincide con la distancia entre marcadores fila: $(\mathbf{y}_i - \mathbf{y}_s)' (\mathbf{y}_i - \mathbf{y}_s) = (\mathbf{a}_i - \mathbf{a}_s)' (\mathbf{a}_i - \mathbf{a}_s)$ Los marcadores fila, coinciden con las coordenadas de las filas en el espacio de las componentes principales: $\boldsymbol{\psi} = \mathbf{Y}\mathbf{V} = \mathbf{A}$ Es óptima la calidad de representación para las filas</p>	<p>B'B=I</p>
<p>A=U</p> <p>Propiedades: En el Biplot de Componentes Principales, la distancia de Mahalanobis entre dos filas de Y, se aproxima mediante la distancia euclídea entre dos marcadores fila, o sea: $(\mathbf{y}_i - \mathbf{y}_s)' \mathbf{S}^{-1} (\mathbf{y}_i - \mathbf{y}_s) \approx (\mathbf{a}_i - \mathbf{a}_s)' (\mathbf{a}_i - \mathbf{a}_s)$ El i-ésimo marcador fila, coincide con las coordenadas del i-ésimo individuo sobre las r primeras componentes principales estandarizadas, o sea: $\mathbf{a}_i = \boldsymbol{\psi}_i = \sqrt{(n-1)} \mathbf{u}_i = \mathbf{y}_i' \mathbf{V} [\sqrt{(n-1)} \mathbf{D}^{-1}]$ Es baja la calidad de representación para las filas</p>	<p>A'A=I</p>

Tabla 1.1.- Propiedades de los Marcadores en los distintos tipos de Biplots según métrica impuesta

5.- INTERPRETACION DE LAS REPRESENTACIONES BILOT

5.1.- INTERPRETACION GEOMETRICA

Una de las principales características de las representaciones Biplot es el hecho de que su interpretación se basa en conceptos vectoriales geométricos sencillos, específicamente en la proyección sobre una representación plana, de las características principales de las filas y columnas de una cierta matriz de datos \mathbf{Y} de orden $(n \times p)$.

En la representación, las filas de \mathbf{A} (matriz de marcadores filas) y las columnas de \mathbf{B} (matriz de marcadores columna) pueden verse como coordenadas de puntos filas y puntos columnas respectivamente en un espacio euclídeo referido a los mismos ejes ortogonales.

Los marcadores \mathbf{b}_j para las columnas, se representan como vectores de tal forma de facilitar la proyección de los marcadores filas \mathbf{a}_i (puntos en la representación) sobre el marcador columna \mathbf{b}_j .

La dirección de los vectores columna \mathbf{b}_j representa la dirección en la que aumentan los valores de la variable correspondiente, y las proyecciones de todos los puntos fila \mathbf{a}_i sobre un vector columna en particular reproducen aproximadamente los elementos de la columna j -ésima en la matriz original, permitiendo al mismo tiempo una ordenación aproximada de los individuos (filas) respecto a esa variable en particular.

Dado que los Biplots permiten aproximar los datos originales ($y_{ij} = \mathbf{a}_i' \mathbf{b}_j = a_{i1} b_{j1} + a_{i2} b_{j2}$) a través de la representación conjunta de los vectores \mathbf{a}_i y \mathbf{b}_j en un espacio bi-dimensional, entonces es posible reconstruir:

- 1- La similitud entre individuos como función inversa a la distancia entre marcadores fila sobre el Biplot, por lo que a mayor distancia existirá una menor similitud.
- 2- La variabilidad y covariabilidad de las variables a través de las longitudes y ángulos de los vectores que representan a las variables.
- 3- Las relaciones entre individuos y variables en términos de producto escalar, o sea a través de las proyecciones de los puntos individuos sobre los vectores que representan a las variables.
- 4- La correlación entre variables a través del ángulo entre los vectores correspondientes.
- 5- La ordenación de los individuos respecto a una variable en particular, a través del orden encontrado en las proyecciones de los marcadores fila sobre el eje determinado por el marcador columna correspondiente a esa variable, tal como se muestra en la figura 1.2.

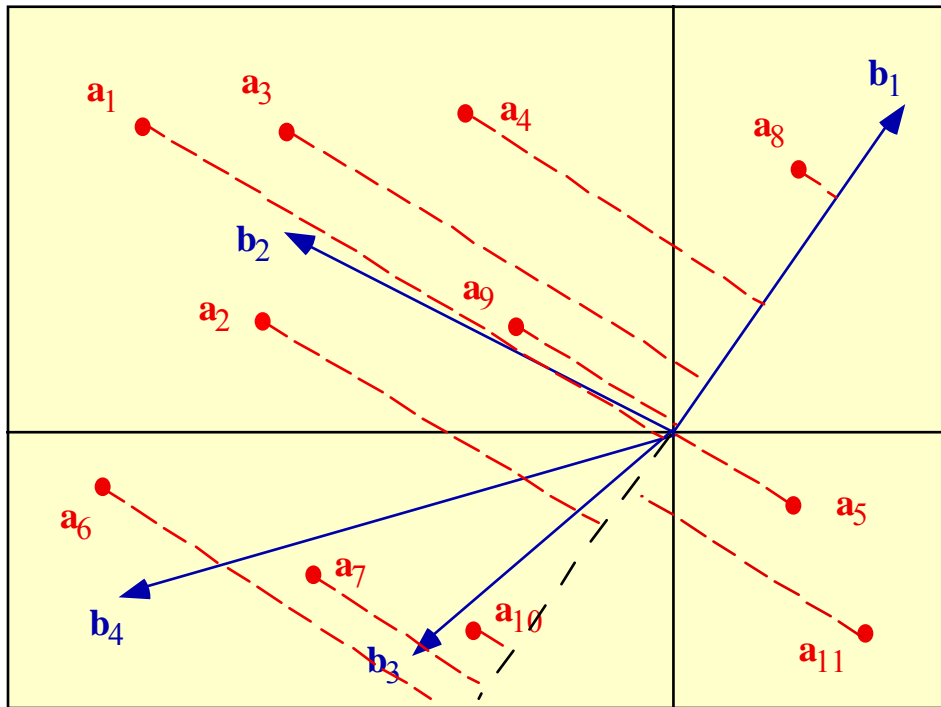


Figura 1.2.- Ordenación de los individuos sobre el Biplot respecto a la variable y_1

- 6- La ordenación de las variables para cada individuo proyectando sobre cada marcador fila a_i el orden de las distintas variables, o sea en forma inversa a la figura 1.2.
- 7- El orden de las medias de las columnas al proyectar los marcadores columna sobre el vector que une el marcador fila promedio con el origen, ya que:

$$y_{i.} \cong a'_i \mathbf{b}_{.} = \| \text{Proy}(\mathbf{a}_i / \mathbf{b}_{.}) \| (\text{signo}) \| \mathbf{b}_{.} \| = \| \text{Proy}(\mathbf{b}_{.} / \mathbf{a}_i) \| (\text{signo}) \| \mathbf{a}_i \|$$

8- El orden de las medias de las filas al proyectar los marcadores fila sobre el vector que une el marcador columna promedio con el origen, ya que:

$$\mathbf{y}_{\cdot j} \cong \mathbf{a}' \cdot \mathbf{b}_j = \|\text{Proy}(\mathbf{b}_j / \mathbf{a}_{\bullet})\| (\text{signo}) \|\mathbf{a}_{\bullet}\| = \|\text{Proy}(\mathbf{a}_{\bullet} / \mathbf{b}_j)\| (\text{signo}) \|\mathbf{b}_j\|$$

9- El orden de la media total al proyectar los marcadores fila (columna) promedio sobre el vector que une el marcador columna (fila) promedio con el origen, ya que:

$$\mathbf{y}_{\bullet\bullet} \cong \mathbf{a}' \cdot \mathbf{b}_{\bullet} = \|\text{Proy}(\mathbf{b}_{\bullet} / \mathbf{a}_{\bullet})\| (\text{signo}) \|\mathbf{a}_{\bullet}\| = \|\text{Proy}(\mathbf{a}_{\bullet} / \mathbf{b}_{\bullet})\| (\text{signo}) \|\mathbf{b}_{\bullet}\|$$

10-En el caso de tablas de dos vías también se pueden aproximar los efectos fila, tal como se muestra en la figura 1.3, ya que:

$$\hat{\alpha}_i \cong \|\text{Proy}(\mathbf{a}_i - \mathbf{a}_{\bullet}) / \mathbf{b}_{\bullet}\| (\text{signo}) \|\mathbf{b}_{\bullet}\|$$

11-En forma inversa a los efectos fila también se pueden aproximar los efectos columna en tablas de dos vías, ya que:

$$\hat{\beta}_j \cong \|\text{Proy}(\mathbf{b}_j - \mathbf{b}_{\bullet}) / \mathbf{a}_{\bullet}\| (\text{signo}) \|\mathbf{a}_{\bullet}\|$$

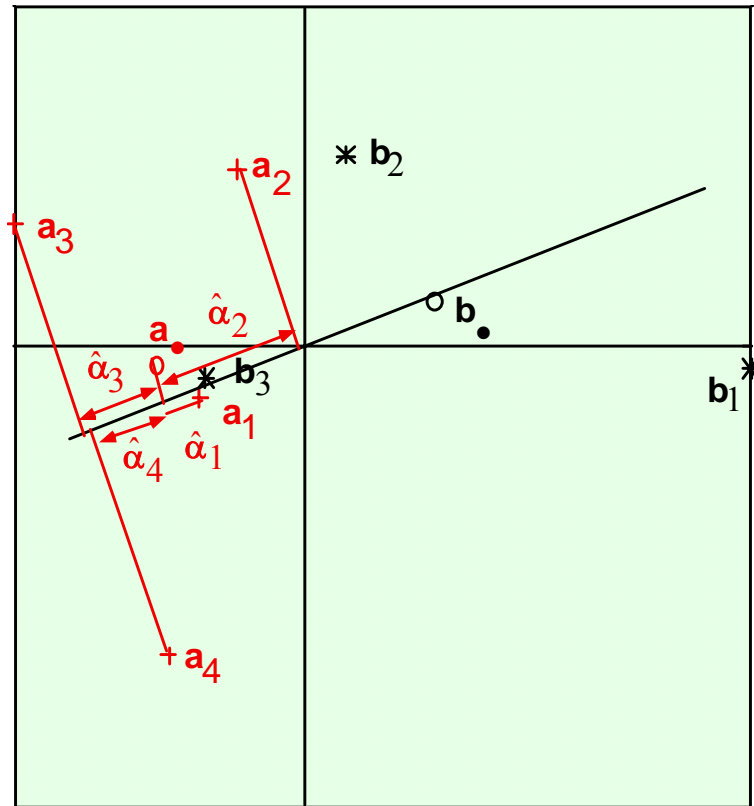


Figura 1.3.- Aproximación de los Efectos Fila sobre el gráfico Biplot

12-La magnitud de los efectos de interacción en tablas de dos vías se pueden aproximar tal como se muestra en la figura 1.4, ya que:

$$(\hat{\alpha}\hat{\beta})_{ij} \cong \left\| \text{Proy}(a_i - a_{\bullet}) / (b_j - b_{\bullet}) \right\| (\text{signo}) \left\| (b_j - b_{\bullet}) \right\|$$

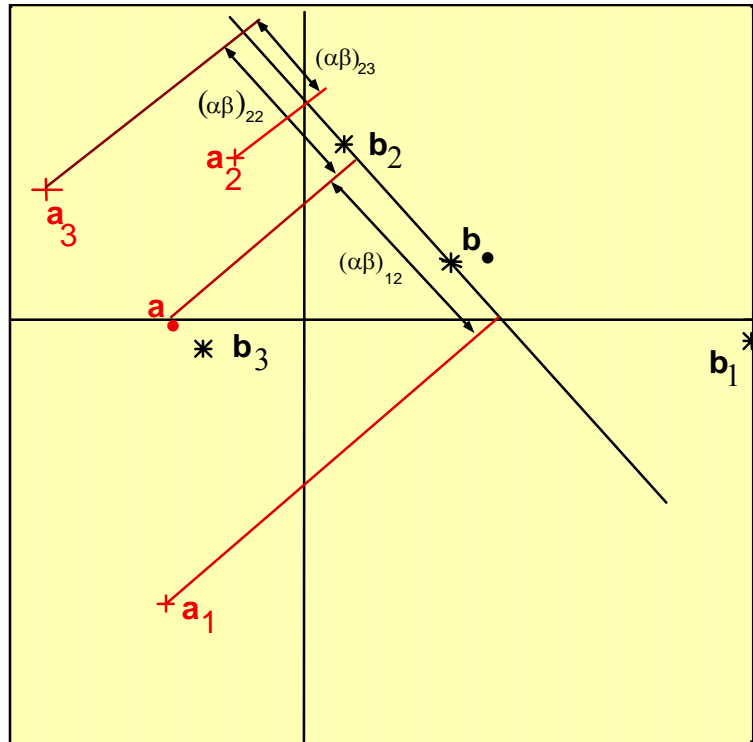


Figura 1.4.- Aproximación de los efectos de interacción sobre el gráfico Biplot

13-Los productos internos entre las filas y entre las columnas de \mathbf{Y} , ya que:

$$y_{ij} \cong \mathbf{a}'_i \mathbf{b}_j = \|\text{Proy}(\mathbf{a}_i / \mathbf{b}_j)\| (\text{signo}) \|\mathbf{b}_j\| = \|\text{Proy}(\mathbf{b}_j / \mathbf{a}_i)\| (\text{signo}) \|\mathbf{a}_i\|$$

De lo antes dicho en cuanto a la interpretación geométrica de las representaciones Biplot, en el caso de la matriz de datos (11x4) presentada en la anterior figura 1.2 y suponiendo los datos centrados por columna, tenemos que:

- Sobre la similitud entre individuos: se observan varios grupos de individuos, el de los individuos 1, 2, 3, 4 y 9, el de los individuos 6, 7 y 10, el de los individuos 5 y 11 y el del individuo 8.
- Sobre la variabilidad de las variables: la variable y_4 es la que tiene mayor variabilidad y la variable y_3 la menor.
- Sobre las relaciones entre individuos y variables: proyectando los puntos fila \mathbf{a}_i sobre el marcador columna \mathbf{b}_1 , se puede aproximar el orden en forma creciente de los individuos respecto a la variable y_1 , así tenemos que el individuo con menor valor para esa variable es el 6, mientras que el de mayor valor es el 8.
- Sobre la correlación entre variables: el ángulo de casi 90° entre \mathbf{b}_1 y \mathbf{b}_2 revela ausencia de correlación entre las variables y_1 e y_2 ; el ángulo mayor de 90° en sentido inverso entre \mathbf{b}_1 y \mathbf{b}_3 , y \mathbf{b}_1 y \mathbf{b}_4 , revela una relación inversa entre las variables y_3 e y_4 respecto a y_1 .
- El producto interno entre marcadores fila y columna permiten reconstruir cada elemento de la matriz de datos, por ejemplo:

$$y_{43} = \mathbf{a}'_4 \mathbf{b}_3 = \| \text{Proy}(\mathbf{a}_4 / \mathbf{b}_3) \| (\text{signo}) \| \mathbf{b}_3 \|$$

o sea que, y_{43} es aproximadamente el módulo de la proyección de \mathbf{a}_4 sobre \mathbf{b}_3 , multiplicado por la longitud de \mathbf{b}_3 .

5.2.- BONDAD DE LA APROXIMACION BILOT

En la interpretación de un Biplot es también 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 es posible obtener su representación exacta sobre el plano.

La calidad global de representación se define en forma similar al coeficiente de determinación en Análisis de Regresión, o sea, como el cociente entre la suma de cuadrados de los elementos de $\mathbf{Y}_{(q)}$ y la suma de cuadrados de los elementos de \mathbf{Y} .

Sin embargo, ese indicador no es suficiente para evaluar los ajustes individuales de cada fila y cada columna, por lo que es importante para efectos de la interpretación, considerar las contribuciones del factor al elemento fila y al elemento columna (calidades de representación), que permiten determinar la parte de la variabilidad de cada elemento explicada por el factor.

La interpretación de esas contribuciones se hace en forma similar al Análisis de Correspondencias, en términos de inercia explicada o varianza absorbida (GALINDO y CUADRAS, 1986; GREENACRE, 1984, 1993).

Suponiendo que todos los elementos tienen peso (masa) unitario, la inercia de una variable (elemento columna) con respecto al centro de gravedad, es igual al cuadrado de la distancia de dicho punto al centro de gravedad.

De allí que, la contribución del factor al elemento columna es:

$$CFE_{col} = \frac{\mathbf{b}_{jk}^2}{\sum_{k=1}^q \mathbf{b}_{jk}^2}$$

Esas contribuciones permitirán conocer cuales variables son las responsables de la ordenación de los individuos sobre las proyecciones en cada uno de los ejes Biplot.

Si los datos están centrados, el centro de gravedad θ coincide con el origen, luego:

$$d^2(\mathbf{b}_j, \theta) = \sum_{k=1}^q b_{jk}^2$$

En ese caso, la CFE_{col} coincide con el cuadrado del coeficiente de correlación, entre la variable j-ésima y el eje k-ésimo.

En el ejemplo de la figura 1.2, tendremos que las variables \mathbf{y}_2 e \mathbf{y}_4 están mejor representadas sobre el eje 1, mientras que las variables \mathbf{y}_1 e \mathbf{y}_3 están mejor representadas sobre el eje 2.

Por otra parte la contribución del factor al elemento fila, se define como el cociente entre el cuadrado de las coordenadas del i-ésimo punto fila sobre el eje α , y la suma de cuadrados de las coordenadas de los n puntos fila sobre los

q ejes ($q < r$), o sea:

$$CFE_{fila} = \frac{\mathbf{a}_{ik}^2}{\sum_{k=1}^q \mathbf{a}_{ik}^2}$$

En este caso, se pierde la interpretación en términos de correlación, siendo mas útil la interpretación en términos de calidad de representación tal como se muestra en la siguiente figura, que representa la parte de la información de un individuo recogida por el eje k-ésimo en un plano, o sea :

$$CLR_{fila/k} = \text{Cos}^2(\alpha_{ik})$$

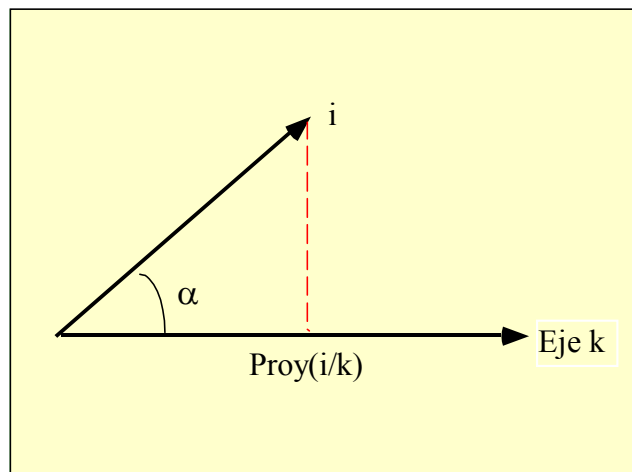


Figura 1.5.- Contribución del Factor al Elemento fila i-ésimo

De allí que, en el mismo ejemplo de la figura 1.2, los individuos 5 y 6 son los mejor representados sobre el eje 1, ya que sus respectivos marcadores \mathbf{a}_i están más cercanos a dicho eje, mientras que en el eje 2 el individuo 8 es el mejor representado.

5.3.- INDICADORES COMPLEMENTARIOS EN LA INTERPRETACION

En la interpretación de un gráfico Biplot se pueden utilizar también, otros indicadores complementarios, que permitan resaltar la importancia que distintos tipos de transformaciones en los datos (centrado, estandarización), tiene sobre las distancias entre individuos y entre variables (VASQUEZ, 1995).

Basándose en descomposiciones de las distancias euclídeas entre individuos o entre variables, o también en descomposiciones relativas a la traza de $(Y'Y)$ donde se encuentra resumida información importante sobre la variabilidad de los datos, tenemos entre esos indicadores los siguientes:

1) Basándose en la desagregación de las distancias del vector fila o columna al origen de coordenadas:

- En el JK-Biplot, se puede aproximar con alta calidad la distancia de cada vector fila al origen de coordenadas.
- Los cosenos directores de un vector fila, se pueden interpretar como una medida de la contribución de la j -ésima variable a la determinación de la posición del individuo sobre el espacio q -dimensional.

2) Basándose en la desagregación en sumandos de las trazas de las matrices $(\mathbf{Y}'\mathbf{Y})$ e $(\mathbf{Y}\mathbf{Y}')$:

- En el JK-Biplot, se puede aproximar la proporción de la distancia del centro de gravedad al origen de coordenadas en la nube original de los individuos, y la proporción de la variabilidad captada por dicha representación.
- Se puede aproximar la proporción de la traza de $(\mathbf{Y}'\mathbf{Y})$ determinada por la distancia del centro de gravedad al origen de coordenadas, y la proporción de la traza de $(\mathbf{Y}'\mathbf{Y})$ explicada por la traza de la matriz de varianzas y covarianzas \mathbf{S} .

CAPITULO II

FORMAS ALTERNATIVAS DE

APROXIMACIONES BILOT

FORMAS ALTERNATIVAS DE APROXIMACIONES BIPLLOT

1.- INTRODUCCION

La fundamentación teórica de los métodos Biplot, se basa en la estrecha relación existente entre la Teoría de Aproximación de Matrices (ECKART & YOUNG, 1936) y el álgebra de la Descomposición de una Matriz en sus Valores y Vectores Singulares, cuyos objetivos en ambos casos son la búsqueda de subespacios de ajuste óptimo en el sentido de los mínimos cuadrados.

Específicamente, la Teoría de Aproximación de Matrices tiene por objeto la aproximación de una matriz dada $\mathbf{Y}_{(r)}$ (de rango r), por otra de rango menor $\mathbf{Y}_{(q)}$ ($q < r$) utilizando para ello el método de los mínimos cuadrados, ya que el rango de $\mathbf{Y}_{(q)}$ permite la factorización:

$$\mathbf{Y}_{(q)} = \mathbf{A}_{(q)} \mathbf{B}'_{(q)}$$

de donde, $\mathbf{Y}_{(q)}$ se puede aproximar estimando $\mathbf{B}_{(q)}$ por mínimos cuadrados para valores prefijados de $\mathbf{A}_{(q)}$.

De allí que, la matriz hallada $\mathbf{Y}_{(q)}$ garantiza la minimización de la suma de cuadrados de los errores cometidos en la aproximación, o sea:

$$\|\mathbf{Y} - \mathbf{Y}_{(q)}\|^2 = \sum_{i=1}^n \sum_{j=1}^p (y_{ij} - y_{ij(q)})^2 = \text{tr}[(\mathbf{Y} - \mathbf{Y}_{(q)})(\mathbf{Y} - \mathbf{Y}_{(q)})']$$

HOUSEHOLDER & YOUNG en 1938 demostraron que esa misma aproximación $\mathbf{Y}_{(q)}$ también se puede hallar a través de la Descomposición en Valores Singulares de la matriz $\mathbf{Y}_{(q)}$, aproximación que conserva las propiedades mínimo cuadráticas.

Una vez estimadas por mínimos cuadrados las matrices $\mathbf{A}_{(q)}$ y $\mathbf{B}_{(q)}$, se puede realizar la Descomposición en Valores Singulares de manera que:

$$\mathbf{Y}_{(q)} = \mathbf{A}_{(q)} \mathbf{B}'_{(q)} = \mathbf{U}_{(q)} \mathbf{D}_{(q)} \mathbf{V}'_{(q)}$$

De donde se puede representar cualquier tipo de Biplot.

De allí que, aún cuando los métodos Biplots se usen generalmente para describir conjuntos de datos multivariantes sin hacer supuestos sobre modelos subyacentes ni distribuciones poblacionales, es posible identificar una aproximación Biplot a través de un ajuste bilineal (GOLLOB, 1968).

Uno de los objetivos en ésta investigación, requiere la consideración de la formulación de los Biplot a través de la aproximación de la matriz \mathbf{Y} , realizando un “ajuste a través de Modelos Bilineales Multiplicativos”, cuyos métodos de estimación se basan en el método de los mínimos cuadrados ó en

el método de máxima verosimilitud y cuya aplicación requiere la consideración de algunos supuestos sobre distribuciones poblacionales.

El ajuste de Biplots a través de los Modelos Bilineales Multiplicativos, se puede expresar en notación matricial como:

$$\mathbf{Y} = \mathbf{AB}' + \mathbf{E} \quad (2.1)$$

Las propiedades mínimo cuadráticas de la Descomposición en Valores Singulares de una matriz, proporciona la base para la aproximación de una cierta matriz de datos \mathbf{Y} , utilizando un método de mínimos cuadrados alternados para estimar las matrices de parámetros (matriz de marcadores fila \mathbf{A} y la matriz de marcadores columna \mathbf{B}) en la factorización Biplot.

Los ajustes por mínimos cuadrados alternados conservan las propiedades inherentes a cada tipo de Biplot, además de la propiedad de unicidad en la factorización (imponiendo restricciones en el espacio de las filas o de las columnas de la matriz de datos) (GABRIEL, 1978).

Como parte del marco teórico necesario para el logro de los objetivos de este trabajo de investigación, se presenta a continuación los fundamentos de las aproximaciones Biplot a través de:

- La Descomposición en Valores Singulares
- La Estimación por Mínimos Cuadrados Alternados
- La estimación en los Modelos Bilineales Multiplicativos como extensiones de los Modelos Lineales Generalizados

2- APROXIMACIONES BIPLLOT A TRAVES DE LA DESCOMPOSICION EN VALORES SINGULARES DE UNA MATRIZ

La Descomposición en Valores Singulares es la descomposición de una matriz \mathbf{Y} , como el producto de tres matrices de forma e interpretación geométrica simples. Se fundamenta en el siguiente teorema:

TEOREMA 1:

Una matriz \mathbf{Y} de orden $(n \times p)$ puede ser expresada como

$$\mathbf{Y}_{(n \times p)} = \mathbf{U}_{(n \times q)} \mathbf{D}_{(q \times q)} \mathbf{V}'_{(q \times p)} = \sum_{k=1}^q \alpha_k \mathbf{u}_k \mathbf{v}'_k$$

Donde

$\mathbf{D} = \text{diag} [\alpha_1, \alpha_2, \dots, \alpha_q]$, $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_q$

$\alpha_k = \text{valores singulares} = (\lambda_k)^{1/2}$ ($k=1, 2, \dots, q$)

$\lambda_k = \text{valores propios de } \mathbf{Y}'\mathbf{Y}$

q es el rango de la matriz \mathbf{Y} ($q < \min(n, p)$)

\mathbf{U} matriz cuyos vectores columna son vectores singulares por la izquierda de $\mathbf{Y}\mathbf{Y}'$ y forman una base ortonormal para las columnas de \mathbf{Y}

\mathbf{V} matriz cuyos vectores columna son vectores singulares por la derecha de $\mathbf{Y}'\mathbf{Y}$ y forman una base ortonormal para las filas de \mathbf{Y}

$\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}$ luego las columnas de \mathbf{U} y \mathbf{V} son ortonormales.

Las aproximaciones Biplot a través de la Descomposición en Valores Singulares de una matriz se fundamentan en los siguientes dos teoremas, los cuales permiten realizar “ajustes bilineales” a través de aproximaciones mínimo cuadráticas:

TEOREMA 2 (HOUSEHOLDER & YOUNG, 1938):

Dada una matriz \mathbf{Y} de orden $(n \times p)$, cuyo rango r es igual a $q < p$, entonces se obtiene el siguiente mínimo:

$$\min \|\mathbf{Y} - \mathbf{H}\| = \sum_{k=q+1}^r \alpha_k^2$$

sí y solo sí:

$$\mathbf{H} = \sum_{k=1}^q \lambda_k \mathbf{u}_k \mathbf{v}'_k$$

TEOREMA 3 (ECKART & YOUNG, 1936):

La mejor aproximación de rango q ($q < p$) para una matriz \mathbf{Y} (de orden $n \times p$ y de rango r) es:

$$\mathbf{Y}_{(r)} \cong \mathbf{Y}_{(q)} = \mathbf{U}_{(q)} \mathbf{D}_{(q)} \mathbf{V}'_{(q)} = \sum_{k=1}^q \alpha_k \mathbf{u}_k \mathbf{v}'_k \quad (2.2)$$

La Descomposición en Valores Singulares de \mathbf{Y} permite la desagregación en q sumandos, cada uno de los cuales es de menor importancia en relación con la parte de la información captada de \mathbf{Y} .

Las propiedades mínimo cuadráticas de la Descomposición en Valores Singulares, permiten obtener:

- La mejor aproximación de rango q para las filas de \mathbf{Y} , la cual coincide con la proyección ortogonal de cada vector fila \mathbf{y}_i sobre el subespacio q -dimensional de mejor ajuste a las filas de \mathbf{Y} .
- La mejor aproximación de rango q para las columnas de \mathbf{Y} , la cual coincide con la proyección ortogonal de cada vector columna \mathbf{y}_j sobre el subespacio q -dimensional de mejor ajuste a las columnas de \mathbf{Y} .
- La mejor aproximación mínimo cuadrática de rango q , para un elemento genérico y_{ij} de \mathbf{Y} :

$$y_{ij(q)} = \sum_{k=1}^q \alpha_k u_{ik} v'_{jk}; (i = 1, 2, \dots, n; j = 1, 2, \dots, p)$$

Es posible determinar un indicador de la calidad de la aproximación de \mathbf{Y} , mediante $\mathbf{Y}_{(q)}$ por la Descomposición en Valores Singulares, en forma similar a un coeficiente de determinación en análisis de regresión, o sea, a través del

porcentaje de la suma de cuadrados de \mathbf{Y} , que se consigue explicar con la aproximación $\mathbf{Y}_{(q)}$.

Dado que la suma de cuadrados de \mathbf{Y} es igual a la traza de $(\mathbf{Y}\mathbf{Y}')$, entonces:

$$\begin{aligned}\sum \mathbf{Y}_{(q)}^2 &= \sum_{k=1}^r \alpha_k^2 \\ \sum \mathbf{Y}_{(q)}^2 &= \sum_{k=1}^q \alpha_k^2 \\ \sum [\mathbf{Y} - \mathbf{Y}_{(q)}]^2 &= \sum_{k=q+1}^r \alpha_k^2\end{aligned}\tag{2.3}$$

Luego la calidad de la aproximación de \mathbf{Y} será:

$$CA = \left(\frac{\sum_{k=1}^q \alpha_k^2}{\sum_{k=1}^r \alpha_k^2} \right) \times 100\tag{2.4}$$

3- APROXIMACIONES BILOT A TRAVES DE MINIMOS CUADRADOS ALTERNADOS

En los métodos Biplot, una vez que se ha aproximado la matriz de datos \mathbf{Y} a través de la Descomposición en Valores Singulares, se realiza una factorización en marcadores fila y columna, tal que:

$$\mathbf{Y}_{(q)} = \mathbf{U}_{(q)} \mathbf{D}_{(q)} \mathbf{V}'_{(q)} = \mathbf{A}_{(q)} \mathbf{B}'_{(q)} \quad (2.5)$$

La cual garantiza la representación gráfica aproximada de \mathbf{Y} , a través de la proyección ortogonal de los vectores de marcadores fila \mathbf{a}_i contenidos en la matriz $\mathbf{A}_{(q)}$ sobre los marcadores \mathbf{b}_j contenidos en la matriz $\mathbf{B}_{(q)}$, de donde cada elemento de \mathbf{Y} puede ser reconstruido aproximadamente a través de la siguiente “forma bilineal”:

$$y_{ij} \cong \mathbf{a}'_{i(q)} \mathbf{b}_{j(q)} \quad (2.6)$$

Esta aproximación corresponde a un **ajuste bilineal**, el cual se puede obtener a través de un ajuste mínimo cuadrático, donde se buscan los subespacios de representación óptima de filas o columnas, dependiendo de si las restricciones se imponen respectivamente sobre la matriz \mathbf{B} o la matriz \mathbf{A} (GOLUB & REINSH, 1970).

Así pues, las matrices de marcadores que permiten construir las aproximaciones Biplot de una matriz de datos \mathbf{Y} son equivalentes a las matrices de parámetros en los siguientes modelos bilineales multiplicativos:

$$\mathbf{Y} = \mathbf{A} \mathbf{B}' + \mathbf{E} \quad (2.7)$$

$$\mathbf{Y}' = \mathbf{B} \mathbf{A}' + \mathbf{E}' \quad (2.8)$$

donde

\mathbf{Y} es la matriz a aproximar de orden $(n \times p)$

\mathbf{A} es una matriz de parámetros de orden $(n \times q)$

\mathbf{B} es una matriz de parámetros de orden $(p \times q)$

\mathbf{E} es una matriz de errores aleatorios de orden $(n \times p)$

Las estimaciones en esos modelos se obtienen al efectuar el ajuste mínimo cuadrático de las columnas de la matriz \mathbf{Y} en el modelo (2.7), o de las filas de \mathbf{Y}' en el modelo (2.8), sobre un subespacio q -dimensional de representación óptima.

Considerando la aproximación Biplot en la forma del ajuste del modelo (2.7) y suponiendo que las coordenadas para las filas de \mathbf{A} han sido prefijadas, entonces el subespacio de mejor ajuste a las columnas de \mathbf{Y} se obtiene minimizando el cuadrado de las distancias entre las columnas \mathbf{y}_j y sus

aproximaciones en el subespacio q-dimensional, tal que:

$$\min\|\mathbf{Y} - \mathbf{AB}'\| = \text{traza}[(\mathbf{Y} - \mathbf{AB}')'(\mathbf{Y} - \mathbf{AB}')] \quad (2.9)$$

La solución mínimo cuadrática viene dada por:

$$\hat{\mathbf{B}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y} \quad (2.10)$$

Luego las filas de \mathbf{B} son los coeficientes de regresión obtenidos de la regresión de cada columna de la matriz \mathbf{Y} sobre las columnas de la matriz \mathbf{A} .

De la misma manera considerando el modelo bilineal (2.8), entonces prefijando los valores de \mathbf{B} podemos obtener los valores de \mathbf{A} por mínimos cuadrados, por lo que ajustar ese modelo es equivalente a realizar un ajuste mínimo cuadrático de las filas de \mathbf{Y}' sobre el subespacio q-dimensional de representación óptima, tal que:

$$\min\|\mathbf{Y}' - \mathbf{BA}'\| = \text{traza}[(\mathbf{Y}' - \mathbf{BA}')'(\mathbf{Y}' - \mathbf{BA}')] \quad (2.11)$$

cuya solución viene dada por:

$$\hat{\mathbf{A}} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{Y} \quad (2.11)$$

O sea que, las filas de **A** se corresponden con los coeficientes de la regresión de las filas de **Y** sobre las columnas de **B**.

Se demuestra para **Y** en rango 1 y en rango 2, que existe relación entre el ajuste de un Biplot a través de un modelo bilineal con estimación por mínimos cuadrados alternados y el ajuste de un Biplot a través de la Descomposición en Valores Singulares (GOLUB & REINSCH, 1970; GABRIEL, 1995; BLAZQUEZ, 1998).

Sin embargo, aún cuando se demuestra matemáticamente que la estimación por mínimos cuadrados alternados converge a la Descomposición en Valores Singulares, esta convergencia puede ser a un mínimo local, como en el caso en que la solución esté próxima a cualquiera de los vectores singulares, en cuyo caso el algoritmo converge a ese valor y no al primero que es el mayor (BLAZQUEZ, 1998).

Las restricciones de ortonormalidad se pueden considerar a posteriori de las estimaciones de **A** ó **B**, realizando la Descomposición en Valores Singulares de la solución final obtenida de las regresiones cruzadas, para luego recalcular de nuevo **A** y **B** como:

$$\mathbf{A} = \mathbf{U}_{(q)} \mathbf{D}_{(q)}^\gamma \quad \mathbf{B} = \mathbf{V}_{(q)} \mathbf{D}_{(q)}^{1-\gamma}$$

El método de mínimos cuadrados alternados permite (para mayores detalles ver BLAZQUEZ, 1998):

- Introducir métricas diferentes de la matriz identidad en el espacio de las filas y las columnas, pudiéndose por lo tanto calcular los Biplots Generalizados, el Análisis de Correspondencias, el Análisis de Correlación Canónica y el Análisis Canónico de Poblaciones.
- Introducir pesos diferentes para cada uno de los elementos de la matriz **Y**, al igual que en el “método criss-cross” de GABRIEL & ZAMIR (1979), dando así la oportunidad de incluir datos faltantes sin tener que suprimir filas y/o columnas.

Por otra parte, este método de estimación tiene la desventaja que no permite el cálculo de los errores de estimación, ya que se parte de una matriz estimada a priori y luego a partir de ella estimar la otra, no constituyendo por ende la matriz de partida una matriz de variables no aleatorias, tal como sucede en el caso del Análisis de Regresión Lineal.

4.- MODELOS LINEALES GENERALIZADOS

4.1.- INTRODUCCION

Hemos visto que, basándose en el teorema de HOUSEHOLDER & YOUNG (1938) cualquier tipo de Biplot puede entenderse como un ajuste a través del siguiente Modelo Bilineal Multiplicativo:

$$\mathbf{Y} = \mathbf{A} \mathbf{B}' + \mathbf{E} \quad \Rightarrow \quad y_{ij} = a'_i b_j$$

Teniendo en cuenta que *los Modelos Bilineales son extensiones de los Modelos Lineales Generalizados, la comprensión de los primeros requiere del conocimiento de los últimos.*

Esta es la razón por la cual, se presenta a continuación una síntesis de los fundamentos teóricos de los Modelos Lineales Generalizados (NELDER & WEDDERBURN, 1972), los cuales constituyen la base para abordar en el próximo apartado la aproximación de los Biplot a través de Modelos Bilineales.

Sin embargo, los lectores que así lo deseen pueden pasar directamente al próximo apartado para no perder el hilo del desarrollo que veníamos haciendo, y leerlo en el momento que así lo requieran.

4.2.- FAMILIA EXPONENCIAL

Se ha investigado que las buenas propiedades de la distribución Normal son compartidas por una clase de distribuciones, las llamadas distribuciones de la familia Exponencial.

Un vector aleatorio \mathbf{y} pertenece a la familia Exponencial si su función de probabilidad tiene la siguiente forma:

$$f(\mathbf{y}; \theta) = s(\mathbf{y}) t(\theta) e^{a(\mathbf{y})b(\theta)} = \exp [a(\mathbf{y}) b(\theta) + c(\theta) + d(\mathbf{y})]$$

siendo

θ un vector de parámetros

a , b , s y t funciones conocidas

$$s(\mathbf{y}) = \exp[d(\mathbf{y})]$$

$$t(\theta) = \exp[c(\theta)]$$

De donde la función de log-verosimilitud L es:

$$L = \log f(\mathbf{y}; \theta) = a(\mathbf{y}) b(\theta) + c(\theta) + d(\mathbf{y}) \quad (2.12)$$

Al obtener las derivadas parciales de L respecto del parámetro θ se demuestra que:

$$E[a(\mathbf{y})] = - c'(\theta) / b'(\theta)$$

$$\text{Var}[a(\mathbf{y})] = [b'(\theta) c'(\theta) - c''(\theta) b'(\theta)] / [b'(\theta)]^3$$

siendo

c' la primera derivada de L respecto de θ

c'' la segunda derivada de L respecto de θ

Si $a(\mathbf{y}) = \mathbf{y}$ se dice que la distribución está en su forma canónica, llamándose a $b(\theta)$ parámetro natural de la distribución (parámetro estructural). Si existen otros parámetros además de θ se consideran como parámetros incidentales y se suponen conocidos.

La familia Exponencial abarca las siguientes distribuciones univariantes: Normal, Poisson, Binomial, Binomial Negativa, Multinomial, Gamma (RAO, 1973; ROHATGI, 1976), presentándose sus características generales en forma canónica en la Tabla 2.1.

DISTRIB	$b(\theta)$	$a(\mathbf{y})$	$c(\theta)$	$d(\mathbf{y})$	$E[a(\mathbf{y})]$	$Var[a(\mathbf{y})]$
Normal	μ/σ^2	y	$-1/2[\mu/\sigma^2 + \log(2\pi\sigma^2)]$	$-y^2/2\sigma^2$	μ	σ^2
Poisson	$\log\lambda$	y	$-\lambda$	$-\log y!$	λ	λ
Binomial	$\log(p/1-p)$	y	$n \log(1-p)$	$\log \binom{n}{y}$	np	$np(1-p)$

Tabla 2.1.- Distribuciones Univariantes de la Familia Exponencial y sus características en forma canónica

4.3.- DEFINICION DE MODELOS LINEALES GENERALIZADOS

Consideremos una clase de modelos basada en la familia de distribuciones Exponencial, a los cuales se les denomina Modelos Lineales Generalizados.

NELDER & WEDDERBURN (1972) fueron los primeros en considerar ésta clase de modelos, posteriormente las ideas de esos autores fueron extendidas por MC CULLAGH & NELDER (1989) y DOBSON (1990).

