

BIPLOTS PARA LOS RESIDUOS DEL SEGUNDO NIVEL DE UN MODELO LINEAL JERARQUICO CON DOS NIVELES

Mario Miguel Ojeda y Sergio Francisco Juárez Cerrillo

Facultad de Estadística e Informática, Universidad Veracruzana, Xalapa, Veracruz, México

RESUMEN

En este trabajo se presenta una formulación del modelo lineal general jerárquico con dos niveles. Este modelo se reformula en el segundo nivel como un modelo lineal general multivariante, obteniéndose una matriz de residuos para este nivel. Con objeto de analizar la matriz de residuos obtenida, se utiliza el biplot, el cual se propone para explorar la presencia de observaciones extremas (atípicas). La respuesta se ilustra con un ejemplo.

Palabras clave: modelos jerárquicos, datos jerárquicos, modelos mixtos, biplots, exploración de residuos multivariantes.

MSC: 62J01

ABSTRACT

This paper a general formulation of a two level hierarchical linear model. In the second level, the model is reformulated as a multivariate linear model, obtaining a matrix of residuals for this level. The biplot is proposed as a graphical tool that helps for multivariate outliers detection. This proposal is illustrated with one example.

Key words: hierarchical models, hierarchical data, linear mixed models, biplots, multivariate outliers detection.

1. INTRODUCCION

Un modelo jerárquico lineal considera una estructura de anidamiento de unidades muestrales dentro de grupos y permite la utilización de variables explicatorias tanto al nivel de las unidades como en el nivel de los grupos, así como la definición de estructuras complejas sobre la matriz de varianzas y covarianzas de los errores aleatorios. Casos particulares de estos modelos son los modelos de componentes de la varianza y covarianza y los modelos de regresión con coeficientes aleatorios, entre otros. Estos modelos se han aplicado en diversas áreas del análisis estadístico, tales como la modelación de rendimientos educativos, el estudio de relaciones econométricas considerando efectos contextuales, el análisis de datos de encuestas que utilizan diseños de muestreo complejos, el análisis de datos longitudinales y, más recientemente, en el meta-análisis y la modelación de series de experimentos. El modelo lineal jerárquico también se le conoce como modelo lineal general de efectos mixtos, modelo lineal general en varios niveles y modelo de componentes de varianza y covarianza. Un tratamiento monográfico y abundantes referencias acerca de estos modelos se puede encontrar en Goldstein (1995) y Bryk y Raudenbush (1992).

En un análisis jerárquico integral, se deben considerar los pasos naturales en el ajuste de modelos estadísticos, a saber: análisis inicial de datos, formulación del modelo, ajuste del modelo y diagnóstico. Tal como en los modelos lineales de un nivel, los residuos son de gran importancia para detectar violaciones de los supuestos involucrados en el modelo, por lo que se necesitan estrategias claras de diagnóstico en el modelaje jerárquico. El analista de datos debe ser cuidadoso con los supuestos en que se basan los modelos lineales jerárquicos, e investigar la razonabilidad de tales supuestos tomando como base los datos.

En este contexto, herramientas para el diagnóstico y análisis de sensibilidad serán bienvenidas en los próximos años. Ver Seltzer (1993) para referencias con respecto a esto. Ojeda (1992, 1993) presenta ideas generales relativas a estos tópicos en el análisis de datos de encuestas usando modelos jerárquicos. En particular, el análisis de residuos requiere de herramientas especiales tanto para un diagnóstico adecuado como para exploración, sin embargo, el desarrollo para detectar outliers ha recibido poca atención. En este trabajo, para explorar los residuos en el segundo nivel, el vector de errores en dicho nivel se reformula como una matriz

de residuos de un modelo lineal multivariante y para detectar posibles observaciones extremas, la estructura del espacio de las hileras y columnas de esta matriz se analiza usando componentes principales y biplots.

2. EL MODELO LINEAL JERARQUICO CON DOS NIVELES

Sea Y la variable respuesta y sean X_1, \dots, X_K las variables explicativas. Estas variables se observan en J grupos con n_j objetos para el respectivo grupo ($j = 1, \dots, J$). Sean: $\mathbf{y}_j = (y_{1j}, y_{2j}, \dots, y_{n_j})^T$ el vector con las observaciones de la variable respuesta en el grupo j , $\mathbf{X}_j = (\mathbf{X}_{1j}, \mathbf{X}_{2j}, \dots, \mathbf{X}_{n_j})^T$ la matriz de diseño, donde $\mathbf{X}_{ij} = (1, X_{1ij}, X_{2ij}, \dots, X_{Kij})^T$, $\beta_j = (\beta_0, \beta_{1j}, \dots, \beta_{Kj})^T$ el vector de parámetros y $\mathbf{e}_j = (e_{1j}, e_{2j}, \dots, e_{n_j})^T$ el vector con los errores en el primer nivel. En este trabajo se usa el modelo lineal

$$\mathbf{y}_j = \mathbf{X}_j \beta_j + \mathbf{e}_j, \quad j = 1, \dots, J \quad (2.1)$$

con $E(\mathbf{e}_j) = \mathbf{0}$ y $\text{var}(\mathbf{e}_j) = \Omega_{(1)j}$. Esta es la ecuación del modelo en el primer nivel. Ahora, supóngase que los β_j 's varían de grupo de modo que pueden ser tratados como aleatorios en un segundo nivel del modelo. Si además se tienen disponibles observaciones en q variables explicatorias Z_1, Z_2, \dots, Z_q a nivel de grupo. Entonces, los β_j 's se deben tratar como aleatorios en un segundo nivel del modelo. Para esto, sean: $\mathbf{Z}_j = \mathbf{I}_{K+1} \otimes (Z_{1j}, Z_{2j}, \dots, Z_{qj})$, donde \otimes denota al producto Kronecker, $\gamma = (\gamma_0^T, \gamma_1^T, \dots, \gamma_K^T)^T$ el vector de parámetros (fijos), donde $\gamma_k = (\gamma_{0k}, \gamma_{1k}, \dots, \gamma_{qk})^T$ y $\mathbf{u}_j = (u_{0j}, u_{1j}, u_{Kj})^T$ el vector con los errores en el segundo nivel en el grupo j . La ecuación del modelo en el segundo nivel es

$$\beta_j = \mathbf{Z}_j \gamma + \mathbf{u}_j, \quad j = 1, \dots, J; \quad (2.2)$$

con $E(\mathbf{u}_j) = \mathbf{0}$ y $\text{var}(\mathbf{u}_j) = \Omega_{(2)}$. Combinando la ecuación (2.1) con la ecuación (2.2), se tiene el modelo lineal general mixto

$$\mathbf{y}_j = \mathbf{X}_j' \mathbf{Z}_j \gamma + \mathbf{X}_j \mathbf{u}_j + \mathbf{e}_j,$$

al cual se le agregan los supuestos $\text{cov}(\mathbf{e}_j, \mathbf{e}_{j'}) = \mathbf{0}$, $\text{cov}(\mathbf{u}_j, \mathbf{u}_{j'}) = \mathbf{0}$, y $\text{cov}(\mathbf{u}_j, \mathbf{e}_{j'}) = \mathbf{0}$. ($j, j' = 1, \dots, J, j \neq j'$). De estos se sigue que

$$E(\mathbf{y}_j | \gamma) = \mathbf{X}_j \mathbf{Z}_j \gamma, \quad \text{var}(\mathbf{y}_j | \gamma) = \mathbf{V}_j = \mathbf{X}_j \Omega_{(2)} \mathbf{X}_j^T + \Omega_{(1)j}.$$