Al igual que en un Modelo Lineal Clásico, en los Modelos Lineales Generalizados se tiene una muestra de n observaciones de una cierta variable respuesta Y que conforma un vector $\mathbf{y} = (y_1, y_2, \dots, y_n)$ y un conjunto de variables predictoras A_1, A_2, \dots, A_k que conforman una matriz \mathbf{A} de orden $(n \times k)$.

La filosofía de los Modelos Lineales Generalizados es fundamentalmente la misma del análisis de regresión y ANOVA; o sea, los valores esperados $E(\mathbf{y})$ de las variables observadas, se relacionan con predictores η que son combinaciones lineales de las variables predictoras A , o sea:

$$\eta = E(\mathbf{y}) = \mathbf{A} \beta$$

Los Modelos Lineales Generalizados se definen formalmente en función de las siguientes componentes:

1) Un conjunto de variables independientes Y_i , cada una de las cuales tiene distribución idéntica de la familia Exponencial con un solo parámetro θ_i (el cual puede variar para cada i), por lo que su función de probabilidad conjunta es:

$$f(Y_1, Y_2, \dots, Y_n; \theta_1, \theta_2, \dots, \theta_n) = \prod_{i=1}^n \exp[a(y_i)b(\theta_i) + c(\theta_i) + d(y_i)] =$$

$$f(Y_i; \theta_i) = \exp\left[\sum_{i=1}^n a(y_i)b(\theta_i) + \sum_{i=1}^n c(\theta_i) + \sum_{i=1}^n d(y_i)\right] \quad (2.13)$$

Los parámetros θ_i no son usualmente de interés, ya que corresponden uno a cada observación muestral, por ello se considera el conjunto de parámetros β_k ($k < n$).

2) Un conjunto de variables predictoras A_1, A_2, \dots, A_k que conforman una matriz \mathbf{A} de orden $(n \times k)$.

3) Un conjunto de parámetros $\beta_1, \beta_2, \dots, \beta_k$ ($k < n$) tal que una combinación de esos k parámetros sea igual a una función del valor esperado μ_i de Y_i , es decir:

$$g[E(y_i)] = g(\mu_i) = \mathbf{a}'_i \boldsymbol{\beta}$$

De donde, la especificación del Modelo Lineal Generalizado (MLG) para un vector $\mathbf{y} = (y_1, y_2, \dots, y_n)'$, con $\boldsymbol{\mu} = E(\mathbf{y})$ y considerando una “función

link” $g(\cdot)$, es la siguiente:

$$\eta = g[E(\mathbf{y})] = g(\mu) = \mathbf{A}\beta \quad (2.14)$$

siendo

$g(\cdot)$ una función monótona y diferenciable denominada “función link” que sirve de enlace entre la variable respuesta “ \mathbf{y} ” y el predictor lineal η .

$\mathbf{A} = [\mathbf{a}'_i] = [a_{ij}]$ una matriz ($n \times k$) de variables predictoras

β un vector ($k \times 1$) de parámetros desconocidos

Los errores de la estimación de “ \mathbf{y} ” son:

$$\text{var}(\mathbf{y}) = \sigma^2 \mathbf{D}[V(\mu) / w]$$

siendo

σ^2 un parámetro desconocido

$\mathbf{D}[V(\mu) / w]$ una matriz diagonal

$V(\cdot)$ una función varianza dada a priori

$w = (w_1, \dots, w_n)$ los pesos que son funciones de los valores ajustados μ .

Las estimaciones de β y σ^2 se obtienen aplicando un proceso iterativo de mínimos cuadrados ponderados de una forma linearizada de la “función link” $g(\cdot)$. A este procedimiento se le puede denominar de Regresiones Lineales Generalizadas.

4.4.- FUNCIONES LINK

En los Modelos Lineales Generalizados las “funciones link” tienen por finalidad relacionar el predictor lineal ($\mathbf{A}\beta$) con el valor esperado de la variable respuesta “ \mathbf{y} ”, luego:

$$\eta = g[E(\mathbf{y})] = \mathbf{A} \beta \quad (2.15)$$

En los Modelos Lineales Clásicos, para variables respuesta continuas distribuidas Normalmente, el predictor lineal η y el valor esperado de “ \mathbf{y} ” son idénticos, o sea:

$$\eta = \mu = \mathbf{A} \beta$$

de donde, la “función link” es la identidad ($g = \mathbf{I}$), lo que permite tanto a η como a μ tomar valores entre $-\infty$ y $+\infty$ sobre la recta real.

En otros modelos donde la variable respuesta también puede ser discreta y con cualquier distribución perteneciente a la familia Exponencial, el predictor η no es siempre lineal, por lo que se hace necesario transformar la variable respuesta “ \mathbf{y} ”, usando para ello una “función link” diferente de la identidad ($g \neq \mathbf{I}$), tal que el predictor sea convertido en lineal, consiguiendo así que tanto η como μ tomen valores entre $-\infty$ y $+\infty$ sobre la recta real.

Por ejemplo, en el caso de variables respuesta con distribución de Poisson ($\mu > 0$), en el análisis de independencia en tablas de dos vías, se transforma a

través de la “función link logarítmica” ($g = \log$), la variable y'_i ordenada en una matriz $\mathbf{Y} = [y'_i]$ tal que :

$$\eta = \log[E(\mathbf{Y})] = \log \mu = \mathbf{A} \beta$$

siendo \mathbf{Y} y \mathbf{A} matrices conocidas.

de donde
$$\mu = g^{-1}(\eta) = e^\eta$$

En el caso de variables respuesta con distribución Binomial ($0 < \mu < 1$), para el análisis de matrices de datos de presencia-ausencia ($\mathbf{Y} = [y'_i]$), se transforma la variable y'_i a través de la “función link logística” $g = \log [\mu / (1-\mu)]$, tal que:

$$\eta = \log [\mu / (1-\mu)] = \mathbf{A} \beta$$

de donde
$$\mu = g^{-1}(\eta) = [e^\eta / (1 + e^\eta)]$$

Cada “función link” conduce a un modelo en particular perteneciente a la clase de los Modelos Lineales Generalizados, que se denomina generalmente como la “función link” utilizada en el ajuste, así tenemos por ejemplo los modelos logit, probit y log-log que utilizan como “funciones link” el logit, el probit y el log-log respectivamente.

En la tabla 2.2 se presentan a continuación las “funciones link” más utilizadas:

NOMBRE	MODELO	F. LINK	PREDICTOR	$\mu = g^{-1}(\eta)$
Identidad	Lineal	I	$\eta = \mu = X\beta$	$\mu = \eta$
Logit	Logístico	$\log [\mu / (1-\mu)]$	$\eta = \log [\mu / (1-\mu)] = X\beta$	$\mu = [e^\eta / (1+e^\eta)]$
Probit	Probit	$\phi^{-1}(\mu)$	$\eta = \phi^{-1}(\mu) = X\beta$	$\mu = \phi(\eta)$
Log-Log	Log-Log	$\log[-\log(1-\mu)]$	$\eta = \log[-\log(1-\mu)] = X\beta$	$\mu = 1 - \exp[\exp(\eta)]$
Log	Logarítmico	$\text{Log}(\mu)$	$\eta = \log(\mu) = X\beta$	$\mu = e^\eta$
Raíz Cuadrada		$\mu^{1/2}$	$\eta = \mu^{1/2} = X\beta$	$\mu = \eta^2$
Inversa		μ^{-1}	$\eta = \mu^{-1} = X\beta$	$\mu = \eta^{-1}$

Tabla 2.2.- Características de las Funciones Link más utilizadas

4.5.- ESTIMACION

En los Modelos Lineales Generalizados las variables respuesta Y tienen distribución idéntica de la familia Exponencial, por lo que la distribución de probabilidad de cada una de ellas expresada en forma canónica tiene la siguiente expresión:

$$f(y_i ; \theta_i) = \exp [y_i b(\theta_i) + c(\theta_i) + d(y_i)]$$

En la estimación de los parámetros de estos modelos se usa el método de máxima verosimilitud o de máxima log-verosimilitud, para lo cual se considera la distribución conjunta de las Y_i , o función de verosimilitud L que es igual a:

$$L = f(y_1, \dots, y_n; \theta_1, \dots, \theta_n) = \prod \exp [y_i b(\theta_i) + c(\theta_i) + d(y_i)]$$

de donde:

$$L = \exp [\sum y_i b(\theta_i) + \sum c(\theta_i) + \sum d(y_i)]; \quad (i=1,2,\dots,n) \quad (2.16)$$

De allí que la función de log-verosimilitud L puede escribirse como:

$$L = \log L = \sum y_i b(\theta_i) + \sum c(\theta_i) + \sum d(y_i) \quad (2.17)$$

donde

$$E(Y_i) = \mu_i = -c'(\theta_i) / b'(\theta_i)$$

siendo

$$g(\mu_i) = \mathbf{a}'_i \beta = \eta_i$$

Una propiedad de la familia de distribuciones Exponencial es que satisfacen las condiciones de regularidad, por lo que el máximo global de la función L viene dado por la solución de las ecuaciones $[\partial L / \partial \theta = 0]$, o equivalentemente $[\partial L / \partial \beta = 0]$ (FISZ, 1963; COX & HINKLEY, 1974).

La primera derivada de la función de log-verosimilitud ($\partial L / \partial \beta$) se define como el score U_j con respecto al parámetro β_j , o sea:

$$U_j = \partial L / \partial \beta_j = \sum_i (\partial L_i / \partial \beta_j); \quad (j=1,2,\dots,k) \quad (2.18)$$

donde

$$L_i = y_i b(\theta_i) + c(\theta_i) + d(y_i)$$

Para obtener U_j se utiliza

$$\begin{aligned} \partial L_i / \partial \beta_j &= (\partial L_i / \partial \theta_i) \cdot (\partial \theta_i / \partial \mu_i) \cdot (\partial \mu_i / \partial \beta_j) \\ &= \{ [(y_i - \mu_i) a_{ij}] / \text{var}(Y_i) \} \cdot (\partial \mu_i / \partial \eta_i) \end{aligned} \quad (2.19)$$

Luego

$$\partial L / \partial \beta_j = U_j = \sum_{i=1}^n \frac{(y_i - \mu_i) a_{ij}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)$$

donde a_{ij} es el j-ésimo elemento de $[\mathbf{a}'_i]$.

En general las ecuaciones $U_j = 0$ ($j = 1, 2, \dots, k$) son no lineales por lo que tienen que ser resueltas por métodos iterativos. Aplicando el método de Newton-Raphson se puede utilizar la t-ésima aproximación para la estimación b de β , dada por:

$$b^t = b^{(t-1)} - (\partial^2 L / \partial \beta_j \partial \beta_k)^{-1} U^{(t-1)} \quad (2.20)$$

A la matriz de segundas derivadas de L se le llama matriz de información y se denota:

$$\mathfrak{I}_{jk} = - [(\partial^2 L / \partial \beta_j \partial \beta_k)]$$

De allí que la ecuación (2.20) se puede escribir como:

$$\mathbf{b}^t = \mathbf{b}^{(t-1)} + [\mathfrak{I}^{(t-1)}] \mathbf{U}^{(t-1)}$$

Multiplicando ambos lados de esa ecuación por $\mathfrak{I}^{(t-1)}$ obtenemos:

$$\mathfrak{I}^{(t-1)} \mathbf{b}^t = \mathfrak{I}^{(t-1)} \mathbf{b}^{(t-1)} + \mathbf{U}^{(t-1)} \tag{2.21}$$

Se demuestra que la matriz de información para los Modelos Lineales Generalizados es igual a :

$$\mathfrak{I}_{jk} = \sum_i [(a_{ij} a_{ik}) / \text{var}(Y_i)] \cdot (\partial \mu_i / \partial \eta_i)^2 = \sum_i (a_{ij} a_{ik}) w_{ii} \tag{2.22}$$

la cual puede escribirse como:

$$\mathfrak{I} = \mathbf{A} \mathbf{W} \mathbf{A} \tag{2.23}$$

siendo \mathbf{W} una matriz diagonal (nxn) con elementos:

$$w_{ii} = \frac{1}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \tag{2.24}$$

La expresión del lado derecho de la ecuación (2.24) es el vector evaluado en $\mathbf{b}^{(t-1)}$ con elementos:

$$\sum_k \sum_i (a_{ij} a_{ik}) / \text{var}(Y_i) \cdot (\partial \mu_i / \partial \eta_i)^2 \mathbf{b}_k^{(m-1)} + \sum_i [(y_i - \mu_i) a_{ij} / \text{var}(Y_i)] \cdot (\partial \mu_i / \partial \eta_i)$$

Esta última expresión puede escribirse como:

$$\mathbf{A}' \mathbf{W} \mathbf{y}^* \tag{2.25}$$

Evaluando μ_i y $(\partial \eta_i / \partial \mu_i)$, en $\beta = \mathbf{b}^{(t-1)}$, entonces \mathbf{y}^* es igual a:

$$y_i^* = \sum_k a_{ik} b_k^{(t-1)} + (y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right) = \eta_i + (y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)$$

De donde, finalmente la ecuación (2.21) puede escribirse como:

$$(\mathbf{A}' \mathbf{W} \mathbf{A}) \mathbf{b}^t = \mathbf{A}' \mathbf{W} \mathbf{y}^*$$

luego $\mathbf{b}^t = (\mathbf{A}' \mathbf{W} \mathbf{A})^{-1} \mathbf{A}' \mathbf{W} \mathbf{y}^*$ (2.26)

lo que coincide con las ecuaciones normales en un Modelo Lineal, obtenido por mínimos cuadrados ponderados, o sea, es equivalente a realizar la regresión de \mathbf{y}^* sobre \mathbf{A} con pesos \mathbf{W} , excepto que debe ser resuelta en forma iterativa porque en general \mathbf{y}^* y \mathbf{W} dependen de \mathbf{b} . La variable dependiente

\mathbf{y}^* es una forma linearizada de la “función link” y los pesos \mathbf{W} funciones de los valores ajustados μ .

Se demuestra así, que las estimaciones máximo-verosímiles de los parámetros β en el predictor lineal η de un Modelo Lineal Generalizado pueden obtenerse por un procedimiento iterativo de mínimos cuadrados ponderados (DOBSON, 1990).

Una aproximación indirecta, aunque no óptima, puede conseguirse transformando los valores de la variable respuesta Y mediante la “función link” y usar como pesos en la estimación por mínimos cuadrados ponderados, la varianza de la inversa de cada una de las observaciones (GABRIEL, GALINDO & VICENTE-VILLARDON, 1998).

El algoritmo se puede resumir de la siguiente manera:

Dados valores iniciales para \mathbf{b}^0 se calculan \mathbf{y}^{*0} y \mathbf{w}^0 efectuando la regresión de \mathbf{y}^{*0} sobre $\mathbf{a}_1, \dots, \mathbf{a}_p$ con pesos \mathbf{w}^0 , de donde \mathbf{b}^1 implicara η^1 . El procedimiento es iterativo hasta que la diferencia entre aproximaciones sucesivas \mathbf{b}^t y \mathbf{b}^{t-1} sea lo suficientemente pequeña, en cuyo caso \mathbf{b}^t se toma como la estimación máximo-verosímil de β .

A través de este método, utilizando la inversa de la matriz de información se pueden obtener las varianzas de los estimadores de la siguiente forma:

$$\begin{aligned}\text{var}(\hat{\beta}) &= \sigma^2 (\mathbf{A}'\mathbf{W}\mathbf{A})^{-1} = \sigma^2 \mathfrak{S}^{-1} \\ \hat{\sigma}^2 &= \mathbf{e}'\mathbf{W}\mathbf{e}/(n-p)\end{aligned}\tag{2.27}$$

siendo: \mathbf{e} = residuos mínimo cuadráticos = $(\mathbf{y}^* - \mathbf{A} \mathbf{b})$

4.6.- BONDAD DEL AJUSTE

En los Modelos Lineales Generalizados se usa para medir la bondad del ajuste la Deviance (NELDER & WEDDERBURN, 1972), la cual se define en función del estadístico de la razón de verosimilitud como:

$$D = 2 \log \lambda = 2 [l(\mathbf{b}_{\max}; \mathbf{y}) - l(\mathbf{b}; \mathbf{y})] \sim \chi^2_{n-k}\tag{2.28}$$

siendo

\mathbf{b} el estimador en el modelo propuesto bajo la hipótesis nula H_0 , con k parámetros ($k < n$)

\mathbf{b}_{\max} el estimador máximo-verosímil del vector completo de parámetros β ($n \times 1$) bajo la hipótesis alternativa H_1

$\lambda = [L(\mathbf{b}_{\max}; \mathbf{y}) - L(\mathbf{b}; \mathbf{y})]$ = estadístico de razón de verosimilitud

$\log \lambda = [L(\mathbf{b}_{\max}; \mathbf{y}) - L(\mathbf{b}; \mathbf{y})]$ = estadístico de razón de log- verosimilitud

Si el valor de D es muy alto, se rechaza hipótesis nula H_0 , por lo que el modelo propuesto es muy pobre para la descripción de los datos. Si un modelo con k parámetros es bueno para describir un conjunto de n observaciones entonces $D \cong n - k$.

Se demuestra que la expresión (2.28) (Mc CULLAGH & NELDER, 1989; DOBSON, 1990) es igual a:

$$D = \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{Ab})' \mathbf{W} (\mathbf{y} - \mathbf{Ab}) = \frac{1}{\sigma^2} \mathbf{e}' \mathbf{W} \mathbf{e} \tag{2.29}$$

Para la distribución Normal, la Deviance es exactamente la suma de residuos al cuadrado, mientras que para la Poisson es el estadístico G^2 definido por BISHOP, FIENBERG & HOLLAND (1975).

En la Tabla 2.3, se presenta la Deviance para varias distribuciones de la familia Exponencial, donde $\mu_i = \lambda_i = p_i = \sum_i y_i/n$

DISTRIBUCION	L(b,y)	DEVIANCE
Normal	$\left[-\frac{1}{2} \sigma^2 \sum (y - \bar{y})^2 - \frac{1}{2} n \log(2\pi\sigma^2) \right]$	$\left[-\frac{1}{\sigma^2} \sum (y - \bar{y})^2 \right]$
Poisson	$\sum [y_i \log \lambda_i - \lambda_i - \log y_i!]$	$2 \left[\sum y_i \log (y_i / \lambda_i) \right]$
Binomial	$\sum \left[y_i \log (p_i / 1 - p_i) + n_i \log (1 - p_i) + \log \binom{n_i}{y_i} \right]$	$2 \left[\sum y_i \log (y_i / p_i) \right]$

Tabla 2.3.- Deviance para algunas Distribuciones de la Familia Exponencial

Otra medida de la Bondad del Ajuste que se puede utilizar en los Modelos

$$\chi^2 = \sum_{i=1}^n (y_i - \mu)^2 / V(\mu) = \mathbf{e}' \mathbf{W} \mathbf{e}$$

$$\mathbf{W} = \frac{\mathbf{1}}{\text{var}(y_i)} \left[\frac{\partial \mu}{\partial \eta_i} \right]$$

Lineales Generalizados, es el estadístico χ^2 de Pearson Generalizado, o sea:

$$(2.30)$$

Donde $V(\mu)$, es la función varianza estimada en cada distribución, o sea:

$$V(\mu) = \sigma^2 \text{ (para la distribución Normal)}$$

$$V(\mu) = \lambda_i \text{ (para la distribución Poisson)}$$

$$V(\mu) = (p_i / 1 - p_i) \text{ (para la distribución Binomial)}$$

Para la distribución Normal, el estadístico χ^2 es la suma del cuadrado de los residuos, mientras que para la Poisson y la Binomial coincide con el estadístico χ^2 de Pearson clásico.

Bajo los supuestos de normalidad de los Modelos Lineales, tanto la Deviance como el estadístico χ^2 de Pearson Generalizado, tienen distribuciones exactas χ^2 bajo la hipótesis nula H_0 .

5- APROXIMACIONES BILOT A TRAVES DE MODELOS BILINEALES

5.1.- INTRODUCCION

Hemos visto que, la aproximación de matrices por otras de bajo rango abarca métodos que han sido descritos por diversos autores como Modelización

Bilineal o Multiplicativa (GOLLOB, 1968; MANDEL, 1961, 1969; BRADU & GABRIEL, 1974, 1978).

La especificación general de los Modelos Bilineales Multiplicativos Clásicos utilizados para el ajuste de una matriz de datos \mathbf{Y} , considerando la misma notación que la utilizada en la factorización Biplot, es la siguiente:

$$\mathbf{Y} = \mathbf{A}\mathbf{B}' + \mathbf{E}$$

donde \mathbf{Y} es la matriz de datos de orden $(n \times p)$, \mathbf{A} y \mathbf{B} son matrices de parámetros de orden $(n \times q)$ y $(p \times q)$ respectivamente, y \mathbf{E} es una matriz de errores aleatorios de orden $(n \times p)$.

Generalizando esa especificación a los Modelos Bilineales Multiplicativos Generalizados, dada una “función link” $g(\cdot)$ tenemos que:

$$g[E(\mathbf{Y})] = \mathbf{A}\mathbf{B}' \quad \Rightarrow \quad y_{ij} = \mathbf{a}'_i \mathbf{b}_j \quad (2.31)$$

Cualquier tipo de los Modelos Bilineales, incluyendo este multiplicativo, son extensiones del Modelo Lineal Generalizado (NELDER y WEDDERBURN, 1972) cuyo predictor hemos visto que tiene la siguiente expresión general:

$$\eta = g[E(\mathbf{y})] = \mathbf{A}'\boldsymbol{\beta} \quad \Rightarrow \quad g(\mu_i) = \mathbf{a}'_i \boldsymbol{\beta} = \sum_j \mathbf{a}_{ij} \beta_j \quad (2.32)$$

donde “ \mathbf{y} ” es un vector columna ($n \times 1$), η es el predictor lineal para $E(\mathbf{y})$, g es la “función link”, \mathbf{A} es una matriz ($n \times k$) de variables predictoras, y β es un vector columna ($k \times 1$) de parámetros.

La diferencia entre el ajuste bilineal (2.31) y el lineal (2.32), es que en el primero se trata de una matriz \mathbf{Y} ($n \times p$) de variables respuesta y en el producto ($\mathbf{A}\mathbf{B}'$) ambas matrices son desconocidas excepto por su rango, mientras que en el segundo modelo se trata de un vector de respuestas \mathbf{y} ($n \times 1$) y un vector de parámetros β ($k \times 1$), siendo la matriz de variables \mathbf{A} conocida a priori.

El ajuste de Biplot a través de Modelos Bilineales Multiplicativos Generalizados, a desarrollar en el próximo capítulo, requiere del manejo y conocimiento de los conceptos teóricos fundamentales para el ajuste de un Modelo Bilineal Clásico o Generalizado. Por esa razón, se presenta a continuación un resumen que permitirá complementar el marco teórico para el desarrollo de los objetivos en esta investigación.

5.2.- AJUSTE DE MODELOS BILINEALES CLASICOS

Un Modelo Bilineal Clásico se puede expresar en forma general de la siguiente forma:

$$y_{ij} = \mu + \alpha_i + \beta_j + (\text{interacción}) + e_{ij} \quad (2.33)$$

siendo

y_{ij} la variable respuesta distribuída Normalmente, agrupada en la matriz **Y** de orden (n x p)

μ el efecto global común

α_i el efecto del factor fila

β_j el efecto del factor columna

(interacción) el efecto de interacción entre filas-columnas o debida a factores externos

e_{ij} el error o perturbación aleatoria

Este tipo de modelos son extensiones de los Modelos Lineales Generalizados ($\mathbf{y} = \mathbf{A} \boldsymbol{\beta} + \mathbf{E}$), cuando la “función link” es la identidad ($\mathbf{g} = \mathbf{I}$) y la distribución de la variable respuesta es Normal, siendo:

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1k} \\ \vdots \\ y_{n1} \\ \vdots \\ y_{nk} \end{bmatrix}; \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_n \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}; \mathbf{A} = \begin{bmatrix} 1 & 1 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \dots & 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{bmatrix}; \mathbf{e} = \begin{bmatrix} e_{11} \\ \vdots \\ e_{1k} \\ \vdots \\ e_{n1} \\ \vdots \\ e_{nk} \end{bmatrix}$$

El término de interacción en el modelo, se describe a través de términos multiplicativos pudiéndose detectar aproximadamente a través de métodos gráficos y por consiguiente a través de un Biplot.

Dependiendo de los términos incluidos en el modelo, aditivos, multiplicativos o ambos tipos, se definen los diferentes tipos de Modelos Bilineales utilizados en el análisis de tablas de dos vías y/o de tablas de contingencia (BISHOP, FIENBERG & HOLLAND, 1975; CHRISTENSEN, 1990). Así tenemos:

1) Los modelos aditivos:

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \quad (2.34)$$

2) Los modelos aditivos extendidos (modelo de Tukey):

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda \alpha_i \beta_j + e_{ij} \quad (2.35)$$

3) Los modelos multiplicativos:

$$y_{ij} = \lambda \gamma_i \delta_j + e_{ij} \quad (2.36)$$

4) Los modelos aditivos – multiplicativos (modelos de Mandel):

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda \gamma_i \delta_j + e_{ij} \quad (2.37)$$

Casos particulares de los modelos de Mandel son:

4.a) El modelo lineal columna:

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda \alpha_i \delta_j + e_{ij} \quad (2.38)$$

4.b) El modelo lineal fila

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda \gamma_i \beta_j + e_{ij} \quad (2.39)$$

El ajuste mínimo cuadrático en tablas de dos vías, a través de modelos mixtos aditivos y multiplicativos, es un proceso en dos etapas: en la primera etapa se ajustan los términos aditivos, y en la segunda etapa se ajustan los términos multiplicativos a los residuos de los ajustes precedentes.

MANDEL (1961, 1969) estima los parámetros por mínimos cuadrados, pero haciendo uso al final en la estimación de la interacción de la Descomposición

en Valores Singulares de los residuos estimados, o sea:

$$y_{ij} = \mu + \alpha_i + \beta_j + \sum_k \alpha^*_k u_k v'_k + e_{ij}$$

donde

α^*_k son los valores singulares generalizados

u_{ik} son los vectores singulares por la izquierda de $(\mathbf{E}\mathbf{E}')$

v_{jk} son los vectores singulares por la derecha de $(\mathbf{E}'\mathbf{E})$

$\mathbf{E} = [e_{ij}]$ es la matriz de residuos de un ajuste mínimo cuadrático para y_{ij} a través de un modelo aditivo.

Teóricamente los coeficientes de α^*_k satisfacen las ecuaciones matriciales:

$$\begin{aligned} (\mathbf{E} \mathbf{E}') u_k &= (\alpha^*_k)^2 u_k \\ (\mathbf{E}' \mathbf{E}) v_k &= (\alpha^*_k)^2 v_k \end{aligned}$$

La Descomposición en Valores Singulares de la matriz \mathbf{E} , siendo $\text{rango}(\mathbf{E})=r$ es igual a:

$$\mathbf{E} = \alpha^*_1 u_1 v'_1 + \dots + \alpha^*_r u_r v'_r \tag{2.40}$$

La aplicación de la Descomposición en Valores Singulares a los residuos en el modelo de Mandel es más evidente al reescribir el mismo como:

$$y_{ij} - \mu - \alpha_i - \beta_j = \sum_{k=1,q} \alpha^*_k u_{ik} v'_{jk} + e_{ij} \tag{2.41}$$

siendo $(q < r)$ y $\alpha^*_1 > \alpha^*_2 > \dots > \alpha^*_q$

Los primeros q términos del miembro derecho en las expresiones (2.40) y (2.41) coinciden. La suma de los últimos $(r-q)$ términos de (2.40), corresponden a la matriz de e_{ij} en el modelo (2.41).

Posteriormente a MANDEL (1961, 1969), BRADU Y GABRIEL en 1978, discuten el rol de la Descomposición en Valores Singulares en el análisis de la estructura multiplicativa de una tabla de dos vías.

5.3.- AJUSTE DE MODELOS BILINEALES GENERALIZADOS

La generalización de los Modelos Bilineales Clásicos la constituyen los Modelos Bilineales Generalizados, cuando las variables contenidas en la matriz \mathbf{Y} tienen cualquier distribución de la familia Exponencial y la “función link” utilizada en el ajuste es diferente de la identidad ($g \neq \mathbf{I}$).

Los Modelos Bilineales Generalizados tienen la siguiente expresión general:

$$\eta_{ij} = g [E(y_{ij})] = \mu + \alpha_i + \beta_j + \text{interacción} \quad (2.42)$$

Algunos casos particulares de los Modelos Bilineales Generalizados son:

- Los Modelos de Regresión y ANOVA, usados cuando las variables observadas y_i tienen distribución Normal y la “función link” utilizada en el ajuste es la “identidad”.
- Los Modelos Log-Lineales, usados en el análisis de tablas de contingencia, cuando la frecuencia de cada celda tiene distribución de Poisson, y la “función link” utilizada en el ajuste es la logarítmica.
- Los Modelos de Regresión Logística, usados en el análisis de tablas con datos dicotómicos (presencia-ausencia), cuya distribución es Binomial, y la “función link” utilizada en el ajuste es la logística.

La estimación en esos modelos es una generalización del método en dos pasos utilizado en los Modelos Bilineales Clásicos y del método de máxima verosimilitud utilizado en los Modelos Lineales Generalizados, o sea: se estiman los parámetros por máxima verosimilitud (ó equivalentemente por mínimos cuadrados ponderados) realizando Regresiones Generalizadas de una forma linearizada de la “función link”.

FALGUEROLLES (1995, 1996) y VAN EEUWIJK (1995) estiman los parámetros de la interacción en el modelo (2.42) considerando la Descomposición en Valores Singulares. El método constituye una generalización al propuesto por MANDEL en 1961, ya que consideran en el modelo distribuciones de la familia Exponencial para la variable respuesta y “funciones link” diferentes de la identidad, o sea:

$$\eta_i = g_i(\mu_i) = \text{modelolineal} + \sum_{k=1}^n \alpha_k^* u_{ik} v'_{jk} \quad (2.43)$$

Proponen un algoritmo alternante de estimación, como una generalización al utilizado en los Modelos Lineales Generalizados:

Suponiendo conocidos los efectos columna en la parte lineal del modelo y los vectores singulares v_{jk} , estiman los parámetros correspondientes al término del efecto fila α_i y al término en la interacción $(\alpha_k^* u_{ik})$ a través de una Regresión Lineal Generalizada sobre las columnas de \mathbf{Y} .

Recíprocamente, conocidos los efectos fila y los vectores singulares u_{ik} , entonces la estimación del parámetro del efecto columna y del término

de interacción ($\alpha^*_k v_{ik}$), se reduce a una Regresión Lineal Generalizada sobre las filas de la matriz de datos \mathbf{Y} .

Finalmente, estimados u_{ik} y v_{ik} entonces se estima α^*_k , lo que provee una medida de la Bondad del Ajuste.

Detalles de ese algoritmo pueden verse en FALGUEROLLES & FRANCIS (1992).

Recientemente, como una generalización a las propuestas de FALGUEROLLES & FRANCIS (1992) y VAN EEUWIJK (1995), GABRIEL & RITZ (1994) y GABRIEL (1995, 1998) proponen para la estimación de los parámetros en los Modelos Bilineales Generalizados, el método iterativo de mínimos cuadrados ponderados de los Modelos Lineales Generalizados, el cual proporciona estimadores óptimos en el caso de modelos cuya variable respuesta tenga distribuciones de la familia Exponencial.

La aplicación de ese método de estimación a matrices de variables, se denomina de Regresiones Bilineales Generalizadas, el cual resulta de una generalización del clásico uso de la Descomposición en Valores Singulares (HOUSEHOLDER & YOUNG, 1938), Regresión con Rango Reducido (RAO, 1964; DAVIES & TSO, 1982; TER BRAAK & LOOMAN, 1991), la técnica Criss-Cross de ajuste con Rango Reducido (GABRIEL & ZAMIR, 1979). GABRIEL (1998) propone además, una extensión a los Modelos Bilineales Segmentados (modelos con suma de segmentos bilineales).

El método de “Regresiones Bilineales Generalizadas”, se basa en la especificación del siguiente modelo general:

$$g(\mu) = \mathbf{X} \Theta \mathbf{Z}' \quad (2.44)$$

donde:

$\mu = E(\mathbf{Y})$ = matriz de esperanzas

$g(.)$ = función link

\mathbf{X} y \mathbf{Z} son matrices dadas a priori

Θ matriz de parámetros desconocida, excepto por su rango r

Algunos casos particulares de ese modelo son:

a) Cuando $\mathbf{X} = \mathbf{I}_n$ y $\mathbf{Z} = \mathbf{I}_m$, y la “función link” es también la identidad, entonces la especificación del modelo anterior, se convierte en:

$$\mu = \mathbf{I}_n \Theta \mathbf{I}_m = \Theta$$

en cuyo caso la solución del ajuste mínimo cuadrático del modelo con rango reducido coincide con la Descomposición en Valores Singulares de \mathbf{Y} (HOUSEHOLDER & YOUNG, 1938).

b) Reescribiendo el Modelo como Bilineal, lo cual es posible por el rango de Θ , ya que se puede factorizar como $\Theta = \mathbf{A}\mathbf{B}'$, entonces:

$$\mu = \mathbf{I}_n \Theta \mathbf{I}_m = \Theta = \mathbf{A}\mathbf{B}' \quad (2.45)$$

Este es el caso de ajuste de un Biplot como Modelo Bilineal, cuyos términos de interacción se pueden detectar aproximadamente a través de un Biplot.

De allí, surge la siguiente especificación general para el modelo a utilizar en el ajuste de los Biplot:

$$g(\mu) = \mathbf{X} \Theta \mathbf{Z}' = \mathbf{X} (\mathbf{A}\mathbf{B}') \mathbf{Z}' = (\mathbf{X}\mathbf{A}) (\mathbf{Z}\mathbf{B})' \quad (2.46)$$

En la estimación de la matriz de parámetros Θ y los errores de estimación de los parámetros, se aplica un proceso iterativo de mínimos cuadrados ponderados, para lo cual se reespecifica el modelo (2.46) concatenando las columnas o las filas de $\mathbf{Y}^{(1)}$, considerando el producto Kroenecker entre matrices⁽²⁾ y la concatenación de columnas de matrices producto⁽³⁾, de la siguiente forma:

Por concatenación de columnas de la matriz \mathbf{Y} como :

$$g\langle \mu \rangle = [\mathbf{Z} \otimes (\mathbf{X}\mathbf{A})] \langle \mathbf{B}' \rangle \quad (2.46a)$$

y por concatenación de las filas de \mathbf{Y} como:

$$g\langle \mu' \rangle = [\mathbf{X} \otimes (\mathbf{Z}\mathbf{B})] \langle \mathbf{A}' \rangle \quad (2.46b)$$

En el caso del modelo (2.45) cuando \mathbf{X} y \mathbf{Z} son matrices identidad, su reespecificación concatenando las columnas de \mathbf{Y} es igual a:

$$g\langle\mu\rangle = [\mathbf{I}_p \otimes (\mathbf{I}_n \mathbf{A})] \langle\mathbf{B}'\rangle = [\mathbf{I}_p \otimes \mathbf{A}] \langle\mathbf{B}'\rangle = [\text{diag}(\mathbf{A})] \langle\mathbf{B}'\rangle \quad (2.45a)$$

y concatenando las filas de \mathbf{Y} , igual a:

$$g\langle\mu'\rangle = [\mathbf{I}_n \otimes (\mathbf{I}_p \mathbf{B})] \langle\mathbf{A}'\rangle = [\mathbf{I}_n \otimes \mathbf{B}] \langle\mathbf{A}'\rangle = [\text{diag}(\mathbf{B})] \langle\mathbf{A}'\rangle \quad (2.45b)$$

(1) $\langle\mathbf{Y}\rangle$ es un vector columna ($n \times 1$) donde se han concatenado las columnas de la matriz \mathbf{Y} de orden ($n \times p$), o sea: $[\langle\mathbf{Y}\rangle]' = [y_{11}, y_{21}, \dots, y_{n1}, y_{12}, y_{22}, \dots, y_{n2}, \dots, y_{1p}, y_{2p}, \dots, y_{np}]'$
 Por su parte $\langle\mathbf{Y}'\rangle$ es también un vector columna ($n \times 1$), pero donde se han concatenado las filas de la matriz \mathbf{Y} de orden ($n \times p$), o sea: $[\langle\mathbf{Y}'\rangle]' = [y_{11}, \dots, y_{1p}, y_{21}, \dots, y_{2p}, \dots, y_{n1}, \dots, y_{np}]'$

(2) El producto Kroenecker entre dos matrices se define como:

$$\mathbf{P}_{(m \times p \times n \times q)} = \mathbf{A}_{(m \times n)} \otimes \mathbf{B}_{(p \times q)} = a_{ij} \mathbf{B} = \mathbf{B} a_{ij}$$

(3) La concatenación de columnas de matrices producto se puede expresar como:

$\langle \mathbf{A} \mathbf{B} \mathbf{C} \rangle = [\mathbf{C}' \otimes \mathbf{A}] \langle \mathbf{B} \rangle$, de donde en el modelo 2.45 tendremos que:

$$\langle g(\mu) \rangle = \langle \mathbf{A} \mathbf{B}' \mathbf{I} \rangle = (\mathbf{I} \otimes \mathbf{A}) \langle \mathbf{B}' \rangle = [\text{diag}(\mathbf{A})] \langle \mathbf{B}' \rangle$$

y en el modelo 2.46 : $\langle g(\mu) \rangle = \langle (\mathbf{X}\mathbf{A}) \mathbf{B}' \mathbf{Z}' \rangle = [\mathbf{Z} \otimes (\mathbf{X}\mathbf{A})] \langle \mathbf{B}' \rangle$

El procedimiento de estimación iterativo, una vez reespecificados los modelos, se resume de la siguiente manera:

El paso t -ésimo se inicializa con la etapa “criss” dando un valor inicial $\hat{\mathbf{A}}^{(t-1)}$, con lo cual para estimar $\langle \hat{\mathbf{B}}^{(t)} \rangle$, se realiza la regresión generalizada (2.46a), ó la (2.45a), según sea el caso.

El paso t-ésimo termina con la estimación de $\langle \hat{\mathbf{A}}^{(t)} \rangle$ en la etapa “cross”, para lo cual se realiza la regresión generalizada (2.46b) ó (2.45b) según sea el caso.

Una vez que se han estimado las matrices de parámetros (\mathbf{AB}') , se puede realizar una representación Biplot, realizando una factorización de acuerdo al tipo de Biplot a representar (CMP, RMP, etc.).

De acuerdo al diagnóstico realizado, a través de la representación Biplot (BRADU & GABRIEL, 1978), si el examen del patrón indica el ajuste de más de un modelo aditivo a los datos, se analiza la posibilidad de reformularlos en un solo modelo pero dividido en k segmentos, tal que cada segmento de rango especificado r_k dependa de sus propias matrices de covariables \mathbf{X}_k y \mathbf{Z}_k , o sea:

$$g(\mu) = \sum_{k=1}^K \mathbf{X}_k \Theta_k \mathbf{Z}_k' = \sum_{k=1}^K \mathbf{X}_k \mathbf{A}_k \mathbf{B}_k' \mathbf{Z}_k' \tag{2.47}$$

Considerando el producto Kroenecker y la concatenación de columnas y filas de una matriz, ello puede reescribirse como:

$$g \langle \mu \rangle = [\mathbf{Z}_1 \otimes (\mathbf{X}_1 \mathbf{A}_1), \dots, \mathbf{Z}_k \otimes (\mathbf{X}_k \mathbf{A}_k)] [\langle \mathbf{B}'_1 \rangle, \dots, \langle \mathbf{B}'_k \rangle]$$

$$g \langle \mu' \rangle = [\mathbf{X}_1 \otimes (\mathbf{Z}_1 \mathbf{B}_1), \dots, \mathbf{X}_k \otimes (\mathbf{Z}_k \mathbf{B}_k)] [\langle \mathbf{A}'_1 \rangle, \dots, \langle \mathbf{A}'_k \rangle]$$

Finalmente, usando Regresión Lineal Generalizada “criss-cross”, se pueden estimar ambos conjuntos de parámetros en cada segmento en forma iterativa.

CAPITULO III

BIPLOT DE REGRESION AJUSTADOS A TRAVES DE MODELOS BILINEALES GENERALIZADOS

BIPLOT DE REGRESION AJUSTADOS A TRAVES DE MODELOS BILINEALES GENERALIZADOS

1- INTRODUCCION

En este capítulo se presenta el desarrollo del ajuste de Biplots a través de Modelos Bilineales Generalizados con fines descriptivos el cual constituye uno de los objetivos primordiales en esta investigación, y que sirve de base para generalizar los resultados en el próximo capítulo al ajuste de Biplots con Información Externa a través de Modelos Bilineales Generalizados.

Para tal fin, se generaliza la especificación de los Modelos Lineales Generalizados (ver apartado 4.3 del capítulo II) de un vector “y” de orden (nx1) para la variable respuesta, a una matriz \mathbf{Y} de orden (nxp) que contiene p variables respuesta, luego cada columna de \mathbf{Y} proviene de un Modelo Lineal Generalizado con la misma “función link”.

El ajuste de Biplots a través de Modelos Bilineales Generalizados con fines descriptivos, permite la aproximación de una cierta matriz de datos \mathbf{Y} de orden (nxp) y de rango q (q < r) a través del ajuste del siguiente modelo bilineal generalizado multiplicativo:

$$\boldsymbol{\eta} = g[E(\mathbf{Y})] = \mathbf{AB}' \quad (3.1)$$

donde η es el predictor (combinación lineal de las variables predictoras) y las matrices de parámetros \mathbf{A} y \mathbf{B} de orden $(n \times q)$ y $(p \times q)$ respectivamente, son desconocidas excepto por su rango q , luego el producto de ellas es equivalente a la factorización Biplot.

Hemos visto en el capítulo II, que es posible obtener una aproximación Biplot de la matriz de esperanzas “link” $g[E(\mathbf{Y})]$ a través de un ajuste iterativo de mínimos cuadrados ponderados, que equivale a la realización de Regresiones Generalizadas alternadas de las columnas de la matriz \mathbf{Y} sobre las columnas de la matriz \mathbf{A} .

El ajuste de los Biplots también puede hacerse en forma separada, realizando regresiones generalizadas para cada fila y cada columna, de donde es posible construir Biplots de Predicción o Regresión (GOWER & HAND, 1996).

Adaptando la terminología de GOWER & HAND (1996), a los Biplots ajustados a través del modelo (3.1) los denominaremos en lo sucesivo “**Biplots de Regresión**” y su interpretación dependerá de la “función link” utilizada en el ajuste.

De allí que, es de suma importancia el análisis de la geometría en los Biplots de Regresión la cual es similar a la geometría en un ajuste lineal (con función link la identidad), o sea, se requiere de la proyección ortogonal de las columnas de la matriz \mathbf{Y} sobre el subespacio óptimo generado por las columnas de \mathbf{A} (suponiendo \mathbf{A} conocida), o recíprocamente generado por las columnas de \mathbf{B} (suponiendo \mathbf{B} conocida).

A continuación se presentarán dos tipos de Biplots de Regresión:

- Los Biplots de Regresión Lineal
- Los Biplots de Regresión no Lineal

Se analizarán no sólo los elementos a considerar en el ajuste, sino también la geometría de cada uno de ellos que es la base para la interpretación final en las representaciones Biplot.

2.- BILOT DE REGRESION LINEAL

2.1.- AJUSTE DE BILOT DE REGRESION LINEAL

El ajuste de un Biplot a través de Regresión Lineal, para variables continuas con distribución Normal, utiliza como “función link” la identidad.

Cuando la “función link” utilizada en el ajuste es la identidad, entonces el modelo a ajustar a las columnas de la matriz de datos \mathbf{Y} para estimar la matriz de parámetros \mathbf{B} , suponiendo la matriz \mathbf{A} conocida, es:

$$\eta = g[E(\mathbf{Y})] = g(\mu) = \mu = \mathbf{AB}' \quad (3.2)$$

Recíprocamente, el modelo a ajustar a las filas de la matriz de datos \mathbf{Y} para estimar la matriz de parámetros \mathbf{A} , suponiendo conocida la matriz \mathbf{B} es:

$$\eta' = g[E(\hat{\mathbf{Y}})] = g(\hat{\mu}) = \mu' = \mathbf{B}\mathbf{A}' \quad (3.3)$$

donde

\mathbf{Y} matriz (n x p) de variables respuestas, cada una con distribución Normal.

\mathbf{A} matriz de variables predictoras (en el modelo 3.2) o de parámetros (en el modelo 3.3).

\mathbf{B} matriz de variables predictoras (en el modelo 3.3) o de parámetros (en el modelo 3.2).

A los Biplots ajustados a través de los modelos (3.2) y (3.3) los denominaremos “*Biplot de Regresión Lineal*”.

Dado que es posible demostrar que estos ajustes Biplot son equivalentes a los obtenidos por mínimos cuadrados alternados, entonces la estimación de las matrices de parámetros \mathbf{A} y \mathbf{B} puede hacerse por el método de mínimos cuadrados alternados (expuesto en el capítulo anterior), o sea prefijando \mathbf{A} para calcular \mathbf{B} y viceversa.

De allí, tenemos que las matrices de coeficientes de regresión correspondientes a las regresiones de \mathbf{Y} sobre \mathbf{A} , y de \mathbf{Y}' sobre \mathbf{B} son respectivamente:

$$\mathbf{B} = (\mathbf{A}'\mathbf{A})^{-1} \mathbf{A}'\mathbf{Y} \quad (3.4)$$

$$\mathbf{A}' = (\mathbf{B}'\mathbf{B})^{-1} \mathbf{B}'\mathbf{Y}' \quad (3.5)$$

Así pues, el procedimiento de ajuste de “Biplot de Regresión Lineal” a través de mínimos cuadrados alternados, permite realizar representaciones Biplot de \mathbf{Y} e \mathbf{Y}' .

Por tanto, cualquiera de las representaciones Biplot (RMP, CMP, SQRT, HJ), pueden entenderse como ajustes de “Biplot de Regresión Lineal”, realizados a través de los Modelos Bilineales Multiplicativos (3.2) y (3.3).

Para obtener representaciones únicas de los Biplot de Regresión Lineal es necesario imponer restricciones, así tenemos que:

a) El ajuste del “Biplot de Regresión Lineal” a través del modelo (3.3), bajo las restricciones $\mathbf{B}'\mathbf{B} = \mathbf{I}$, conlleva a la estimación de la siguiente matriz de parámetros o coeficientes de regresión:

$$\mathbf{A}' = (\mathbf{B}'\mathbf{B})^{-1} \mathbf{B}'\mathbf{Y}' = \mathbf{B}'\mathbf{Y}'$$

Realizando la Descomposición en Valores Singulares de la solución final, para obtener la factorización Biplot, tendremos que:

$$\hat{\eta}' = \mathbf{B}\hat{\mathbf{A}}' = \mathbf{V}\mathbf{D}\mathbf{U}'$$

Siendo:

$$\mathbf{D} = \text{diag} [\alpha_1, \alpha_2, \dots, \alpha_q], \quad \alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_q$$

$$\alpha_k = \text{valores singulares} = (\lambda_k)^{1/2} \quad (k=1,2,\dots,q)$$

$$\lambda_k = \text{valores propios de } \mathbf{Y}'\mathbf{Y}$$

q es el rango de la matriz \mathbf{Y} ($q < \min(n,p)$)

\mathbf{U} matriz cuyos vectores columna son vectores singulares por la izquierda de $\mathbf{Y}\mathbf{Y}'$ y forman una base ortonormal para las columnas de \mathbf{Y}

\mathbf{V} matriz cuyos vectores columna son vectores singulares por la derecha de $\mathbf{Y}'\mathbf{Y}$ y forman una base ortonormal para las filas de \mathbf{Y}

$\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}$ luego las columnas de \mathbf{U} y \mathbf{V} son ortonormales.

De donde, las matrices de marcadores fila y columna serán:

$$\mathbf{A} = \mathbf{U}\mathbf{D} = \mathbf{Y}\mathbf{B} \qquad \mathbf{B} = \mathbf{V} \qquad (3.6)$$

Luego las columnas de la matriz de marcadores columna \mathbf{B} , bajo la restricción impuesta, corresponden a los vectores propios ortonormalizados asociados a los q mayores valores propios de $(\mathbf{Y}'\mathbf{Y})$ y serán los que definan las direcciones del subespacio q -dimensional de mejor ajuste a la nube de puntos fila de la matriz \mathbf{Y} (VASQUEZ, 1995).

Se observa que las matrices obtenidas en (3.6) se corresponden a las obtenidas en la factorización RMP-Biplot (JK-Biplot).

En consecuencia, la aproximación del **RMP-Biplot de Regresión Lineal** para la fila i -ésima de la matriz de datos \mathbf{Y} será:

$$\mathbf{y}_{i(q)} = \text{Pr oy}(\mathbf{y}_i / \mathbf{V}_{(q)}) = \sum_{k=1}^q (\mathbf{y}'_i \mathbf{V}_k) \mathbf{V}_k \quad (3.7)$$

De donde, el ajuste bilineal para cada elemento fila de la matriz \mathbf{Y} se obtiene por:

$$y_{ij(q)} \cong (\mathbf{a}_{i(q)})' \mathbf{b}_{j(q)} = \sum_{k=1}^q (\mathbf{y}'_i \mathbf{V}_k) v_{jk} \quad (3.8)$$

Además, los marcadores verifican las siguientes propiedades:

- El producto escalar (con la métrica identidad) de las filas de la matriz \mathbf{Y} coincide con el producto escalar entre marcadores filas, o sea:
 $\mathbf{Y}\mathbf{Y}' = \mathbf{A}\mathbf{A}'$.
- La distancia euclídea entre filas de \mathbf{Y} coincide con la distancia entre marcadores fila, o sea: $(\mathbf{y}_i - \mathbf{y}_s)' (\mathbf{y}_i - \mathbf{y}_s) = (\mathbf{a}_i - \mathbf{a}_s)' (\mathbf{a}_i - \mathbf{a}_s)$
- Los marcadores fila coinciden con las coordenadas de las filas en el espacio de las componentes principales: $\boldsymbol{\psi} = \mathbf{Y}\mathbf{V} = \mathbf{A}$
- La calidad de representación para las filas es óptima.

La medida del desfase de las aproximaciones del “RMP-Biplot de Regresión Lineal” será:

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^p (\text{residuos})^2 &= \sum_i \sum_j (y_{ij} - \mathbf{a}'_{i(q)} \mathbf{b}_{j(q)})^2 = \|\mathbf{Y}\|^2 - \sum_{k=1}^q \mathbf{V}'_k (\mathbf{Y}'\mathbf{Y}) \mathbf{V}_k = \\ &= \text{traza}(\mathbf{Y}'\mathbf{Y}) - \sum_{k=1}^q \lambda_k = \sum_{k=q+1}^p \lambda_k \end{aligned}$$

b) El ajuste del “Biplot de Regresión Lineal” a través del modelo (3.2), bajo las restricciones $\mathbf{A}'\mathbf{A} = \mathbf{I}$, conlleva a la estimación de la siguiente matriz de parámetros:

$$\mathbf{B} = (\mathbf{A}'\mathbf{A})^{-1} \mathbf{A}'\mathbf{Y} = \mathbf{A}'\mathbf{Y}$$

Realizando la Descomposición en Valores Singulares de la solución final, para obtener la factorización Biplot, tendremos que:

$$\hat{\mu} = \mathbf{A}\hat{\mathbf{B}}' = \mathbf{U}\mathbf{D}\mathbf{V}'$$

De donde, las matrices de marcadores fila y columna, serán:

$$\mathbf{A} = \mathbf{U} \qquad \mathbf{B} = \mathbf{V}\mathbf{D} = \mathbf{A}'\mathbf{Y} \qquad (3.9)$$

Luego las columnas de la matriz de marcadores columna \mathbf{A} , bajo la restricción impuesta ($\mathbf{A}'\mathbf{A} = \mathbf{I}$), corresponden a los vectores propios ortonormalizados asociados a los q mayores valores propios de la matriz $(\mathbf{Y}\mathbf{Y}')$ y serán los que

definan las direcciones del subespacio q-dimensional de mejor ajuste a la nube de puntos columna de la matriz \mathbf{Y} .

Se observa que las matrices obtenidas en (3.9), se corresponden a las obtenidas en la factorización CMP-Biplot (GH-Biplot).

En consecuencia, la aproximación del “*CMP-Biplot de Regresión Lineal*” para la columna j-ésima de la matriz de datos \mathbf{Y} , será:

$$\mathbf{y}_{j(q)} = \text{Proy}(\mathbf{y}_j / \mathbf{U}_{(q)}) = \sum_{k=1}^q (\mathbf{y}'_j \mathbf{U}_k) \mathbf{U}_k \tag{3.10}$$

De donde, el ajuste bilineal de cada elemento de la matriz se obtiene por:

$$y_{ij(q)} \cong (\mathbf{a}_{i(q)})' \mathbf{b}_{j(q)} = \sum_{k=1}^q (\mathbf{y}'_j \mathbf{U}_k) u_{ik} \tag{3.11}$$

Además, los marcadores verifican:

- El producto escalar de las columnas de \mathbf{Y} coincide con el producto escalar de los marcadores columna, o sea: $\mathbf{y}_j' \mathbf{y}_j \approx \mathbf{b}_j' \mathbf{b}_j$.
- La longitud al cuadrado de los vectores que representan a los marcadores columna, aproximan la covarianza entre las variables correspondientes, de donde: $\mathbf{b}_j' \mathbf{b}_j = s_{jj} = s_j^2$

- La longitud de los marcadores columna aproxima la desviación estándar de las variables.
- El coseno del ángulo entre dos marcadores columna aproxima la correlación entre variables correspondientes:

$$\text{Cos}(\mathbf{b}_j, \mathbf{b}_k) \approx \text{Corr}(\mathbf{y}_j, \mathbf{y}_k).$$
- Las coordenadas de la matriz de marcadores columna \mathbf{B} en el Biplot de Componentes Principales, son equivalentes a los pesos de las variables a lo largo de los ejes principales.
- La calidad de representación de las columnas es óptima.

La medida del desfase de las aproximaciones del “CMP-Biplot de Regresión Lineal”, será:

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^p (\text{residuos})^2 &= \sum_i \sum_j (y_{ij} - \mathbf{a}'_{i(q)} \mathbf{b}_{j(q)})^2 = \|\mathbf{Y}\|^2 - \sum_{k=1}^q \mathbf{U}'_k (\mathbf{Y}\mathbf{Y}') \mathbf{U}_k = \\ &= \text{traza}(\mathbf{Y}\mathbf{Y}') - \sum_{k=1}^q \lambda_k = \sum_{k=q+1}^p \lambda_k \end{aligned}$$

2.2.- GEOMETRIA DE LOS BIPLLOT DE REGRESION LINEAL

Al igual que en el caso de los modelos de regresión lineal, el ajuste de los “Biplots de Regresión Lineal” a través de mínimos cuadrados alternados tiene la siguiente interpretación geométrica:

Cada vector de la matriz de datos \mathbf{Y} (vector fila \mathbf{y}_i o vector columna \mathbf{y}_j) se considera respectivamente como un punto en el espacio euclídeo p -dimensional y n -dimensional.

Para un valor dado de un vector \mathbf{b}_j de la matriz de parámetros columna \mathbf{B} , el vector de valores ajustados $\hat{\boldsymbol{\mu}}_j = \mathbf{A} \mathbf{b}_j$ es un punto en el mismo espacio n -dimensional de las columnas de \mathbf{Y} .

El valor del vector de parámetros columna varía de muestra a muestra, por lo que se debe seleccionar la solución tal que:

$$\hat{\boldsymbol{\mu}}_j = \mathbf{A} \hat{\mathbf{b}}_j$$

esté lo más cerca posible al valor observado \mathbf{y}_j , en el sentido de la mínima distancia euclídea, o sea:

$$\min \sum (\mathbf{y}_j - \hat{\boldsymbol{\mu}}_j)^2$$

Recíprocamente, para un valor dado de un vector \mathbf{a}_i de la matriz de parámetros fila \mathbf{A} , el vector de valores ajustados $\hat{\boldsymbol{\mu}}'_i = \mathbf{a}'_i \mathbf{B}$ es un punto en el mismo

espacio p -dimensional de las filas de \mathbf{Y} y la selección de la mejor solución para el vector de parámetros fila se halla en la misma forma que para el vector de parámetros columna.

En el caso por ejemplo, del modelo $\eta_i = \mathbf{a}'_i \boldsymbol{\beta}$ con una variable predictora y un parámetro, la solución pertenece al conjunto de todos los vectores $(\mathbf{a} \boldsymbol{\beta})$, para $-\infty < \boldsymbol{\beta} < +\infty$, es decir todos los puntos sobre la línea que parte del origen en la dirección del vector “ \mathbf{a} ”, y la solución mínimo cuadrática será aquella más cercana a “ \mathbf{y} ”, la cual se halla trazando la perpendicular (YP) sobre $(\mathbf{a} \boldsymbol{\beta})$, tal como se muestra en la figura 3.1 (McCULLAGH & NELDER, 1989, p. 72):

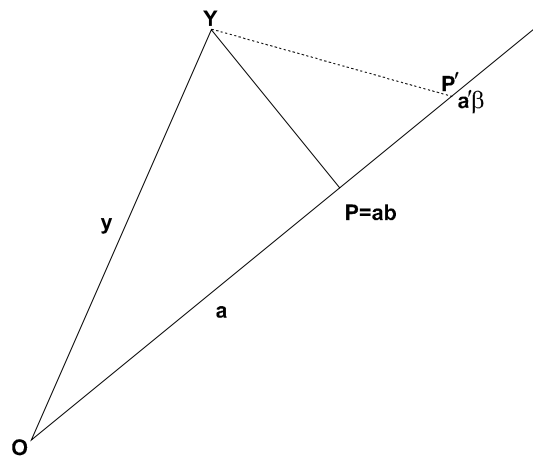


Figura 3.1. Geometría en un ajuste mínimo cuadrático
(Modelo con 1 variable y 1 parámetro)

El vector de valores ajustados OP ($P = \mathbf{a}\mathbf{b}$, siendo \mathbf{b} el estimador mínimo cuadrático del coeficiente de regresión), coincide con la proyección ortogonal

de “y” sobre el subespacio de “a”, siendo el punto sobre la línea “a” más cercano a “y”.

En el caso, de un modelo de regresión con dos variables predictoras y un parámetro ($\eta = \mathbf{a}_1\beta_1 + \mathbf{a}_2\beta_2$), la geometría del ajuste dependerá de la relación entre las variables (McCULLAGH & NELDER, 1989, pp. 74-76), tal como se observa en la figura 3.2:

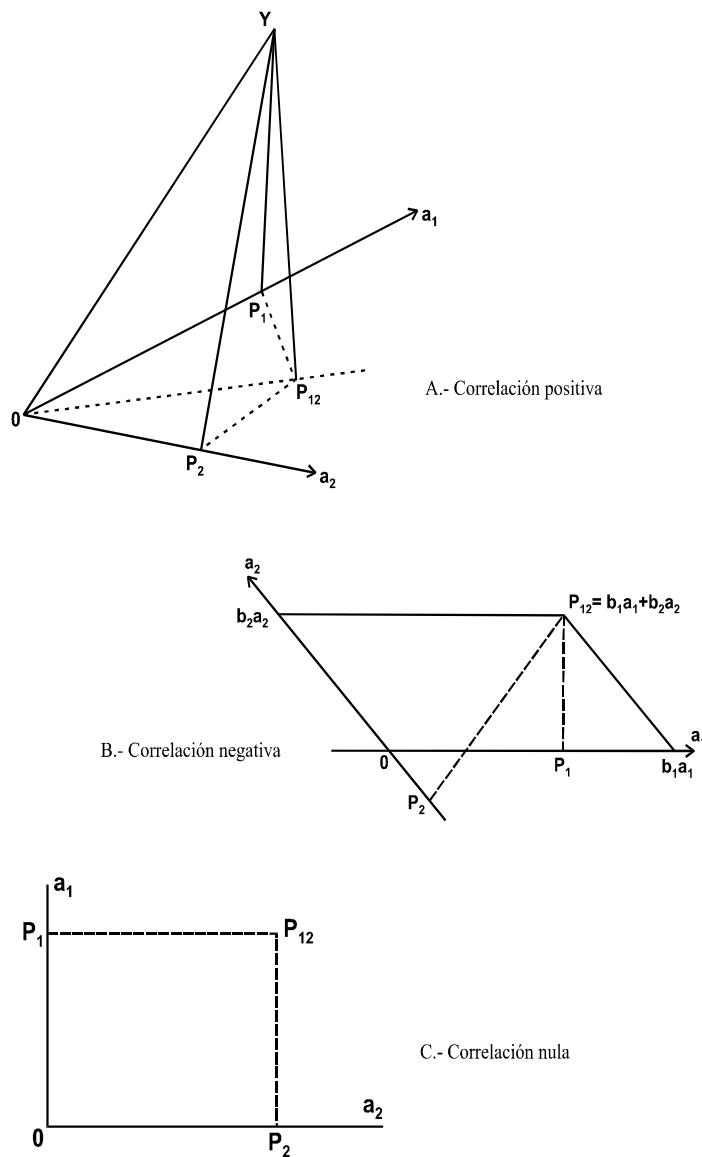


Figura 3.2. Geometría en un ajuste mínimo cuadrático dependiendo de la correlación entre variables. (Modelo con 2 variables y 1 parámetro)

El vector de valores ajustados OP_{12} ($OP_{12}=a_1b_1+a_2b_2$, siendo b_1 y b_2 los estimadores mínimo cuadráticos del vector de coeficientes de regresión β), coincide con la proyección ortogonal de “ y ” sobre el plano $(\mathbf{a}_1, \mathbf{a}_2)$ siendo el punto sobre el plano más cercano a “ y ”.

En el caso del ajuste de los “Biplots de Regresión Lineal”, el análisis de su geometría requiere la consideración de todos los vectores fila o columna de la matriz \mathbf{Y} , tal como se puede ver en GOWER & HAND (1996) y VICENTE-VILLARDON y GALINDO (1998).

Siguiendo a esos autores, la geometría en el ajuste de los “Biplots de Regresión Lineal” utilizando el modelo (3.2), es equivalente a realizar un ajuste para cada columna de la matriz \mathbf{Y} para aproximar cada vector columna \mathbf{y}_j a partir de las coordenadas de los individuos contenidas en la matriz \mathbf{A} , además los coeficientes de regresión coinciden con las coordenadas de cada variable sobre el Biplot, y se estiman utilizando la fórmula:

$$\mathbf{b}_j = (\mathbf{A}'\mathbf{A})^{-1} \mathbf{A}' \mathbf{y}_j \quad (3.12)$$

Llamando \mathcal{L} al espacio generado por las columnas de \mathbf{A} y ajustando el plano de regresión, se obtiene una superficie de respuesta lineal \mathcal{H} , como se puede observar en la figura 3.3.

El conjunto de puntos de \mathcal{H} , que aproxima un cierto valor de la variable y_j , está constituido por la intersección del plano normal \mathcal{N} en ese valor particular de y_j y la intersección con el plano de regresión \mathcal{H} .

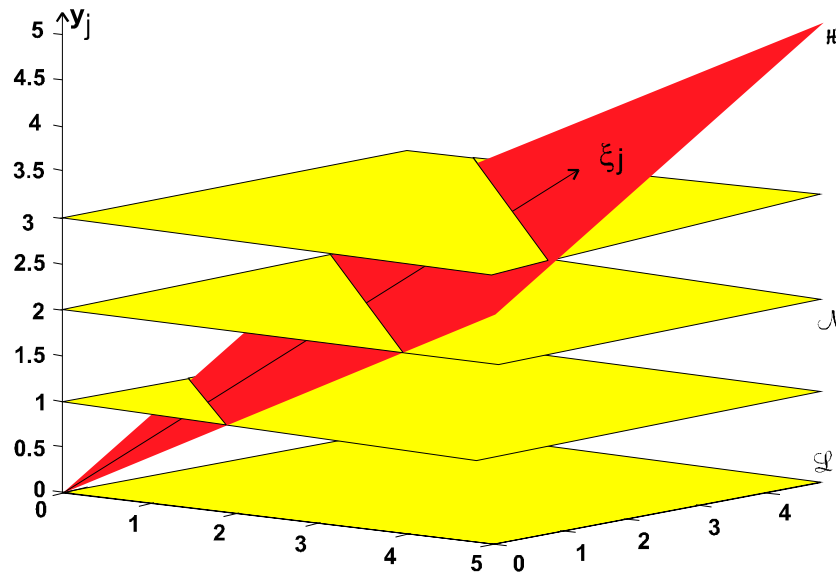


Figura 3.3. Superficie de respuesta lineal en el ajuste de Biplots de Regresión Lineal

De donde, el conjunto de puntos de \mathcal{H} que aproxima un cierto valor de la variable y_j , está sobre una recta en \mathcal{H} que parte del origen y que pasa por el centro de la intersección de cada plano normal sobre la superficie \mathcal{H} . Para distintos valores a predecir de la variable y_j , se obtienen rectas paralelas en el plano de regresión \mathcal{H} .

A la dirección de \mathcal{H} normal a todas esas rectas paralelas se le llama ξ_j , y es el eje de referencia para la predicción de los valores de la variable y_j .

Por su parte, en la figura 3.4 se puede observar que el conjunto de puntos en el espacio \mathcal{L} generado por las columnas de \mathbf{A} , que predicen diferentes valores de la variable y_j , está también sobre rectas paralelas (proyecciones para cada valor de y_j según 3.10).

A la dirección normal a esas rectas paralelas en \mathcal{L} se le llama eje Biplot β_j (coeficiente de la regresión de y_j sobre \mathbf{a}_i), y coincide con la proyección de ξ_j sobre \mathcal{L} .

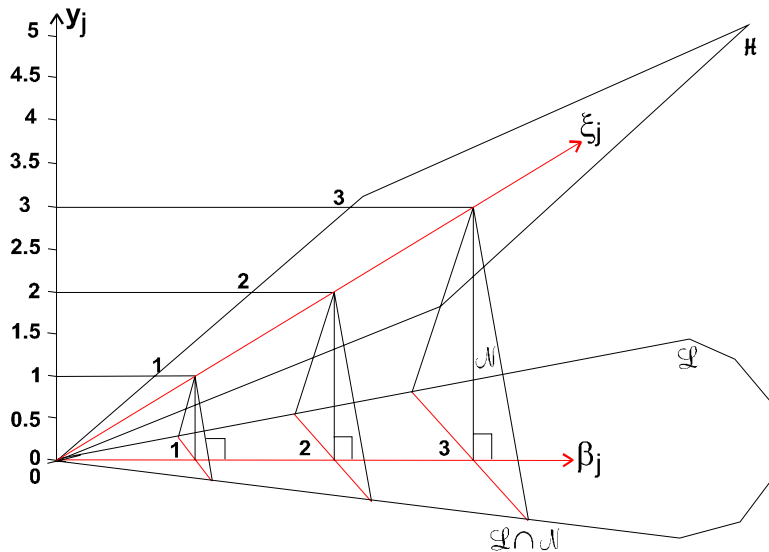


Figura 3.4. Geometría en los Biplots de Regresión Lineal

Las calidades de representación de las variables y_j se pueden aproximar entonces, a través de la regresión para cada columna de la matriz de datos \mathbf{Y} .

En líneas muy generales, ajustar un plano de regresión a las columnas de la matriz \mathbf{Y} conlleva a la obtención de una superficie de respuesta lineal, cuya proyección sobre un subespacio de baja dimensión es también lineal.

En forma recíproca, ajustar un plano de regresión a las columnas de la matriz \mathbf{B} , utilizando el modelo (3.3) para ajustar cada fila de \mathbf{Y} , conlleva también a la obtención de una superficie de respuesta lineal que al proyectarla en un subespacio de baja dimensión es también lineal.

Dada la geometría en el ajuste de los “Biplot de Regresión Lineal”, la interpretación de sus representaciones gráficas se hace en la misma forma que en los Biplots Clásicos, o sea, en términos de proyección de los marcadores fila \mathbf{a}_i sobre los marcadores columna \mathbf{b}_j , o también en forma inversa (véase para mayor detalle el apartado 5 del capítulo I).

Como la “función link” utilizada en el ajuste es la identidad, las proyecciones de los marcadores fila sobre los marcadores columna, permiten una ordenación de los individuos respecto a cada una de las variables consideradas en el ajuste.

Al mismo tiempo, los marcadores fila son equivalentes a los coeficientes de regresión en Análisis de Regresión Múltiple, por lo que se pueden interpretar en forma similar, o sea, por cada unidad de incremento en el marcador \mathbf{b}_j la variable correspondiente \mathbf{y}_j se incrementa en la misma magnitud.

El elemento y_{ij} se aproxima a través del módulo de la proyección de \mathbf{a}_i sobre \mathbf{b}_j , multiplicado por la longitud de \mathbf{b}_j , o sea:

$$y_{ij} = \mathbf{a}_i' \mathbf{b}_j = \|\text{Proy}(\mathbf{a}_i / \mathbf{b}_j)\| \|\mathbf{b}_j\|$$

3- BILOT DE REGRESION NO LINEAL

3.1.- AJUSTE DE BILOT DE REGRESION NO LINEAL

Consideremos ahora el caso, cuando los elementos de la matriz de datos \mathbf{Y} tienen cualquier distribución de la familia Exponencial.

Supongamos que, los valores esperados de las y_{ij} denotados por $\mu_{ij} = E(y_{ij})$, están relacionados con predictores η_{ij} a través de “funciones link” diferentes de la identidad ($g(\mu_{ij}) \neq \mathbf{I}$), o sea:

$$\eta_{ij} = E(y_{ij}) = g(\mu_{ij}) = \mathbf{a}_i' \mathbf{b}_j$$

Matricialmente la expresión general del modelo lineal será:

$$g(\mu) = \eta = \mathbf{A} \mathbf{B}' \quad (3.13)$$

A los Biplot ajustados a través de modelos de este tipo los denominaremos “*Biplot de Regresión no Lineal*”.

BRADU & GABRIEL (1978) y BLAZQUEZ (1998) demuestran que dotando de estructuras particulares a los marcadores **A** y **B**, se pueden obtener cualquiera de los Modelos Bilineales Generalizados (aditivos y/o con términos multiplicativos) cuya interacción se puede representar a través de un Biplot.

La estimación de las matrices de parámetros **A** y **B** se puede hacer de varias formas:

- A través de Regresiones Generalizadas Alternadas (BLAZQUEZ, 1998) realizando estimaciones separadas para cada columna de **Y** bajo el supuesto de independencia de los individuos, y que los parámetros de cada una de las variables son independientes, o sea, que la respuesta de los individuos a cada una de las variables es independiente.
- A través de Regresión Bilineal Generalizada (GABRIEL, 1998) estimando conjuntamente todas las columnas de la matriz **Y**, para lo cual se concatenan las columnas de dicha matriz, por lo que no es necesaria la suposición de parámetros independientes para cada una de las variables.
- A través de Regresión Bilineal Generalizada, estimando conjuntamente y en forma simultánea todas las filas y columnas de la matriz **Y**. Este método alternativo de estimación conjunta de las matrices de parámetros **A** y **B**, constituye uno de los aportes de ésta investigación (ver mayores detalles en el capítulo V).

3.2.- GEOMETRIA DE LOS BIPLLOT DE REGRESION NO LINEAL

Para interpretar la representación gráfica de los “Biplots de Regresión no Lineal” se requiere conocer la geometría del ajuste de los mismos, la cual depende de la “función link” utilizada para tal fin.

Así tenemos por ejemplo, que en el caso del ajuste de “Biplot de Probabilidad” a matrices \mathbf{Y} para datos de presencia-ausencia (los cuales se utilizan para predecir la probabilidad p de ocurrencia de un determinado suceso), tenemos que expresar la probabilidad como una combinación lineal ($\eta = g(p) = \mathbf{AB}'$) lo cual es inconsistente con las leyes de la probabilidad, ya que la probabilidad varía entre (0,1) y η lo hace entre $(-\infty, +\infty)$.

Ese problema se puede resolver usando alguna “función link” que haga la conversión de la escala de probabilidad (0,1) a la escala $(-\infty, +\infty)$, como por ejemplo:

1. La función logística o Logit: $g_1(p) = \log [p / (1 - p)]$ característica de los modelos logísticos cuya interpretación se hace en términos del log de odds-ratio.
2. La función inversa normal o Probit: $g_2(p) = \Phi^{-1}(p)$.
3. La función complementaria log-log: $g_3(p) = \log [-\log (1 - p)]$

Esas tres “funciones link”, se caracterizan por ser funciones continuas y crecientes en el intervalo (0,1), las cuales tienen como patrón común curvas de

respuestas sigmoidales (no lineales) similares, tal como se observa en la figura 3.5:

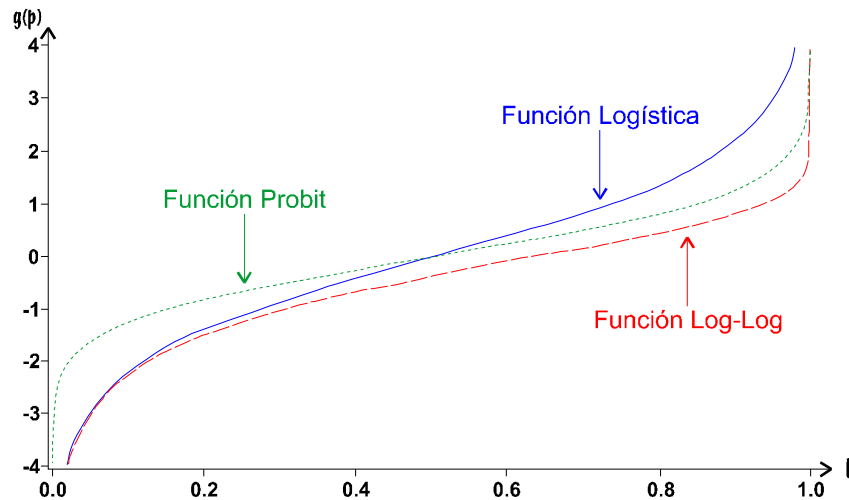


Figura 3.5. Gráfico comparativo entre distintas Funciones Link

Del análisis de la Figura 3.5 se puede concluir que:

“La geometría de un Biplot de Probabilidad ($\eta = g[E(\mathbf{Y})= p] = \mathbf{AB}'$) para datos de presencia-ausencia, será similar para cualquiera de esas tres “funciones link”, razón por la cual conociendo la geometría de uno de ellos, se puede generalizar a los otros”. La diferencia estriba solo en la estimación de los parámetros del Biplot” (véase para mayores detalles el apartado 2.1.2 del capítulo V).

Consideremos el caso de los Biplots de Probabilidad ajustados a través del siguiente modelo logístico (VICENTE-VILLARDON & GALINDO, 1998; VICENTE-VILLARDON, GALINDO & CARDENAS, 2000):

$$p = e^{\eta} / (1 + e^{\eta})$$

Tal ajuste genera una superficie de respuesta logística \mathcal{H} . Su geometría se analiza en un gráfico bi-dimensional complementándolo con un eje de predicción, tal como se presenta en la figura 3.6.

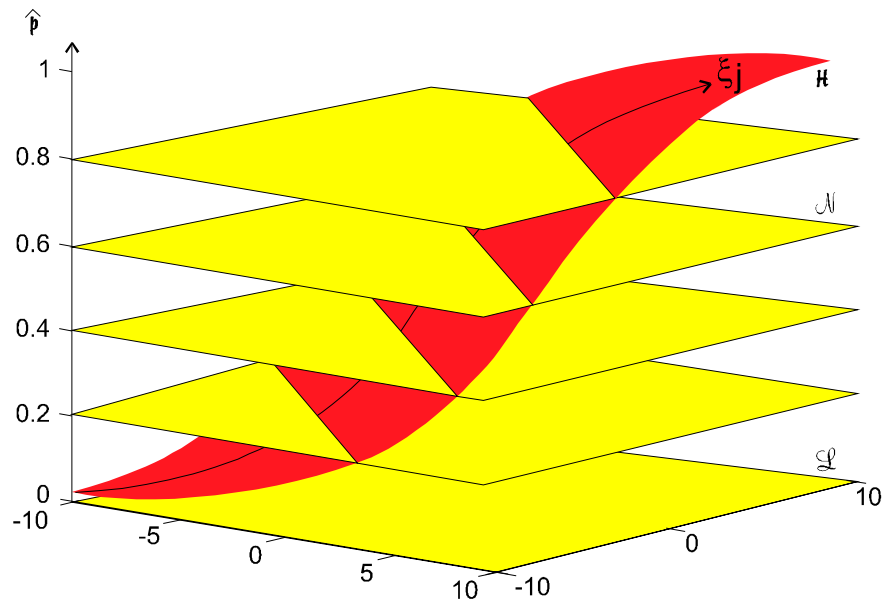


Figura 3.6. Superficie de respuesta logística en el ajuste de Biplots de Probabilidad

Al igual que en el ajuste de los Biplot de Regresión Lineal, el conjunto de puntos en la superficie de respuesta sigmoideal \mathcal{H} que predicen un valor específico de la probabilidad, son rectas paralelas para las distintas probabilidades.

Esas rectas, se producen por la intersección de los planos normales \mathcal{N} al eje de predicción de las probabilidades, planos que pasan por los puntos correspondientes a las predicciones deseadas.

Como se ve en la figura 3.7, el conjunto de puntos de \mathcal{L} que predicen valores concretos de las probabilidades, se encuentran sobre rectas paralelas. La dirección normal a todas esas rectas denominada β_j , y es la proyección de la curva sigmoideal ξ_j sobre \mathcal{L} y es el correspondiente eje Biplot de predicción lineal β_j .

La dirección de ese eje β_j , es la determinada por el vector de parámetros estimados (b_{j1}, b_{j2}) , de donde el producto escalar entre marcadores fila y columna $(a_{ik} \text{ } b_{jk})$ aproxima, salvo un factor de escala, la probabilidad esperada P_{ij} .

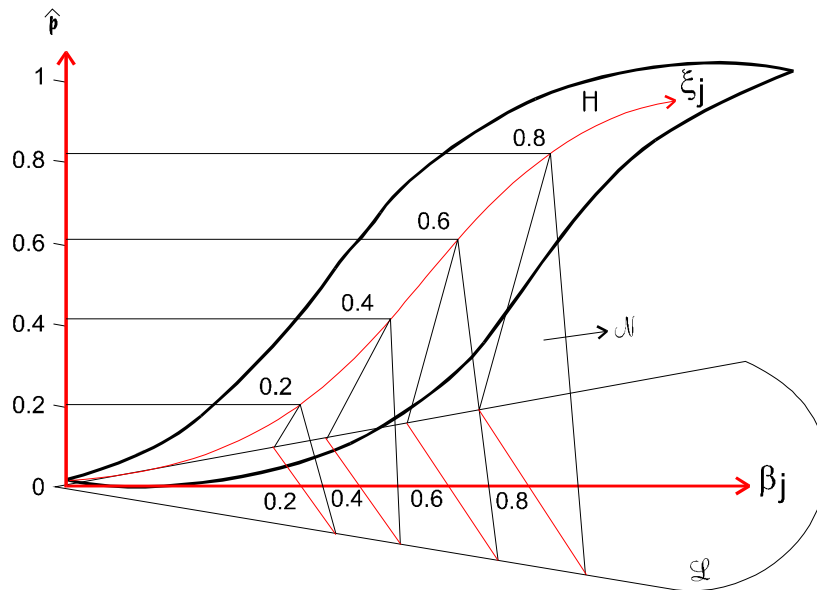


Figura 3.7. Geometría de los Biplots de Probabilidad Logística

De lo antes expuesto, se concluye que la predicción en los “Biplot Logísticos de Probabilidad” se realiza en la misma forma que en los “Biplot de Regresión Lineal”, a pesar de que la respuesta ajustada no sea lineal sino sigmoideal, ya que ***la proyección de la curva de respuesta no lineal sobre un subespacio en baja dimensión es siempre lineal***, sin embargo la escala de predicción en el eje Biplot no se encuentra igualmente espaciada (VICENTE-VILLARDON & GALINDO, 1998).

En el caso del ajuste de otros Biplots de Probabilidad, a través de modelos Probit o log-log, utilizando respectivamente las “funciones link” probit y log-log, también generan curvas de respuesta sigmoideales no lineales al igual que los Biplots Logísticos de Probabilidad (como se puede ver en la figura 3.5).

Por esa razón, ***todos los “Biplots de Probabilidad” tienen la misma geometría*** y por lo tanto las proyecciones de sus respectivas curvas de respuesta sobre el subespacio de mejor ajuste, ***generan ejes Biplot de predicción lineal***, al igual que en el caso de los “Biplots de Regresión Lineal”.

Considerando el ajuste de ***cualquier otro “Biplot de Regresión no Lineal”***, utilizando “funciones link” como la logarítmica, raíz cuadrada e inversa, aún cuando las curvas de respuesta que generan esos ajustes sean no lineales, las proyecciones de dichas curvas sobre el subespacio de mejor ajuste serán siempre lineales como en el caso de los “Biplots de Probabilidad”, ***obteniendo siempre ejes Biplot de predicción lineal***.

Una vez analizada la geometría de los “Biplots de Regresión no Lineal” la interpretación se hace en la forma usual, o sea, en términos de producto escalar de un marcador fila por un marcador columna, y de proyección de marcadores fila sobre marcadores columna y viceversa.

La relación entre las diferentes variables puede verse según el ángulo que forman, de donde la predicción se hará para ambas variables en el mismo sentido (ángulo agudo), en sentido contrario (ángulo obtuso), o en forma independiente (ángulo recto).

En el caso de los “Biplots de Probabilidad”, por ejemplo, el producto interno entre marcadores fila y columna aproxima la probabilidad esperada de la variable y_j , la cual se encuentra representada por el marcador b_j sobre los ejes de la representación Biplot. Los parámetros estimados para cada una de las variables determinan las direcciones de los ejes Biplot. La proyección de cada uno de los puntos fila sobre el segmento que representa a cada una de las variables, permite obtener la probabilidad estimada de presencia de la variable para cada individuo.

Para facilitar la interpretación de las proyecciones, se puede incluir una escala en los vectores que representan a las variables o indicar los puntos que predicen el valor 0.5, ya que ello permite la predicción directa de presencia-ausencia de los diferentes caracteres en estudio.