Para escribir en notación matricial compacta al modelo, se definen las matrices

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}, \quad \mathbf{X} = \bigoplus_{j=1}^J \mathbf{X}_j, \quad \mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_J \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_J \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_J \end{bmatrix},$$

donde \bigoplus denota suma directa. Finalmente, se tiene el modelo lineal mixto

$$\mathbf{y} = \mathbf{XZ}\gamma + \mathbf{Xu} + \mathbf{e}, \quad (2.3)$$

con $E(\mathbf{y} | \gamma) = \mathbf{XZ}\gamma$ y $\text{var}(\mathbf{y} | \gamma) = \mathbf{V} = \bigoplus_{j=1}^J \mathbf{V}_j$. Usualmente se agrega el supuesto de normalidad en el modelo de la siguiente forma: $\mathbf{e}_j \sim N(\mathbf{0}, \Omega_{(1)j})$ y $\mathbf{u}_j \sim N(\mathbf{0}, \Omega_{(2)})$.

3. EL BIPLLOT

El biplot (Gabriel, 1971) es una representación gráfica de una matriz \mathbf{X} $n \times p$ de rango 2, por medio de vectores bidimensionales $\mathbf{a}_1, \dots, \mathbf{a}_n$ para sus hileras (marcas de hilera) y vectores bidimensionales $\mathbf{b}_1, \dots, \mathbf{b}_p$ para sus columnas (marcas de columna). Esto se realiza mediante la factorización $\mathbf{X} = \mathbf{AB}^T$, donde \mathbf{A} y \mathbf{B} son matrices de rango dos y los \mathbf{a}_i 's y \mathbf{b}_j 's son las hileras \mathbf{A} y \mathbf{B} , respectivamente. Los \mathbf{a}_i 's se asignan a las de \mathbf{X} y lo mismo se hace para los \mathbf{b}_j 's con las columnas de \mathbf{X} . Estos vectores se grafican en un sistema coordenado rectangular y la gráfica que se obtiene se llama biplot de \mathbf{X} . La factorización $\mathbf{X} = \mathbf{AB}^T$ se conoce como factorización de rango completo de \mathbf{X} , ver Searle (1982) pág. 194. En esta factorización se tiene

$\mathbf{X}\mathbf{X}^T = \mathbf{A}(\mathbf{B}^T\mathbf{B})\mathbf{A}^T$ y $\mathbf{X}^T\mathbf{X} = \mathbf{B}(\mathbf{A}^T\mathbf{A})\mathbf{B}^T$, por lo que los productos internos de las hileras (columnas) de \mathbf{X} se representan por los respectivos productos internos entre las hileras de \mathbf{A} (\mathbf{B}) en la métrica $\mathbf{B}^T\mathbf{B}$ ($\mathbf{A}^T\mathbf{A}$). Por ejemplo, si se desea que los productos internos entre las hileras de \mathbf{X} estén representados por las hileras de \mathbf{A} , esto es $\mathbf{X}\mathbf{X}^T = \mathbf{A}\mathbf{A}^T$, entonces, se debe tener $\mathbf{B}^T\mathbf{B} = \mathbf{I}$; cuando se hace esto, se puede mostrar que la métrica $(\mathbf{X}\mathbf{X}^T)^{-}$ se induce por las columnas de \mathbf{X} , es decir, se tiene $\mathbf{B}\mathbf{B}^T = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-}\mathbf{X}$. Por otra parte, se requiere lo mismo para las columnas de \mathbf{X} , esto es $\mathbf{X}^T\mathbf{X} = \mathbf{B}\mathbf{B}^T$ se debe tener $\mathbf{A}^T\mathbf{A} = \mathbf{I}$, y al hacer esto se obtiene $\mathbf{A}\mathbf{A}^T = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T$. En general, si se desea que los productos internos entre las hileras de \mathbf{X} estén dados en la métrica Φ , esto es $\mathbf{X}\Phi\mathbf{X}^T = \mathbf{A}\mathbf{A}^T$, entonces se debe imponer la restricción $\mathbf{B}^T\Phi\mathbf{B} = \mathbf{I}$, y cualquier inversa generalizada $(\mathbf{X}\Phi\mathbf{X}^T)^{-}$ de $\mathbf{X}\Phi\mathbf{X}^T$ se puede usar como métrica para las columnas de \mathbf{X} . De manera similar, se impone la métrica Ψ para las columnas de \mathbf{X} , esto es $\mathbf{X}^T\Psi\mathbf{X} = \mathbf{B}\mathbf{B}^T$, se debe tener $\mathbf{A}^T\Psi\mathbf{A} = \mathbf{I}$ y cualquier inversa generalizada $(\mathbf{X}^T\Psi\mathbf{X})^{-}$ de $\mathbf{X}^T\Psi\mathbf{X}$ se puede usar como métrica para las hileras de \mathbf{X} .

Un procedimiento útil para construir el biplot de cualquier matriz \mathbf{X} de rango 2, es por medio de su *descomposición en valores singulares* (DVS), definida por

$$\mathbf{X} = \mathbf{L}\Delta\mathbf{M}^T = \sum_{k=1}^2 \delta_k \mathbf{l}_k \mathbf{m}_k^T \quad (3.1)$$

donde $\mathbf{L} = [\mathbf{l}_1 \ \mathbf{l}_2]$, $\mathbf{M} = [\mathbf{m}_1 \ \mathbf{m}_2]$ y $\Delta = \text{diag}(\delta_1, \delta_2)$ con $\mathbf{L}^T\mathbf{L} = \mathbf{M}^T\mathbf{M} = \mathbf{I}$ y $\delta_1 \geq \delta_2$. Las columnas de \mathbf{L} son los eigenvectores ortonormalizados de $\mathbf{X}\mathbf{X}^T$ y las columnas de \mathbf{M} son los eigenvectores ortonormalizados de $\mathbf{X}^T\mathbf{X}$, estas columnas se llaman los vectores singulares izquierdos y derechos de \mathbf{X} , respectivamente. Los δ_k 's son las raíces cuadradas positivas de los eigenvectores de $\mathbf{X}\mathbf{X}^T$ y $\mathbf{X}^T\mathbf{X}$, estos números se llaman los valores singulares de \mathbf{X} . Luego, haciendo