La bondad de ajuste para cada variable puede utilizarse como indicador de la calidad de representación de las variables, de donde las variables mal

representadas tendrán segmentos más largos que las bien representadas, lo que contrasta con la interpretación habitual de los Biplots donde sucede lo contrario.

Por ejemplo, en la figura 3.8, donde se aproxima una matriz (4x3) las probabilidades estimadas se obtienen proyectando cada uno de los 4 puntos (individuos) sobre el segmento que representa a cada una de las 3 variables, si la proyección cae en el centro del segmento corresponde a la predicción 0.5.

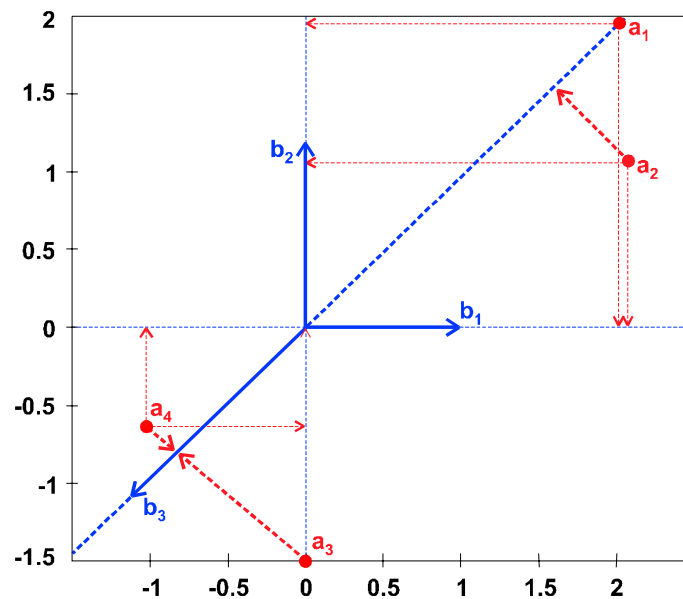


Figura 3.8 Representación del Biplot de Probabilidad Logístico

Si la predicción es menor de 0.5 es indicador de ausencia de ese carácter variable, y si es mayor a 0.5 de presencia de ese carácter.

Así tenemos, de las proyecciones en la figura 3.8, presencia para los individuos 1, 2 y 3 del carácter 1, para los individuos 1 y 2 del carácter 2 y para los individuos 3 y 4 del carácter 3. En todos los otros casos hay indicios de ausencia.

En el caso de otros “Biplots de Regresión no Lineal”, como por ejemplo los que utilizan en el ajuste la “función link” logarítmica, de mucha aplicación en el análisis de independencia en tablas de 2 ó más vías, la representación Biplot del logaritmo de las frecuencias ($\log f_{ij} \cong \mathbf{a}'_i \mathbf{b}_j$) se puede interpretar en términos de independencia completa, múltiple o parcial, o en términos de dependencia.

El diagnóstico se puede utilizar para la diagnosis de modelos subyacentes en los datos (GABRIEL, GALINDO & VICENTE-VILLARDON, 1998). Así tenemos por ejemplo, que si sobre la representación Biplot los marcadores fila \mathbf{a}_i y columna \mathbf{b}_j , se sitúan sobre líneas perpendiculares, entonces se puede interpretar como que las clasificaciones de las filas son independientes de las de las columnas.

CAPITULO IV

BIPLOT DE REGRESION CON INFORMACION EXTERNA

BILOT DE REGRESION CON INFORMACION EXTERNA

1.- INTRODUCCION

En este capítulo, se aborda el objetivo principal de este trabajo de investigación: la descripción de una matriz de datos \mathbf{Y} de orden $(n \times p)$, individuos \times variables, a través de su representación Biplot en un espacio bi-dimensional, pero restringiendo los ejes para que sean combinaciones lineales de variables externas contenidas en las matrices \mathbf{X} $(n \times m)$ y \mathbf{Z} $(p \times s)$.

La información que aportan las variables externas, puede usarse informalmente para añadir información subjetivamente al análisis de los resultados, lo que puede conducir a interpretaciones más objetivas.

La matriz \mathbf{X} puede ser un vector unitario, una matriz de variables dummy o una matriz de variables continuas (por ejemplo variables demográficas) que caractericen a los individuos de la matriz de datos \mathbf{Y} .

Recíprocamente, la matriz \mathbf{Z} también puede ser un vector unitario, una matriz de diseño u otra matriz de variables explicativas (por ejemplo, variables medidas sobre los mismos individuos en otro contexto diferente, pero relacionado al de las variables en la matriz \mathbf{Y}), tal que capturen las relaciones

entre las columnas de \mathbf{Y} . Si no se utilizan matrices de información externa, entonces simplemente \mathbf{X} y \mathbf{Z} se suponen matrices identidad.

Se pretende con ésta investigación proponer una técnica en el sentido del Análisis Canónico de Correspondencias (TER BRAAK, 1986), en el cual se relaciona una comunidad biológica contenida en una matriz de abundancias \mathbf{Y} de orden $(n \times p)$, “ n lugares y p especies”, con una matriz \mathbf{X} de orden $(n \times m)$, “ n lugares y m variables ambientales”.

El Análisis Canónico de Correspondencias es una técnica de Análisis Directo del Gradiente, que resulta de una generalización del Análisis de Correspondencias, en el que los ejes en la representación se restringen tal que las coordenadas de los lugares sean combinaciones lineales de las variables ambientales \mathbf{X} , o sea el análisis equivale a realizar una regresión utilizando el modelo $\mathbf{Y} = \mathbf{C}\mathbf{D}' = (\mathbf{X}\mathbf{A})\mathbf{D}'$, donde para valores fijados de \mathbf{C} y por lo tanto de \mathbf{A} , se estima \mathbf{D} por mínimos cuadrados alternados. En la misma forma, fijando los valores de \mathbf{D} se estima \mathbf{A} . Finalmente las coordenadas de las filas de \mathbf{C} se calculan como $\mathbf{C} = \mathbf{X}\mathbf{A}$.

Imponiendo a posteriori las restricciones de ortogonalidad, entonces las coordenadas contenidas en la matriz estimada \mathbf{D} son las medias ponderadas de las contenidas en la matriz \mathbf{C} , mientras que las contenidas en \mathbf{C} son el resultado de calcular primero una media ponderada \mathbf{D} y hacer después una regresión utilizando las columnas de \mathbf{X} .

TER BRAAK utiliza para sus representaciones las medias ponderadas C como coordenadas para el Análisis Canónico de Correspondencias.

Los resultados se presentan en forma de un diagrama de dispersión, donde los lugares y las especies se representan por puntos y las variables ambientales por vectores, pudiéndose proyectar por ende los puntos especies sobre cada vector, para obtener una ordenación de los mismos respecto a la variable ambiental que representa ese vector. Cuando las variables ambientales tienen vectores sobre la representación más largos, entonces ello implica mayor correlación con la variación de las especies.

El gráfico Biplot obtenido en un Análisis Canónico de Correspondencias, permite visualizar patrones de variación en la composición de la comunidad biológica, los cuales pueden explicarse a través de las variables ambientales permitiendo a su vez la ordenación de las especies de acuerdo a las mismas. En la interpretación de la ordenación de las especies se utilizan los coeficientes canónicos y las correlaciones estructurales del Análisis de Correlación Canónica (HOTTELING, 1936). El Biplot resultante se puede complementar con los coeficientes de regresión, tal que nos permita estimar la influencia de cada variable ambiental sobre la composición de las especies.

TER BRAAK (1986) demostró que el Análisis Canónico de Correspondencias, es equivalente a un GH-Biplot (GABRIEL, 1971) de la matriz de abundancia promedio en el lugar i -ésimo, ponderada por el valor de la variable ambiental en el mismo lugar.

Así pues, el Análisis Canónico de Correspondencias es equivalente a un GH-Biplot Generalizado de la matriz de medias ponderadas de la abundancia de las especies, con respecto a las variables ambientales.

Es posible utilizar un procedimiento similar para cualquier otro tipo de Biplot (JK y HJ). Una aplicación utilizando el HJ-Biplot puede verse en FERNANDEZ-GOMEZ (1995).

En el caso que nos concierne del ajuste de Biplot de Regresión con Información Externa a través del modelo: $\mathbf{Y} = (\mathbf{X}\mathbf{A})(\mathbf{B}'\mathbf{Z}') = \mathbf{C}\mathbf{D}'$, dicha especificación es similar a la utilizada en el Análisis Canónico de Correspondencias $\mathbf{Y} = \mathbf{X}\mathbf{A}\mathbf{D}' = \mathbf{C}\mathbf{D}'$, tan solo que ahora tanto las coordenadas para las filas como para las columnas, son combinaciones lineales de las variables externas \mathbf{X} y \mathbf{Z} respectivamente.

De allí que en los Biplot de Regresión con Información Externa el ajuste resulta en una extensión del Análisis Canónico de Correlaciones (TER BRAAK, 1986), y por ende del Análisis de Correspondencias cuando se imponen en el mismo restricciones lineales (BOCKENHOLT & TAKANE, 1994), de donde del ajuste se podrá realizar cualquier representación Biplot, las cuales permitirán visualizar patrones de variación en la composición de los individuos que pueden explicarse a través de las variables externas \mathbf{X} , y en la estructura de las variables que pueden explicarse a través de las variables externas \mathbf{Z} .

En ésta investigación, se persigue con la inclusión a priori de las matrices de información externa \mathbf{X} y \mathbf{Z} , una mejor explicación del comportamiento de los individuos y las relaciones entre las variables, contenidas en la matriz de datos \mathbf{Y} , además de explicar la interrelación entre los conjuntos de variables contenidas en las matrices \mathbf{Y} , \mathbf{X} y \mathbf{Z} .

Para tal fin, es necesario seleccionar los ejes Biplot de manera tal, que sean combinaciones lineales de las variables externas. De esa manera, en el gráfico Biplot se pueden visualizar patrones de variación en la caracterización de los individuos, los cuales pueden explicarse a través de las variables externas, permitiendo además ordenar los individuos de acuerdo a las mismas.

La representación Biplot buscada, se consigue aplicando en la aproximación del “Biplot de Regresión con Información Externa”, una generalización al método iterativo de estimación usado en los Modelos Lineales Generalizados, e interpretando la ordenación de los individuos de acuerdo a los coeficientes de regresión del Análisis de Regresión Múltiple y a las correlaciones estructurales del Análisis de Correlación Canónica.

Para el desarrollo de este objetivo, hemos considerado y generalizado, adecuando para este fin, las ideas expuestas por DENIS (1991), CHOULAKIAN (1996), BLAZQUEZ (1998) y GABRIEL (1998), autores que dirigen sus contribuciones al ajuste de los Biplots con el objetivo fundamental de utilizarlos en la diagnosis de Modelos Bilineales subyacentes en los datos.

DENIS (1991) y CHOULAKIAN (1996), proponen una generalización a los modelos propuestos por FALGUEROLLES (1995, 1996) y VAN EEUWIJK (1995).

El prototipo del Modelo Bilineal propuesto por DENIS en 1991 y CHOULAKIAN en 1996, es del tipo aditivo-multiplicativo con información externa introducida a través de matrices de datos complementarios \mathbf{X} y \mathbf{Z} .

El modelo está constituido por una parte lineal y una parte bilineal de rango reducido. La información externa está directamente incorporada en la parte lineal del modelo y representa la porción de los datos que está explicada por la misma. Mientras que, la parte bilineal modela la interacción entre filas y columnas, explicando el resto de la estructura oculta en los datos.

La especificación general de ese modelo, es la siguiente:

$$g(\mu) = \mathbf{X}\mu\mathbf{Z}' + \alpha\mathbf{Z}' + \mathbf{X}\beta' + \gamma\theta\delta' \quad (4.1)$$

El primer término, es la parte de la información contenida en los datos explicada tanto por \mathbf{X} como por \mathbf{Z} , mientras que, el segundo y tercer término representan la parte de la información que es explicada por \mathbf{X} (para las filas) o por \mathbf{Z} (para las columnas).

Respecto al cuarto y último término, representan la parte de la interacción en los datos que no puede ser explicada ni por \mathbf{X} ni por \mathbf{Z} (CHOULAKIAN,

1996), o que de alguna forma puede ser interpretada a través de la información externa contenida en las matrices \mathbf{X} y \mathbf{Z} (DENIS, 1991).

Otra diferencia entre esos dos autores, está en la “función link” utilizada en el ajuste y por ende en el método de estimación de los parámetros.

DENIS (1991) usa como “función link” la identidad y supone que la matriz \mathbf{Y} es de rango completo, por lo que el método de estimación que utiliza es el de mínimos cuadrados alternados, equivalente en este caso, a la realización simultánea de un Análisis de Regresión y un Análisis de Componentes Principales ya que el término bilineal está en rango reducido.

TAKANE & SHIBAYAMA (1991) proponen un modelo aditivo sin el término bilineal, suponiendo distribución Normal y “función link” la identidad. El procedimiento de estimación que utilizan es similar al de MANDEL (1961), estimando sucesivamente por mínimos cuadrados cada uno de los términos en el modelo, para finalmente realizar un Análisis de Componentes Principales de cada término para la interpretación de los mismos.

Por su parte, CHOULAKIAN (1996) utiliza como “función link” la función logarítmica y también la identidad, y aplica en la estimación de los parámetros una generalización al método de Newton-Raphson utilizado en los Modelos Lineales Generalizados.

BLAZQUEZ en 1998 extiende las ideas de DENIS (1991) para ajustar los Biplot con Información Externa con fines descriptivos, utilizando en el ajuste Modelos Bilineales Clásicos del tipo multiplicativo. Para la estimación asume que la “función link” es la identidad y que la matriz \mathbf{Y} no es de rango completo, finalmente realiza la factorización Biplot tal que las coordenadas para las filas y columnas sean combinación lineal de las variables externas.

A tal fin reespecifica el modelo (4.1) como:

$$E(\mathbf{Y}) = \mathbf{X} \Theta \mathbf{Z}' = \mathbf{X} (\mathbf{A}\mathbf{B}') \mathbf{Z}' = (\mathbf{X}\mathbf{A}) (\mathbf{B}'\mathbf{Z}') \quad (4.2)$$

Donde la matriz $\Theta = \mathbf{A}\mathbf{B}'$

Estima los coeficientes de las combinaciones lineales contenidos en las matrices \mathbf{A} y \mathbf{B} , a través de un procedimiento iterativo de mínimos cuadrados ponderados alternados, como una generalización a la regresión con rango reducido (DAVIES & TSO, 1982). A posteriori, para representar el Biplot realiza la Descomposición en Valores Singulares de la matriz Θ (TER BRAAK & LOOMAN, 1991).

Por otra parte, GABRIEL en 1998 propone un modelo para el ajuste de Biplots a través de Modelos Bilineales Multiplicativos (clásicos y generalizados), en cuya especificación se considera la dependencia de

columnas y filas de la matriz \mathbf{Y} (no de rango completo) sobre matrices de covariables \mathbf{X} y \mathbf{Z} respectivamente (ambas de rango completo), o sea:

$$g(\mu) = \mathbf{X} \mathbf{A} \mathbf{B}' \mathbf{Z}' \quad (4.3)$$

La estimación la hace a través del método de “Regresión Bilineal Generalizada” (anteriormente expuesto en el apartado 5.3 del capítulo II), el cual es una generalización del método "criss-cross" para datos Gaussianos (GABRIEL & ZAMIR, 1979).

La adecuación de las propuestas de esos autores, a los objetivos de ésta investigación, persigue fundamentalmente el ajuste de los “*Biplots de Regresión con Información Externa para fines descriptivos*”, tal que se obtengan ejes (combinación lineal de las variables externas) que permitan la ordenación de los individuos, en función de las variables contenidas en las matrices \mathbf{X} y \mathbf{Z} .

2- BILOT DE REGRESION CON INFORMACION EXTERNA

Consideraremos la aproximación de la matriz de datos \mathbf{Y} de orden $(n \times p)$ y de rango r , por una de bajo rango ($q < r$) a través de su factorización Biplot, tal que las coordenadas para las filas y columnas en la factorización sean

combinaciones lineales de variables externas contenidas en matrices \mathbf{X} ($n \times m$) y \mathbf{Z} ($p \times s$) respectivamente.

Para el ajuste de los “Biplots de Regresión con Información Externa”, se propone en esta investigación utilizar uno de los siguientes modelos:

$$g[E(\mathbf{Y})=\mu] = \mathbf{X} \Theta \mathbf{Z}' = \mathbf{X} (\mathbf{A}\mathbf{B}') \mathbf{Z}' \quad (4.4)$$

$$g[E(\mathbf{Y}')=\mu'] = \mathbf{Z} \Theta' \mathbf{X}' = \mathbf{Z} (\mathbf{B}\mathbf{A}') \mathbf{X}' \quad (4.5)$$

siendo:

$g(\mu)$ una forma linearizada de la “función link”

$\Theta = \mathbf{A}\mathbf{B}'$ la matriz de parámetros (desconocida excepto por su rango r)

$\mathbf{C} = \mathbf{X}\mathbf{A}$ las coordenadas para las filas en el Biplot

$\mathbf{D} = \mathbf{Z}\mathbf{B}$ las coordenadas para las columnas.

Esos modelos especifican la matriz de esperanzas $g(\mu)$ solo por su rango r , y la dependencia de sus filas y columnas de matrices de variables externas \mathbf{X} y \mathbf{Z} respectivamente. Por ejemplo, \mathbf{X} puede tratarse de un vector o matriz unitaria, una matriz de variables dummy o variables continuas con información adicional sobre los sujetos, mientras que \mathbf{Z} puede tratarse de un vector o matriz unitaria, una matriz de diseño o de variables medidas en otro momento o contexto.

Si el rango de la matriz de parámetros Θ es ($q < r$), entonces se puede factorizar como $\Theta = \mathbf{AB}'$, siendo \mathbf{A} de orden ($m \times q$) y \mathbf{B} de orden ($s \times q$).

Para simplificar la notación, en lugar de $g[E(\mathbf{Y})=\mu]$ nos referiremos en lo sucesivo a \mathbf{Y} , entendiéndose que corresponde a una forma linearizada de la “función link”.

Considerando la aproximación Biplot en la forma del ajuste del modelo (4.4), y suponiendo que las coordenadas para las filas de \mathbf{A} han sido prefijadas, entonces el subespacio de mejor ajuste a las columnas de \mathbf{Y} se obtiene al ajustar ese modelo.

De la misma manera, considerando el modelo bilineal (4.5) entonces prefijando los valores de \mathbf{B} , podemos obtener los valores de \mathbf{A} por regresiones generalizadas alternadas, por lo que ajustar ese modelo es equivalente a realizar el ajuste de las filas de \mathbf{Y} sobre el subespacio q -dimensional de representación óptima.

Para la estimación por regresiones alternadas de los coeficientes de las combinaciones lineales contenidas en las matrices \mathbf{A} y \mathbf{B} , se realizan algunas operaciones convenientemente en los modelos (4.4) y (4.5), así tenemos que:

La estimación de la matriz \mathbf{B} se puede hacer post-multiplicando el modelo (4.4) por la matriz $[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}]$, de donde:

$$\mathbf{YZ}(\mathbf{Z}'\mathbf{Z})^{-1} = (\mathbf{XA})\mathbf{B}'$$

Fijando \mathbf{A} , se obtiene:

$$\hat{\mathbf{B}}' = (\mathbf{A}'\mathbf{X}'\mathbf{X}\mathbf{A})^{-1} \mathbf{A}'\mathbf{X}'\mathbf{Y}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$$

Para estimar la matriz \mathbf{A} , se post-multiplican ambos miembros en (4.5) por la matriz $[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]$, reespecificando el modelo como:

$$\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{Z}\mathbf{B}) \mathbf{A}'$$

Fijando \mathbf{B} , obtenemos:

$$\hat{\mathbf{A}}' = (\mathbf{B}'\mathbf{Z}'\mathbf{Z}\mathbf{B})^{-1} \mathbf{B}'\mathbf{Z}'\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Las estimaciones en esos modelos, para los coeficientes de las combinaciones lineales contenidos en las matrices \mathbf{A} y \mathbf{B} respectivamente, se pueden obtener también a través del método de “Regresiones Bilineales Generalizadas”, de las columnas o filas de la matriz \mathbf{Y} respectivamente, sobre un subespacio q -dimensional de representación óptima.

Para ello los modelos se reespecifican concatenando las columnas de la matriz \mathbf{Y} y las filas de las matrices \mathbf{A} y \mathbf{B} en el sentido de GABRIEL (GABRIEL, 1998). En efecto:

1) La concatenación de columnas y filas de matrices permite reescribir una matriz \mathbf{Y} de orden $(n \times p)$ como un vector columna $(n \times 1)$, denotándolo $\langle \mathbf{Y} \rangle$

cuando se han concatenado las columnas de la matriz \mathbf{Y} , o $\langle \mathbf{Y}' \rangle$ cuando se han concatenado las filas de la matriz \mathbf{Y} , o sea:

$$\mathbf{Y} = \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{np} \end{bmatrix} \Rightarrow \langle \mathbf{Y}' \rangle = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n} \\ \vdots \\ y_{n1} \\ \vdots \\ y_{np} \end{bmatrix}; \langle \mathbf{Y} \rangle = \begin{bmatrix} y_{11} \\ \vdots \\ y_{n1} \\ \vdots \\ y_{1n} \\ \vdots \\ y_{np} \end{bmatrix}$$

2) Teniendo en cuenta que el producto Kroenecker entre dos matrices se define como:

$$\mathbf{P}_{(mp \times nq)} = [\mathbf{A}_{(m \times n)} \otimes \mathbf{B}_{(p \times q)}] = [a_{ij} \mathbf{B}] = [\mathbf{B} a_{ij}]$$

En el modelo 4.4 tendremos que:

$$[\mathbf{Z}_{(p \times s)} \otimes (\mathbf{XA})_{(n \times q)}] = [z_{ij} (\mathbf{XA})]$$

Cuando $\mathbf{Z} = \mathbf{I}$ entonces:

$$[\mathbf{I}_p \otimes (\mathbf{XA})] = [I_{ij} (\mathbf{XA})]$$

Y si además $\mathbf{X} = \mathbf{I}_n$ entonces:

$$[\mathbf{I}_p \otimes (\mathbf{I}_n \mathbf{A})] = [\text{diag} (\mathbf{A})]$$

3) La concatenación de columnas del producto de varias matrices, permite expresar como un vector columna dicho producto utilizando la siguiente relación:

$$\langle \mathbf{A} \mathbf{B} \mathbf{C} \rangle = [\mathbf{C}' \otimes \mathbf{A}] \langle \mathbf{B} \rangle$$

En el modelo 4.4 tendremos que:

$$g \langle \mu \rangle = \langle (\mathbf{X}\mathbf{A}) \mathbf{B}' \mathbf{Z}' \rangle = [\mathbf{Z} \otimes (\mathbf{X}\mathbf{A})] \langle \mathbf{B}' \rangle$$

Cuando \mathbf{X} e \mathbf{Y} son matrices identidad entonces:

$$g \langle \mu \rangle = \langle \mathbf{A} \mathbf{B}' \mathbf{I} \rangle = [\mathbf{I} \otimes \mathbf{A}] \langle \mathbf{B}' \rangle = [\text{diag}(\mathbf{A})] \langle \mathbf{B}' \rangle$$

Ahora bien, considerando las definiciones y relaciones anteriores, entonces los modelos (4.4) y (4.5) se pueden reespecificar de la siguiente manera:

Concatenando las columnas de la matriz \mathbf{Y} y las filas de la matriz \mathbf{A} como:

$$g \langle \mu \rangle = [\mathbf{Z} \otimes (\mathbf{X}\mathbf{A})] \langle \mathbf{B}' \rangle \tag{4.6}$$

y concatenando las filas de \mathbf{Y} y \mathbf{A} como:

$$g \langle \mu' \rangle = [\mathbf{X} \otimes (\mathbf{Z}\mathbf{B})] \langle \mathbf{A}' \rangle \tag{4.7}$$

Cuando en el modelo (4.6) \mathbf{X} y \mathbf{Z} son matrices identidad, entonces:

$$g \langle \mu \rangle = [\mathbf{I}_p \otimes (\mathbf{I}_n \mathbf{A})] \langle \mathbf{B}' \rangle = [\mathbf{I}_p \otimes \mathbf{A}] \langle \mathbf{B}' \rangle = [\text{diag}(\mathbf{A})] \langle \mathbf{B}' \rangle \quad (4.6a)$$

y en el caso del modelo (4.7) entonces:

$$g \langle \mu' \rangle = [\mathbf{I}_n \otimes (\mathbf{I}_p \mathbf{B})] \langle \mathbf{A}' \rangle = [\mathbf{I}_n \otimes \mathbf{B}] \langle \mathbf{A}' \rangle = [\text{diag}(\mathbf{B})] \langle \mathbf{A}' \rangle \quad (4.7a)$$

Una vez estimadas las matrices de parámetros \mathbf{A} y \mathbf{B} , el problema del ajuste de los “Biplots de Regresión con Información Externa” a través del modelo (4.6) se reduce a cómo factorizar:

$$g(\hat{\mu}) = \mathbf{X} \hat{\mathbf{A}} \hat{\mathbf{B}}' \mathbf{Z}'$$

Recíprocamente, se puede proceder en forma similar en el caso del modelo (4.7).

El problema de la estimación simultánea de las matrices \mathbf{A} y \mathbf{B} en esos modelos, o de la estimación en forma separada de \mathbf{A} y \mathbf{B} utilizando “Regresión Bilineal Generalizada”, se aborda con detalle en el capítulo V.

En este capítulo, se contempla sólo el problema de la factorización. Una solución es usar la Descomposición en Valores Singulares de $\Theta = \mathbf{A} \mathbf{B}'$.

En el caso del modelo (4.4) se tiene que:

$$g(\hat{\mu}) = \mathbf{X}(\mathbf{UDV}')\mathbf{Z}'$$

y como en los métodos Biplot, se puede definir la factorización de la siguiente manera:

$$\hat{\mathbf{A}} = \mathbf{XUD}^{\gamma} \qquad \hat{\mathbf{B}} = \mathbf{ZVD}^{1-\gamma} \qquad (4.8)$$

de donde, dependiendo del valor dado a γ (0, 1, $\frac{1}{2}$), se pueden obtener respectivamente el RMP, CMP, SQRT Biplot (GABRIEL, 1971, 1981), y además seleccionando los marcadores tal que cada matriz contenga la matriz diagonal \mathbf{D} se puede obtener el RCMP-Biplot (GALINDO, 1985, 1986).

En los “Biplot de Regresión”, además de poder representar cada tipo de Biplot, las matrices de marcadores \mathbf{A} y \mathbf{B} en la factorización pueden entenderse como marcadores de distintas representaciones Biplot para distintas regresiones.

De allí que, la interpretación de los “Biplot de Regresión” dependerá no solo del “tipo de Biplot”, sino también del “coeficiente de regresión” representado y de la “función link” utilizada en el ajuste (por ser \mathbf{Y} una forma linearizada de la función link).

Así pues, los gráficos de los “Biplot de Regresión con Información Externa” implican diferentes representaciones para distintos coeficientes de regresión,

cuyas interpretaciones dependen del tipo de Biplot y de la “función link” utilizada en el ajuste, permitiéndonos detectar aproximadamente la interacción con variables externas no implícitamente consideradas en la matriz de datos, aunque no nos permita separar en el gráfico su influencia directa, lo que se podría hacer utilizando modelos vectoriales (BLAZQUEZ, 1998).

Las diferentes representaciones de los “Biplot con Información Externa”, así como sus correspondientes interpretaciones, son el objeto de los dos apartados siguientes.

3.- REPRESENTACIONES DE LOS BIPLLOT DE REGRESION CON INFORMACION EXTERNA

Consideraremos los “Biplot de Regresión con Información Externa” ajustados a través de los modelos (4.4) y (4.5):

$$g(\mu) = \mathbf{X} \Theta \mathbf{Z}' = \mathbf{X} (\mathbf{A}\mathbf{B}') \mathbf{Z}' \quad (4.4)$$

$$g(\mu') = \mathbf{Z} \Theta' \mathbf{X}' = \mathbf{Z} (\mathbf{B}\mathbf{A}') \mathbf{X}' \quad (4.5)$$

Reordenando las matrices convenientemente en esas especificaciones, se pueden obtener varias representaciones Biplot, considerando diferentes regresiones cuyos coeficientes de regresión se pueden representar a través de diferentes Biplot de Regresión.

3.1.- REPRESENTACION BILOT DE LOS COEFICIENTES DE REGRESION ($AD' \cong R_{yx}$)

Reordenando el modelo (4.4), de la siguiente manera:

$$g(\mu) = \mathbf{X} (\mathbf{A} \mathbf{B}' \mathbf{Z}') = \mathbf{X} (\mathbf{A} \mathbf{D}') \quad (4.9)$$

donde

$$\mathbf{D} = \mathbf{Z} \mathbf{B}$$

Suponiendo las matrices \mathbf{Y} , \mathbf{X} y \mathbf{Z} estandarizadas (centradas y reducidas) tal que las variables tengan media cero y varianza unitaria, entonces (4.9) es equivalente a realizar una regresión generalizada de las columnas de la matriz de datos \mathbf{Y} sobre las columnas de la matriz de variables externas \mathbf{X} para estimar los coeficientes de regresión (\mathbf{AD}'), o sea:

$$(\hat{\mathbf{A}}\mathbf{D}') = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

donde

$$\mathbf{X}'\mathbf{X} = \mathbf{R}_{xx}$$

$$\mathbf{Y}'\mathbf{Y} = \mathbf{R}_{yy}$$

$$\mathbf{X}'\mathbf{Y} = \mathbf{R}_{yx}$$

entonces

$$\hat{\mathbf{A}}\mathbf{D}' = \mathbf{R}_{xx}^{-1}\mathbf{R}_{yx}$$

De donde la matriz de parámetros (\mathbf{AD}') es equivalente a la matriz de correlaciones (\mathbf{R}_{yx}) ponderada por la inversa de la matriz de correlaciones (\mathbf{R}_{xx}).

Una vez estimada la matriz de coeficientes de regresión (\mathbf{AD}'), se puede realizar la representación Biplot de los mismos, previa factorización en marcadores fila y columna (Véase apartado 4.1).

Luego, la obtención de la representación Biplot de los coeficientes de regresión ($\mathbf{AD}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$) estimados a través de la regresión $\mathbf{Y} = \mathbf{X}(\mathbf{AD}')$ es equivalente a la obtención del Biplot de \mathbf{Y} .

De la representación Biplot de la matriz de correlación ponderada (\mathbf{AD}'), se pueden obtener indicadores de la correlación entre el conjunto de variables \mathbf{Y} y el conjunto de variables externas \mathbf{X} .

En la figura 4.1 se presenta la representación Biplot de (\mathbf{AD}'), para una matriz \mathbf{Y} de orden (5x2). En dicha representación, proyectando los marcadores fila \mathbf{a}_i , sobre los marcadores columna \mathbf{d}_j , se puede obtener una ordenación de las variables externas \mathbf{X} de acuerdo a su relación con \mathbf{Y} .

Por ejemplo, para la segunda columna de la matriz de datos \mathbf{Y} , el orden en forma creciente de los valores que toman las filas \mathbf{a}_i , se puede aproximar a través de la proyección $\text{Proy}(\mathbf{a}_i / \mathbf{d}_2)$, de donde $\mathbf{a}_1 < \mathbf{a}_3 < \mathbf{a}_4 < \mathbf{a}_5 < \mathbf{a}_2$.

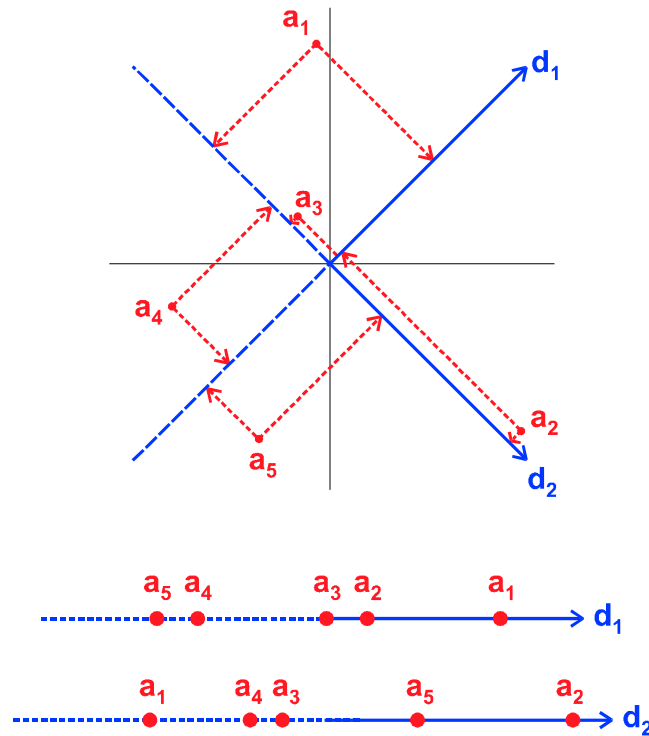


Figura 4.1 Representación del Biplot de Regresión Lineal para una matriz (5x2) ajustado a través del modelo $g(\mu)=X(A B'Z')=X(AD')$

Luego, de las proyecciones $\text{Proy}(\mathbf{a}_i / \mathbf{d}_2)$ se puede concluir que: la mayor correlación de la variable \mathbf{Y}_2 es con la variable externa \mathbf{X}_2 , y la menor con la variable \mathbf{X}_1 .

Por otra parte, la matriz de correlaciones ponderada (\mathbf{AD}') se puede interpretar como el coseno cuadrado del ángulo formado entre \mathbf{Y} y su proyección (predicción), que es igual a:

$$\hat{\mathbf{Y}} = \mathbf{X}(\hat{\mathbf{A}}\mathbf{D}')$$

Si el ángulo de proyección es agudo, los coeficientes de regresión ponderados contenidos en la matriz \mathbf{AD}' serán positivos, luego habrán indicios de correlación positiva entre \mathbf{Y} y \mathbf{X} . Si el ángulo es obtuso, entonces los coeficientes de regresión serán negativos existiendo correlación negativa entre \mathbf{X} e \mathbf{Y} . Finalmente si el ángulo de proyección es recto, los coeficientes serán nulos lo que indicará la no correlación entre \mathbf{X} e \mathbf{Y} .

Restringiendo las columnas de (\mathbf{AD}') para que sean ortonormales, entonces la proyección de los puntos fila de \mathbf{X} sobre (\mathbf{AD}') , aproximarán los incrementos porcentuales en las variables \mathbf{Y} para incrementos unitarios en las variables \mathbf{X} correspondientes, o sea, en este caso su interpretación es similar a los coeficientes de regresión del Análisis de Regresión Múltiple.

3.2.- REPRESENTACION BIPLLOT DE LOS COEFICIENTES DE REGRESION ($\mathbf{BC}' \cong R_{y'z}$)

Reescribiendo el modelo (4.5) de la siguiente manera:

$$g(\mu') = \mathbf{Z}(\mathbf{B A}' \mathbf{X}') = \mathbf{Z}(\mathbf{B C}') \quad (4.10)$$

donde $\mathbf{C} = \mathbf{X A}$

De esa forma (4.10) equivale a realizar la regresión generalizada de las filas de la matriz de datos \mathbf{Y} sobre las columnas de la matriz de variables externas \mathbf{Z} (suponiendo estandarizadas las variables en los conjuntos \mathbf{Y} y \mathbf{Z}), de donde la matriz de coeficientes de regresión (\mathbf{BC}') será equivalente a la matriz de correlaciones $(\mathbf{R}_{y'z})$ ponderada por la inversa de la matriz de correlaciones (\mathbf{R}_{zz}) :

$$(\hat{\mathbf{B}}\hat{\mathbf{C}}') = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}' = \mathbf{R}_{zz}^{-1}\mathbf{R}_{y'z}$$

Realizando la factorización Biplot se puede representar el Biplot de (\mathbf{BC}') , el cual es equivalente al Biplot de \mathbf{Y}' en (4.10). De donde, a través de la representación del “Biplot de Regresión” de los coeficientes ponderados (\mathbf{BC}') se pueden obtener indicadores de las relaciones entre los perfiles de los individuos en la matriz \mathbf{Y} y las variables externas contenidas en la matriz \mathbf{Z} .

En la representación Biplot de (\mathbf{BC}') , proyectando los vectores estimados \mathbf{b}_i de la matriz \mathbf{B} , sobre los vectores de la matriz \mathbf{C} ($\mathbf{C} = \mathbf{XA}$), se puede obtener una ordenación de los perfiles de los individuos, de acuerdo a su relación con las variables externas \mathbf{Z} , en la misma forma que en la figura 4.1.

Restringiendo las columnas de (\mathbf{BC}') para que sean ortonormales, entonces la proyección de los puntos fila de \mathbf{Z} sobre (\mathbf{BC}') aproximarán en forma similar a los coeficientes de regresión, los incrementos porcentuales en los perfiles de los individuos para incrementos unitarios en las variables \mathbf{Z} .

3.3.- REPRESENTACION BIPLLOT DE LOS COEFICIENTES DE REGRESION ($AB' \cong R_{(y'z),x}$)

Si en el modelo (4.4) se post-multiplican ambos miembros por la matriz $[Z(Z'Z)^{-1}]$, entonces tenemos que:

$$[g(\mu) Z (Z'Z)^{-1}] = X (AB') \tag{4.11}$$

El primer miembro es comparable a la transpuesta de la matriz de correlaciones entre los perfiles de los individuos en Y y las variables en Z (siendo Y una forma linearizada de la función link) la cual denotaremos $R'_{y'z} = [g(\mu)Z(Z'Z)^{-1}]$, entonces el modelo se puede escribir de la siguiente manera:

$$R'_{y'z} = X (AB')$$

En este caso, en el análisis se relacionan en forma indirecta los perfiles de los individuos en la matriz Y y las variables en las matrices X y Z .

Se puede hacer la representación Biplot de la matriz de parámetros (AB') en forma equivalente al Biplot de $[g(\mu) Z (Z'Z)^{-1}]$ en (4.11).

La estimación de (AB') resulta equivalente a la realización de la regresión generalizada de las filas de la matriz $R'_{y'z}$ sobre las columnas de la matriz X , de donde:

$$(\hat{A}\hat{B}') = (X'X)^{-1} X'YZ(Z'Z)^{-1} = (X'X)^{-1} (X'R'_{y'z}) = R_{xx}^{-1} R_{(y'z),x}$$

Al igual que en las anteriores representaciones Biplot, se pueden obtener a través de esos coeficientes de regresión:

- Indicadores de las correlaciones entre la matriz $\mathbf{R}_{y'z}$ con el conjunto de variables externas \mathbf{X} .
- Ordenación de las correlaciones ponderadas $\mathbf{R}'_{y'z}$, de acuerdo a su relación con las variables externas \mathbf{X} , en la misma forma que se hizo en la figura 4.1 para los coeficientes de regresión \mathbf{AD}' .
- Aproximación de los incrementos porcentuales en $\mathbf{R}'_{y'z}$, para incrementos unitarios en las variables \mathbf{X} .

3.4.- REPRESENTACION BILOT DE LOS COEFICIENTES DE REGRESION ($\mathbf{BA}' \cong \mathbf{R}_{(yx),z}$)

Si en el modelo (4.5) se post-multiplican ambos miembros por la matriz $[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]$, entonces:

$$[g(\hat{\mu}) \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = \mathbf{Z}(\mathbf{BA}') \quad (4.12)$$

El primer miembro en esa expresión lo designamos $\mathbf{R}'_{yx} = [g(\hat{\mu})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]$.

Estimando los coeficientes (\mathbf{BA}') en (4.12) se puede realizar la representación Biplot de los mismos en forma equivalente al Biplot de \mathbf{R}'_{yx} en (4.12).

La estimación de los coeficientes de regresión, será equivalente a la realización de la regresión generalizada de las filas de la matriz \mathbf{R}'_{yx} sobre las columnas de la matriz \mathbf{Z} , de donde:

$$(\hat{\mathbf{B}}\hat{\mathbf{A}}') = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{R}'_{yx}) = \mathbf{R}_{zz}^{-1}\mathbf{R}_{(yx),z}$$

A través de ese Biplot se pueden obtener:

- Indicadores de las relaciones entre las matrices \mathbf{R}'_{yx} y \mathbf{Z} , ponderadas por las correlaciones inversas de las variables en \mathbf{Z} .
- Ordenamiento de los indicadores contenidos en la matriz \mathbf{R}'_{yx} de acuerdo a su relación con las variables \mathbf{Z} , a través de las proyecciones de los marcadores \mathbf{a}_i sobre los marcadores \mathbf{b}_j .
- Aproximación de los incrementos porcentuales en \mathbf{R}'_{yx} para incrementos unitarios en \mathbf{Z} .

4.- INTERPRETACION DE LOS BIPLLOT DE REGRESION CON INFORMACION EXTERNA

Consideremos el ajuste de los “Biplot de Regresión con Información Externa” a través de cualquiera de los modelos especificados en (4.4) y (4.5), donde se consideran tres conjuntos de variables en el análisis (\mathbf{Y} , \mathbf{X} y \mathbf{Z}).

Hemos visto que, las matrices de marcadores **A** y **B** en los “Biplot de Regresión con Información Externa” se pueden entender como marcadores de los Biplot de los coeficientes de regresión estimados a través del ajuste de alguno de los modelos (4.9), (4.10), (4.11) y (4.12).

El ajuste de cualquiera de esos “Biplot de Regresión” es comparable a los Análisis de Regresión Múltiple y de Correlación Canónica (HOTTELING, 1936).

El objetivo fundamental en ambos análisis que es relacionar dos conjuntos de variables se logra: en el Análisis de Regresión Múltiple a través de la aproximación mínimo cuadrática a bajo rango de la “matriz de coeficientes de regresión”, y en el Análisis de Correlación Canónica a través de la aproximación mínimo cuadrática a bajo rango de la “matriz de correlaciones” entre los dos conjuntos de variables en estudio.

La aproximación mínimo cuadrática de esas matrices, se puede hallar a través de la Descomposición en Valores Singulares Generalizada (HOUSEHOLDER & YOUNG, 1938; GREENACRE, 1984, 1988), pudiéndose hacer en consecuencia una representación Biplot, en cuya interpretación se relacionan los marcadores filas y columnas con las “correlaciones estructurales” del Análisis de Correlación Canónica, y con los “coeficientes de regresión” del Análisis de Regresión Múltiple.

La diferencia fundamental entre ambos análisis estriba en que, mientras en el Análisis de Correlación Canónica se dan papeles simétricos a los dos

conjuntos de variables en estudio, en el Análisis de Regresión Múltiple se les da papeles asimétricos.

En el caso del ajuste de los “Biplots de Regresión con Información Externa”, existen tres conjuntos de variables en el análisis (Y , X , Z), sin embargo, hemos visto antes en el apartado 3, que los modelos utilizados en el ajuste se pueden reescribir reordenando las matrices de forma tal, de centrar el interés en relacionar sólo dos conjuntos de variables, dando papeles simétricos a ambos.

Luego, dicho análisis es similar a un Análisis de Correlación Canónica (HOTTELING, 1936; MARDIA, KENT & BIBBY, 1979), pudiéndose hallar las variables canónicas y las correlaciones estructurales, elementos ambos a considerar en la interpretación de las diferentes representaciones, de acuerdo a los diferentes tipos de Biplot.

En resumen, las coordenadas de los “Biplot de Regresión con Información Externa”, tienen interpretaciones relacionadas con los coeficientes de regresión del Análisis de Regresión Múltiple, y con los coeficientes de correlación “inter-set” e “intra-set” del Análisis de Correlación Canónica (TER BRAAK, 1990).

En el siguiente apartado, se detalla la interpretación de las coordenadas de los “Biplot de Regresión con Información Externa”, sólo para la representación de los coeficientes de regresión estimados a través del modelo (4.9), pudiéndose generalizar las conclusiones que se obtienen a todas las otras representaciones.

4.1.- BILOT DE LOS COEFICIENTES DE REGRESION ($\mathbf{AD}' \cong \mathbf{R}_{yx}$)

Consideremos el modelo (4.9):

$$g(\mu) = \mathbf{X} \Theta \mathbf{Z}' = \mathbf{X} (\mathbf{AD}')$$

donde $\mathbf{D} = \mathbf{ZB}$

El interés de este análisis, está en relacionar sólo los dos conjuntos de variables $\mathbf{Y}_{(n \times p)}$ y $\mathbf{X}_{(n \times m)}$, los cuales representan respectivamente y en forma simétrica “p” y “m” variables medidas sobre el mismo conjunto de “n” individuos.

Vimos en el anterior apartado 3.1, que la estimación a través de regresión generalizada de los parámetros contenidos en la matriz (\mathbf{AD}') , es igual a:

$$(\hat{\mathbf{AD}}) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} = \mathbf{R}_{xx}^{-1} \mathbf{R}_{yx}$$

En el ajuste de los “Biplot de Regresión con Información Externa” a través del modelo (4.9), se les da papeles simétricos a los dos conjuntos de variables (\mathbf{Y} y \mathbf{X}), por lo que dicho análisis es similar a un Análisis de Correlación Canónica entre \mathbf{Y} y \mathbf{X} (HOTTELING, 1936; MARDIA, KENT & BIBBY, 1979).

El objetivo en el Análisis de Correlación Canónica, es hallar las variables canónicas (combinación lineal de las originales) para cada conjunto de

variables, tal que su correlación canónica sea máxima. Ese objetivo se logra a través de la aproximación mínimo cuadrática a bajo rango de la matriz de correlaciones ponderadas (\mathbf{AD}') entre los conjuntos de variables \mathbf{Y} y \mathbf{X} .

La aproximación mínimo cuadrática de (\mathbf{AD}') es de la forma (\mathbf{AD}') = \mathbf{AB}' , de donde el óptimo se obtiene minimizando la siguiente expresión:

$$\left\| \mathbf{R}_{YY}^{-1/2} (\mathbf{AD}') \mathbf{R}_{XX}^{-1/2} - (\mathbf{R}_{YY}^{-1/2} \mathbf{A})(\mathbf{R}_{XX}^{-1/2} \mathbf{B})' \right\|^2$$

Por el Teorema de ECKART & YOUNG (1936), ese mínimo también se puede obtener a través de la Descomposición en Valores Singulares Ordinaria de la siguiente matriz:

$$[\mathbf{R}_{yy}^{-1/2} (\mathbf{AD}') \mathbf{R}_{xx}^{-1/2}] = \mathbf{U} \mathbf{D}_\alpha \mathbf{V}' \quad (4.13)$$

lo que equivale a realizar la Descomposición en Valores Singulares Generalizados de (\mathbf{AD}'), referida a las métricas \mathbf{R}_{yy}^{-1} y \mathbf{R}_{xx}^{-1} en los espacios columna y fila respectivamente (GREENACRE, 1984), siendo \mathbf{D}_α la matriz diagonal que contiene los valores singulares, \mathbf{U} la matriz de vectores singulares por la izquierda, \mathbf{V} la matriz de vectores singulares por la derecha ($\mathbf{U}'\mathbf{U}=\mathbf{V}'\mathbf{V}=\mathbf{I}$), o sea:

$$(\mathbf{AD}') = \mathbf{ND}_\alpha \mathbf{M}'$$

siendo

$$\mathbf{N}' \mathbf{R}_{yy}^{-1} \mathbf{N} = \mathbf{M}' \mathbf{R}_{xx}^{-1} \mathbf{M} = \mathbf{I}$$

$$\mathbf{N} = \mathbf{R}_{yy}^{1/2} \mathbf{U} \Rightarrow \mathbf{R}_{yy}^{-1/2} \mathbf{N} = \mathbf{U}$$

$$\mathbf{M} = \mathbf{R}_{xx}^{1/2} \mathbf{V} \Rightarrow \mathbf{R}_{xx}^{-1/2} \mathbf{M} = \mathbf{V}$$

(4.14)

siendo las columnas de las matrices \mathbf{N} y \mathbf{M} los vectores singulares generalizados por la izquierda y la derecha respectivamente y $\mathbf{D}_\alpha = [\text{diag}(\alpha_k^2)]$ la matriz diagonal que contiene los valores singulares generalizados.

Las variables canónicas serán iguales a:

$$\mathbf{Y}^* = \mathbf{Y} \mathbf{R}_{yy}^{-1/2} \mathbf{U} = \mathbf{Y} \mathbf{R}_{yy}^{-1} \mathbf{N}$$

(4.15)

$$\mathbf{X}^* = \mathbf{X} \mathbf{R}_{xx}^{-1/2} \mathbf{V} = \mathbf{X} \mathbf{R}_{xx}^{-1} \mathbf{M}$$

La correlación canónica entre esas variables coincide con los elementos α_k de la diagonal de la matriz \mathbf{D}_α .

La interpretación de las variables canónicas \mathbf{Y}^* y \mathbf{X}^* es similar a la de las componentes principales, por lo que es útil considerar las contribuciones (correlaciones) de cada variable original a las canónicas (contribución del elemento al factor en los Biplot) y viceversa, o sea la contribución de cada

variable canónica a las variables originales (contribuciones del factor al elemento).

Teniendo en cuenta la descomposición de la matriz (\mathbf{AD}') en (4.14), entonces se pueden hallar las coordenadas para las filas y columnas en la factorización Biplot a través de:

$$(\mathbf{AD}') = \mathbf{NDM}' = \mathbf{ND}_\alpha^\gamma \mathbf{D}_\alpha^{1-\gamma} \mathbf{M}' \quad (4.16)$$

Llamando respectivamente \mathbf{A}^* y \mathbf{B}^* , a las coordenadas estimadas para los marcadores fila y columna en esa factorización Biplot, y sustituyendo las matrices \mathbf{N} y \mathbf{M} por sus respectivas expresiones en (4.14), entonces:

$$\mathbf{A}^* = \mathbf{N D}_\alpha^\gamma = (\mathbf{R}_{yy}^{1/2} \mathbf{U D}_\alpha^\gamma) \quad (4.17)$$

$$\mathbf{B}^* = \mathbf{M D}_\alpha^{1-\gamma} = (\mathbf{R}_{xx}^{1/2} \mathbf{V D}_\alpha^{1-\gamma})$$

De allí que, las filas de las matrices de marcadores \mathbf{A}^* y \mathbf{B}^* , contienen las coordenadas de los dos grupos de variables respecto a las variables canónicas.

Las ponderaciones canónicas son $\mathbf{R}_{yy}^{-1/2} \mathbf{A}^* = \mathbf{U D}_\alpha$ y $\mathbf{R}_{xx}^{-1/2} \mathbf{B}^* = \mathbf{V D}_\alpha^{1-\gamma}$, las cuales se interpretan en forma similar a los coeficientes de regresión estandarizados en regresión múltiple, y su estimación presenta los mismos problemas en el caso de existir multicolinealidad, siendo recomendable en ese caso abstenerse de su interpretación.

Por otra parte, se observa que la estructura de las matrices \mathbf{A}^* y \mathbf{B}^* , es equivalente a las matrices de “correlaciones estructurales” definidas en el Análisis de Correlaciones Canónicas, ya que:

Las *correlaciones Intraset*, se definen como las correlaciones entre variables observadas y canónicas del mismo conjunto, o sea:

$$\text{Corr}(\mathbf{Y}, \mathbf{Y}^*) = (\mathbf{S}_{yy}^{-1/2} \mathbf{S}_{yy^*} \mathbf{S}_{y^*y^*}^{-1/2})$$

Pero como $\mathbf{S}_{yy} = \mathbf{S}_{y^*y^*} = \mathbf{I}$ entonces:

$$\text{Corr}(\mathbf{Y}, \mathbf{Y}^*) = \mathbf{S}_{yy^*} = \mathbf{Y}'\mathbf{Y}^*$$

Sustituyendo allí, la expresión de \mathbf{Y}^* en (4.15), tenemos que:

$$\text{Corr}(\mathbf{Y}, \mathbf{Y}^*) = \mathbf{Y}'\mathbf{Y}^* = \mathbf{Y}'\mathbf{Y} \mathbf{R}_{yy}^{-1} \mathbf{N} = \mathbf{R}_{yy} \mathbf{R}_{yy}^{-1} \mathbf{N} = \mathbf{N} = \mathbf{A}^* \quad (\text{cuando } \gamma=0)$$

Procediendo de igual forma se demuestra que:

$$\text{Corr}(\mathbf{X}, \mathbf{X}^*) = \mathbf{X}'\mathbf{X}^* = \mathbf{X}'\mathbf{X} \mathbf{R}_{xx}^{-1} \mathbf{M} = \mathbf{R}_{xx} \mathbf{R}_{xx}^{-1} \mathbf{M} = \mathbf{M} = \mathbf{B}^* \quad (\text{cuando } \gamma=1)$$

De donde:

$$\text{Corr}(\mathbf{y}_j, \mathbf{y}^*_k) = \mathbf{n}_k \tag{4.18}$$

$$\text{Corr}(\mathbf{x}_j, \mathbf{x}^*_k) = \mathbf{m}_k$$

El cuadrado del coeficiente de “correlación intraset” para una variable específica de un conjunto dado, se interpreta como la proporción de la varianza de tal variable, explicada por una variable canónica del mismo conjunto.

Las *correlaciones Interset*, se definen como las correlaciones entre variables observadas y canónicas de distintos conjuntos, o sea:

$$\begin{aligned} \text{Corr}(\mathbf{Y}, \mathbf{X}^*) &= \mathbf{Y}'\mathbf{X}^* = \mathbf{Y}'\mathbf{X} \mathbf{R}_{xx}^{-1/2}\mathbf{V} = \mathbf{R}_{yx} \mathbf{R}_{xx}^{-1/2}\mathbf{V} = \mathbf{R}_{yx} \mathbf{R}_{xx}^{-1/2}\mathbf{R}_{xx}^{-1/2}\mathbf{M} \\ \text{Corr}(\mathbf{Y}, \mathbf{X}^*) &= \mathbf{R}_{yx} \mathbf{R}_{xx}^{-1}\mathbf{M} = \mathbf{N} \mathbf{D}_\alpha \mathbf{M}'\mathbf{R}_{xx}^{-1} \mathbf{M} = \mathbf{N} \mathbf{D}_\alpha = \mathbf{A}^* \quad (\text{cuando } \gamma=1) \end{aligned}$$

Recíprocamente:

$$\text{Corr}(\mathbf{X}, \mathbf{Y}^*) = \mathbf{X}'\mathbf{Y}^* = \mathbf{R}_{yx} \mathbf{R}_{yy}^{-1/2}\mathbf{U} = \mathbf{M}\mathbf{D}_\alpha = \mathbf{B}^* \quad (\text{cuando } \gamma=0)$$

De donde:

$$\text{Corr}(\mathbf{y}_j, \mathbf{x}^*_k) = \mathbf{R}_{yx} \mathbf{m}_k = r_k \mathbf{n}_k \tag{4.19}$$

$$\text{Corr}(\mathbf{x}_j, \mathbf{y}^*_k) = \mathbf{R}_{yx} \mathbf{n}_k = r_k \mathbf{m}_k$$

Siendo

$$r_k = (\alpha_k^2)^{1/2}$$

$$\mathbf{D}_\alpha = [\text{diag}(\alpha_k^2)]$$

La interpretación de las “correlaciones interset” es similar al caso de los coeficientes de correlación múltiple, el cuadrado de ellas nos da la proporción de la varianza que se puede predecir mediante la variable canónica del otro

dominio, cantidad que es similar a la Redundancia en el Análisis de Redundancia (VAN DER WOLLEMBERG, 1977).

El Análisis de Redundancia, también llamado Regresión con Rango Reducido (IZENMAN, 1975), o Análisis de Componentes Principales sobre Variables Instrumentales (RAO, 1964), tiene el mismo propósito fundamental que en Análisis de Regresión Multivariante, imponiendo la restricción de que la matriz de parámetros a estimar tenga rango reducido, y estimando los valores de \mathbf{Y} a partir de las variables canónicas.

Se observa de (4.18) y (4.19) que las correlaciones “interset” e “intraset” solo difieren por un factor de escala, que es precisamente el coeficiente de correlación canónica contenido en la matriz \mathbf{D} .

Dependiendo de la factorización realizada, y en cual matriz de marcadores (\mathbf{A}^* o \mathbf{B}^*), aparezca la matriz diagonal \mathbf{D}_α , se tratará de las “correlaciones interset” (cuando la matriz de marcadores contengan a \mathbf{D}_α), o de las “correlaciones intraset” (si no se incluye a \mathbf{D}_α).

Dando valores a γ en la factorización Biplot (4.17), se pueden realizar representaciones para cada tipo de Biplot, interpretando los resultados de acuerdo a la estructura de las matrices de marcadores fila \mathbf{A}^* y marcadores columna \mathbf{B}^* en cada uno.

Así tenemos:

Para $\gamma = 0$ se obtiene la representación **CMP-Biplot** de la matriz (\mathbf{AD}') la cual está relacionada al Análisis de Correlación Canónica, en el sentido que:

- Los marcadores fila, aproximan la matriz de “*correlaciones intraset*” entre las variables contenidas en la matriz de datos \mathbf{Y} y las variables canónicas \mathbf{Y}^* . El cuadrado de estas correlaciones, es aproximadamente igual a la proporción de la varianza de \mathbf{Y} , que se puede predecir por la variable canónica \mathbf{Y}^* .
- Los marcadores columna, aproximan la matriz de “*correlaciones interset*” entre las variables canónicas \mathbf{Y}^* y las variables del conjunto de variables \mathbf{X} . El cuadrado de estas correlaciones es aproximadamente igual a la proporción de la varianza de \mathbf{X} , que se puede predecir a través de las variables canónicas \mathbf{Y}^* .

Cuando $\gamma = 1$ se obtiene la representación **JK-Biplot** de la matriz (\mathbf{AD}') , donde:

- Los marcadores fila que conforman la matriz \mathbf{A}^* , aproximan la matriz de “*correlaciones interset*” entre variables en la matriz \mathbf{Y} y las variables canónicas \mathbf{X}^* . El cuadrado de dichas correlaciones, aproxima la proporción de varianza de \mathbf{Y} que se puede predecir a través de las variables canónicas \mathbf{X}^* .

- Los marcadores columna que conforman la matriz \mathbf{B}^* , permiten aproximar la matriz de “*correlaciones intraset*” de las variables \mathbf{X} con las variables canónicas \mathbf{X}^* .

Cuando en (4.17) $\gamma = 1/2$, se obtiene la representación correspondiente al **SQRT- Biplot** de la matriz (\mathbf{AD}') . En este tipo de Biplot no es de fundamental importancia la interpretación de los marcadores, ya que el objetivo que se persigue con ésta representación, es generalmente para la diagnosis de modelos.

Finalmente el **RCMP-Biplot (HJ-Biplot)**, se obtiene cuando en (4.17) la factorización Biplot elegida corresponde a:

$$\mathbf{A}^* = (\mathbf{R}_{yy}^{1/2} \mathbf{U} \mathbf{D}_\alpha) \quad \mathbf{B}^* = (\mathbf{R}_{xx}^{1/2} \mathbf{V} \mathbf{D}_\alpha)$$

En este caso, los marcadores contenidos en esas matrices se pueden interpretar de la siguiente manera:

- Los marcadores fila, aproximan la matriz de “*correlaciones interset*” entre variables en el conjunto \mathbf{Y} y las variables canónicas \mathbf{X}^* . Su cuadrado aproxima, la proporción de varianza de \mathbf{Y} que se puede predecir a través de las variables canónicas \mathbf{X}^* .
- Los marcadores columna, aproximan la matriz de “*correlaciones interset*” entre variables canónicas \mathbf{Y}^* y variables externas en el conjunto \mathbf{X} . Su cuadrado es la proporción de varianza de las variables en \mathbf{X} explicada por las variables canónicas \mathbf{Y}^* .

4.2.- BILOT DE LOS COEFICIENTES DE REGRESION ($\mathbf{BC}' \cong \mathbf{R}_{y'z}$)

Considerando el modelo (4.10):

$$g(\mu') = \mathbf{Z} (\mathbf{B} \mathbf{A}' \mathbf{X}') = \mathbf{Z} (\mathbf{B} \mathbf{C}') \quad (4.19)$$

donde $\mathbf{C} = \mathbf{XA}$

La estimación de los coeficientes de regresión (\mathbf{BC}') es igual a:

$$(\hat{\mathbf{B}}\hat{\mathbf{C}}') = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}' = \mathbf{R}_{zz}^{-1}\mathbf{R}_{y'z}$$

La aproximación Biplot de (\mathbf{BC}') a través de la Descomposición en Valores Singulares Generalizada implica que:

$$(\mathbf{BC}') = \mathbf{ND}_\alpha \mathbf{M}' = \mathbf{ND}_\alpha^\gamma \mathbf{D}_\alpha^{1-\gamma} \mathbf{M}$$

siendo

$$\mathbf{N} = \mathbf{R}_{y'y'}^{1/2} \mathbf{U}$$

$$\mathbf{M} = \mathbf{R}_{zz}^{1/2} \mathbf{V}$$

De allí que, las matrices de los marcadores fila y columna en la factorización Biplot serán:

$$\mathbf{A}^* = (\mathbf{R}_{y'y'}^{1/2} \mathbf{U} \mathbf{D}_\alpha^\gamma) \quad \mathbf{B}^* = (\mathbf{R}_{zz}^{1/2} \mathbf{V} \mathbf{D}_\alpha^{1-\gamma}) \quad (4.20)$$

Esas matrices tienen interpretación de acuerdo a las correlaciones estructurales en el Análisis de Correlaciones Canónicas, como ya hemos visto.

Dependiendo del valor de γ en la factorización Biplot (4.20), se pueden realizar representaciones para cada tipo de Biplot, interpretando los resultados de acuerdo a la estructura de las matrices de marcadores fila \mathbf{A}^* y marcadores columna \mathbf{B}^* en cada uno de ellos, de la siguiente manera:

Para $\gamma = 0$ se obtiene la representación **CMP-Biplot (GH-Biplot)** de la matriz (\mathbf{BC}') , en la cual:

- Los marcadores fila, aproximan la matriz de “*correlaciones intraset*” entre el perfil de los individuos en la matriz de datos \mathbf{Y} y las variables canónicas \mathbf{Y}'^* .
- Los marcadores columna, aproximan la matriz de “*correlaciones interset*” entre las variables \mathbf{Y}'^* y las variables del conjunto \mathbf{Z} .

Para $\gamma = 1$ se obtiene la representación **RMP- Biplot (JK-Biplot)** de la matriz (\mathbf{BC}') , en la cual:

- Los marcadores fila que conforman la matriz \mathbf{A}^* , aproximan la matriz de “*correlaciones interset*” entre los perfiles de los individuos en \mathbf{Y} y las variables canónicas \mathbf{Z}^* .

- Los marcadores columna que conforman la matriz \mathbf{B}^* , permiten aproximar la matriz de “*correlaciones intraset*” de las variables \mathbf{Z} y \mathbf{Z}^* .

Cuando $\gamma = 1/2$ en (4.20) se obtiene la representación **SQRT- Biplot**. La interpretación de los marcadores fila y columna, como se ha dicho anteriormente, no tiene mucha importancia, ya que el uso de este tipo de representación es fundamentalmente en la diagnosis de modelos.

Cuando en (4.20) la factorización Biplot elegida corresponde a:

$$\mathbf{A}^* = (\mathbf{R}_{y'y'}^{1/2} \mathbf{U} \mathbf{D}_\alpha) \quad \mathbf{B}^* = (\mathbf{R}_{zz}^{1/2} \mathbf{V} \mathbf{D}_\alpha)$$

Entonces se obtiene la representación del **RCMP-Biplot (HJ-Biplot)**. En tipo de Biplot tenemos que:

- Los marcadores fila, aproximan la matriz de “*correlaciones interset*” entre el perfil de los individuos en \mathbf{Y} y las variables \mathbf{Z}^* .
- Los marcadores columna, aproximan la matriz de “*correlaciones interset*” entre variables canónicas \mathbf{Y}'^* y las variables externas en el conjunto \mathbf{Z} .

4.3.- BILOT DE LOS COEFICIENTES DE REGRESION (\mathbf{AB}')

Considerando el modelo (4.11):

$$\mathbf{R}'_{y'z} = \mathbf{X} (\mathbf{AB}')$$

Se puede hacer la representación Biplot de la matriz de parámetros (\mathbf{AB}'), cuya estimación se obtiene realizando la regresión de las columnas de $\mathbf{R}'_{y'z}$ sobre las columnas de la matriz \mathbf{X} , de donde:

$$(\hat{\mathbf{A}}\hat{\mathbf{B}}') = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{R}'_{y'z}) = \mathbf{R}_{xx}^{-1}\mathbf{R}_{(y'z),x}$$

La aproximación Biplot de (\mathbf{AB}'), a través de la Descomposición en Valores Singulares Generalizada implica que:

$$\mathbf{AB}' = \mathbf{N}\mathbf{D}_\alpha^\gamma\mathbf{D}_\alpha^{1-\gamma}\mathbf{M}'$$

siendo

$$\mathbf{N} = \mathbf{R}_{y'z,y'z}^{1/2}\mathbf{U}$$

$$\mathbf{M} = \mathbf{R}_{xx}^{1/2}\mathbf{V}$$

De donde, las matrices de marcadores para filas \mathbf{A}^* y para columna \mathbf{B}^* en la factorización Biplot serán:

$$\mathbf{A}^* = (\mathbf{R}_{y'z,y'z}^{1/2}\mathbf{U}\mathbf{D}_\alpha^\gamma) \quad \mathbf{B}^* = (\mathbf{R}_{xx}^{1/2}\mathbf{V}\mathbf{D}_\alpha^{1-\gamma}) \quad (4.21)$$

Esas matrices de marcadores tienen diferentes interpretaciones, de acuerdo al valor de γ escogido para la factorización Biplot ($\gamma = 0, 1, 1/2$), y por consiguiente del tipo de Biplot a representar: el CMP-Biplot, el RMP-Biplot y el SQRT-Biplot, respectivamente. Además el RCMP-Biplot cuando en cada una de las matrices de marcadores aparece la matriz diagonal \mathbf{D}_α .

La interpretación de las matrices \mathbf{A}^* y \mathbf{B}^* se puede hacer en forma general de la siguiente forma:

Cuando en la matriz de marcadores fila \mathbf{A}^* “no aparece” la matriz diagonal \mathbf{D}_α , entonces se puede aproximar a través de esos marcadores la matriz de “correlaciones intraset” entre las variables contenidas en la matriz $\mathbf{R}_{y'z}$ y las variables canónicas $\mathbf{R}^*_{y'z}$.

Si por el contrario en la matriz de marcadores fila \mathbf{A}^* “aparece” la matriz diagonal \mathbf{D}_α , en ese caso los marcadores fila aproximan sobre la representación la matriz de “correlaciones interset” entre las variables en la matriz $\mathbf{R}_{y'z}$ y las variables canónicas \mathbf{X}^* .

Si en la matriz de marcadores columna \mathbf{B}^* “aparece” la matriz diagonal \mathbf{D}_α se puede aproximar a través de esos marcadores la matriz de “correlaciones interset” entre variables canónicas $\mathbf{R}^*_{y'z}$ y las variables externas en el conjunto \mathbf{X} .

Si por el contrario en la matriz de marcadores columna \mathbf{B}^* “no aparece” la matriz diagonal D_α , se puede aproximar a través de ellos la matriz de “correlaciones intraset” entre las variables \mathbf{X} y las variables canónicas \mathbf{X}^* .

4.4.- BIPLLOT DE LOS COEFICIENTES DE REGRESION (\mathbf{BA}')

Considerando el modelo (4.12):

$$[\mathbf{g}(\hat{\mu}) \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}] = \mathbf{Z} (\mathbf{BA}')$$

Entonces realizando la regresión de \mathbf{R}'_{yx} sobre \mathbf{Z} [$\mathbf{R}'_{yx} = \mathbf{g}(\hat{\mu}) \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$], el estimador de la matriz (\mathbf{BA}') será igual a:

$$(\hat{\mathbf{B}}\hat{\mathbf{A}}') = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{Z}'\mathbf{Z})^{-1} (\mathbf{Z}'\mathbf{R}'_{yx}) = \mathbf{R}_{zz}^{-1} \mathbf{R}_{(yx),z}$$

La aproximación Biplot de (\mathbf{BA}'), a través de la Descomposición en Valores Singulares Generalizada de dicha matriz conduce a la siguiente factorización Biplot:

$$\mathbf{A}^* = (\mathbf{R}_{(yx)(yx)})^{1/2} \mathbf{U} \mathbf{D}_\alpha^\gamma \quad \mathbf{B}^* = (\mathbf{R}_{zz})^{1/2} \mathbf{V} \mathbf{D}_\alpha^{1-\gamma} \quad (4.22)$$

Esas matrices de marcadores, tienen diferentes interpretaciones de acuerdo al valor de γ escogido para la factorización Biplot ($\gamma = 0, 1, 1/2$), y por consiguiente del tipo de Biplot a representar: el CMP-Biplot, el RMP-Biplot

y el SQRT-Biplot, respectivamente. Además el RCMP-Biplot cuando en cada una de las matrices de marcadores aparece la matriz diagonal \mathbf{D}_α .

La interpretación de las matrices \mathbf{A}^* y \mathbf{B}^* se puede hacer en forma general de la siguiente forma:

Cuando en la matriz de marcadores fila \mathbf{A}^* “no aparece” la matriz diagonal D_α , entonces se pueden aproximar a través de ellos la matriz de “correlaciones intraset” entre las variables contenidas en la matriz \mathbf{R}_{yx} y las variables canónicas \mathbf{R}^*_{yx}

Si en la matriz de marcadores fila \mathbf{A}^* “aparece” la matriz diagonal D_α , entonces los marcadores fila aproximan la matriz de “correlaciones interset” entre las variables en el conjunto \mathbf{R}_{yx} y las variables canónicas \mathbf{Z}^* .

Cuando en la matriz de marcadores columna \mathbf{B}^* “aparece” la matriz diagonal D_α , entonces los marcadores aproximan la matriz de “correlaciones interset” entre las variables canónicas del conjunto \mathbf{R}_{yx} y las variables externas en el conjunto \mathbf{Z} .

Si en la matriz de marcadores columna \mathbf{B}^* “no aparece” la matriz diagonal D_α , los marcadores respectivos aproximan la matriz de “correlaciones intraset” entre las variables \mathbf{Z} y \mathbf{Z}^* .

En el caso de los “**Biplots de Regresión No Lineal con Información Externa**”, se debe considerar además en la interpretación de los marcadores la “función link” utilizada en el ajuste. Por ejemplo en el ajuste de Biplots Logísticos el producto interno entre marcadores aproxima la probabilidad de éxito, mientras que en los Biplot Logarítmicos aproximan el log de odds-ratio.

Las anteriores representaciones de los “Biplot de Regresión con Información Externa”, con las matrices de marcadores de acuerdo al tipo de Biplot, se resumen en la tabla 4.1

REPRESENTACION BILOT	TIPOS DE BILOT	MARCADORES
Coefficientes de Regresión (AD´)	GH (CMP)	$G = A^* = (R_{yy}^{1/2}U)$ $H = B^* = (R_{xx}^{1/2}VD_{\alpha})$
	JK (RMP)	$J = A^* = (R_{yy}^{1/2}UD_{\alpha})$ $K = B^* = (R_{xx}^{1/2}V)$
	HJ (RCMP)	$J = A^* = (R_{yy}^{1/2}UD_{\alpha})$ $H = B^* = (R_{xx}^{1/2}VD_{\alpha})$
	SQRT	$A^* = (R_{yy}^{1/2}UD_{\alpha}^{1/2})$ $B^* = (R_{xx}^{1/2}VD_{\alpha}^{1/2})$
Coefficientes de Regresión (BC´)	GH (CMP)	$G = A^* = (R_{y'y'}^{1/2}U)$ $H = B^* = (R_{zz}^{1/2}VD_{\alpha})$
	JK (RMP)	$J = A^* = (R_{y'y'}^{1/2}UD_{\alpha})$ $K = B^* = (R_{zz}^{1/2}V)$
	HJ (RCMP)	$J = A^* = (R_{y'y'}^{1/2}UD_{\alpha})$ $H = B^* = (R_{zz}^{1/2}VD_{\alpha})$
	SQRT	$A^* = (R_{y'y'}^{1/2}UD_{\alpha}^{1/2})$ $B^* = (R_{zz}^{1/2}VD_{\alpha}^{1/2})$
Coefficientes de Regresión (AB´)	GH (CMP)	$G = A^* = (R_{y'z,y'z}^{1/2}U)$ $H = B^* = (R_{xx}^{1/2}VD_{\alpha})$
	JK (RMP)	$J = A^* = (R_{y'z,y'z}^{1/2}UD_{\alpha})$ $K = B^* = (R_{xx}^{1/2}V)$
	HJ (RCMP)	$J = A^* = (R_{y'z,y'z}^{1/2}UD_{\alpha})$ $H = B^* = (R_{xx}^{1/2}VD_{\alpha})$
	SQRT	$A^* = (R_{y'z,y'z}^{1/2}UD_{\alpha}^{1/2})$ $B^* = (R_{xx}^{1/2}VD_{\alpha}^{1/2})$
Coefficientes de Regresión (BA´)	GH (CMP)	$G = A^* = (R_{yx,yx}^{1/2}U)$ $H = B^* = (R_{zz}^{1/2}VD_{\alpha})$
	JK (RMP)	$J = A^* = (R_{yx,yx}^{1/2}UD_{\alpha})$ $K = B^* = (R_{zz}^{1/2}V)$
	HJ (RCMP)	$J = A^* = (R_{yx,yx}^{1/2}UD_{\alpha})$ $H = B^* = (R_{zz}^{1/2}VD_{\alpha})$
	SQRT	$A^* = (R_{yx,yx}^{1/2}UD_{\alpha}^{1/2})$ $B^* = (R_{zz}^{1/2}VD_{\alpha}^{1/2})$

Tabla 4.1.- Diferentes Representaciones de los “Biplot de Regresión con Información Externa”

5.- PROPIEDADES DE LOS MARCADORES EN LAS REPRESENTACIONES DE LOS BIPLLOT DE REGRESION CON INFORMACION EXTERNA

En los “Biplots de Regresión”, los marcadores para filas y columnas conservan las propiedades inherentes a cada tipo de Biplot, independientemente de la representación que se haga.

Sabemos que los distintos tipos de Biplot Clásicos (GABRIEL, 1971), no tienen propiedades óptimas, en lo que se refiere a la calidad de representación de las de las filas y columnas de la matriz de datos \mathbf{Y} .

Sin embargo, el RCMP-Biplot (GALINDO, 1985, 1986), por el contrario si tiene propiedades óptimas, en lo que se refiere a la calidad de representación tanto de las filas como de las columnas de la matriz de datos \mathbf{Y} .

Así tenemos, resumiendo las propiedades de acuerdo a cada tipo de Biplot las siguientes:

En el CMP-Biplot (GH-Biplot):

1) Es baja la calidad de representación para las filas, y óptima la calidad de representación para las columnas.

2) En el Biplot de Componentes Principales, la distancia de Mahalanobis entre dos filas de \mathbf{Y} , se aproxima mediante la distancia euclídea entre dos marcadores fila, o sea: $(\mathbf{y}_i - \mathbf{y}_s)' \mathbf{S}^{-1} (\mathbf{y}_i - \mathbf{y}_s) \approx (\mathbf{a}_i - \mathbf{a}_s)' (\mathbf{a}_i - \mathbf{a}_s)$

3) El i -ésimo marcador fila, coincide con las coordenadas del i -ésimo individuo sobre las r primeras componentes principales estandarizadas, o sea: $\mathbf{a}_i = \psi_i = \sqrt{(n-1)} \mathbf{u}_i = \mathbf{y}_i' \mathbf{V}[\sqrt{(n-1)} \mathbf{D}^{-1}]$

4) El producto escalar de las columnas de \mathbf{Y} , coincide con el producto escalar de los marcadores columna, o sea: $\mathbf{y}_j' \mathbf{y}_k \approx \mathbf{b}_j' \mathbf{b}_k$

5) La longitud al cuadrado de los vectores que representan a los marcadores columna, aproximan la covarianza entre las variables correspondientes, de donde: $\mathbf{b}_j' \mathbf{b}_j = s_{jj} = s_j^2$

6) La longitud de los marcadores columna, aproxima la desviación estándar de las variables

7) El coseno del ángulo entre dos marcadores columna, aproxima la correlación entre variables correspondientes: $\text{Cos}(\mathbf{b}_j, \mathbf{b}_k) \approx \text{Corr}(\mathbf{y}_j, \mathbf{y}_k)$

8) Las coordenadas de la matriz de marcadores columna \mathbf{B} en el Biplot de Componentes Principales, son equivalentes a los pesos de las variables a lo largo de los ejes principales.

En el **RMP-Biplot (JK-Biplot)** tenemos que:

1) Es óptima la calidad de representación para las filas, y mala la calidad de representación para las columnas.

2) El producto escalar, con la métrica identidad, de las filas de \mathbf{Y} coincide con el producto escalar entre marcadores filas: $\mathbf{Y}\mathbf{Y}' = \mathbf{A}\mathbf{A}'$

3) La distancia euclídea entre filas de \mathbf{Y} , coincide con la distancia entre marcadores fila, o sea: $(\mathbf{y}_i - \mathbf{y}_s)'(\mathbf{y}_i - \mathbf{y}_s) = (\mathbf{a}_i - \mathbf{a}_s)'(\mathbf{a}_i - \mathbf{a}_s)$

4) Los marcadores fila, coinciden con las coordenadas de las filas en el espacio de las componentes principales: $\boldsymbol{\psi} = \mathbf{Y}\mathbf{V} = \mathbf{A}$

5) Las coordenadas para las columnas, coinciden con las proyecciones de los ejes originales en el espacio de las componentes principales: $\mathbf{I}_p\mathbf{V} = \mathbf{V} = \mathbf{B}$

6) La similitud entre columnas, se aproxima utilizando la inversa de la matriz de dispersión entre individuos

En el **RCMP-Biplot (HJ-Biplot)** tenemos que:

1) Es óptima la calidad de representación tanto para las filas como para las columnas.

- 2) El producto escalar, con la métrica identidad, de las filas de \mathbf{Y} coincide con el producto escalar entre marcadores filas: $\mathbf{Y}\mathbf{Y}' = \mathbf{A}\mathbf{A}'$
- 3) La distancia euclídea entre filas de \mathbf{Y} , coincide con la distancia entre marcadores fila: $(\mathbf{y}_i - \mathbf{y}_s)'(\mathbf{y}_i - \mathbf{y}_s) = (\mathbf{a}_i - \mathbf{a}_s)'(\mathbf{a}_i - \mathbf{a}_s)$
- 4) Los marcadores fila, coinciden con las coordenadas de las filas en el espacio de las componentes principales: $\boldsymbol{\psi} = \mathbf{Y}\mathbf{V} = \mathbf{A}$
- 5) El producto escalar de las columnas de \mathbf{Y} , coincide con el producto escalar de los marcadores columna, o sea: $\mathbf{y}_j' \mathbf{y}_j \approx \mathbf{b}_j' \mathbf{b}_j$
- 6) La longitud al cuadrado de los vectores que representan a los marcadores columna, aproximan la covarianza entre las variables correspondientes, de donde: $\mathbf{b}_j' \mathbf{b}_j = s_{jj} = s_j^2$
- 7) La longitud de los marcadores columna, aproxima la desviación estándar de las variables
- 8) El coseno del ángulo entre dos marcadores columna, aproxima la correlación entre variables correspondientes: $\text{Cos}(\mathbf{b}_j, \mathbf{b}_k) \approx \text{Corr}(\mathbf{y}_j, \mathbf{y}_k)$
- 9) Las coordenadas de la matriz de marcadores columna \mathbf{B} en el Biplot de Componentes Principales, son equivalentes a los pesos de las variables a lo largo de los ejes principales.

Para finalizar, en la tabla 4.2, se resumen las distintas interpretaciones de los marcadores fila y columna, desde el punto de vista del Análisis de Correlación Canónica, del Análisis de Regresión Múltiple, y de acuerdo a las propiedades inherentes a cada tipo de Biplot.

Propiedades de los Marcadores	
<p>Marcadores Fila Tipo 1: $A^* = (R_{TT}^{1/2}U)$ siendo $T = Y, Y', (Y'Z)$ o (YX)</p> <ul style="list-style-type: none"> ✓ Baja calidad de representación. ✓ Aproximan la matriz de Correlaciones Intraset entre los siguientes pares de variables: $(Y, Y^*), (Y', Y'^*), (R_{y'z}, R^*_{y'z}), (R_{yx}, R^*_{yx})$. ✓ La raíz cuadrada de la distancia euclídea entre marcadores fila aproxima la raíz cuadrada de la distancia de Mahalanobis entre puntos fila de T. 	
<p>Marcadores Fila Tipo 2: $A^* = (R_{TT}^{1/2}UD_\alpha)$ siendo $T = Y, Y', (Y'Z)$ o (YX)</p> <ul style="list-style-type: none"> ✓ Alta calidad de representación. ✓ Aproximan la matriz de Correlaciones Interset entre los siguientes pares de variables: $(Y, X^*), (Y', Z^*), (R_{y'z}, X^*), (R_{yx}, Z^*)$. ✓ Aproximan la proporción de varianza de Y, que se puede predecir a través de las variables canónicas X^* o Z^*. ✓ La distancia euclídea entre marcadores fila aproxima la distancia entre puntos fila de T. ✓ Permiten situar en el gráfico regiones de predicción para las variables T. 	
<p>Marcadores Columna Tipo 1: $B^* = (R_{ww}^{1/2}VD_\alpha)$ siendo $w = X$ o Z</p> <ul style="list-style-type: none"> ✓ Alta calidad de representación. ✓ Aproximan la matriz de Correlaciones Interset entre los siguientes pares de variables: $(X, Y^*), (Z, Y'^*)$. ✓ Aproximan la proporción de varianza de las variables X o Z que se puede predecir por las variables canónicas Y^*. <ul style="list-style-type: none"> ✓ Similar a los coeficientes de determinación R^2. ✓ Equivalentes a la contribución del factor al elemento. 	
<p>Marcadores Columna Tipo 2: $B^* = (R_{ww}^{1/2}V)$, siendo $w = X$ o Z</p> <ul style="list-style-type: none"> ✓ Baja calidad de representación. ✓ Aproximan la matriz de Correlaciones Intraset entre los siguientes pares de variables: $(X, X^*), (Z, Z^*)$. ✓ Es posible superponer sobre el gráfico una escala de predicción de los valores originales de las variables T. 	