$$\mathbf{A} = \mathbf{L}\Delta^\alpha, \quad \mathbf{B} = \mathbf{M}\Delta^{1-\alpha}, \quad \alpha \in [0,1], \quad (3.2)$$

obtenemos $\mathbf{X} = \mathbf{A}\mathbf{B}^T$, y, por lo tanto, el respectivo biplot de \mathbf{X} . Generalmente el rango de \mathbf{X} es mayor que 2, cuando esto sucede, se puede construir un biplot aproximado de \mathbf{X} usando para (3.2) la matriz

$$\mathbf{X}_{(2)} = \sum_{k=1}^2 \delta_k \mathbf{l}_k \mathbf{m}_k^T, \quad \text{donde ahora } \mathbf{X} = \sum_{k=1}^r \delta_k \mathbf{l}_k \mathbf{m}_k^T \text{ denota la DVS de } \mathbf{X} \text{ y } r \text{ es el rango de } \mathbf{X}. \mathbf{X}_{(2)} \text{ es la matriz que}$$

se obtiene omitiendo las últimas $r-2$ matrices $\delta_k \mathbf{l}_k \mathbf{m}_k^T$ y resulta ser la matriz de rango 2 que mejor aproxima \mathbf{X} en el sentido de mínimos cuadrados; es decir, es la matriz que minimiza la norma de Frobenius.

$$\|\mathbf{X} - \mathbf{X}'\|^2 = \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - x'_{ij})^2,$$

entre todas las matrices \mathbf{X}' de rango 2 o menos (x_{ij} y x'_{ij} denotan el ij -elemento de \mathbf{X} y \mathbf{X}' , respectivamente). Una medida de la bondad de ajuste está dada por

$$\frac{\|\mathbf{X}_{(2)}\|^2}{\|\mathbf{X}\|^2} = (\delta_1^2 + \delta_2^2) / \sum_{k=1}^r \delta_k^2.$$

Este es el teorema de aproximación de rango inferior de Eckart y Young, ver Gabriel (1971) o Greenacre (1984).

Cuando \mathbf{X} es una matriz de datos multivariantes (hileras asociadas a individuos y columnas asociadas a variables) centrada por su vector de medias, dos biplots útiles en un contexto de exploración multivariante se obtienen haciendo $\alpha = 0$ y $\alpha = 1$ en (3.2). Para $\alpha = 1$ se obtiene

$$\mathbf{J} = [\mathbf{l}_1 \ \mathbf{l}_2] \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix} \text{ y } \mathbf{K} = [\mathbf{m}_1 \ \mathbf{m}_2],$$

conocido como \mathbf{JK}^T - biplot. Para $\alpha = 0$ e introduciendo el factor \sqrt{n} , se obtiene s

$$\mathbf{G} = \sqrt{n} [\mathbf{I}_1 \ \mathbf{I}_2] \text{ y } \mathbf{H} = \left(1/\sqrt{n}\right) [\mathbf{m}_1 \ \mathbf{m}_2] \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix},$$

conocido como el \mathbf{GH}^T - biplot. El \mathbf{JK}^T - biplot aproxima las distancias euclidianas entre los individuos, es decir, los productos internos entre las hileras de \mathbf{X} están dados en la métrica \mathbf{I} . Este biplot coincide con la gráfica de puntajes de los dos primeros componentes principales de \mathbf{X} basados en su matriz de varianzas y covarianzas \mathbf{S} . Si \mathbf{X} está estandarizada, entonces el \mathbf{JK}^T - biplot es equivalente al gráfico de un análisis de componentes principales basado en su matriz de correlación. En el \mathbf{GH}^T - biplot se tiene

$$\mathbf{HH}^T = (1/n) \sum_{k=1}^2 \delta_k^2 \mathbf{m}_k \mathbf{m}_k^T, \text{ y, puesto que } \mathbf{S} = (1/n) \sum_{k=1}^r \delta_k^2 \mathbf{m}_k \mathbf{m}_k^T, \text{ tenemos que (aplicando el teorema de aproximación de rango inferior) } \mathbf{HH}^T \text{ es la matriz de rango 2 que mejor se aproxima a } \mathbf{S} \text{ en el sentido de mínimos cuadrados, con bondad de ajuste } (\delta_1^4 + \delta_2^4) / \sum_{k=1}^r \delta_k^4.$$

Entonces, la matriz de varianzas y covarianzas de \mathbf{X} se aproxima por \mathbf{HH}^T , esto es, las longitudes de las hileras de \mathbf{H} aproximan las desviaciones estándar de las variables en \mathbf{X} ; y los cosenos de los ángulos entre estas aproximan las respectivas correlaciones. Más aun, las distancias entre las hileras de \mathbf{G} aproximan las distancias de Mahalanobis entre los individuos, con bondad de ajuste $2/r$, Gabriel (1971). En resumen, en el \mathbf{GH}^T - biplot la métrica entre las columnas de \mathbf{X} está dada por \mathbf{I} y la métrica de las hileras está dada por $(\mathbf{X}^T \mathbf{X})^{-1}$. Si en el \mathbf{GH}^T - biplot se omiten las marcas de hilera, se obtiene una gráfica llamada el \mathbf{h} - plot. Otra elección para $\alpha = 1/2$, la cual proporciona un biplot útil como herramienta de diagnóstico para modelos de rango 2 en tablas de contingencia de dos entradas (Bradu y Gabriel, 1978).