Tabla 4.2.- Interpretación de los Marcadores en los “Biplot de Regresión con Información Externa”

CAPITULO V

ESTIMACION MAXIMO VEROSIMIL EN BILOT DE REGRESION

ESTIMACION MAXIMO VEROSIMIL EN BILOT DE REGRESION

1.- INTRODUCCION.-

En el capítulo III apartado 3.1, se propone utilizar para el ajuste de los “Biplot de Regresión” a través de Modelos Bilineales Multiplicativos Generalizados, la siguiente especificación general:

$$g(\mu) = \Theta = \mathbf{A} \mathbf{B}' \quad (5.1)$$

siendo:

\mathbf{Y} la matriz de datos de orden $(n \times p)$ constituída por n filas o individuos y p columnas o variables con distribuciones de la familia Exponencial.

$\mu = E(\mathbf{Y})$ la matriz de esperanzas de \mathbf{Y} .

$g(\cdot)$ la función link utilizada en el ajuste.

Θ la matriz $(n \times p)$ de parámetros, desconocida excepto por su rango $q < r$, lo que permite su factorización en el producto de las matrices \mathbf{A} y \mathbf{B} , matrices de orden $(n \times q)$ y $(q \times p)$ respectivamente .

El ajuste (5.1) involucra el problema de la factorización de la matriz Θ para realizar la representación Biplot, problema que involucra a su vez la estimación de las matrices de parámetros \mathbf{A} y \mathbf{B} .

Si consideramos el hecho de que las p variables en la matriz de datos \mathbf{Y} tienen distribuciones de la familia Exponencial, entonces esa matriz puede estar formada por distintos tipos de datos: dicotómicos (presencias-ausencias con distribución Binomial), frecuencias (con distribución Poisson), cuantitativos (con distribución Normal), etc.

Los ajustes de los “Biplot de Regresión” no son invariantes a transformaciones en los datos: por traslación (centrado), o por traslación y cambios de escala (estandarización).

La aproximación de una matriz a través del ajuste de un Biplot, tiene como finalidad el análisis de la variabilidad de los datos, para lo cual se requiere el centrado de los mismos por columna para eliminar de las variables el efecto del promedio, de manera tal, que la recta de ajuste quede dominada por la mayor dispersión de las variables y no por el alejamiento respecto del centro de gravedad de la nube de individuos.

Un análisis Biplot con datos centrados, es equivalente a un ajuste mínimo cuadrático a través del siguiente modelo bilineal:

$$y_{ij} = \beta_j + \mathbf{a}'_i \mathbf{b}_j + e_{ij}$$

El ajuste de la parte lineal en ese modelo conduce a la estimación de:

$$\hat{\beta}_j = \bar{x}_{.j}$$

De donde:

$$y_{ij} - \bar{X}_{.j} = \mathbf{a}_i' \mathbf{b}_j$$

Lo que equivale a la aproximación Biplot de la matriz de datos centrados por columna, cuya filosofía es equivalente a la de un Análisis de Componentes Principales.

Si en un Análisis de Componentes Principales para datos de frecuencias se transforman las mismas centrándolas, entonces ello es equivalente a realizar un Análisis Factorial de Correspondencias eliminando el mayor valor propio igual a la unidad, ya que el mismo explica una dispersión nula lo cual no tiene sentido en este tipo de análisis.

En el ajuste de los “Biplot de Regresión Logísticos” (véase para mayor detalle el apartado 3.2 en el capítulo III) para datos de presencia-ausencia con valores 0 ó 1 respectivamente, el centrado de los mismos no tiene sentido.

Por esa razón, en el modelo utilizado es necesario considerar el término constante o independiente β_j , el cual hemos visto representa a la media de las columnas en la matriz de datos. Sin la inclusión en el modelo de ese término constante, la predicción de las probabilidades para valores iguales a cero será siempre igual a $1/2$, lo cual no tiene porque ser así.

De allí que, en el ajuste de los “Biplot de Regresión” al igual que en el Análisis de Regresión Múltiple, el modelo a utilizar para ajustar a los datos

debe considerar la posibilidad de inclusión de un término constante o independiente, para el caso de datos no centrados.

Comparando el ajuste de los “Biplot de Regresión” a través del modelo (5.1) al utilizado en Análisis de Regresión Múltiple, en este modelo la matriz Θ es desconocida excepto por su rango q , lo cual permite su factorización en el producto de matrices desconocidas (\mathbf{AB}'). Este es un problema de estimación similar al de los Modelos Bilineales Generalizados Multiplicativos (véase para mayores detalles el apartado 5.3 en el capítulo II).

Los métodos principales de estimación utilizados hasta el presente para el ajuste de Modelos Bilineales Multiplicativos abarcan:

- a) El método de Mínimos Cuadrados Alternados para el ajuste de Modelos Bilineales Clásicos, método que no permite el cálculo de los errores de estimación (BLAZQUEZ, 1998).
- b) El método iterativo de Regresión Bilineal Generalizada, método que si permite la estimación de los errores de estimación (GABRIEL, 1998).

Ambos métodos son alternantes, se supone a priori conocida una de las matrices de parámetros (\mathbf{A} ó \mathbf{B}) para poder estimar la otra matriz, o sea la estimación se hace en “**forma separada**” para cada matriz, suponiendo conocida la otra.

A continuación, para el ajuste de los “Biplot de Regresión” a través de Modelos Bilineales Multiplicativos Generalizados, se propone un método de “**estimación conjunta**” que proporciona estimaciones máximo-verosímiles simultáneas de las matrices de parámetros **A** y **B**, método que contempla la posibilidad de inclusión de un término independiente en el modelo, dependiendo del tipo de datos en la matriz **Y**.

Paralelamente, se generaliza para el ajuste de los “Biplot de Regresión”, el método de estimación en “forma separada” de Regresiones Bilineales Generalizadas (GABRIEL, 1998). Esta generalización constituye un método alternativo al de “estimación conjunta”, que ayuda a resolver los problemas computacionales que se pueden presentar en la inversión de matrices de orden muy alto.

2.- ESTIMACION MAXIMO VEROSIMIL CONJUNTA EN BIPLOTS DE REGRESION

Consideremos el siguiente modelo general para el ajuste de los “Biplots de Regresión”:

$$g(\mu) = \mathbf{AB}' = \mathbf{a}'_0 \mathbf{b}_0 + \mathbf{a}'_1 \mathbf{b}_1 + \dots + \mathbf{a}'_q \mathbf{b}_q = \sum_{i,j,k} (a'_{ik} b_{jk}) \quad (5.2)$$

para $(i = 1,2,\dots,n; j = 1,2,\dots,p; k = 0,1,2,\dots,q)$

Donde

$$\mu = E(\mathbf{Y})$$

$$[\mathbf{Y}]_{(n \times p)} = [y_{ij}]$$

$$[\mathbf{A}]_{(n \times q)} = [\mathbf{a}_k] = [a_{ik}]$$

$$[\mathbf{B}]_{(q \times p)} = [\mathbf{b}_k] = [b_{jk}]$$

El primer sumando en (5.2) permite la inclusión de un término independiente (o constante) en el modelo, siendo $\mathbf{a}_o = 1$ un valor constante dado a priori y $\mathbf{b}_o = 0$ solo en los modelos sin término independiente.

Así pues, en la especificación general del modelo se diferencian dos tipos de sumandos, un primer tipo que corresponde al primer término en (5.2) y que permite incluir el término independiente a través del producto $(\mathbf{a}'_o \mathbf{b}_o)$, suponiendo a priori un valor constante para \mathbf{a}_o igual a la unidad, y un segundo tipo que incluye todos los términos en (5.2) excepto el primero y que no supone valores a priori ni para \mathbf{a} ni para \mathbf{b} .

Se puede entonces reespecificar nuevamente el modelo diferenciando los dos tipos de términos, o sea:

$$g(\mu) = \mathbf{A}\mathbf{B}' = \mathbf{a}'_o \mathbf{b}_o + (\mathbf{a}'_1 \mathbf{b}_1 + \dots + \mathbf{a}'_q \mathbf{b}_q) \quad (5.3)$$

de donde, para $(i = 1, 2, \dots, n; j = 1, 2, \dots, p; k = 1, 2, \dots, q)$, tenemos que:

$$g(\mu) = \mathbf{a}'_o \mathbf{b}_o + \sum_{i,j,k} \mathbf{a}'_{ik} \mathbf{b}_{jk}$$

Para la obtención “conjunta o simultánea” de los estimadores máximo-verosímiles de las matrices de parámetros **A** y **B** en los “Biplots de Regresión”, se propone una generalización del método iterativo de Newton-Raphson utilizado en los Modelos Lineales Generalizados (ver apartado 4.5 del capítulo II), uniendo las regresiones separadas para cada columna y fila de **Y** en (5.2) en una sola tal como lo hace BAKER en 1992.

2.1.- MODELOS SIN TERMINO INDEPENDIENTE

Un caso particular es el modelo sin término independiente cuando $\mathbf{b}_0 = 0$, tal que:

$$g(\mu) = \mathbf{AB}' = \mathbf{a}'_1 \mathbf{b}_1 + \mathbf{a}'_2 \mathbf{b}_2 + \dots + \mathbf{a}'_q \mathbf{b}_q \tag{5.4}$$

($i = 1,2,\dots,n$; $j = 1,2,\dots,p$; $k = 1,2,\dots,q$)

En el caso del ajuste del Biplot de Regresión a través del modelo (5.4) para 2 parámetros ($k = 1,2$), las matrices **A** y **B** tienen la siguiente forma:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ \vdots & \vdots \\ a_{i1} & a_{i2} \\ \vdots & \vdots \\ a_{n1} & a_{n2} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} \\ \vdots & \vdots \\ b_{j1} & b_{j2} \\ \vdots & \vdots \\ b_{p1} & b_{p2} \end{bmatrix}$$

Se propone el siguiente procedimiento de estimación máximo-verosímil “conjunta” para los parámetros **A** y **B** en el modelo sin término independiente,

reescribiendo las matrices **A** y **B** como vectores columna concatenando las filas de esas matrices y designándolas $\langle \mathbf{A}' \rangle$ y $\langle \mathbf{B}' \rangle$, o sea:

$$\langle \mathbf{A}' \rangle = \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{i1} \\ a_{i2} \\ \vdots \\ a_{n1} \\ a_{n2} \end{bmatrix}_{(2nx1)} \qquad \langle \mathbf{B}' \rangle = \begin{bmatrix} b_{11} \\ b_{12} \\ \vdots \\ b_{j1} \\ b_{j2} \\ \vdots \\ b_{p1} \\ b_{p2} \end{bmatrix}_{(2px1)}$$

Aplicando el método de Newton-Raphson tenemos el siguiente sistema de ecuaciones:

$$\begin{bmatrix} \langle \hat{\mathbf{B}}' \rangle \\ \langle \hat{\mathbf{A}}' \rangle \end{bmatrix}_{(t+1)} = \begin{bmatrix} \langle \hat{\mathbf{B}}' \rangle \\ \langle \hat{\mathbf{A}}' \rangle \end{bmatrix}_{(t)} - [\mathfrak{I}(\mathbf{B}, \mathbf{A})]_{(t)}^{-1} [\mathbf{U}(\mathbf{B}, \mathbf{A})]_{(t)}$$

La matriz de información $\mathfrak{I}(\mathbf{B}, \mathbf{A})$ de orden $(2p+2n) \times (2p+2n)$, está constituida por las siguientes cuatro grandes sub-matrices:

$$\mathfrak{I}(\mathbf{B}, \mathbf{A}) = \begin{bmatrix} \mathfrak{I}(\mathbf{B}) = \left[\partial^2 L / \partial \mathbf{b}^2 \right]_{(2px2p)} & \mathfrak{I}(\mathbf{BA}) = \left[\partial^2 L / \partial \mathbf{b} \partial \mathbf{a} \right]_{(2px2n)} \\ \mathfrak{I}(\mathbf{AB}) = \left[\partial^2 L / \partial \mathbf{a} \partial \mathbf{b} \right]_{(2nx2p)} & \mathfrak{I}(\mathbf{A}) = \left[\partial^2 L / \partial \mathbf{a}^2 \right]_{(2nx2n)} \end{bmatrix}$$

donde, para simplificar la notación $\partial^2 L = \partial^2 \log L$ designa a la segunda derivada de la función de log-verosimilitud.

$$\mathfrak{I}(\mathbf{BA}) = \begin{bmatrix} \ddots & & \vdots & & \vdots & & \ddots \\ \dots & \begin{bmatrix} \frac{\partial^2 L}{\partial b_{j_1} \partial a_{i_1}} & \frac{\partial^2 L}{\partial b_{j_1} \partial a_{i_2}} \\ \frac{\partial^2 L}{\partial b_{j_2} \partial a_{i_1}} & \frac{\partial^2 L}{\partial b_{j_2} \partial a_{i_2}} \end{bmatrix} & \dots & \begin{bmatrix} \frac{\partial^2 L}{\partial b_{j_1} \partial a_{h_1}} & \frac{\partial^2 L}{\partial b_{j_1} \partial a_{h_2}} \\ \frac{\partial^2 L}{\partial b_{j_2} \partial a_{h_1}} & \frac{\partial^2 L}{\partial b_{j_2} \partial a_{h_2}} \end{bmatrix} & \dots \\ \dots & \vdots & \ddots & \vdots & \ddots & \vdots & \dots \\ \dots & \begin{bmatrix} \frac{\partial^2 L}{\partial b_{i_1} \partial a_{i_1}} & \frac{\partial^2 L}{\partial b_{i_1} \partial a_{i_2}} \\ \frac{\partial^2 L}{\partial b_{i_2} \partial a_{i_1}} & \frac{\partial^2 L}{\partial b_{i_2} \partial a_{i_2}} \end{bmatrix} & \dots & \begin{bmatrix} \frac{\partial^2 L}{\partial b_{i_1} \partial a_{h_1}} & \frac{\partial^2 L}{\partial b_{i_1} \partial a_{h_2}} \\ \frac{\partial^2 L}{\partial b_{i_2} \partial a_{h_1}} & \frac{\partial^2 L}{\partial b_{i_2} \partial a_{h_2}} \end{bmatrix} & \dots \\ \dots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \\ \ddots & & \vdots & & \vdots & & \ddots \end{bmatrix}$$

Al igual que $\mathfrak{I}(\mathbf{BA})$, la matriz $\mathfrak{I}(\mathbf{AB})$ es una matriz de productos cruzados.

El vector $\mathbf{U}(\mathbf{B},\mathbf{A})$ está constituido por los siguientes 2 subvectores:

$$\mathbf{U}(\mathbf{B}, \mathbf{A}) = \begin{bmatrix} \mathbf{U}\langle \mathbf{B}' \rangle \\ \mathbf{U}\langle \mathbf{A}' \rangle \end{bmatrix}_{(2p+2n) \times 1} = \begin{bmatrix} \partial \mathcal{L} / \partial b_{jk} \\ \partial \mathcal{L} / \partial a_{ik} \end{bmatrix}$$

Considerando que el orden de la matriz $\mathfrak{I}(\mathbf{B},\mathbf{A})$ es muy grande, se hacen los siguientes supuestos (BAKER, 1992) para la estimación máximo verosímil conjunta de los parámetros \mathbf{A} y \mathbf{B} :

a) Se supone que los “n” individuos se seleccionan en forma aleatoria, por lo que se pueden considerar independientes los parámetros filas a_{ik} , de donde los productos cruzados entre pares de individuos serán nulos ya que están

incorrelacionados, o sea, para $i = 1, 2, \dots, n$:

$$(\partial^2 L / \partial a_{i1} \partial a_{i2}) = 0 \quad (i \neq i'); \quad (\partial^2 L / \partial a_{ik}^2) \neq 0 \quad (i = i')$$

Bajo este supuesto, la submatriz $\mathfrak{J}(\mathbf{A})$ se convierte en una matriz diagonal, cuyos elementos están constituidos por “n” submatrices de orden (2x2) cada una:

$$\mathfrak{J}(\mathbf{A}) = \text{diag} \left[\begin{array}{cc} \left[\partial^2 L / \partial a_{i1}^2 \right] & \left[\partial^2 L / \partial a_{i1} \partial a_{i2} \right] \\ \left[\partial^2 L / \partial a_{i2} \partial a_{i1} \right] & \left[\partial^2 L / \partial a_{i2}^2 \right] \end{array} \right]_{2n \times 2n}$$

b) Se suponen independientes los parámetros de los individuos y las variables, luego:

$$(\partial^2 L / \partial a_{ik} \partial b_{jk}) = 0 \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, p; k = 1, 2)$$

De allí que bajo este supuesto, las submatrices $\mathfrak{J}(\mathbf{B}, \mathbf{A})$ e $\mathfrak{J}(\mathbf{A}, \mathbf{B})$ se convierten en matrices nulas:

$$[\mathfrak{J}(\mathbf{B}, \mathbf{A})] = [\mathfrak{J}(\mathbf{A}, \mathbf{B})] = [\mathbf{O}]$$

c) Se suponen independientes los parámetros b_{jk} de cada una de las variables, luego los productos cruzados entre pares de variables serán nulos pudiendo

existir covariación entre los parámetros de una misma variable, o sea:

$$(\partial^2 L / \partial \boldsymbol{\beta}_{jk} \partial \boldsymbol{\beta}_{j'k}) = 0 \quad (j \neq j'); \quad (\partial^2 L / \partial \boldsymbol{\beta}_{jk}^2) \neq 0 \quad (j = j')$$

De allí que bajo este supuesto, la submatriz $\mathfrak{I}(\mathbf{B})$ se convierte en una matriz diagonal cuyos elementos están constituídos por “p” submatrices de orden (2x2) cada una ($j = 1, 2, \dots, p$), o sea:

$$[\mathfrak{I}(\mathbf{B})] = \text{diag} \left[\begin{array}{cc} [\partial^2 L / \partial \boldsymbol{\beta}_{j1}^2] & [\partial^2 L / \partial \boldsymbol{\beta}_{j1} \partial \boldsymbol{\beta}_{j2}] \\ [\partial^2 L / \partial \boldsymbol{\beta}_{j2} \partial \boldsymbol{\beta}_{j1}] & [\partial^2 L / \partial \boldsymbol{\beta}_{j2}^2] \end{array} \right]_{(2p \times 2p)}$$

Este supuesto, es equivalente a suponer que la respuesta de los individuos a cada una de las variables es independiente. El mismo supuesto también se hace en el caso del ajuste de cada columna de la matriz \mathbf{Y} por separado.

Bajo los tres supuestos anteriores, la matriz de información $\mathfrak{I}(\mathbf{B}, \mathbf{A})$ se convierte en una matriz diagonal, cuyos elementos son también matrices diagonales, haciéndose por consiguiente más manejable su inversión desde el punto de vista computacional, ya que:

$$\mathfrak{I}(\mathbf{B}, \mathbf{A}) = \begin{bmatrix} \mathfrak{I}(\mathbf{B}) & \mathbf{0} \\ \mathbf{0} & \mathfrak{I}(\mathbf{A}) \end{bmatrix}$$

Luego a través de un procedimiento iterativo aplicando el método de Newton-Raphson, se obtiene finalmente el siguiente sistema de $(2p + 2n)$ ecuaciones:

$$\langle \hat{\mathbf{B}}' \rangle_{(t+1)} = \langle \hat{\mathbf{B}}' \rangle_t - [\mathfrak{J}(\mathbf{B})]_t^{-1} [\mathbf{U}(\mathbf{B})]_t \tag{5.5}$$

$$\langle \hat{\mathbf{A}}' \rangle_{(t+1)} = \langle \hat{\mathbf{A}}' \rangle_t - [\mathfrak{J}(\mathbf{A})]_t^{-1} [\mathbf{U}(\mathbf{A})]_t$$

En ese sistema de ecuaciones, corresponde un par de ecuaciones para cada variable j -ésima (para $k = 1, 2$) y un par de ecuaciones para cada individuo i -ésimo, o sea:

para cada variable ($j = 1, 2, \dots, p$):

$$\hat{b}_{j1}^{(t+1)} = \hat{b}_{j1}^{(t)} - (\partial^2 L / \partial b_{j1}^2)^{-1} (\partial L / \partial b_{j1}) - (\partial^2 L / \partial b_{j1} \partial b_{j2})^{-1} (\partial L / \partial b_{j2})$$

$$\hat{b}_{j2}^{(t+1)} = \hat{b}_{j2}^{(t)} - (\partial^2 L / \partial b_{j2}^2)^{-1} (\partial L / \partial b_{j2}) - (\partial^2 L / \partial b_{j2} \partial b_{j1})^{-1} (\partial L / \partial b_{j1})$$

para cada individuo ($i = 1, 2, \dots, n$)

$$\hat{a}_{i1}^{(t+1)} = \hat{a}_{i1}^{(t)} - (\partial^2 L / \partial a_{i1}^2)^{-1} (\partial L / \partial a_{i1}) - (\partial^2 L / \partial a_{i1} \partial a_{i2})^{-1} (\partial L / \partial a_{i2})$$

$$\hat{a}_{i2}^{(t+1)} = \hat{a}_{i2}^{(t)} - (\partial^2 L / \partial a_{i2} \partial a_{i1})^{-1} (\partial L / \partial a_{i1}) - (\partial^2 L / \partial a_{i2}^2)^{-1} (\partial L / \partial a_{i2})$$

La resolución de ese sistema de ecuaciones requiere la sustitución de todas las derivadas, cuyas expresiones dependen de sí el “Biplot de Regresión” a ajustar es lineal o no lo es, cuyos desarrollos se presentan en el próximo apartado.

El procedimiento de estimación se inicializa suponiendo valores iniciales simultáneamente para \mathbf{B}^0 y \mathbf{A}^0 . En la primera iteración se obtienen nuevos valores estimados \mathbf{B}^1 y \mathbf{A}^1 , los cuales se utilizan en la segunda iteración como valores iniciales y así se procede sucesivamente hasta que la diferencia entre aproximaciones obtenidas de las iteraciones sucesivas (t-1) y (t) sea lo suficientemente pequeña, en cuyo caso se toman esas aproximaciones como las estimaciones máximo-verosímiles de $\langle \mathbf{B}' \rangle$ y $\langle \mathbf{A}' \rangle$.

En la práctica, se puede comenzar con valores iniciales provenientes de una aproximación mínimo cuadrática, para lo cual los valores de $\hat{\mathbf{A}}^0$ se pueden aproximar a través de algún método de Análisis de Proximidades, como por ejemplo una ordenación de los individuos mediante Coordenadas Principales.

Una de las ventajas de utilizar el procedimiento iterativo de Newton-Raphson, es que se pueden obtener fácilmente las varianzas asintóticas de los estimadores así como también los estadísticos Deviance y χ^2 para verificar la bondad del ajuste.

Así tenemos que para muestras grandes, los errores de estimación asintóticos vienen dados por la inversa de la matriz de información $\mathfrak{I}(\mathbf{B}, \mathbf{A})$ (para mayores detalles ver el apartado 4.5 del Capítulo II), o sea:

$$\text{Var} \begin{bmatrix} \langle \hat{\mathbf{B}}' \rangle \\ \langle \hat{\mathbf{A}}' \rangle \end{bmatrix} = \sigma^2 [\mathfrak{I}(\mathbf{B}, \mathbf{A})]^{-1}$$

Donde

$$\hat{\sigma}^2 = \{ \langle \mathbf{e}' \rangle [\text{diag} \langle \mathbf{W} \rangle] \langle \mathbf{e} \rangle \} / [2nx2p - r(2n + 2p - r)]$$

$$r = \text{rango}(\Theta) = \text{rango}(\mathbf{AB}')$$

Siendo

\mathbf{e} = vector de residuos en la estimación

$\mathbf{W} = [w_{ii}] = (1/\text{var}(\mathbf{y}_i))[\partial\mu_i / \partial\eta_i]$ = matriz de ponderaciones o pesos

La Deviance por su parte será:

$$D = (1/\sigma^2)(\langle \mathbf{e}' \rangle [\text{diag} \langle \mathbf{W} \rangle] \langle \mathbf{e} \rangle)$$

y el estadístico χ^2 será:

$$\chi^2 = \langle \mathbf{e}' \rangle [\text{diag} \langle \mathbf{W} \rangle] \langle \mathbf{e} \rangle$$

2.1.1.- BIPLOTS DE REGRESION LINEAL

Supongamos que se desea ajustar el “Biplot de Regresión Lineal” a través del modelo (5.4), considerando que \mathbf{Y} se distribuye normalmente y que la “función link” es la identidad.

Para la estimación conjunta de las matrices de parámetros \mathbf{A} y \mathbf{B} , de acuerdo al sistema de ecuaciones (5.5), se requiere el cálculo de las derivadas de la función de verosimilitud.

Bajo el supuesto que \mathbf{Y} tiene distribución Normal, entonces:

$$f(y, \mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}[(y - \mu)/\sigma]^2\right\}$$

ya que μ depende de los parámetros del modelo, entonces:

$$\mu_{ij} = f(a_{ik}, b_{jk})$$

Luego, la función de probabilidad conjunta o función de verosimilitud será función de los parámetros \mathbf{A} y \mathbf{B} , siendo su expresión igual a:

$$L(\mathbf{A}, \mathbf{B}) = \prod_i \prod_j \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}[(y_{ij} - \mu_{ij})/\sigma]^2\right\}$$

De allí que, la función de log-verosimilitud es igual a:

$$L = \log L = -n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_i \sum_j (y_{ij} - \mu_{ij})^2 \tag{5.6}$$

Los estimadores máximo verosímiles se obtienen de la resolución del siguiente sistema de ecuaciones:

$$\frac{\partial L}{\partial b_{jk}} = \sum_i \left(\frac{\partial L}{\partial \mu_{ij}} \frac{\partial \mu_{ij}}{\partial \eta_{ij}} \frac{\partial \eta_{ij}}{\partial b_{jk}} \right) \tag{5.7}$$

$$\frac{\partial \mathcal{L}}{\partial a_{ik}} = \sum_j \left(\frac{\partial \mathcal{L}}{\partial \mu_{ij}} \frac{\partial \mu_{ij}}{\partial \eta_{ij}} \frac{\partial \eta_{ij}}{\partial a_{ik}} \right) \quad (5.8)$$

Llamando $\eta_{ij} = g(\mu_{ij}) = \mu_{ij} = \sum_i \sum_j a_{ik} b'_{jk} \quad (k = 1,2)$

Entonces $\frac{\partial \eta_{ij}}{\partial b_{jk}} = a_{ik} \quad \frac{\partial \eta_{ij}}{\partial a_{ik}} = b_{jk} \quad (5.9)$

y escribiendo:

$$g'(\mu_{ij}) = \frac{\partial \eta_{ij}}{\partial \mu_{ij}} = 1$$

tenemos que:

$$\frac{1}{g'(\mu_{ij})} = \frac{\partial \mu_{ij}}{\partial \eta_{ij}} = 1 \quad (5.10)$$

De allí que, para la resolución de ese sistema de ecuaciones se obtiene a continuación las derivadas correspondientes:

a) Derivadas respecto de b_{jk} $(k = 1,2; j = 1, 2, \dots, p)$

Hallando la $(\partial \mathcal{L} / \partial \mu_{ij})$ en (5.6) y luego sustituyendo en (5.7) esa expresión

más las expresiones (5.9) y (5.10), tenemos que:

$$\frac{\partial L}{\partial b_{jk}} = \frac{1}{\sigma^2} \sum_i (y_{ij} - \mu_{ij}) a_{ik}$$

Para hallar $(\partial^2 L / \partial b_{jk}^2)$ se considera el hecho que la matriz de Información de Fisher es igual a: $\mathfrak{I}^{-1} = [\partial^2 L / \partial b_{jk}^2]^{-1}$, siendo su inversa igual a: $\mathfrak{I} = -E[\partial^2 L / \partial^2_{jk}]$, lo que permite reducir los cálculos ya que $E(Y_{ij}) = \mu_{ij}$.

Considerando además en la obtención de las segundas derivadas, que la primera derivada es el producto de una constante por la sumatoria de un producto, tenemos finalmente dado el supuesto c) que:

Para $j = j'$

$$\left[\frac{\partial^2 L}{\partial b_{jk}^2} \right] = \frac{1}{\sigma^2} \sum_i a_{ik}^2$$

Para $j \neq j'$

$$\left[\frac{\partial^2 L}{\partial b_{jk} \partial b_{j'k}} \right] = 0$$

Mientras que para $k \neq k'$

$$\left[\frac{\partial^2 L}{\partial b_{jk} \partial b_{jk'}} \right] = \frac{1}{\sigma^2} \sum_i a_{ik} a_{ik'}$$

b) Derivadas respecto de a_{ik} ($k = 1, 2; i = 1, 2, \dots, n$)

Hallando la $(\partial L / \partial \mu_{ij})$ en (5.6) y luego sustituyendo en (5.8) esa expresión, más las expresiones (5.9) y (5.10), tenemos que:

$$\frac{\partial L}{\partial a_{ik}} = \frac{1}{\sigma^2} \sum_j (y_{ij} - \mu_{ij}) b_{jk}$$

Luego hallando las segundas derivadas, y considerando el supuesto a) tenemos que:

Para $i = i'$

$$\left[\frac{\partial^2 L}{\partial a_{ik}^2} \right] = \frac{1}{\sigma^2} \sum_j b_{jk}^2$$

Para $i \neq i'$

$$\left[\frac{\partial^2 L}{\partial a_{ik} \partial a_{i'k}} \right] = 0$$

Para $k \neq k'$

$$\left[\frac{\partial^2 L}{\partial a_{ik} \partial a_{ik'}} \right] = \frac{1}{\sigma^2} \sum_j b_{jk} b_{jk'}$$

Se demuestra en el apartado 4.5 del capítulo II, que este procedimiento para obtener los estimadores máximo-verosímiles de \mathbf{A} y \mathbf{B} , es equivalente a la obtención a través de un procedimiento iterativo de mínimos cuadrados ponderados (DOBSON, 1990).

2.1.2.- BIPLOTS DE REGRESION NO LINEAL

En el caso de los “Biplots de Regresión no Lineal” ajustados a través del modelo (5.4), la estimación máximo-verosímil conjunta de los parámetros \mathbf{A} y \mathbf{B} de acuerdo al sistema de ecuaciones (5.5), requiere del cálculo de las derivadas de la función de log-verosimilitud L .

Por ejemplo, en el caso de datos de presencia-ausencia, donde las columnas de \mathbf{Y} tienen distribución Binomial, tenemos:

$$f(y, p) = \binom{n}{y} p^y (1-p)^{n-y}$$

Como la probabilidad p depende de los parámetros del modelo, entonces:

$$p_{ij} = f(a_{ik}, b_{jk}) \quad (k = 1, 2)$$

Entonces la función de probabilidad conjunta o función de verosimilitud será función de los parámetros \mathbf{A} y \mathbf{B} , siendo su expresión:

$$L(\mathbf{A}, \mathbf{B}) = \prod_i \prod_j \binom{n_{ij}}{y_{ij}} p_{ij}^{y_{ij}} (1-p_{ij})^{n_{ij}-y_{ij}}$$

De donde la función de log-verosimilitud será:

$$L = \sum_i \sum_j \left\{ \log \binom{n_{ij}}{y_{ij}} + y_{ij} \log p_{ij} + (n_{ij} - y_{ij}) \log(1 - p_{ij}) \right\} \quad (5.11)$$

Los estimadores máximo verosímiles se obtienen de la resolución del siguiente sistema de ecuaciones:

$$\frac{\partial L}{\partial b_{jk}} = \sum_i \left(\frac{\partial L}{\partial p_{ij}} \frac{\partial p_{ij}}{\partial \eta_{ij}} \frac{\partial \eta_{ij}}{\partial b_{jk}} \right) \quad (5.12)$$

$$\frac{\partial L}{\partial a_{ik}} = \sum_j \left(\frac{\partial L}{\partial p_{ij}} \frac{\partial p_{ij}}{\partial \eta_{ij}} \frac{\partial \eta_{ij}}{\partial a_{ik}} \right) \quad (5.13)$$

Llamando $\eta_{ij} = g(p_{ij}) = \sum_i \sum_j a_{ik} b'_{jk} \quad (k = 1,2)$

Entonces: $\frac{\partial \eta_{ij}}{\partial b_{jk}} = a_{ik} \quad \frac{\partial \eta_{ij}}{\partial a_{ik}} = b_{jk} \quad (5.14)$

Y escribiendo $g'(p_{ij})$ por :

$$g'(p_{ij}) = \frac{\partial \eta_{ij}}{\partial p_{ij}} \quad \Rightarrow \quad \frac{1}{g'(p_{ij})} = \frac{\partial p_{ij}}{\partial \eta_{ij}} \quad (5.15)$$

Finalmente se obtienen las derivadas requeridas en la resolución del sistema de ecuaciones, o sea:

a) Derivadas respecto de b_{jk} (k=1,2; j = 1, 2, …,p)

Hallando la $(\partial L/ \partial p_{ij})$ en (5.11), y luego sustituyendo en (5.12) esa expresión e incluyendo las expresiones (5.14) y (5.15), tenemos que:

$$\frac{\partial L}{\partial b_{jk}} = \sum_i \left[\frac{y_{ij} - n_{ij} p_{ij}}{p_{ij}(1 - p_{ij})} \frac{1}{g'(p_{ij})} a_{ik} \right]$$

Para hallar $(\partial^2 L/ \partial b_{jk}^2)$ se considera el hecho que la matriz de Información de Fisher es igual a: $\mathfrak{I}^{-1} = [\partial^2 L/ \partial b_{jk}^2]^{-1}$, siendo su inversa igual a: $\mathfrak{I} = -E[\partial^2 L/ \partial b_{jk}^2]$, lo que permite reducir los cálculos ya que $E(Y_{ij}) = n p_{ij}$.

Considerando además en la obtención de las segundas derivadas, que la primera derivada es la sumatoria de un producto, tenemos finalmente dado el supuesto c) que:

Para $j = j'$

$$\left[\frac{\partial^2 L}{\partial b_{jk}^2} \right] = \sum_i \frac{n_{ij}}{p_{ij}(1 - p_{ij}) [g'(p_{ij})]^2} a_{ik}^2$$

Para $j \neq j'$

$$\left[\frac{\partial^2 L}{\partial b_{jk} \partial b_{j'k}} \right] = 0$$

Para $k \neq k'$

$$\left[\frac{\partial^2 L}{\partial b_{jk} \partial b_{jk'}} \right] = \sum_i \frac{n_{ij}}{p_{ij}(1-p_{ij})(g'(p_{ij}))^2} a_{ik} a_{ik'}$$

b) Derivadas respecto de a_{ik} ($k = 1, 2; i = 1, 2, \dots, n$)

Hallando la $(\partial L / \partial p_{ij})$ en (5.11) y luego sustituyendo en (5.13) esa expresión, más las expresiones (5.14) y (5.15), tenemos que:

$$\frac{\partial L}{\partial a_{ik}} = \sum_j \left[\frac{y_{ij} - n_{ij} p_{ij}}{p_{ij}(1-p_{ij})} \frac{1}{g'(p_{ij})} b_{jk} \right]$$

Luego hallando las segundas derivadas, y considerando el supuesto a) tenemos que:

Para $i = i'$

$$\left[\frac{\partial^2 L}{\partial a_{ik}^2} \right] = \sum_j \frac{n_{ij}}{p_{ij}(1-p_{ij})[g'(p_{ij})]^2} b_{jk}^2$$

Para $i \neq i'$

$$\left[\frac{\partial^2 L}{\partial \alpha_{ik} \partial \alpha_{i'k}} \right] = 0$$

Para $k \neq k'$

$$\left[\frac{\partial^2 L}{\partial \alpha_{ik} \partial \alpha_{ik'}} \right] = \sum_j \frac{n_{ij}}{p_{ij}(1-p_{ij})(g'(p_{ij}))^2} b_{jk} b_{jk'}$$

En esas derivadas, $g'(p)$ se determina considerando la “función link” utilizada en el ajuste del “Biplot de Regresión no Lineal”.

Así tenemos que, considerando la “función link logit” para los Biplots Logísticos, la “función probit” para los Biplots Probit y la “función link log-log” para los Biplots log-log, entonces:

1- En el ajuste de Biplots Logísticos tenemos:

$$\eta = g(p) = \log(p / 1-p) = \log p - \log(1-p)$$

de donde la primera derivada respecto de p es igual a:

$$g'(p) = \partial[\log p - \log(1-p)] = \frac{1}{p} + \frac{1}{(1-p)} = \frac{1}{p(1-p)}$$

2- En el ajuste de los Biplots Probit tenemos:

$$\eta = g(p) = \text{probit } p = \phi^{-1}(p) = \xi = \frac{y - np}{np(1 - p)}$$

$$g'(p) = \partial \left[\frac{y - np}{np(1 - p)} \right] = \partial \left[\frac{y}{np(1 - p)} - \frac{1}{(1 - p)} \right] = \frac{-y(1 - 2p) - np^2}{np^2(1 - p)^2}$$

3- En el ajuste de los Biplots log-log tenemos:

$$\eta = g(p) = \log [-\log (1 - p)]$$

$$g'(p) = \partial [\log(-\log(1 - p))] = \frac{1}{(1 - p)\log(1 - p)}$$

2.2.- MODELO CON TERMINO INDEPENDIENTE

El procedimiento de ajuste en este caso es similar al caso de los modelos sin término independiente, pero diferenciado los dos tipos de términos en la especificación del modelo (5.3), o sea:

$$g(\mu) = \mathbf{A} * \mathbf{B} * \mathbf{a}' = \mathbf{a}'_0 \mathbf{b}_0 + (\mathbf{a}'_1 \mathbf{b}_1 + \dots + \mathbf{a}'_q \mathbf{b}_q)$$

Como $\mathbf{a}_0 = 1$ entonces:

$$g(\mu) = \mathbf{A} * \mathbf{B} * \mathbf{a}' = \mathbf{b}_0 + (\mathbf{a}'_1 \mathbf{b}_1 + \dots + \mathbf{a}'_q \mathbf{b}_q)$$

de donde, para ($i = 1,2,\dots,n$; $j = 1,2,\dots,p$; $k = 0,1,2,\dots,q$)

$$g(\mu) = \mathbf{A}^* \mathbf{B}^{*'} = b_o + \sum_{l,j,k}^q a'_{ik} b_{jk} = \mathbf{A}_0 \mathbf{B}'_0 + \mathbf{A} \mathbf{B}' = \mathbf{1}_n \mathbf{B}'_o + \mathbf{A} \mathbf{B}'$$

Para la obtención conjunta de los estimadores máximo-verosímiles de \mathbf{A}^* y \mathbf{B}^* en los “Biplot de Regresión” ajustados a través de un modelo con término independiente (\mathbf{A}^* y \mathbf{B}^* en este caso corresponden a *matrices de parámetros no a matrices de marcadores fila y columna*), se generalizan adecuándolos a este contexto los procedimientos de estimación: de Newton-Raphson utilizado en los Modelos Lineales Generalizados y el de Regresión Bilineal Segmentada (GABRIEL, 1998).

En el caso del ajuste del Biplot de Regresión a través del modelo (5.3), para 3 parámetros ($k = 0, 1, 2$), las matrices \mathbf{A}^* y \mathbf{B}^* tienen la siguiente forma:

$$\mathbf{A}^* = [\mathbf{A}_0 \quad \mathbf{A}] = \begin{bmatrix} a_{10} & a_{11} & a_{12} \\ \vdots & \vdots & \vdots \\ a_{i0} & a_{i1} & a_{i2} \\ \vdots & \vdots & \vdots \\ a_{n0} & a_{n1} & a_{n2} \end{bmatrix}_{(nx3)} \Rightarrow \langle \mathbf{A}^{*'} \rangle = \begin{bmatrix} a_{10} \\ a_{11} \\ a_{12} \\ \vdots \\ \vdots \\ a_{n0} \\ a_{n1} \\ a_{n2} \end{bmatrix} = \begin{bmatrix} 1 \\ a_{11} \\ a_{12} \\ \vdots \\ \vdots \\ 1 \\ a_{n1} \\ a_{n2} \end{bmatrix}_{(3nx1)}$$

$$\mathbf{B}^* = [\mathbf{B}_0 \quad \mathbf{B}] = \begin{bmatrix} b_{10} & b_{11} & b_{12} \\ \vdots & \vdots & \vdots \\ b_{j0} & b_{j1} & b_{j2} \\ \vdots & \vdots & \vdots \\ b_{p0} & b_{p1} & b_{p2} \end{bmatrix}_{(px3)} \Rightarrow \langle \mathbf{B}^{*'} \rangle = \begin{bmatrix} b_{10} \\ b_{11} \\ b_{12} \\ \vdots \\ b_{p0} \\ b_{p1} \\ b_{p2} \end{bmatrix}_{(3px1)}$$

En la estimación de las matrices de parámetros, se generaliza el procedimiento propuesto en el modelo sin término independiente y aplicando el método de Newton-Raphson, tenemos:

$$\begin{bmatrix} \langle \hat{\mathbf{B}}^{*'} \rangle \\ \langle \hat{\mathbf{A}}^{*'} \rangle \end{bmatrix}_{(t+1)} = \begin{bmatrix} \langle \mathbf{B}^{*'} \rangle \\ \langle \hat{\mathbf{A}}^{*'} \rangle \end{bmatrix}_{(t)} - [\mathfrak{I}(\mathbf{B}^*, \mathbf{A}^*)]_{(t)}^{-1} [\mathbf{U}(\mathbf{B}^*, \mathbf{A}^*)]_{(t)}$$

La matriz de información $\mathfrak{I}(\mathbf{B}^*, \mathbf{A}^*)$ en este caso, es de orden $[(3p+3n) \times 3p+3n]$ y está constituida por las siguientes cuatro submatrices:

$$\mathfrak{I}(\mathbf{B}^*, \mathbf{A}^*) = \begin{bmatrix} \mathfrak{I}(\mathbf{B}^*) & \mathfrak{I}(\mathbf{B}^* \mathbf{A}^*) \\ \mathfrak{I}(\mathbf{A}^* \mathbf{B}^*) & \mathfrak{I}(\mathbf{A}^*) \end{bmatrix}$$

Cada una de esas submatrices, tiene la misma expresión que las correspondientes submatrices $\mathfrak{I}(\mathbf{B})$, $\mathfrak{I}(\mathbf{A})$, $\mathfrak{I}(\mathbf{BA})$, $\mathfrak{I}(\mathbf{AB})$ en el modelo sin término independiente, tan solo que ahora el orden y los subíndices en ellas varían ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, p$ y $k = 0, 1, 2$).

Haciendo los mismos supuestos que en el modelo sin término independiente, entonces tendremos el siguiente sistema de ecuaciones (para $k = 0, 1, 2$):

$$\begin{aligned} \langle \hat{\mathbf{B}}^{*'} \rangle_{(t+1)} &= \langle \mathbf{B}^{*'} \rangle_{(t)} - [\mathfrak{I}(\mathbf{B}^*)]_{(t)}^{-1} [\mathbf{U}(\mathbf{B}^*)]_{(t)} \\ \langle \hat{\mathbf{A}}^{*'} \rangle_{(t+1)} &= \langle \hat{\mathbf{A}}^{*'} \rangle_{(t)} - [\mathfrak{I}(\mathbf{A}^*)]_{(t)}^{-1} [\mathbf{U}(\mathbf{A}^*)]_{(t)} \end{aligned} \tag{5.16}$$

De donde las varianzas asintóticas de los estimadores serán:

$$\begin{aligned} \text{Var}(\langle \hat{\mathbf{B}}^{*'} \rangle) &= \sigma^2 [\mathfrak{J} \langle \hat{\mathbf{B}}^{*'} \rangle]^{-1} \\ \text{Var}(\langle \hat{\mathbf{A}}^{*'} \rangle) &= \sigma^2 [\mathfrak{J} \langle \hat{\mathbf{A}}^{*'} \rangle]^{-1} \\ \hat{\sigma}^2 &= \{ \langle \mathbf{e}' \rangle [\text{diag} \mathbf{W}] \langle \mathbf{e} \rangle \} / [(3p \times 3n) - r(3p + 3n - r)] \\ r &= \text{rango}(\Theta) = \text{rango}(\mathbf{A} * \mathbf{B}^{*'}) \end{aligned}$$

El sistema de ecuaciones (5.16) se resuelve sustituyendo las derivadas correspondientes, para luego iterativamente, partiendo de los valores iniciales $\langle \mathbf{B}^{0*'} \rangle$ y $\langle \mathbf{A}^{0*'} \rangle$ obtener las estimaciones definitivas de $\langle \mathbf{B}^{*'} \rangle$ y $\langle \mathbf{A}^{*'} \rangle$.

En la obtención de las derivadas hay que tener en cuenta que, el primer elemento en cada una de las n submatrices (3x1) de $\langle \mathbf{A}^{*'} \rangle$ es igual a uno, por lo que las derivadas con respecto a a_{ik} son nulas cuando $k = 0$, mientras que las derivadas con respecto a b_{jk} son iguales a uno cuando $k = 0$, o sea:

como:

$$\frac{\partial \mathcal{L}}{\partial b_{jk}} = \sum_i \left(\frac{\partial \mathcal{L}}{\partial \mu_{ij}} \frac{\partial \mu_{ij}}{\partial \eta_{ij}} \frac{\partial \eta_{ij}}{\partial b_{jk}} \right) \tag{5.17}$$

$$\frac{\partial \mathcal{L}}{\partial a_{ik}} = \sum_j \left(\frac{\partial \mathcal{L}}{\partial \mu_{ij}} \frac{\partial \mu_{ij}}{\partial \eta_{ij}} \frac{\partial \eta_{ij}}{\partial a_{ik}} \right) \tag{5.18}$$

Llamando: $\eta_{ij} = g(\mu_{ij}) = \mu_{ij} = \sum_i \sum_j a_{ik} b'_{jk} \quad (k = 0,1,2)$

Entonces:

$$\begin{aligned} \frac{\partial \eta_{ij}}{\partial b_{jk}} &= a_{ik}; (k \neq 0) & \frac{\partial \eta_{ij}}{\partial a_{ik}} &= b_{jk}; (k \neq 0) \\ \frac{\partial \eta_{ij}}{\partial b_{jk}} &= 1; (k = 0) & \frac{\partial \eta_{ij}}{\partial a_{ik}} &= 0; (k = 0) \end{aligned} \quad (5.19)$$

Resolviendo las correspondientes derivadas en (5.17) y (5.18), tenemos que:

a) Derivadas respecto de b_{jk} ($k = 0, 1, 2; j = 1, 2, \dots, p$)

Procediendo en la misma forma que en el apartado 2.1.1 de los “*Biplot de Regresión Lineal*”, tendremos:

$$\frac{\partial L}{\partial b_{jk}} = \frac{1}{\sigma^2} \sum_i (y_{ij} - \mu_{ij}) a_{ik}; (k \neq 0)$$

$$\frac{\partial L}{\partial b_{jk}} = \frac{1}{\sigma^2} \sum_i (y_{ij} - \mu_{ij}); (k = 0)$$

Para $j = j'$

$$\left[\frac{\partial^2 L}{\partial b_{jk}^2} \right] = \frac{1}{\sigma^2} \sum_i a_{ik}^2; (k \neq 0)$$

$$\left[\frac{\partial^2 L}{\partial b_{jk}^2} \right] = \frac{1}{\sigma^2} \sum_i 1 = \frac{n}{\sigma^2}; (k = 0)$$

Para $j \neq j'$

$$\left[\frac{\partial^2 L}{\partial b_{jk} \partial b_{j'k}} \right] = 0; (j \neq j')$$

Por otra parte tenemos que:

$$\left[\frac{\partial^2 L}{\partial b_{jk} \partial b_{jk'}} \right] = \frac{1}{\sigma^2} \sum_i a_{ik} a_{ik'}; (k \neq k')$$

$$\left[\frac{\partial^2 L}{\partial b_{jk} \partial b_{jk}} \right] = \frac{1}{\sigma^2} \sum_i 1.1 = \frac{n}{\sigma^2}; (k = k')$$

Mientras que en los “***Biplots de Regresión no Lineal***” (en el caso de datos de presencia-ausencia), procediendo en la misma forma que en el apartado 2.1.2, tendremos:

$$\frac{\partial L}{\partial b_{jk}} = \sum_i \left[\frac{y_{ij} - n_{ij} p_{ij}}{p_{ij} (1 - p_{ij})} \frac{1}{g'(p_{ij})} a_{ik} \right]; (k \neq 0)$$

$$\frac{\partial L}{\partial b_{jk}} = \sum_i \left[\frac{y_{ij} - n_{ij} p_{ij}}{p_{ij} (1 - p_{ij})} \frac{1}{g'(p_{ij})} \right]; (k = 0)$$

Para $j = j'$

$$\left[\frac{\partial^2 L}{\partial b_{jk}^2} \right] = \sum_i \frac{n_{ij}}{p_{ij} (1 - p_{ij}) [g'(p_{ij})]^2} a_{ik}^2; (k \neq 0)$$

$$\left[\frac{\partial^2 L}{\partial b_{jk}^2} \right] = \sum_i \frac{n_{ij}}{p_{ij} (1 - p_{ij}) [g'(p_{ij})]^2}; (k = 0)$$

Para $j \neq j'$

$$\left[\frac{\partial^2 L}{\partial b_{jk} \partial b_{j'k}} \right] = 0$$

Por otra parte considerando el parámetro estimado tenemos que:

$$\left[\frac{\partial^2 L}{\partial b_{jk} \partial b_{j'k}} \right] = \sum_i \frac{n_{ij}}{p_{ij}(1-p_{ij})(g'(p_{ij}))^2} a_{ik} a_{ik'}; (k \neq k')$$

$$\left[\frac{\partial^2 L}{\partial b_{jk} \partial b_{j'k}} \right] = \sum_i \frac{n_{ij}}{p_{ij}(1-p_{ij})(g'(p_{ij}))^2}; (k = k')$$

b) Derivadas respecto de a_{ik} ($k = 0, 1, 2; i = 1, 2, \dots, n$)

En los “*Biplots de Regresión Lineal*” tenemos:

$$\frac{\partial \mathcal{L}}{\partial a_{ik}} = \frac{1}{\sigma^2} \sum_j (y_{ij} - \mu_{ij}) b_{jk}; (k \neq 0)$$

$$\frac{\partial \mathcal{L}}{\partial a_{ik}} = 0; (k = 0)$$

Para $i = i'$

$$\left[\frac{\partial^2 L}{\partial a_{ik}^2} \right] = \frac{1}{\sigma^2} \sum_j b_{ij}^2; (k \neq 0)$$

$$\left[\frac{\partial^2 L}{\partial a_{ik}^2} \right] = 0; (k = 0)$$

Para $i \neq i'$

$$\left[\frac{\partial^2 L}{\partial \hat{a}_{ik} \partial \hat{a}_{i'k}} \right] = 0$$

Y por último considerando el parámetro estimado tenemos que:

$$\left[\frac{\partial^2 L}{\partial \hat{a}_{ik} \partial \hat{a}_{ik'}} \right] = \frac{1}{\sigma^2} \sum_j b_{jk} b_{jk'}; (k \neq k')$$

Mientras que en los “***Biplots de Regresión no Lineal***” (para datos de presencia-ausencia) tenemos que:

$$\frac{\partial L}{\partial \hat{a}_{ik}} = \sum_j \left[\frac{y_{ij} - n_{ij} p_{ij}}{p_{ij}(1-p_{ij})} \frac{1}{g'(p_{ij})} b_{jk} \right]; (k \neq 0)$$

$$\frac{\partial L}{\partial \hat{a}_{ik}} = 0; (k = 0)$$

Para $i = i'$

$$\left[\frac{\partial^2 L}{\partial \hat{a}_{ik}^2} \right] = \sum_j \frac{n_{ij}}{p_{ij}(1-p_{ij}) [g'(p_{ij})]^2} b_{jk}^2; (k \neq 0)$$

$$\left[\frac{\partial^2 L}{\partial \hat{a}_{ik}^2} \right] = 0; (k = 0)$$

Para $i \neq i'$

$$\left[\frac{\partial^2 L}{\partial \hat{a}_{ik} \partial \hat{a}_{i'k}} \right] = 0$$

Y dependiendo del parámetro estimado tenemos que:

$$\left[\frac{\partial^2 L}{\partial \hat{a}_{ik} \partial \hat{a}_{ik'}} \right] = \sum_j \frac{n_{ij}}{p_{ij}(1-p_{ij})(g'(p_{ij}))^2} b_{jk} b_{jk'}; (k \neq k')$$

3.- ESTIMACION MAXIMO-VEROSIMIL SEPARADA EN BIPLOTS DE REGRESION

El problema de la estimación cuando el orden de la matriz de información $\mathfrak{I}(\mathbf{B}, \mathbf{A})$ es muy alto, y por ende la inversión de dicha matriz es muy laboriosa computacionalmente, se puede solucionar utilizando en la estimación de los parámetros un procedimiento iterativo con 2 etapas en cada iteración, tal como lo hacen en otro contexto autores como BLAZQUEZ (1998), GABRIEL (1998) y BIRBAUN (1968).

A continuación se propone un procedimiento de estimación que resulta en una generalización de lo expuesto en el apartado 2 de este capítulo, y de la metodología utilizada por los autores citados.

3.1.- *MODELOS SIN TERMINO INDEPENDIENTE*

Utilizando el modelo sin término independiente, anteriormente formulado en el modelo (5.4):

$$g(\mu) = \mathbf{a}'_1 \mathbf{b}_1 + \mathbf{a}'_2 \mathbf{b}_2 + \dots + \mathbf{a}'_q \mathbf{b}_q$$

Se puede realizar un procedimiento de estimación en dos etapas de la siguiente manera:

En la primera etapa, se estiman los parámetros de las variables b_{jk} suponiendo conocidas las coordenadas de las filas a_{ik} , y utilizando la siguiente ecuación correspondiente al sistema de ecuaciones (5.5) del apartado 2.1:

$$\langle \mathbf{B}' \rangle_{(t+1)} = \langle \mathbf{B}' \rangle_t - [\mathfrak{Z}(\mathbf{B})]_t^{-1} [\mathbf{U}(\mathbf{B})]_t$$

O sea que:

$$\hat{b}_{j_1}^{(t+1)} = \hat{b}_{j_1}^{(t)} - \left(\partial^2 L / \partial b_{j_1}^2 \right)^{-1} \left(\partial L / \partial b_{j_1} \right) - \left(\partial^2 L / \partial b_{j_1} \partial b_{j_2} \right)^{-1} \left(\partial L / \partial b_{j_2} \right)$$

$$\hat{b}_{j_2}^{(t+1)} = \hat{b}_{j_2}^{(t)} - \left(\partial^2 L / \partial b_{j_2}^2 \right)^{-1} \left(\partial L / \partial b_{j_2} \right) - \left(\partial^2 L / \partial b_{j_2} \partial b_{j_1} \right)^{-1} \left(\partial L / \partial b_{j_1} \right)$$

Ese sistema de ecuaciones, es equivalente a realizar la siguiente regresión generalizada:

$$g < \mu > = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A} \end{bmatrix} < \mathbf{B}' > = [\mathbf{diag} \mathbf{A}] < \mathbf{B}' > = (\mathbf{I}_p \otimes \mathbf{A}) < \mathbf{B}' >$$

Donde

$$[\mathbf{I}]_p = [\mathbf{diag} \mathbf{1}_p]$$

$$[\mathbf{I}_p \otimes \mathbf{A}] = [\mathbf{I}_{ij} \mathbf{A}] = [\mathbf{diag} \mathbf{A}]$$

Una vez obtenidas esas estimaciones, se procede a *la etapa 2* para estimar las coordenadas de los individuos a_{jk} , suponiendo conocidos los parámetros de las variables b_{jk} , a través de la siguiente ecuación:

$$\langle \hat{\mathbf{A}}' \rangle_{(t+1)} = \langle \hat{\mathbf{A}}' \rangle_t - [\mathfrak{J}(\mathbf{A})]_t^{-1} [\mathbf{U}(\mathbf{A})]_t$$

De donde:

$$\begin{aligned} \hat{a}_{i1}^{(t+1)} &= \hat{a}_{i1}^{(t)} - (\partial^2 L / \partial a_{i1}^2)^{-1} (\partial L / \partial a_{i1}) - (\partial^2 L / \partial a_{i1} \partial a_{i2})^{-1} (\partial L / \partial a_{i2}) \\ \hat{a}_{i2}^{(t+1)} &= \hat{a}_{i2}^{(t)} - (\partial^2 L / \partial a_{i2} \partial a_{i1})^{-1} (\partial L / \partial a_{i1}) - (\partial^2 L / \partial a_{i2}^2)^{-1} (\partial L / \partial a_{i2}) \end{aligned}$$

La resolución a ese sistema de ecuaciones, es equivalente a realizar la regresión generalizada siguiente:

$$g \langle \mu' \rangle = \begin{bmatrix} \mathbf{B} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{B} \end{bmatrix} \langle \mathbf{A}' \rangle = [\mathbf{diagB}] \langle \mathbf{A}' \rangle = (\mathbf{I}_n \otimes \mathbf{B}) \langle \mathbf{A}' \rangle$$

El procedimiento de estimación en 2 etapas, se realiza iterativamente hasta que converja.

Los errores de los estimadores se calculan de la siguiente forma:

$$\text{Var}(\langle \hat{\mathbf{B}}' \rangle) = \sigma^2 [\mathfrak{I}(\mathbf{B})]^{-1}$$

$$\text{Var}(\langle \hat{\mathbf{A}}' \rangle) = \sigma^2 [\mathfrak{I}(\mathbf{A})]^{-1}$$

donde:

$$\hat{\sigma}^2 = \{ \langle \mathbf{e}' \rangle [\text{diag} \langle \mathbf{W} \rangle] \langle \mathbf{e} \rangle \} / [2nx2p - r(2n + 2p - r)]$$

$$r = \text{rango}(\Theta) = \text{rango}(\mathbf{A}\mathbf{B}')$$

El método aquí propuesto, es una alternativa:

- a) Al método de los Mínimos Cuadrados Alternados, ya que permite unir las regresiones separadas para cada columna y cada fila de la matriz de datos original \mathbf{Y} ($n \times p$) en dos regresiones generalizadas, una para estimar los parámetros b_{jk} correspondientes a las “p” columnas, y la otra para estimar los parámetros a_{ik} de las “n” filas de \mathbf{Y} . Tiene además la ventaja que permite el cálculo de los errores de los estimadores.
- b) Al procedimiento IRLS (Iteratively Reweighted Least Square) propuesto por GABRIEL en 1998 (extensión al método criss-cross), ya que bajo los supuestos a) y c) de independencia de los individuos e independencia de los parámetros de las variables, el paso criss es equivalente a la etapa 1, y el paso cross a la etapa 2.

3.2.- MODELOS CON TERMINO INDEPENDIENTE

Se puede generalizar el procedimiento de estimación separada máximo verosímil de los parámetros para modelos sin término independiente, al ajuste de “Biplot de Regresión” a través de modelos con término independiente, o sea:

$$g(\mu) = \mathbf{a}'_0 \mathbf{b}_0 + \mathbf{a}'_1 \mathbf{b}_1 + \mathbf{a}'_2 \mathbf{b}_2 + \dots + \mathbf{a}'_q \mathbf{b}_q$$

Como $\mathbf{a}_0 = 1$ entonces:

$$g(\mu) = \mathbf{b}_0 + \sum_{i,j,k} a_{ik} b_{jk} = \mathbf{A}_o \mathbf{B}'_o + \mathbf{AB}' = \mathbf{1}_n \mathbf{B}'_o + \mathbf{AB}'$$

para $i = 1,2, \dots, n$; $j = 1,2, \dots, p$; $k = 0,1, \dots, q$.

Utilizando la misma notación que en el apartado 2.2 de este capítulo, la anterior especificación se puede reescribir en notación matricial como:

$$g(\mu) = \mathbf{A} * \mathbf{B}' = \mathbf{I}_n \mathbf{A}_o \mathbf{B}'_o \mathbf{I}_p + \mathbf{I}_n \mathbf{AB}' \mathbf{I}_p = \mathbf{I}_n \mathbf{1}_n \mathbf{B}'_o \mathbf{I}_p + \mathbf{I}_n \mathbf{AB}' \mathbf{I}_p$$

o como:

$$g(\mu') = \mathbf{B} * \mathbf{A}' = \mathbf{I}_p \mathbf{B}_o \mathbf{A}'_0 \mathbf{I}_n + \mathbf{I}_p \mathbf{BA}' \mathbf{I}_n = \mathbf{I}_p \mathbf{B}_o \mathbf{1}'_n \mathbf{I}_n + \mathbf{I}_p \mathbf{BA}' \mathbf{I}_n$$

donde $\mathbf{A}_o = \mathbf{1}_n$

El procedimiento de estimación en dos etapas será:

Etapas 1: Se estiman los parámetros de las variables b_{jk} , suponiendo conocidas las coordenadas de las filas a_{ik} , a través del sistema de ecuaciones:

$$\langle \hat{\mathbf{B}}^{*'} \rangle_{(t+1)} = \langle \hat{\mathbf{B}}^{*'} \rangle_t - [\mathfrak{I}(\mathbf{B}^*)]_t^{-1} [\mathbf{U}(\mathbf{B}^*)]_t$$

Etapas 2: Se estiman las coordenadas de los individuos a_{ik} , suponiendo conocidos los parámetros de las variables b_{jk} a través del siguiente sistema de ecuaciones:

$$\langle \hat{\mathbf{A}}^{*'} \rangle_{(t+1)} = \langle \hat{\mathbf{A}}^{*'} \rangle_t - [\mathfrak{I}(\mathbf{A}^*)]_t^{-1} [\mathbf{U}(\mathbf{A}^*)]_t$$

La resolución de esos sistemas de ecuaciones, es equivalente al procedimiento de Regresión Bilineal Segmentada propuesto por GABRIEL (1998), ya que:

La primera etapa es equivalente a realizar la siguiente regresión generalizada:

$$g(\mu) = \mathbf{A}^* \mathbf{B}^{*'} = \mathbf{I}_n \mathbf{1}_n \mathbf{B}'_o \mathbf{I}_p + \mathbf{I}_n \mathbf{A} \mathbf{B}' \mathbf{I}_p$$

lo que es a su vez igual a:

$$g < \mu > = \begin{bmatrix} \mathbf{1}_n & \cdots & \mathbf{0} & \mathbf{A} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{1}_n & \mathbf{0} & \cdots & \mathbf{A} \end{bmatrix} \begin{bmatrix} < \mathbf{B}_0' > \\ < \mathbf{B}' > \end{bmatrix} =$$

$$g < \mu > = [(\mathbf{I}_p \otimes \mathbf{1}_n); (\mathbf{I}_p \otimes (\mathbf{I}_n \mathbf{A}))] < \mathbf{B}^{*'} >$$

Y la segunda etapa, es equivalente a realizar la regresión generalizada siguiente:

$$g(\mu') = \mathbf{B}^* \mathbf{A}^{*'} = \mathbf{I}_p \mathbf{B}_0 \mathbf{1}_n' \mathbf{I}_n + \mathbf{I}_p \mathbf{B} \mathbf{A}' \mathbf{I}_n$$

que es igual a:

$$g < \mu' > = \begin{bmatrix} \mathbf{B}_0 & \cdots & \mathbf{0} & \mathbf{B} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{B}_0 & \mathbf{0} & \cdots & \mathbf{B} \end{bmatrix} \begin{bmatrix} < \mathbf{A}'_0 = \mathbf{1}_n' > \\ < \mathbf{A}' > \end{bmatrix}$$

de donde :

$$g < \mu' > = [(\mathbf{I}_n \otimes (\mathbf{I}_p \mathbf{B}_0)); (\mathbf{I}_n \otimes (\mathbf{I}_p \mathbf{B}))] < \mathbf{1}_n', \mathbf{A}' >$$

4.- ESTIMACION EN BIPLOTS DE REGRESION CON INFORMACION EXTERNA

El procedimiento de estimación de los parámetros en los Biplots de Regresión en forma conjunta o separada, se puede generalizar a los Biplots de Regresión con Información Externa ajustados a través del siguiente modelo:

$$g(\mu) = \mathbf{X} \mathbf{A} \mathbf{B}' \mathbf{Z}' = \mathbf{X} \Theta \mathbf{Z}' \quad (5.20)$$

siendo

$\mathbf{X}_{(n \times m)}$ y $\mathbf{Z}_{(p \times s)}$ las matrices de variables externas

$\Theta_{(m \times q)} = \mathbf{A}_{(m \times q)} \mathbf{B}'_{(q \times s)}$ la matriz de parámetros

Las matrices \mathbf{A} y \mathbf{B} se describen en la misma forma que en el apartado 2.1 para los modelos sin término independiente y que \mathbf{A}^* y \mathbf{B}^* en el apartado 2.2 para los modelos con término independiente.

El sistema de ecuaciones (5.5) obtenido en el apartado 2.1 para los “*modelos sin término independiente*”, se puede generalizar para incluir las matrices de variables externas, pudiéndose hacer el ajuste de los “*Biplot de Regresión*” en forma conjunta o separada, realizando respectivamente las siguientes regresiones generalizadas simultáneamente o en forma separada:

$$\begin{aligned} g < \mu > &= [\mathbf{Z} \otimes (\mathbf{X} \mathbf{A}^*)] < \mathbf{B}^{*'} > \\ g < \mu' > &= [\mathbf{X} \otimes (\mathbf{Z} \mathbf{B}^*)] < \mathbf{A}^{*'} > \end{aligned}$$

En el caso del ajuste de “*Biplot de Regresión con Información Externa*”, a través de “*modelos con término independiente*”, se generaliza el sistema de ecuaciones (5.16), lo que es equivalente a realizar las siguientes regresiones generalizadas en forma conjunta o en forma separada:

$$\begin{aligned} g < \mu > &= [(\mathbf{Z} \otimes (\mathbf{X} \mathbf{1}_n)); (\mathbf{Z} \otimes (\mathbf{X} \mathbf{A}))] < \mathbf{B}_0 \quad \mathbf{B}' > \\ g < \mu' > &= [(\mathbf{X} \otimes (\mathbf{Z} \mathbf{B}_0)); (\mathbf{X}' \otimes (\mathbf{Z} \mathbf{B}))] < \mathbf{1}_n \quad \mathbf{A}' > \end{aligned}$$

5.- PROPIEDADES DE LOS ESTIMADORES MAXIMO VEROSIMILES EN LOS BIPLLOT DE REGRESION

En el procedimiento de estimación máximo verosímil de los parámetros en los Biplots de Regresión se distinguen dos tipos de parámetros: incidentales y estructurales.

El número de parámetros estructurales (b_{jk}) es fijo y depende del número de variables en estudio, mientras que el número de parámetros incidentales (a_{ik}) depende del tamaño de la muestra, luego su número se incrementa en la medida que el tamaño de la muestra aumente.

La propiedad de consistencia es difícil que se cumpla, en el caso de la estimación simultánea de los parámetros incidentales y estructurales, ya que el número de parámetros incidentales crece en la medida que el tamaño “n” de la muestra tienda a infinito (NEYMAN & SCOTT, 1948).

Para obtener estimadores consistentes, ANDERSEN (1972) propone primero la obtención de estadísticos suficientes para los parámetros a_{ik} , para luego utilizar posteriormente un procedimiento de estimación máximo verosímil condicional para los parámetros b_{jk} .

Si el número de individuos “n” es grande respecto al número “p” de columnas, una vez obtenidas las estimaciones de los b_{jk} vía máxima verosimilitud condicional, entonces las estimaciones de los a_{ik} se pueden obtener por el procedimiento de máxima verosimilitud clásico.

Por otra parte, en la estimación de los parámetros en el ajuste de “Biplot de Regresión”, también está presente el problema de la Multicolinealidad por tratarse de un análisis multivariante donde todas las variables varían conjuntamente.

La Multicolinealidad desestabiliza la estimación de los parámetros, razón por la cual esas estimaciones sólo se deben utilizar para las representaciones Biplot en la obtención de un método de ordenación, pero absteniéndose de la interpretación de los marcadores si la multicolinealidad es muy alta.

Aún cuando, la estimación conjunta de los parámetros no garantiza la consistencia de los parámetros, la metodología propuesta en esta investigación no persigue el uso de los resultados obtenidos en la forma inferencial tradicional.

El método de estimación propuesto, se vale de las buenas propiedades que los estimadores deben tener en otros contextos, sin embargo el uso de las

aproximaciones Biplot se hace en la forma clásica, aunque su aproximación se haya hecho a través de una forma no tradicional.

De allí que, las estimaciones obtenidas para los parámetros solo se utilizan en esta investigación para la obtención de los gráficos Biplot y la consecuente aproximación y exploración de las características sobresalientes en los datos.

No obstante, dichas estimaciones no son válidas para el uso en sí de los valores obtenidos y la consiguiente contrastación de hipótesis, tal como se hace en la inferencia estadística clásica.

6.- ESTIMACION PARA DATOS MIXTOS

Si en la matriz de datos \mathbf{Y} , cada una de las p variables tiene una distribución de probabilidades diferente, el procedimiento de estimación máximo-verosímil en forma conjunta y separada de los parámetros en los Biplot de Regresión, *se puede tratar de generalizar* para la estimación de los parámetros de las p variables por separado.

Dicha generalización se apoyaría en el supuesto de independencia entre individuos, ya que de esa forma la función de verosimilitud será el producto de las funciones de probabilidad de cada una de las variables.

7.- ESTIMACION MAXIMO VEROSIMIL CONJUNTA EN BIPLOTS DE REGRESION: UNA APLICACION

Para evaluar la metodología propuesta de estimación máximo verosímil conjunta en los Biplots de Regresión Logísticos, se utiliza el ejemplo de GOWER & HAND (1996), quienes formalizan el Análisis de Correspondencias Múltiples desde el punto de vista del Análisis de Proximidades (Multidimensional Scaling). Se comparan sus resultados con los aquí obtenidos, desde el punto de vista descriptivo, a través de las predicciones obtenidas.

El ejemplo consta de una matriz de datos de orden (20x4), la misma utilizada por JONGMAN, TER BRAAK & VAN TONGEREN (1987) y que se presenta en la tabla 5.1. Las 20 filas corresponden a granjas de la isla de Terschelling y las 4 columnas a las siguientes 4 variables categóricas con sus correspondientes categorías o niveles:

- Nivel de Humedad con 4 categorías, de menor a mayor: H1, H2, H3, H4.
- Tipo de Granja con 4 categorías: granja normal (GN), granja biológica (GB), granja hobby (GH), granja conservacionista de la naturaleza (GC).
- Uso de la Granja con 3 categorías: para producción (U1), uso intermedio (U2), para pasto (U3).
- Nivel de Abonado con 5 categorías, de menor a mayor: A0, A1, A2, A3, A4 (el nivel A0 corresponde a ningún abono).

Granja	Humedad	Granja-Tipo	Granja-Uso	Abono
1	H1	GN	U2	A4
2	H1	GB	U2	A2
3	H2	GN	U2	A4
4	H2	GN	U2	A4
5	H1	GH	U1	A2
6	H1	GH	U2	A2
7	H1	GH	U3	A3
8	H4	GH	U3	A3
9	H3	GH	U1	A1
10	H2	GB	U1	A1
11	H1	GB	U3	A1
12	H3	GN	U2	A2
13	H4	GN	U2	A3
14	H4	GC	U3	A0
15	H4	GC	U2	A0
16	H4	GN	U3	A3
17	H2	GC	U1	A0
18	H1	GC	U1	A0
19	H4	GC	U1	A0
20	H4	GC	U1	A0

Tabla 5.1.- Matriz de Datos (20x4)
(tomada de GOWER & HAND, 1996, p. 75)

De la tabla original se deriva la siguiente matriz indicadora **Y** de orden (20x16):

Granj a	H 1	H 2	H 3	H 4	G N	G B	G H	G C	U 1	U 2	U 3	A 0	A 1	A 2	A 3	A 4
1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1
2	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0
3	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	1
4	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	1
5	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0
6	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0
7	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0
8	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0
9	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0	0
10	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0
11	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0
12	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0	0
13	0	0	0	1	1	0	0	0	0	1	0	0	0	0	1	0
14	0	0	0	1	0	0	0	1	0	0	1	1	0	0	0	0
15	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	0
16	0	0	0	1	1	0	0	0	0	0	1	0	0	0	1	0
17	0	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0
18	1	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0
19	0	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0
20	0	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0

Tabla 5.2.- Matriz Indicadora de Datos **Y**

Se ajusta a los datos un Biplot de Regresión no Lineal, del tipo Logístico para cada uno de las categorías, usando un modelo con término independiente (3 parámetros) y una función link logit, o sea:

Para $i = 1, \dots, n$; $j = 1, \dots, p$; $k = 0, 1, 2$

$$g(\mu) = \mathbf{A} * \mathbf{B}^{*'} = b_o + \sum_{l,j,k}^q a'_{ik} b_{jk} = \mathbf{A}_0 \mathbf{B}'_0 + \mathbf{A} \mathbf{B}' = \mathbf{1}_n \mathbf{B}'_o + \mathbf{A} \mathbf{B}'$$

donde

$$\mu = p = [e^\eta / (1 + e^\eta)]$$

$$\eta = \mathbf{A} * \mathbf{B}^{*'} = \mathbf{b}_o + \mathbf{a}'_1 \mathbf{b}_1 + \mathbf{a}'_2 \mathbf{b}_2$$

$$g(\mu) = \text{logit}(p) = \log [p / (1-p)]$$

$$\text{logit}(p) = b_{j0} + a_{i1} b_{j1} + a_{i2} b_{j2}$$

En la estimación de los parámetros del modelo, se utiliza el siguiente sistema de ecuaciones, asumiendo que los tres supuestos citados en el apartado 2.1 se cumplen:

$$\langle \hat{\mathbf{B}}^{*'} \rangle_{(t+1)} = \langle \hat{\mathbf{B}}^{*'} \rangle_{(t)} - [\mathfrak{I}(\mathbf{B}^*)]_{(t)}^{-1} [\mathbf{U}(\mathbf{B}^*)]_{(t)}$$

$$\langle \hat{\mathbf{A}}^{*'} \rangle_{(t+1)} = \langle \hat{\mathbf{A}}^{*'} \rangle_{(t)} - [\mathfrak{I}(\mathbf{A}^*)]_{(t)}^{-1} [\mathbf{U}(\mathbf{A}^*)]_{(t)}$$

o sea, para $j = 1, 2, \dots, p$ y $k = 0, 1, 2$:

$$\begin{aligned} \hat{b}_{j_0}^{(t+1)} &= \hat{b}_{j_0}^{(t)} - (\partial^2 L / \partial b_{j_0}^2)^{-1} (\partial L / \partial b_{j_0}) - (\partial^2 L / \partial b_{j_0} \partial b_{j_1})^{-1} (\partial L / \partial b_{j_1}) - (\partial^2 L / \partial b_{j_0} \partial b_{j_2})^{-1} (\partial L / \partial b_{j_2}) \\ \hat{b}_{j_1}^{(t+1)} &= \hat{b}_{j_1}^{(t)} - (\partial^2 L / \partial b_{j_1}^2)^{-1} (\partial L / \partial b_{j_1}) - (\partial^2 L / \partial b_{j_1} \partial b_{j_0})^{-1} (\partial L / \partial b_{j_0}) - (\partial^2 L / \partial b_{j_1} \partial b_{j_2})^{-1} (\partial L / \partial b_{j_2}) \\ \hat{b}_{j_2}^{(t+1)} &= \hat{b}_{j_2}^{(t)} - (\partial^2 L / \partial b_{j_2}^2)^{-1} (\partial L / \partial b_{j_2}) - (\partial^2 L / \partial b_{j_2} \partial b_{j_1})^{-1} (\partial L / \partial b_{j_1}) - (\partial^2 L / \partial b_{j_2} \partial b_{j_0})^{-1} (\partial L / \partial b_{j_0}) \end{aligned}$$

y para $i = 1, 2, \dots, n$:

$$\begin{aligned} \hat{a}_{i_0}^{(t+1)} &= \hat{a}_{i_0}^{(t)} - (\partial^2 L / \partial a_{i_0}^2)^{-1} (\partial L / \partial a_{i_0}) - (\partial^2 L / \partial a_{i_0} \partial a_{i_1})^{-1} (\partial L / \partial a_{i_1}) - (\partial^2 L / \partial a_{i_0} \partial a_{i_2})^{-1} (\partial L / \partial a_{i_2}) \\ \hat{a}_{i_1}^{(t+1)} &= \hat{a}_{i_1}^{(t)} - (\partial^2 L / \partial a_{i_1}^2)^{-1} (\partial L / \partial a_{i_1}) - (\partial^2 L / \partial a_{i_1} \partial a_{i_2})^{-1} (\partial L / \partial a_{i_2}) - (\partial^2 L / \partial a_{i_1} \partial a_{i_0})^{-1} (\partial L / \partial a_{i_0}) \\ \hat{a}_{i_2}^{(t+1)} &= \hat{a}_{i_2}^{(t)} - (\partial^2 L / \partial a_{i_2} \partial a_{i_1})^{-1} (\partial L / \partial a_{i_1}) - (\partial^2 L / \partial a_{i_2}^2)^{-1} (\partial L / \partial a_{i_2}) - (\partial^2 L / \partial a_{i_2} \partial a_{i_0})^{-1} (\partial L / \partial a_{i_0}) \end{aligned}$$

Se sustituyen las derivadas, considerando que las variables tienen la distribución binomial, tal como se indicó en el apartado 2.1.2.

Para la resolución del sistema de ecuaciones se elaboró un programa computacional en MATLAB.

Los valores iniciales $\langle \mathbf{A}^{0*'} \rangle$ se obtuvieron mediante la ordenación de las filas de la matriz \mathbf{a} a través de un Análisis de Coordenadas Principales, aplicado a la matriz de similitudes obtenida del coeficiente de Jaccard. Con estos valores se calculan los iniciales para $\langle \mathbf{B}^{0*'} \rangle$ y luego se inicia el proceso iterativo de Newton-Raphson hasta que converja.

En las tablas 5.3 y 5.4, aparecen respectivamente las estimaciones de los parámetros contenidos en las matrices de marcadores fila (\mathbf{A}^*) y columna (\mathbf{B}^*) del Biplot de Regresión.

Granja	a_{i0}	a_{i1}	a_{i2}
1	1	24.613	-0.372
2	1	9.479	15.229
3	1	30.695	-14.968
4	1	30.695	-14.968
5	1	-6.288	20.85
6	1	8.628	19.924
7	1	-2.804	14.822
8	1	-1.832	15.001
9	1	-23.689	6.767
10	1	-14.749	3.637
11	1	-5.12	10.462
12	1	21.625	12.473
13	1	7.162	-7.074
14	1	-7.167	-12.841
15	1	-1.933	-15.647
16	1	-0.645	-4.128
17	1	-11.666	-11.127
18	1	-12.955	-6.025
19	1	-15.558	-16.954
20	1	-15.558	-16.954

Tabla 5.3.- Parámetros estimados para la matriz de marcadores fila \mathbf{A}^*

Variable	b_{j0}	b_{j1}	b_{j2}
H1	-0.953	0.015	0.121
H2	-1.809	0.038	-0.082
H3	-3.564	0.077	0.133
H4	-1.142	-0.081	-0.136
GN	-1.749	0.303	-0.146
GB	-2.109	-0.034	0.079
GH	-6.285	-0.202	0.477
GC	-4.382	-0.229	-0.485
U1	-4.426	-0.571	0.069
U2	-0.378	0.515	-0.118
U3	-1.186	-0.029	0.04
A0	-4.382	-0.229	-0.485
A1	-4.077	-0.259	0.122
A2	-6.754	0.189	0.423
A3	-1.434	-0.001	0.034
A4	-5.477	0.277	-0.181

Tabla 5.4.- Parámetros estimados para la matriz \mathbf{B}^*

Para medir la Bondad del Ajuste, se usa la Deviance y el porcentaje de Bien Clasificados.

Para el cálculo de la Deviance, se compara el modelo logístico utilizado, con un modelo alternativo con un solo parámetro (el término independiente), o sea:

$$p_{ij} = \exp b_{j0} / (1 + \exp b_{j0})$$

Los resultados se presentan en la tabla 5.5.

Variable	Deviance	p valor	% B.C ⁽¹⁾	C.R.(x1000) ⁽¹⁾
H1	8.053	0.0045	80	315
H2	2.658	0.103	90	197
H3	3.404	0.065	90	318
H4	6.87	0.0088	75	422
GN	3.731	0.0534	95	743
GB	2.239	0.1346	85	148
GH	17.552	0	100	825
GC	17.046	0	100	947
U1	1.532	0.2158	100	868
U2	2.352	0.1251	100	889
U3	0.966	0.3257	75	65
A0	17.046	0	100	947
A1	3.668	0.0555	90	524
A2	15.855	0.0001	100	851
A3	0.649	0.4206	80	32
A4	1.533	0.2156	100	935

⁽¹⁾ % B.C. = Porcentaje de Bien Clasificados C.R.= Calidad de la Representación

Tabla 5.5.- Bondad de Ajuste para el Biplot de Regresión Logístico

La Bondad del Ajuste de la solución final, calculada como Porcentaje de Bien Clasificados en el Biplot, es 91.25%.

Del análisis de los resultados de la tabla 5.5, se observa que la Deviance no es significativa (p valor > 0.10) para las variables GB, U1, U2, U3, A3, y A4. Sin embargo, los porcentajes de granjas Bien Clasificados son altos para todas las variables, por lo que la predicción de presencia-ausencia de cada variable para cada una de las granjas, es bastante aproximada.

La Deviance en términos relativos ($\times 1000$) para cada variable, representa las contribuciones relativas o calidad de representación de las variables. En la tabla 5.5 aparecen dichas calidades de representación (C.R) acumuladas para los dos primeros ejes.