En un contexto más general, para obtener el biplot de \mathbf{X} que proporcione los productos internos de sus hileras en la métrica Φ y los productos internos de sus columnas en la métrica Ψ , se usará la *descomposición en valores singulares generalizada* (DVSG) de \mathbf{X} , definida como sigue: sean Ψ y Φ matrices $n \times n$ y $p \times p$ positivas definidas, respectivamente. Considérese la eigendescomposición de Ψ y la DVS de $\Psi^{1/2} \mathbf{X} \Phi^{1/2}$,

$$\Psi = \mathbf{V} \Lambda \mathbf{V}^T, \ \Psi^{1/2} \mathbf{X} \Phi^{1/2} = \mathbf{P} \Theta \mathbf{Q}^T;$$

donde \mathbf{V} es una matriz ortonormal y $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$; las columnas de \mathbf{V} son los eigenvectores ortonormales Ψ asociados a los eigenvalores λ_j 's. $\mathbf{P}^T \mathbf{P} = \mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ y $\Theta = \text{diag}(\theta_1 \dots \theta_r)$ con $\theta_1 \geq \dots \geq \theta_r$ definidos como en (3.1). Ahora, sean $\mathbf{R} = \Psi^{-1/2} \mathbf{P}$ y $\mathbf{S} = \Phi^{-1/2} \mathbf{Q}$, se tiene

$$\mathbf{X} = \mathbf{R} \Theta \mathbf{S}^T = \sum_{k=1}^r \theta_k \mathbf{r}_k \mathbf{s}_k^T, \tag{3.3}$$

donde $\mathbf{R}^T \Psi \mathbf{R} = \mathbf{S}^T \Phi \mathbf{S} = \mathbf{I}$; esto es, las columnas \mathbf{r}_j 's y \mathbf{s}_j 's de \mathbf{R} y \mathbf{S} son ortonormales con respecto a Φ y Ψ respectivamente. La expresión (3.3) define la DVSG de \mathbf{X} en la métrica Φ para las hileras y la métrica Ψ para las columnas. La generalización del teorema de aproximación de rango inferior es:

$$\mathbf{X}_{(s)} = \sum_{k=1}^s \theta_k \mathbf{r}_k \mathbf{s}_k^T$$

para $t \leq r$ es la aproximación de mínimos cuadrados generalizados de \mathbf{X} , en el sentido de que minimiza

$$\text{tr}[\Psi(\mathbf{X} - \mathbf{A})\Phi(\mathbf{X} - \mathbf{A})^T],$$

entre todas las matrices \mathbf{A} de rango t o menor. La bondad de ajuste de $\mathbf{X}_{(s)}$ a \mathbf{X} está dada por $\sum_{k=1}^s \theta_k^2 / \sum_{k=1}^r \theta_k^2$.

Entonces, para construir el biplot de \mathbf{X} con la métrica Φ para las hileras y la métrica Ψ para las columnas, se definen las matrices

$$\mathbf{E} = [\mathbf{r}_1 \quad \mathbf{r}_2] \begin{bmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{bmatrix}^\alpha = [\theta_1^\alpha \mathbf{r}_1 \quad \theta_2^\alpha \mathbf{r}_2],$$

$$\mathbf{F} = [\mathbf{s}_1 \quad \mathbf{s}_2] \begin{bmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{bmatrix}^{1-\alpha} = [\theta_1^{1-\alpha} \mathbf{s}_1 \quad \theta_2^{1-\alpha} \mathbf{s}_2], \alpha \in [0,1];$$

y obtenemos el respectivo biplot de \mathbf{X} . En la discusión anterior, se ha seguido la idea de Greenacre (1984) pág. 346.

4. BILOT DE LOS RESIDUOS EN EL SEGUNDO NIVEL

Consideremos el vector de errores \mathbf{u} en el segundo nivel en el modelo completo dado en (2.3); este vector se expresará del siguiente modo: sean

$$\mathbf{B} = \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_J^T \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_J \end{bmatrix}, \Gamma = [\gamma_0 \gamma_1 \dots \gamma_K], \mathbf{U} = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_J^T \end{bmatrix} = \begin{bmatrix} u_{01} & u_{11} & \dots & u_{K1} \\ u_{02} & u_{12} & \dots & u_{K2} \\ \vdots & \vdots & & \vdots \\ u_{0J} & u_{1J} & \dots & u_{KJ} \end{bmatrix},$$

entonces, tenemos

$$\mathbf{B} = \mathbf{Z}\Gamma + \mathbf{U},$$

por lo que se pueden expresar los errores en el segundo nivel como una matriz de errores multivariantes. En la matriz \mathbf{U} la hilera j correspondiente a j -ésimo grupo y la columna k está asociada con el k -ésimo parámetro en el modelo (2.1). En esta sección \mathbf{U} denota la matriz de errores estimada, es decir la matriz de residuos, y se considera como una matriz de datos multivariantes y se usa el biplot para explorarla; una aplicación similar del biplot se hace en Smith y Cornell (1993): Con este objetivo, se usará la idea de Gnanadesikan y Kettenring (1972), ver también Barnett y Lewis (1994). Gnanadesikan y Kettenring (1972) sugieren calcular estadísticos univariados que midan la contribución de las observaciones en la estructura multivariada de los datos. Específicamente, proponen usar las distancias en la clase de las formas cuadráticas

$$d_\xi(\mathbf{u}_j, \bar{\mathbf{u}}) = (\mathbf{u}_j - \bar{\mathbf{u}})^T \Sigma^\xi (\mathbf{u}_j - \bar{\mathbf{u}}), \xi \in \mathbb{R},$$

donde Σ es la matriz de varianzas y covarianzas de los datos y $\bar{\mathbf{u}}$ es el vector de medias. Así, para $\xi = -1$, si Σ es no singular, se tiene

$$d_{-1}(\mathbf{u}_j, \bar{\mathbf{u}}) = (\mathbf{u}_j - \bar{\mathbf{u}})^T \Sigma^{-1} (\mathbf{u}_j - \bar{\mathbf{u}}).$$

Esta cantidad, la distancia de Mahalanobis, mide la contribución de las observaciones en la orientación de los últimos componentes principales (cp's), por lo que es útil para detectar observaciones que caen lejos de la nube de puntos. Para $\xi = 0$ se tiene

$$d_{-1}(\mathbf{u}_j, \bar{\mathbf{u}}) = (\mathbf{u}_j - \bar{\mathbf{u}})^T (\mathbf{u}_j - \bar{\mathbf{u}}),$$

la distancia euclideana, la cual es sensible a observaciones que inflan la escala y por lo tanto la varianza y las covarianzas. Para $\xi = 1$, se tiene

$$d_1(\mathbf{u}_j, \bar{\mathbf{u}}) = (\mathbf{u}_j - \bar{\mathbf{u}})^T \Sigma (\mathbf{u}_j - \bar{\mathbf{u}}),$$

esta cantidad es útil para determinar cuáles residuos tienen mayor influencia en la orientación y escala de los primeros cp's de Σ . También sugieren usar las distancias entre los \mathbf{u} 's

$$d_\xi(\mathbf{u}_i, \mathbf{u}_j) = (\mathbf{u}_i - \mathbf{u}_j)^T \Sigma^\xi (\mathbf{u}_i - \mathbf{u}_j), \quad \xi \in \mathbb{R},$$

Para explorar gráficamente los residuos, se introducirá en el biplot de \mathbf{U} la métrica para las hileras dada por \mathbf{S}^ξ , donde $\xi \in \mathbb{R}$ y $\mathbf{S} = (1/J)\mathbf{U}^T\mathbf{U}$ es la estimación de la matriz de varianzas y covarianzas de \mathbf{U} . Para la métrica de las columnas se usará \mathbf{I} , lo que proporciona $\mathbf{U}^T\mathbf{U} = \mathbf{J}\mathbf{S}$. Si la DVS de \mathbf{U} es

$$\mathbf{U} = \mathbf{L}\Delta\mathbf{M}^T = \sum_{i=1}^r \delta_i \mathbf{l}_i \mathbf{m}_i^T$$

y puesto que $\bar{\mathbf{u}} = \mathbf{0}$, se tiene entonces que la métrica es $d_\xi(\mathbf{u}_j, \mathbf{0}) = \mathbf{u}_j^T \mathbf{S}^\xi \mathbf{u}_j$. Para $\xi = 1$ resulta

$$\mathbf{u}_j^T \mathbf{S} \mathbf{u}_j = \sum_{i=1}^r \delta_i^2 (\mathbf{m}_i^T \mathbf{u}_j),$$

la cual, efectivamente, pondera más a los primeros cp's. Para $\xi = -1$, se tiene

$$\mathbf{u}_j^T \mathbf{S}^{-1} \mathbf{u}_j = \sum_{i=1}^r \delta_i^{-2} (\mathbf{m}_i^T \mathbf{u}_j),$$

la cual ahora pondera más a los últimos cp's. Para $\xi = 0$, se tiene

$$\mathbf{u}_j^T \mathbf{S}^0 \mathbf{u}_j = \mathbf{u}_j^T \mathbf{u}_j = \|\mathbf{u}_j\|^2,$$

la longitud cuadrada de \mathbf{u}_j . De modo que los primeros cp's son sensibles a observaciones que inflan la varianza o las covarianzas (o correlaciones, si es que los datos están estandarizados). Por otro lado, los últimos cp's son sensibles a observaciones que agregan dimensiones espurias a los datos u ocultan singularidades, por lo que, resultan de especial interés los primeros y los últimos cp's. Siguiendo el procedimiento descrito en la sección anterior con $\Psi = \mathbf{I}$ y $\Phi_\xi = \mathbf{S}^\xi = (1/J^\xi)(\mathbf{U}^T\mathbf{U})^\xi$, se tiene que

$$\mathbf{S} = (1/J)(\mathbf{M}\Delta^2\mathbf{M}^T), \quad \mathbf{S}^\xi = (1/J^\xi)(\mathbf{M}\Delta^{2\xi}\mathbf{M}^T), \quad (\mathbf{S}^\xi)^{1/2} = (1/\sqrt{J^\xi})(\mathbf{M}\Delta^{2\xi}\mathbf{M}^T),$$

por lo que

$$\mathbf{U}(\mathbf{S}^\xi)^{1/2} = (1/\sqrt{J^\xi})(\mathbf{L}\Delta^{1+\xi}\mathbf{M}^T) = (1/\sqrt{J^\xi})\sum_{i=1}^r \delta_i^{1+\xi} \mathbf{l}_i \mathbf{m}_i^T$$

es la DVS de $\mathbf{U}(\mathbf{S}^\xi)^{1/2}$. Ahora, sean $\mathbf{R}_\xi = (1/\sqrt{J^\xi})\mathbf{L}$ y $\mathbf{S}_\xi = (\mathbf{S}^\xi)^{-1/2}\mathbf{M} = \sqrt{J^\xi}(\mathbf{M}\Delta^{-\xi})$, entonces la DVSG de \mathbf{U} está dada por

$$\mathbf{U} = \mathbf{R}_\xi \Delta^{1+\xi} \mathbf{S}_\xi^T = \sum_{i=1}^r \delta_i^{1+\xi} \mathbf{r}_{\xi i} \mathbf{s}_{\xi i}^T.$$

Finalmente, el biplot de la matriz \mathbf{U} , con la métrica $\Psi = \mathbf{I}$ para las columnas y la métrica $\Phi_\xi = \mathbf{S}^\xi = (1/J^\xi)(\mathbf{U}^T\mathbf{U})^\xi$ para las hileras, es

$$\mathbf{E} = \begin{bmatrix} \mathbf{r}_{\xi 1} & \mathbf{r}_{\xi 1} \\ \mathbf{r}_{\xi 1} & \mathbf{r}_{\xi 1} \end{bmatrix} \begin{bmatrix} \delta_1^{1+\xi} & 0 \\ 0 & \delta_2^{1+\xi} \end{bmatrix}^\alpha = \begin{bmatrix} \delta_1^{(1+\xi)\alpha} \mathbf{r}_{\xi 1} & \delta_2^{(1+\xi)\alpha} \mathbf{r}_{\xi 2} \end{bmatrix} = (1/\sqrt{J^\xi}) \begin{bmatrix} \delta_1^{(1+\xi)\alpha} \mathbf{l}_1 & \delta_2^{(1+\xi)\alpha} \mathbf{l}_2 \end{bmatrix},$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{s}_{\xi 1} & \mathbf{s}_{\xi 1} \\ \mathbf{s}_{\xi 1} & \mathbf{s}_{\xi 1} \end{bmatrix} \begin{bmatrix} \delta_1^{1+\xi} & 0 \\ 0 & \delta_2^{1+\xi} \end{bmatrix}^{1-\alpha} = \begin{bmatrix} \delta_1^{(1+\xi)(1-\alpha)} \mathbf{s}_{\xi 1} & \delta_2^{(1+\xi)(1-\alpha)} \mathbf{s}_{\xi 2} \end{bmatrix} = \sqrt{J^\xi} \begin{bmatrix} \delta_1^{1-\alpha-\alpha\xi} \mathbf{m}_1 & \delta_2^{1-\alpha-\alpha\xi} \mathbf{m}_2 \end{bmatrix},$$

donde $\alpha \in [0,1]$ y $\xi \in \mathbb{R}$. Los biplots que se usarán son los que se obtienen fijando a $\alpha = 1$ y variando $\xi = -1$, $\xi = 1$ y $\xi = 0$. Expresando estos biplots en términos de la DVS de \mathbf{U} , se tiene

$$\begin{aligned} \xi = -1, & \quad \mathbf{G} = \sqrt{J} \begin{bmatrix} \mathbf{I}_1 & \mathbf{I}_2 \end{bmatrix}, & \quad \mathbf{H} = (1/\sqrt{J}) \begin{bmatrix} \delta_1 \mathbf{m}_1 & \delta_2 \mathbf{m}_2 \end{bmatrix}; \\ \xi = 1, & \quad \mathbf{E} = (1/\sqrt{J}) \begin{bmatrix} \delta_1^2 \mathbf{I}_1 & \delta_2^2 \mathbf{I}_2 \end{bmatrix}, & \quad \mathbf{F} = \sqrt{J} \begin{bmatrix} \delta_1^{-1} \mathbf{m}_1 & \delta_2^{-1} \mathbf{m}_2 \end{bmatrix}; \\ \xi = 0, & \quad \mathbf{J} = \begin{bmatrix} \delta_1 \mathbf{I}_1 & \delta_2 \mathbf{I}_2 \end{bmatrix}, & \quad \mathbf{K} = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 \end{bmatrix}. \end{aligned}$$