Las variables mal representadas, son las que coinciden en la representación del Biplot de Regresión Logístico con los segmentos de mayor tamaño, como A3, U3, H2 y GB, mientras que las variables bien representadas aparecen con segmentos de menor longitud, como GC, U1, U2, A0, A2 y A4 (tal como se muestra en la figura 5.1). Esto contrasta con la interpretación habitual de los Biplots Clásicos, en los que las variables bien representadas tienen segmentos grandes, mientras que las mal representadas aparecen en torno al origen.

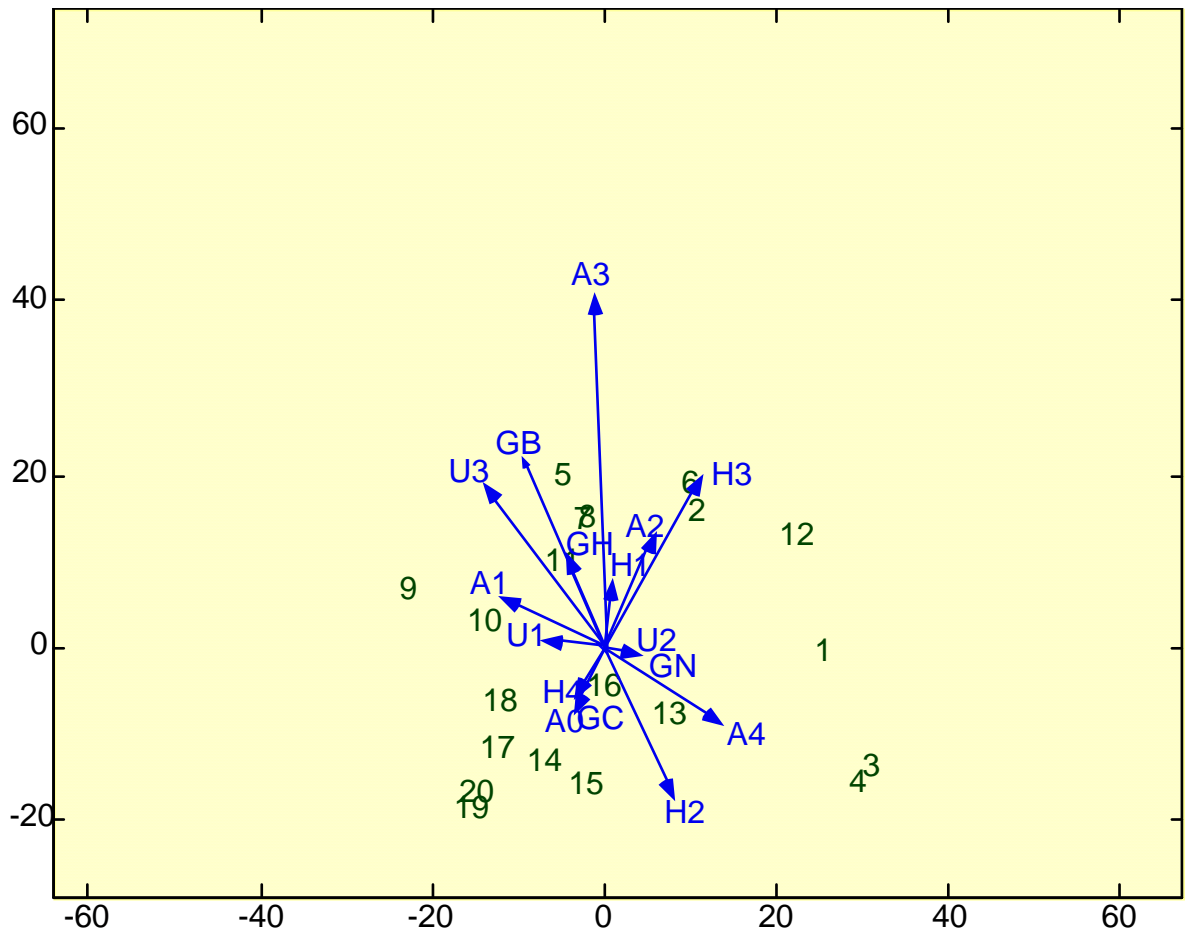


Figura 5.1.- Representación Simultánea del Biplot de Regresión Logístico

La dirección de los ejes Biplot β_j , sobre la representación gráfica de la figura 5.1, está determinada por los parámetros b_{jk} estimados.

La asociación entre las distintas variables puede aproximarse a través del ángulo que forman. Así tenemos, fuerte asociación positiva entre las variables H3 y A2; U2, GN y A4; GC, A0 y H4, por lo que la predicción se realiza en el mismo sentido. Fuerte asociación negativa entre las variables H3 y H4, A1 y A4, U1 y U2, luego la predicción se hace en sentido contrario. Ausencia de

asociación entre las variables H3 y A4, H1 y GN, GN y H4, H3 y A1, luego sus predicciones son independientes.

El producto escalar entre marcadores fila \mathbf{a}_i y marcadores columna \mathbf{b}_j , que aparecen en las tablas 5.3 y 5.4, aproxima salvo un factor de escala la probabilidad p_{ij} . Sin embargo, se puede abreviar ese procedimiento introduciendo en los ejes Biplot escalas de predicción, que permitan visualmente predecir la presencia o ausencia de determinado carácter o categoría.

Para ello, se considera la geometría de los Biplot de Regresión No Lineales (véase figura 3.6, apartado 3.2, capítulo III), en términos de proyecciones de subespacios (GOWER & HAND, 1996; VICENTE-VILLARDON y GALINDO, 1998), donde los β_j o ejes Biplot de predicción, son rectas normales que pasan por el origen y las proyecciones de los puntos que marcan los hiperplanos \mathcal{N}_j al cortar a los ejes ξ_j en el espacio original \mathcal{H} , y que pertenecen a la intersección $\mathcal{L} \cap \mathcal{N}_j$.

Sin embargo, las escalas inducidas por las intersecciones no son las proyecciones sobre \mathcal{L} , por lo que es necesario determinar las coordenadas de esas proyecciones sobre el Biplot, para obtener una escala de predicción sobre β_j .

En esa escala, por tratarse de datos dicotómicos, se marcarán solo los valores correspondientes a los cuartiles, ya que así al proyectar un punto fila por

encima del cuartil 2 (o mediana) se predice presencia del carácter, y en caso contrario se predice ausencia.

Se divide así el espacio sobre la representación en dos regiones de predicción para cada variable, una para la presencia y otra para la ausencia.

Como los parámetros para cada eje Biplot se presentan en escalas muy diferentes a las de las obtenidas para las coordenadas de los individuos, se transforman de la siguiente manera:

Se parte de un punto cualquiera (y_1, y_2) sobre el eje Biplot, representado a través de la recta que pasa por el punto $(0,0)$ y el punto (b_{j1}, b_{j2}) , y que analíticamente es igual a: $y_2 = (b_{j2}/ b_{j1}) y_1$.

Como el modelo utilizado es igual a: $\text{logit}(p) = b_{j0} + b_{j1} y_1 + b_{j2} y_2$, entonces sustituyendo y_2 por la expresión anterior tenemos:

$$\text{logit}(p) = b_{j0} + b_{j1} y_1 + b_{j2} (b_{j2}/ b_{j1}) y_1$$

de donde despejando obtenemos:

$$y_1 = [\text{logit}(p) - b_{j0}] b_{j1} / (b_{j1}^2 + b_{j2}^2) \quad y_2 = [\text{logit}(p) - b_{j0}] b_{j2} / (b_{j1}^2 + b_{j2}^2)$$

Las coordenadas buscadas se obtienen sustituyendo en esas expresiones a “p”, por los valores correspondientes a los tres cuartiles, o sea: 0.25, 0.50 y 0.75.

Por ejemplo, para el punto que predice la probabilidad 0.5, entonces $\text{logit}(0.5) = 0$, luego sustituyendo en las fórmulas anteriores obtenemos en la tabla 5.6 las coordenadas para el cuartil 2 de cada variable:

Variable	y_1	y_2
H1	0.96	7.76
H2	8.42	-18.16
H3	11.62	20.07
H4	-3.69	-6.20
GN	4.68	-2.26
GB	-9.69	22.52
GH	-4.73	11.17
GC	-3.49	-7.39
U1	-7.64	0.92
U2	0.70	-0.16
U3	-14.09	19.43
A0	-3.49	-7.39
A1	-12.88	6.07
A2	5.95	13.31
A3	-0.001	42.14
A4	13.86	-9.05

Tabla 5.6.- Escalas de Predicción para el cuartil 2, sobre la representación Biplot

Una vez obtenidas las coordenadas para la escala de predicción sobre cada eje Biplot, se hace la representación gráfica correspondiendo el extremo de cada eje al cuartil 2, tal como aparece en la figura 5.1.

La probabilidad estimada de presencia, de cada categoría para cada una de las 20 granjas, se puede predecir en la forma clásica proyectando cada uno de los puntos fila sobre el eje Biplot β_j . Si la proyección cae por encima del cuartil 2, entonces la estimación de p es $p \geq 0.5$, o sea se predice presencia de la categoría, en caso contrario se predice ausencia.

Algunas de las variables, generalmente las mal representadas, tienen todas sus predicciones correspondientes a ausencias y son las que coinciden en el gráfico con los segmentos de mayor tamaño, tal como sucede con A3, U3, H2 y GB. Para estas variables no es recomendable, la utilización de las predicciones obtenidas a través de las proyecciones sobre dichos segmentos.

A efectos de una predicción más aproximada, en la siguiente figura 5.2 se hacen marcas para la predicción en los cuartiles 1, 2 y 3. El extremo inferior y superior de cada eje Biplot corresponde respectivamente a los cuartiles 1 y 3, mientras que la marca central corresponde al cuartil 2.

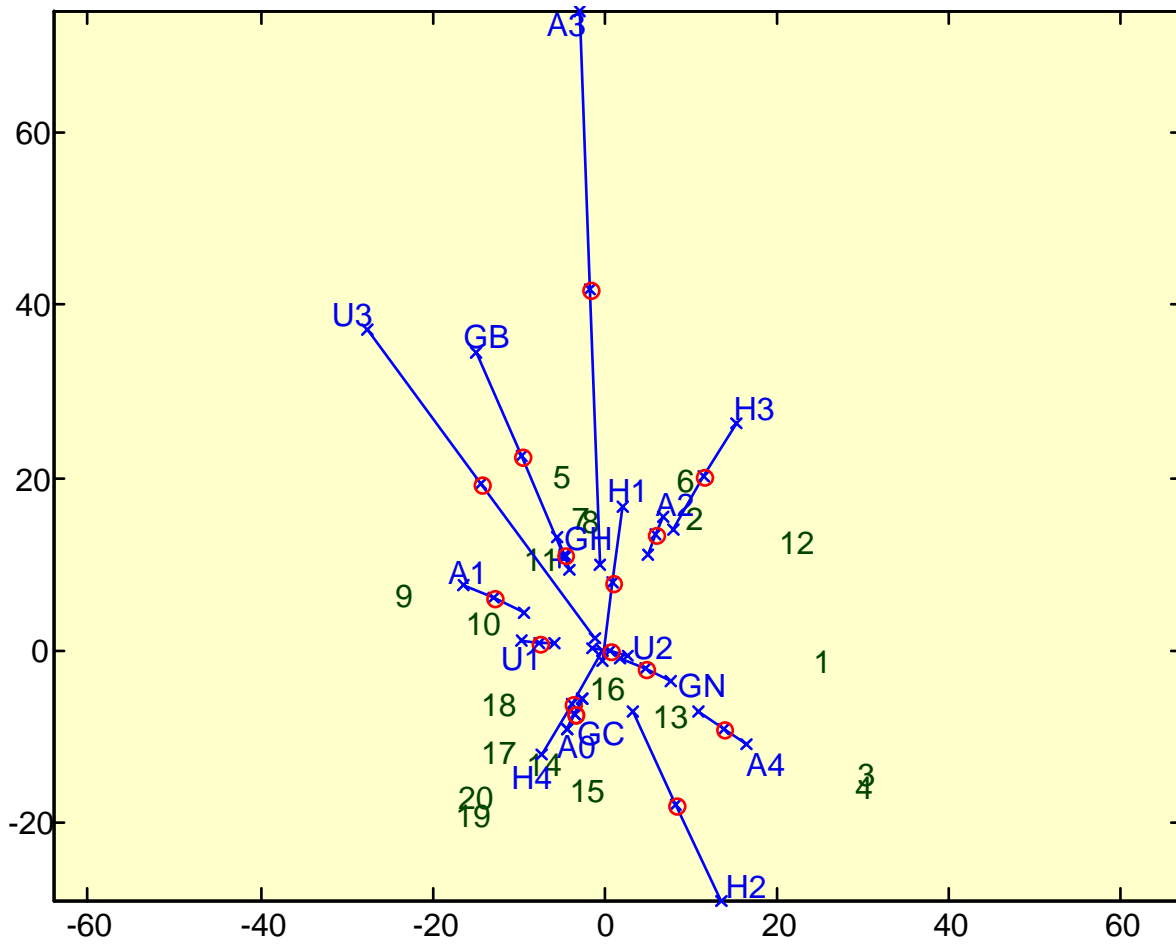


Figura 5.2.- Representación de los cuartiles sobre el Biplot de Regresión Logístico

Las conclusiones de esas figuras gráficas son similares a las obtenidas por GOWER & HAND (1996) y por VICENTE-VILLARDON y GALINDO (1998), reflejando la estructura de la matriz de datos de la Tabla 5.1, así tenemos que:

- Las granjas normales (GN) aparecen ubicadas en la región con niveles de humedad moderados (H2), alto uso de abono (A4) y uso intermedio U2.
- Las granjas biológicas (GB) y las granjas para hobby (GH) aparecen en la región de bajo uso de abono (A1), humedad baja (H1) y uso para la producción (U1) y para pasto (U3).
- Las granjas conservacionistas (GC) están en la región no abonada (A0) con alta humedad (H4).

De lo antes dicho, se puede concluir que los resultados son claramente concordantes con los obtenidos por GOWER & HAND (1996), aún cuando las metodologías de estimación utilizadas sean diferentes.

La metodología aquí utilizada permite considerar en el modelo la distribución probabilística, a través de la cual se han generado los datos. Sin embargo, aunque la estimación de los parámetros, considere la distribución probabilística los resultados obtenidos no se usan en la forma inferencial tradicional.

El método de estimación propuesto se vale de las buenas propiedades que los estimadores deben tener en otros contextos, no obstante, dichas estimaciones no son válidas para el uso en sí de los valores obtenidos y la consiguiente contrastación de hipótesis, tal como se hace en la inferencia estadística clásica.

De allí que, las estimaciones obtenidas para los parámetros, solo se utilizan en esta investigación, para la obtención de los gráficos Biplot y su utilización en la forma clásica, o sea en forma descriptiva, para la aproximación y exploración de las características sobresalientes en los datos.

El método propuesto tiene además la ventaja de que permite incluir el término independiente en el modelo, para calcular la probabilidad aproximada en el punto (0,0).

Por otra parte la estimación en forma simultánea de los parámetros del modelo reduce considerablemente la labor computacional, al sintetizar el sistema de ecuaciones en uno, que resulta más fácil desde el punto de vista operativo.

CONCLUSIONES

1. Es posible ajustar un Biplot a través de Modelos Bilineales Multiplicativos Generalizados, utilizando en la estimación de los parámetros un procedimiento iterativo similar al de los Modelos Lineales Generalizados.

2. El ajuste de un Biplot a través de Modelos Bilineales Multiplicativos Generalizados, equivale al ajuste de un plano de regresión a cada una de las columnas de la matriz de datos, lo que conlleva a la obtención de una superficie de respuesta no necesariamente lineal, siendo sin embargo, siempre lineal la proyección de dicha superficie sobre un subespacio de baja dimensión.

3. La geometría en los Biplots ajustados a través de Modelos Bilineales Multiplicativos Generalizados, es similar a la geometría en un ajuste lineal a través de mínimos cuadrados, por lo que se puede hacer la interpretación de sus representaciones gráficas en la misma forma que en los Biplots Clásicos, pero considerando la función link utilizada en el ajuste.

4. Se puede generalizar el ajuste de los Biplots a través de Modelos Bilineales al ajuste de los Biplots con Información Externa, restringiendo los ejes para que sean combinaciones lineales de variables externas, de manera tal de obtener interpretaciones más objetivas a la luz de la información que aportan esas variables.
5. Las coordenadas de los Biplot con Información Externa, tienen interpretaciones relacionadas con los coeficientes de regresión del Análisis de Regresión Múltiple y con los coeficientes de correlación interser e intraser del Análisis de Correlación Canónica.
6. La interpretación de las variables canónicas en los Biplots con Información Externa es similar a la de las componentes principales.
7. Los “Biplot de Regresión con Información Externa”, implican diferentes representaciones de acuerdo al tipo de regresión al cual es equivalente.

8. En el ajuste de los “Biplot de Regresión”, al igual que en el Análisis de Regresión Múltiple, el modelo a utilizar para ajustar a los datos debe considerar la posibilidad de inclusión de un término constante o independiente (para el caso de datos que no puedan ser centrados).

9. En el ajuste de los Biplots a través de Modelos Bilineales Multiplicativos, se pueden obtener los estimadores de los parámetros en la regresión generalizando el método iterativo de Newton-Raphson utilizado en los Modelos Lineales Generalizados. Es necesario hacer en la estimación los supuestos de independencia de los individuos, independencia de los parámetros de los individuos y las variables, e independencia de los parámetros de cada una de las variables. Se podría considerar la generalización para el caso en que las variables tengan cada una distribución diferente de la familia exponencial.

10. La estimación de los parámetros en el ajuste de los Biplots a través de Modelos Bilineales Multiplicativos se vale de las buenas propiedades que los estimadores deben tener en otros contextos, sin embargo esas estimaciones sólo son válidas para usarlas en las representaciones Biplot en la forma clásica, o sea, para la aproximación y exploración de las características sobresalientes en los datos y no para el uso en sí de los valores obtenidos y la consiguiente contrastación de hipótesis, tal como se hace en la inferencia estadística clásica.

BIBLIOGRAFIA

ANDERSEN, E. B. (1972): 'The numerical solution of a set of conditional estimation equations. *The Journal of the Royal Statistical Society, Series B*, **34**, 42-54.

BENZECRI, J. P. (1973/1982): L'Analyse des Données. Tomo I: *La Taxinomie*. Tomo II: *L'Analyse des Correspondances*. Dunod. París.

BLAZQUEZ, A. (1998): *Análisis Biplot basado en Modelos Lineales Generalizados*. Tesis Doctoral. Universidad de Salamanca, España.

BAKER, Frank B. (1992): *Item Response Theory*. Marcel Dekker, Inc. New York.

BISHOP, Y. M. M.; FIENBERG, S. E. & HOLLAND, P. W. (1975): *Discrete Multivariate Analysis: Theory and Practice*. The Massachusetts Institute of Technology. Massachusetts.

BIRNBAUN, A. (1968): 'Some latent trait models and their use in inferring and examinee's ability'. In F. M. Lord and M. R. Novick (eds.), *Statistical Theories of Mental Test Scores*, Addison-Wesley, 397-472

BOCKENHOLT, U. & TAKANE, Y. (1994): 'Linear Constraints in Correspondence Analysis'. *Correspondence Analysis in the Social Sciences. Recent Development and Applications*. (M. Greenacre and J. Blasius, eds.). Academic Press. London.

BRADU, D. & GABRIEL, K. R. (1974): 'Simultaneous statistical inference on interactions in two-way analysis of variance'. *Journal of American Statistical Association*, **29**, 428-36.

BRADU, D. & GABRIEL, K. R. (1978): 'The Biplot as a Diagnostic Tool for Models of Two-Way Tables'. *Technometrics* **20**, (1): 47-68.

CHOULAKIAN, V. (1996): 'Generalized Bilinear Models'. *Psychometrika* **61**, (2), 271- 283.

CHRISTENSEN, Ronald (1990): *Log-Linear Models*. Springer-Verlag.

COX & HINKLEY (1974): *Theoretical Statistics*. Chapman & Hall, London.

DAVIES, P.T. & TSO, M. K. S. (1982): 'Procedures for reduced-rank regression'. *Applied Statist.*, **31**, 244-255.

DENIS, J. B (1991): 'Ajustements de Modelles Lineaires et Bilineaires sous Contraintes Lineaires avec Donnes Manquantes'. *Statistique Applique*, XXXIX (2), 5-24.

DIAZ-LENO, M. S. (1995): *Los Métodos Biplot como herramienta de diagnóstico en la modelización de datos multidimensionales*. Tesis Doctoral. Universidad de Salamanca.

DOBSON, A. (1990): *An Introduction to Generalized Linear Models*. Chapman & Hall, London.

ECKART, C. & YOUNG, G. (1936): 'The approximation of one matrix by another of lower rank'. *Psychometrika*, **1**, 211-18.

FALGUEROLLES, A. (1995): 'Generalized Bilinear Models and Generalized Biplots: Some Examples'. *Publications du Laboratoire de Statistique et Probabilites*. Université Paul Sabatier. Toulouse.

FALGUEROLLES, A. (1996): Empirical Detection of Influential Observations in Generalized Bilinear Analyses. *Sydney International Statistical Congress*.

FALGUEROLLES, A. de & FRANCIS, B. (1992): 'Algorithmic approaches for fitting bilinear models'. (I. Dodge and J. Whittaker eds.). *COMPSTAT 92, Computational Statistics*, **1**, 470-475. Heidelberg: Physica-Verlag.

FERNANDEZ-GOMEZ, M. J. (1995): *Contribuciones al Análisis Multivariante Directo del Gradiente mediante estudio combinado de configuraciones espaciales*. Tesis Doctoral. Universidad de Salamanca, España.

FISZ, M (1963): *Probability Theory and mathematical statistics. Tercera Edición*. John Wiley, New York.

GABRIEL, K. R (1971): 'The Biplot-graphic display of matrices with applications to principal component analysis'. *Biometrika* **58**, 453-467.

GABRIEL, K. R. (1978): 'Least square approximation of matrices by additive and multiplicative models'. *J. R. Statist. Soc.B*, **40**, 2, 186-196.

GABRIEL, K. R. (1981a): Biplot. In *Encyclopedia of Statistical Sciences*, Vol. I (eds. S. Kotz, N. L. Johnson, and C. Read), pp. 262-265. Wiley. New York.

GABRIEL, K. R (1981b): Biplot Display of Multivariate Matrices for inspection of Data and Diagnosis. In *Interpreting Multivariate Data*, (V. Barnett, ed.), London: John Wiley.

GABRIEL, K. R. (1995): 'Generalised Bilinear Regression'. Internal Report. University of Rochester.

GABRIEL, K. R. (1998): 'Generalised Bilinear Regression'. *Biometrika*, **85**, 3, 689-700.

GABRIEL, K. R. & ZAMIR, S. (1979): 'Lower Rank Approximation of Matrices by Additive and Multiplicative Models'. *J. R. Statist. Soc. B*. **40(2)**: 186-196.

GABRIEL, K. R. & ODOROFF, C. L. (1990): 'Biplots in biomedical research'. *Statistics in Medicine* **9**, 469-485.

GABRIEL, K. R. & RITZ, J. M. (1994): 'Generalised Bilinear Fitting with application to modelling and biplot'. Salamanca Statistics Seminar II, 1-2 July. Universidad de Salamanca.

GABRIEL, K. R.; GALINDO, M. P. y VICENTE-VILLARDON, J. L. (1998): 'Use of Biplots to diagnose Independence Models in Three-Way Contingency Tables'. (M. Greenacre and J. Blasius, eds.). *Visualization of Categorical Data*. Academic Press. London

GALINDO , M. P. (1985): *Contribuciones a la representación simultánea de datos multidimensionales*. Tesis Doctoral. Universidad de Salamanca.

GALINDO , M. P. (1986): 'Una alternativa de representación simultánea: HJ -Biplot' . *Questio* **10** (1), 13-23.

GALINDO , M. P. y CUADRAS, C. (1986): 'Una extensión del método biplot a su relación con otras técnicas'. *Publicación de bioestadística y biomatemática. Universidad de Barcelona*. Nº 17.

GALINDO, M. P.; VICENTE-VILLARDON, J. L. y cols (1986): Representación "Bimodel" de la evolución clínica de pacientes afectos de la enfermedad por síndrome tóxico". *Actas de las VI Jornadas toxicológicas españolas*.

GALINDO, M. P.; LORENTE, F.; ROMO, A. y MARTIN, M. P. (1987): 'Inspección de matrices de datos multivariantes utilizando el método HJ-Biplot: Aplicación a un problema médico'. *Cuadernos de Bioestadística y sus Aplicaciones Informáticas*. Vol. **5**, nº 1: 88-101.

GARCIA-GALEGON, A. C.; TRUJILLANO, R. & VICENTE-TAVERA, S. ; et.al (1995): "Influence of consolidation and hydrofugation treatments of the physical properties of Avila natural and artificially aged granite: a statistical approach". Article in the book "*Degradation and conservation of granitic rocks in monuments*". Ed. By European Commision, Dg Xii D-I.

GOLLOB, H. (1968): 'A statistical model wich combines features of factor analytic and analysis of variance techniques'. *Psichometrika*, **33**: 73-115.

GOLUB, G. H. & REINSCH, C. H. (1970): 'Singular value descomposition and least squares solution'. *Numer. Math.*, **14**, 403-20.

GOWER, J. C. (1992): 'Generalized Biplots'. *Biometrika* **79**, 475-493.

GOWER, J. C. & HARDING, S. (1988): 'Nonlinear Biplots'. *Biometrika*, **75**, 445-455.

GOWER, J. C. & HAND, D. J. (1996): *Biplots*. Chapman & Hall. London.

GRAYBILL, F. A. (1983): *Matrices with applications in Statistics*. 2a. ed. University of California.

GREENACRE, M. J. (1984): *Theory and applications of correspondence analysis*. Academic Press. London.

GREENACRE, M. J. (1988): 'Correspondence Analysis of multivariate data by weighted least square'. *Biometrika*, **75**, 457-467.

GREENACRE, M. J. (1993): *Correspondence analysis in Practice*. Academic Press. London.

HOTTELING, H. (1936): 'Simplified Calculation of Principal Components'. *Psychometrika*, **1**: 27-35.

HOUSEHOLDER, A. S. & YOUNG, G. (1938): 'Matrix Approximation and Latent Roots'. *American Mathematics Monthly*, **45**: 165-171.

IZENMAN, A. J. (1975): 'Reduced-Rank regression for the Multivariate Linear Model'. *J. Mult. Anal.*, **5**, 248-264.

JONGMAN, R. H. G.; TER BRAAK, G. J. F. & VAN TONGEREN, O. F. R. (Editores) (1987): *Data Analysis in Community and landscape ecology*. Pudoc Wageningen.

KRUSKAL, J. B. & WISH, M. W. (1978): *Multidimensional Scaling*, Sage University Paper series on Quantitative Applications in the Social Sciences, series nº 07-011. Sage Publications, Beverly Hills.

MANDEL, J. (1961): 'Non-additivity in Two-way Analysis of Variance'. *J. Amer. Statist. Assn.*, **56**: 878-888.

MANDEL, J. (1969): 'The partitioning of interaction in analysis of variance'. *J. Nat. Bur. Stand* **73 B**, 309-328.

MARDIA, K. V.; KENT, J. T. & BIBBY, J. M. (1979). *Multivariate Analysis*. Academic Press. London.

McCULLAGH, P & NELDER, J.A. (1989): *Generalized Linear Models (Second Edition)*. Chapman & Hall, London.

NELDER, J. A., & WEDDERBURN, R. W. (1972): 'Generalized Linear Models'. *Journal of the Royal Statistical Society A*, **135**, 370-384.

NEYMAN, J. & SCOTT, E. L. (1948): 'Consistent estimates based on partially consistent observations'. *Econometrika*, **16 (1)**, 1-32.

PEDRAZ, C. y GALINDO, M. P. y cols. (1986): 'Study of socio-cultural factors influencing the decision to breast-feed instead of bottle-feed'. *Arch. Pediat.* **36**: 469-477.

PEREZ-MELLADO, V. & GALINDO, M. P. (1986): 'Biplot graphic display of Iberian and North African populations of Podarcis'. *Studies in Herpetology, Rocek Z. (ed)*. 197-200.

RAO, C. R. (1964): 'The use and interpretation of principal component analysis in applied research'. *Shankya A*, **26**, 329-58.

RIVAS, J. C. et. al. (1993): 'HJ-biplot analysis applied to enological parameters in the geographical classification of young red wines'. *American Journal of Enology and Viniculture*. Vol. **44(3)**: 302-308.

ROHATGI, V.K. (1976): *An Introduction to Probability Theory and Mathematical Statistics*. John Wiley & Sons, New York.

SANTOS, C.; et. al. (1991): 'Characterization of young red wines by application of HJ-Biplot Analysis to anthocianin profiles'. *Journal of Agricultural and food chemistry*, **39**: 1086-1090.

STRAUSS, J. S. & 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.

TAKANE, Y. & SHIBAYAMA, T. (1991): 'Principal Component Analysis with external information on both subjects and variables'. *Psychometrika*, **56**, 97-120.

TER BRAAK, C. J. F. (1986): 'Canonical Correspondence Analysis: a new eigenvector technique for Multivariate Direct Gradient Analysis'. *Ecology* **67** (5), 1167-1179.

TER BRAAK, C. J. F. (1990): 'Interpreting Canonical Correlation Analysis through Biplots of Structure Correlation and Weight'. *Psychometrika* **55** (3), 516-531.

TER BRAAK, C. J. F. & LOOMAN, C. W. N (1991): 'Biplots in Reduced-Rank Regression'. *Biometrical Journal*, **8**, 983-1003.

TSIANCO, M. C. & GABRIEL, K. R. (1981): 'Modeling Temperature Data: An Illustration of the use of Biplot and Bimodels in non-linear modeling'. University of Rochester. *Statist. Tech. Rep.* 81/15.

TSIANCO, M. C. & 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.

VAN EEUWIJK, F. (1995): 'Multiplicative Interaction in Generalized Linear Models'. *Biometrics*, **51**, 1017-32

VAN DER WOLLEMBERG, A. L. (1977): 'Redundancy Analysis. An alternative for Canonical Correlation Analysis'. *Psychometrika* , **42**, 207-219.

VASQUEZ, M. (1995): *Aportaciones al Análisis Biplot: Un enfoque algebraico*. Tesis Doctoral. Universidad de Salamanca.

VASQUEZ, M.; RAMIREZ, G.; LOPEZ, M.; GALINDO, P.; BOSH, V. y VAZQUEZ, J. (1999): 'Determinación de puntos atípicos en estudios longitudinales mediante el análisis biplot de datos sin transformaciones previas'. *Anales Venezolanos de Nutrición*, **12(1)**: 10-15.

VICENTE-TAVERA, S. (1992): *Las Técnicas de Representación de Datos Multidimensionales en el Estudio del Índice de Producción Industrial en la C. E. E.* Tesis Doctoral. Universidad de Salamanca.

VICENTE, S. y GALINDO, M. P.; et. al. (1993): 'Análisis gráfico y descripción estructural de la variabilidad de cultivos en Castilla-León'. *Investigación Agraria. Economía*. Vol. **8(3)**: 317-329.

VICENTE-VILLARDON, J. L. (1992): *Una alternativa a las técnicas Factoriales Clásicas basada en una generalización de los métodos Biplot*. Tesis Doctoral. Universidad de Salamanca.

VICENTE-VILLARDON, J. L. y GALINDO, M. P. (1998): Biplot Externo para datos presencia-ausencia basado en superficies logísticas de respuesta. Dpto. de Estadística. Universidad de Salamanca.

VICENTE-VILLARDON, J. L., GALINDO, M. P. y CARDENAS, O. (2000). 'Biplot para Datos Binarios basado en Modelos Logísticos de Respuesta'. XXV Congreso Nacional de Estadística e Investigación Operativa. Servicio de Publicaciones, Universidad de Vigo, 269-270

APENDICES

APENDICE 1

NOTACION UTILIZADA

$\mathbf{Y} = [y_{ij}]$ = matriz de orden $(n \times p)$, $(i = 1, 2, \dots, n ; j = 1, 2, \dots, p)$

n = número de individuos u objetos

p = número de variables

\mathbf{y} = vector de observaciones

\mathbf{y}_i = vector fila i -ésima de $\mathbf{Y} = [y_{i1}, y_{i2}, \dots, y_{ip}]$

\mathbf{y}_j = vector columna j -ésima de $\mathbf{Y} = [y_{1j}, y_{2j}, \dots, y_{nj}]'$

y_{ij} = elemento genérico ij -ésimo de la matriz \mathbf{Y}

$\mathbf{Y}' = \mathbf{Y}'$ = matriz traspuesta

$\mathbf{y}' = \mathbf{y}'$ = vector traspuesto

$\langle \mathbf{Y} \rangle$ = vector columna $(n \times 1)$ donde se han concatenado las columnas de la matriz \mathbf{Y}

$\langle \mathbf{Y}' \rangle$ = vector columna $(n \times 1)$ donde se han concatenado las filas de la matriz \mathbf{Y}

\mathbf{I}_n = matriz identidad de orden $(n \times n)$ con unos sobre la diagonal principal y ceros en cualquier otro lugar

\mathbf{J}_n = matriz unitaria de orden $(n \times n)$ con todos los elementos iguales a uno

$\mathbf{1}_n$ = vector columna unidad de orden $(n \times 1)$

$\mathbf{0}_n$ = vector columna nulo de orden $(n \times 1)$

\mathbf{A} = matriz de marcadores fila de orden $(n \times q) = [\mathbf{a}_i]$

\mathbf{B} = matriz de marcadores columna de orden $(p \times q) = [\mathbf{b}_j]$

\mathbf{S}_{yy} = matriz de varianzas-covarianzas de la matriz \mathbf{Y}

\mathbf{R}_{yy} = matriz de correlaciones de la matriz \mathbf{Y}

\mathbf{R}_{yx} = matriz de correlaciones entre las matrices \mathbf{Y} y \mathbf{X}

$p = \text{Prob}(\mathbf{Y} = y_{ij})$

$\mathbf{A}_{(m \times n)} \otimes \mathbf{B}_{(p \times q)} = a_{ij} \mathbf{B}$ = Producto Kroenecker entre dos matrices

$\hat{\mathbf{A}}$ = matriz de marcadores fila estimada

$\mathbf{Y}_{(q)}$ = aproximación de la matriz \mathbf{Y} de rango q ($q < r$)

$\mathbf{A}_{(q)}$ = aproximación de la matriz \mathbf{A} de rango q

$\mathbf{B}_{(q)}$ = aproximación de la matriz \mathbf{B} de rango q

k = número de ejes en la aproximación = número de parámetros para cada variable, ($k=1,2,\dots, q$)

α, β, \dots = parámetros poblacionales

Ω, Φ = matrices de métricas

$\text{Proy}(\mathbf{a} / \mathbf{b})$ = proyección ortogonal del vector \mathbf{a} sobre el vector \mathbf{b}

APENDICE 2

DEFINICIONES Y RELACIONES UTILIZADAS

1 - DEFINICIONES BASICAS

$\mathbf{1}'\mathbf{y} = \mathbf{y}'\mathbf{1} =$ suma respectivamente de los elementos de los vectores \mathbf{y} e \mathbf{y}'

$\mathbf{y}'\mathbf{y} = \sum_{(i=1,n)} y_i^2 =$ producto escalar de un vector por sí mismo

$\|\mathbf{y}'\| = [\sum y_i^2]^{1/2} =$ norma del vector \mathbf{y}

$\|\mathbf{y}'\|^2 = \mathbf{y}'\mathbf{y} =$ longitud del vector \mathbf{y}

$d(\mathbf{y},\mathbf{x}) = \|\mathbf{y}' - \mathbf{x}'\| = [\sum (y_i - x_i)^2]^{1/2} =$ distancia euclídea entre los vectores \mathbf{x} e \mathbf{y}

$d^2(\mathbf{y},\mathbf{x}) = \|\mathbf{y}' - \mathbf{x}'\|^2 = [\sum (y_i - x_i)^2]$

$\|\mathbf{y}' - \mathbf{x}'\|^2 = \|\mathbf{y}'\|^2 + \|\mathbf{x}'\|^2 - 2\|\mathbf{y}'\| \cdot \|\mathbf{x}'\| \cos \theta$

$\text{var}(\mathbf{y}) = (1/n)\mathbf{y}'\mathbf{y} - [(1/n)\mathbf{1}'\mathbf{y}]^2$

$(\mathbf{a}'\mathbf{b}) =$ producto interno entre vectores de marcadores fila y columna =

$\|\text{Proy}(\mathbf{a} / \mathbf{b})\| \|\mathbf{b}\| = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta$

$\mathbf{t}_{p \times 1} = \mathbf{Y}'\mathbf{1}_n =$ vector de totales para las p variables de la matriz \mathbf{Y}

$\bar{\mathbf{y}} = (1/n)\mathbf{Y}'\mathbf{1}_n =$ vector(promedio)

$\bar{\mathbf{Y}} = [\mathbf{I} - \mathbf{J}/n]\mathbf{Y} =$ matriz(centrada)

$\mathbf{S} = [1/(n-1)]\bar{\mathbf{Y}}'\bar{\mathbf{Y}} =$ matriz(varianzas – covarianzas)

$\bar{\mathbf{Y}}_c = \bar{\mathbf{Y}}[\text{diag}(\mathbf{S})]^{-1} =$ matriz(es tan darizada)

$\mathbf{R} = [1/(n-1)]\bar{\mathbf{Y}}_c'\bar{\mathbf{Y}}_c =$ matriz(correlacion)

$[\mathbf{I} - (1/n) \mathbf{1}'\mathbf{1}] = [\mathbf{I} - (1/n)\mathbf{J}] =$ matriz de centrado

$\mathbf{Y}'\mathbf{Y} =$ matriz de productos cruzados

$\mathbf{S}_{yy} =$ matriz de varianzas-covarianzas de la matriz \mathbf{Y}

$\mathbf{R}_{yy} =$ matriz de correlaciones de \mathbf{Y}

$\mathbf{R}_{yx} =$ matriz de correlaciones entre matrices \mathbf{Y} y $\mathbf{X} = (\mathbf{S}_{yy})^{-1} (\mathbf{S}_{yx}) (\mathbf{S}_{xx})^{-1}$

2 – DESCOMPOSICION EN VALORES SINGULARES

$\mathbf{Y} = \mathbf{U} \mathbf{D} \mathbf{V}' = \sum_k \alpha_k \mathbf{u}_k \mathbf{v}'_k =$ descomposición en valores singulares de la matriz \mathbf{Y}

$\alpha_k =$ valores singulares de $[\mathbf{Y}\mathbf{Y}']$ o de $[\mathbf{Y}'\mathbf{Y}] = \sqrt{\lambda_k}$

$\lambda_k =$ valores propios de $[\mathbf{Y}\mathbf{Y}']$ o de $[\mathbf{Y}'\mathbf{Y}]$

$\mathbf{u}_k =$ vectores singulares por la izquierda

$\mathbf{v}_k =$ vectores singulares por la derecha

$\mathbf{U} =$ matriz cuyas columnas son los vectores singulares de $[\mathbf{Y}\mathbf{Y}']$

$\mathbf{V} =$ matriz cuyas columnas son los vectores singulares de $[\mathbf{Y}'\mathbf{Y}]$

$\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}$

3 – PRODUCTO KROENECKER

$C_{(mp \times nq)} = \text{Producto Kroenecker} = \mathbf{A}_{(m \times n)} \otimes \mathbf{B}_{(p \times q)} = a_{ij} \mathbf{B} = \mathbf{B} a_{ij}$

$\langle \mathbf{X} \mathbf{A} \mathbf{B} \rangle = [\mathbf{B}' \otimes \mathbf{X}] \langle \mathbf{A}' \rangle =$ relación para la concatenación de columnas del producto de las matrices \mathbf{X} , \mathbf{A} , y \mathbf{B} .

$\langle \mathbf{X} \mathbf{A} \mathbf{I}_n \rangle = [\mathbf{I}_n \otimes \mathbf{X}] \langle \mathbf{A}' \rangle$

$[\mathbf{I}_n \otimes \mathbf{X}] = [\text{diag}(\mathbf{X})]$ = matriz diagonal con la matriz \mathbf{X} sobre la diagonal principal

$[\mathbf{Z} \otimes (\mathbf{X}\mathbf{A})] = [z_{ij}(\mathbf{X}\mathbf{A})]$

$\langle \mathbf{I}_n \mathbf{A} \mathbf{B}' \mathbf{I}_m \rangle = [\mathbf{I}_m \otimes (\mathbf{I}_n \mathbf{A})] \langle \mathbf{B}' \rangle = [\mathbf{I}_m \otimes \mathbf{A}] \langle \mathbf{B}' \rangle = [\text{diag}(\mathbf{A})] \langle \mathbf{B}' \rangle$

$\langle (\mathbf{X}\mathbf{A})(\mathbf{B}'\mathbf{Z}') \rangle = [\mathbf{Z} \otimes (\mathbf{X}\mathbf{A})] \langle \mathbf{B}' \rangle$