Por lo mencionado anteriormente, también es de utilidad considerar a los dos últimos cp's en los biplots

$$\begin{aligned} \xi = -1, & \quad \mathbf{G} = \sqrt{J} \begin{bmatrix} \mathbf{I}_{r-1} & \mathbf{I}_r \end{bmatrix}, & \quad \mathbf{H} = (1/\sqrt{J}) \begin{bmatrix} \delta_{r-1} \mathbf{m}_{r-1} & \delta_r \mathbf{m}_r \end{bmatrix}; \\ \xi = 0, & \quad \mathbf{J} = \begin{bmatrix} \delta_{r-1} \mathbf{I}_{r-1} & \delta_r \mathbf{I}_r \end{bmatrix}, & \quad \mathbf{K} = \begin{bmatrix} \mathbf{m}_{r-1} & \mathbf{m}_r \end{bmatrix}; \\ \xi = 1, & \quad \mathbf{E} = (1/\sqrt{J}) \begin{bmatrix} \delta_{r-1}^2 \mathbf{I}_{r-1} & \delta_r^2 \mathbf{I}_r \end{bmatrix}, & \quad \mathbf{F} = \sqrt{J} \begin{bmatrix} \delta_{r-1}^{-1} \mathbf{m}_{r-1} & \delta_r^{-1} \mathbf{m}_r \end{bmatrix} \end{aligned}$$

Nótese que para $\xi = -1$ se obtiene el \mathbf{GH}^T - *biplot* de \mathbf{U} , y para $\xi = 0$ se obtiene el \mathbf{JK}^T - *biplot*, por lo que en el procedimiento anterior se han obtenido los conocidos biplots propuestos por Gabriel (1971). Cuando $\xi = 1$, se tiene un biplot en el cual \mathbf{FF}^T aproxima a \mathbf{S}^{-1} . Por otra parte, nótese que fijando $\alpha = 0$ y haciendo $\xi = -1$ se obtiene también el \mathbf{GH}^T - *biplot*; en $\xi = 1$, \mathbf{FF}^T aproxima a JS y $\xi = 0$, \mathbf{FF}^T aproxima a $\mathbf{U}^T \mathbf{U}$. Si en lugar de centrar a la matriz \mathbf{U} se estandariza (residuos estandarizados) entonces la métrica que se introduce para las hileras es $\Phi_\xi = \mathbf{R}^\xi = (1/J^\xi)(\mathbf{U}^T \mathbf{U})^\xi$, y ahora el énfasis es en la estructura de correlación más que en la varianza y covarianza.

5. EJEMPLO

El Centro de Genética Forestal de la Universidad Veracruzana está realizando un estudio de mejoramiento de plantaciones de *Pinus Patula*, para lo que se ha hecho un ensayo de progenies. Para ilustrar la metodología presentada, se seleccionaron 37 familias de baja variabilidad en la última etapa de su crecimiento (una desviación para la altura final entre 2.9 y 4.6). El objetivo general que se plantea es el estudio de la variabilidad entre los patrones de crecimiento de familias, para lo cual se utiliza el perfil estimado de crecimiento por familia sobre 8 semanas. Cabe hacer notar que algunas familias solo se midieron 6 ó 7 semanas. Como una primera aproximación al problema de modelaje, se postula el siguiente modelo

$$Y_{ij} = \beta_{0j} + \beta_{1j} t_{ij} + \beta_{2j} t_{ij}^2 + \varepsilon_{ij},$$

con $i = 1, \dots, n_j = 6 \text{ ó } 7 \text{ u } 8$; $j = 1, \dots, 37$ y t_{ij} es el tiempo en el que fue observada la altura Y_{ij} . En el segundo nivel se ajustó Un *modelo de medias*

$$\beta_{0j} = \gamma_{00i} + u_{0j},$$

$$\beta_{1j} = \gamma_{01} + u_{1j},$$

$$\beta_{2j} = \gamma_{02} + u_{2j}.$$

El ajuste de este modelo se realizó utilizando el paquete ML3 (Proser *et al.*, 1991) y se obtuvo la estimación de los efectos fijos y aleatorios en el modelo, así como su desviación estándar, los cuales se presentan en la Tabla 1. Nótese que en este problema de curvas de crecimiento, el primer nivel del modelo lo determinan las familias y el segundo nivel está determinado por los tiempos de observación para cada familia.

Tabla 1. Estimaciones de los efectos fijos y aleatorios.
Entre paréntesis se dan las desviaciones estándar.

Efecto	EFECTOS FIJOS		EFECTOS ALEATORIOS	
	Constante	4.205 (0.195)	0.353 (0.340)	
Tiempo	- 1.444 (0.140)	- 0.469 (0.232)	0.447 (0.168)	
Tiempo cuadrado	0.345 (0.026)	0.128 (0.042)	- 0.103 (0.031)	0.022 (0.006)

Con la matriz de residuos en el segundo nivel \mathbf{U} ; se obtuvo la matriz de componentes de varianza y covarianza (estimada) y se siguió el procedimiento presentado en la sección anterior. En la Figura 1 se muestra el h -plot de \mathbf{U} , se puede observar que el parámetro asociado al tiempo cuadrado no varía mucho (relativamente) con respecto a los parámetros de la constante y el tiempo; por lo que se puede considerar fijo en el primer nivel del modelo. En la Figura 2, se presenta el \mathbf{GH}^T - *biplot*, donde se puede ver que las familias 18 y 19 presentan un comportamiento atípico. En las Figuras 3 y 4, se presentan los \mathbf{JK}^T - *biplot* con los ejes 1 y 2 y 2 y 3, respectivamente. También se han graficado elipses de confianza al 95 %. En estas gráficas, se corrobora el comportamiento atípico de las familias 18 y 29; sin embargo, no parece que estas agreguen una tercera dimensión espuria en la estructura de covarianza de \mathbf{U} , ya que su comportamiento no discrepa en el tercer eje. En las Figuras 5, 6 y 7, se muestran los diagramas de dispersión, con elipses de confianza al 95 %, entre los residuos obtenidos. También se detecta el comportamiento extremo de las familias 18 y 29, y se puede ver que la variabilidad del parámetro asociado al tiempo cuadrado es pequeña.

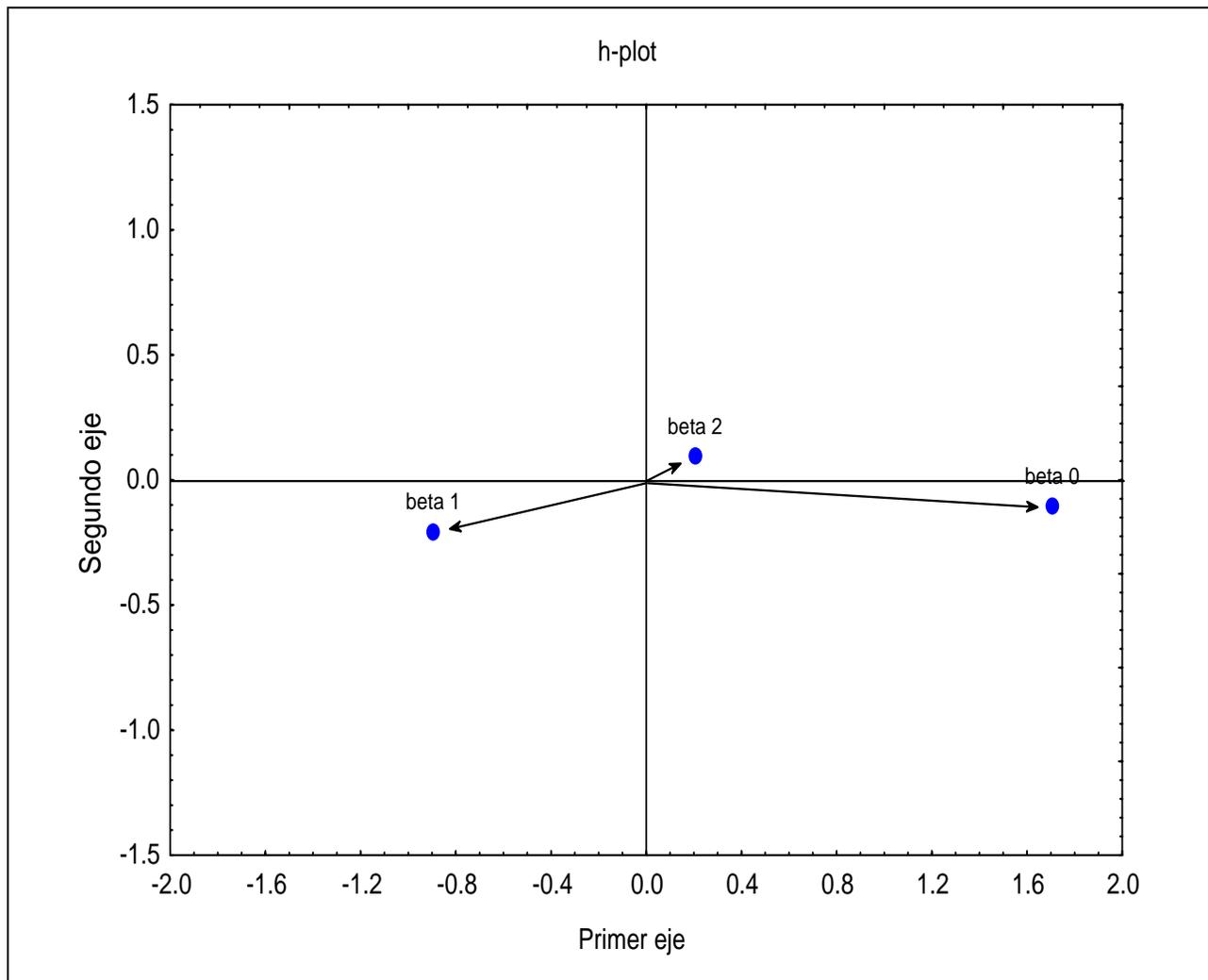


Figura 1.

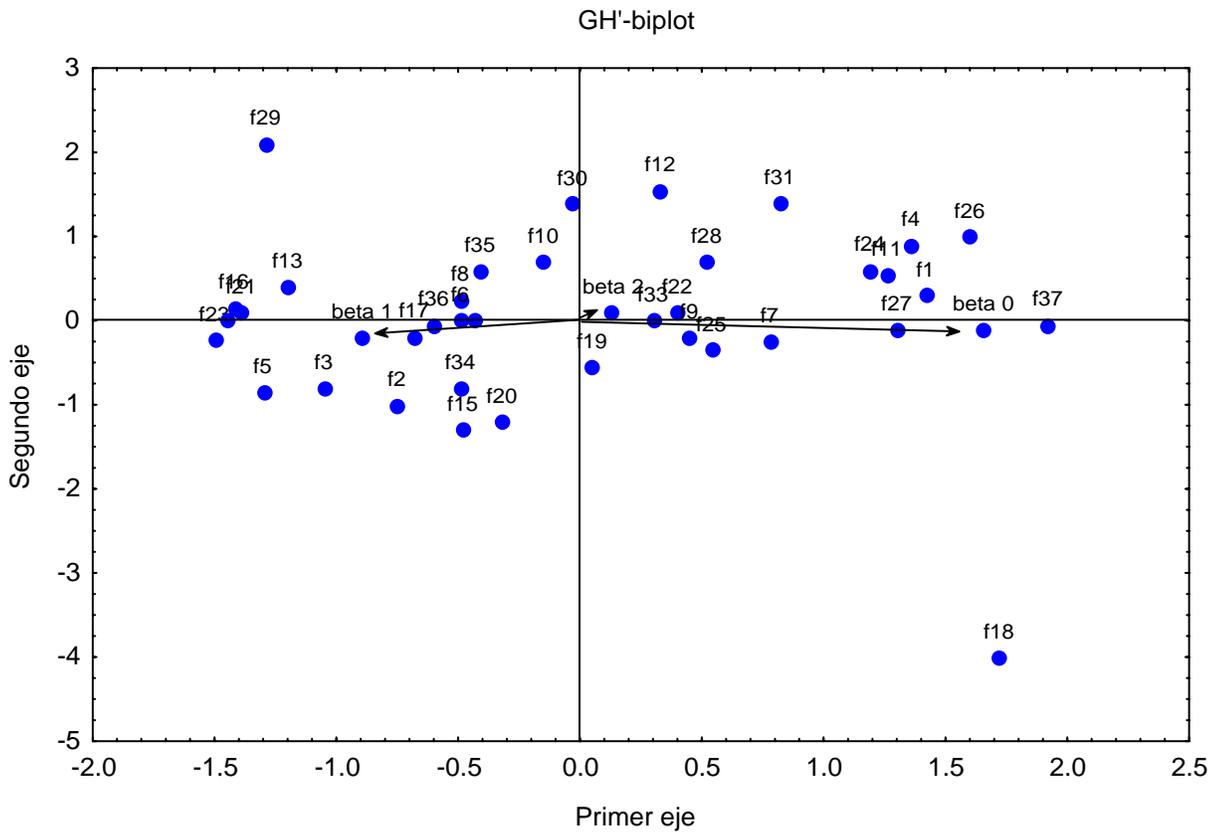


Figura 2.

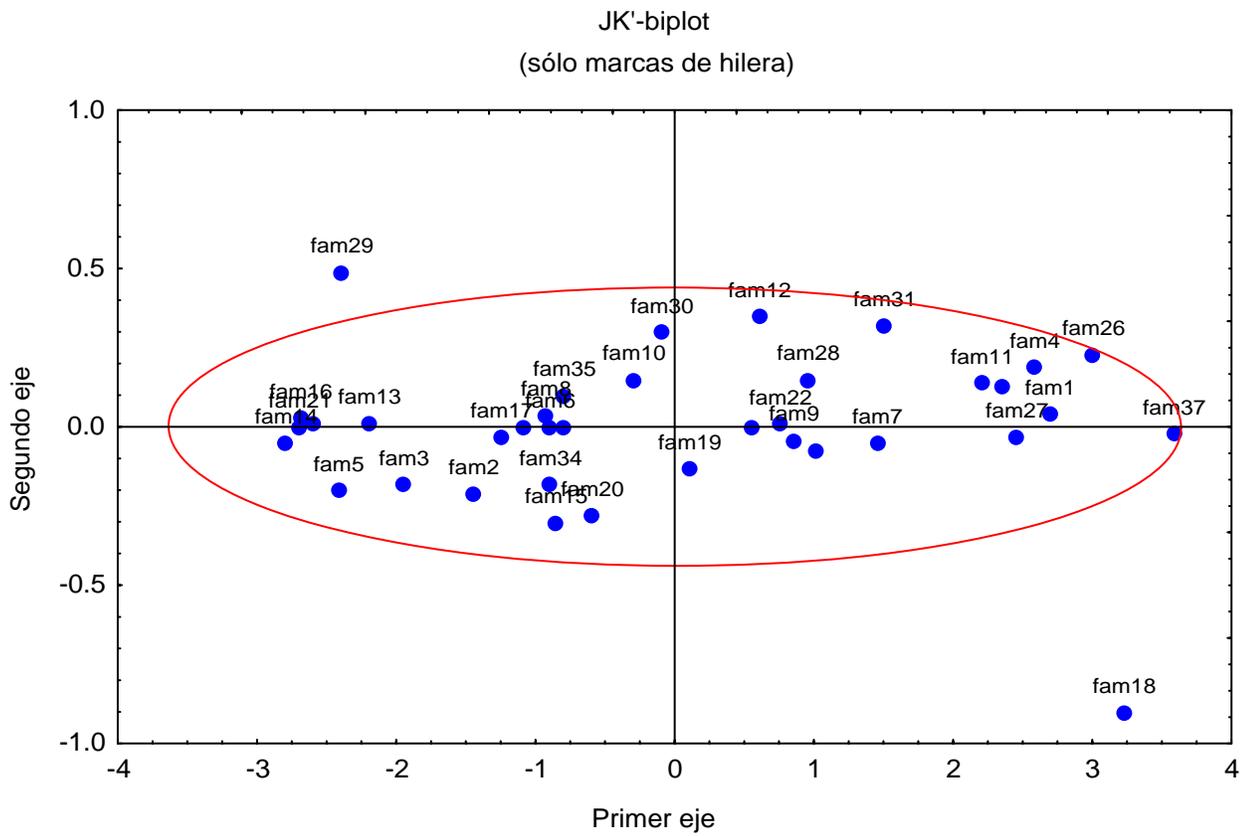


Figura 3.

JK'-biplot
(sólo marcas de hilera)

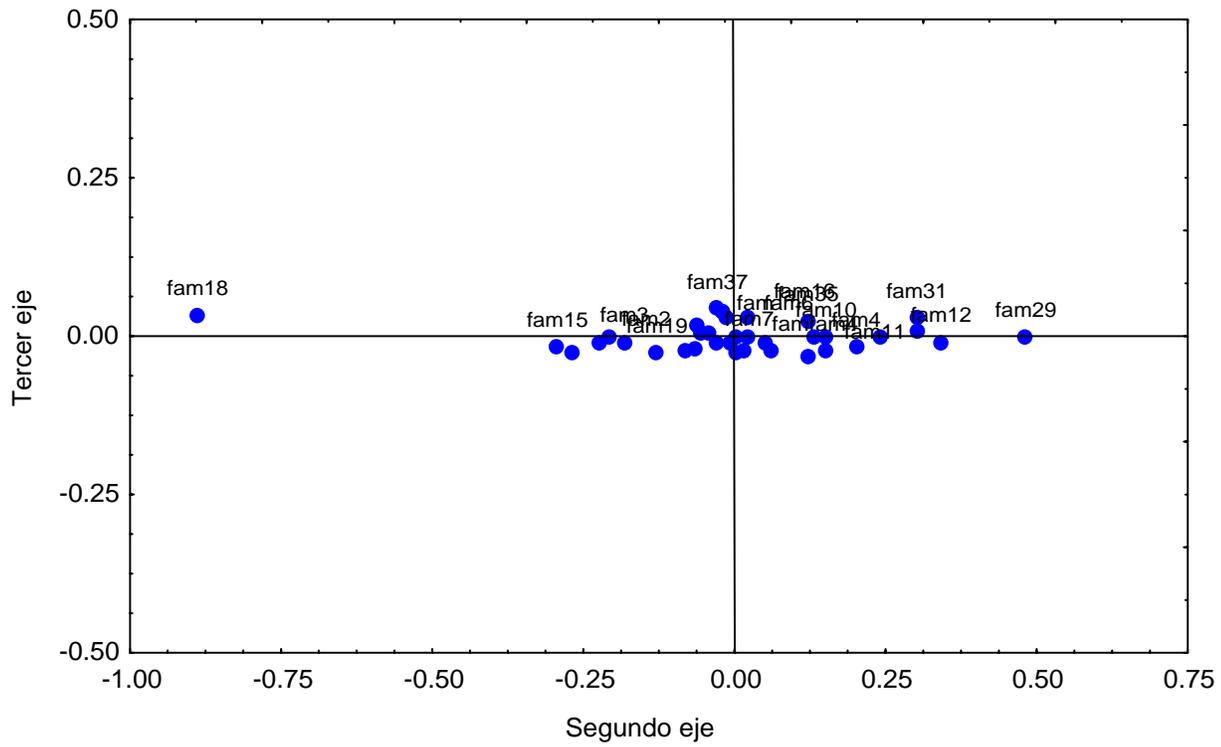


Figura 4.

Diagrama de dispersión
de los residuales

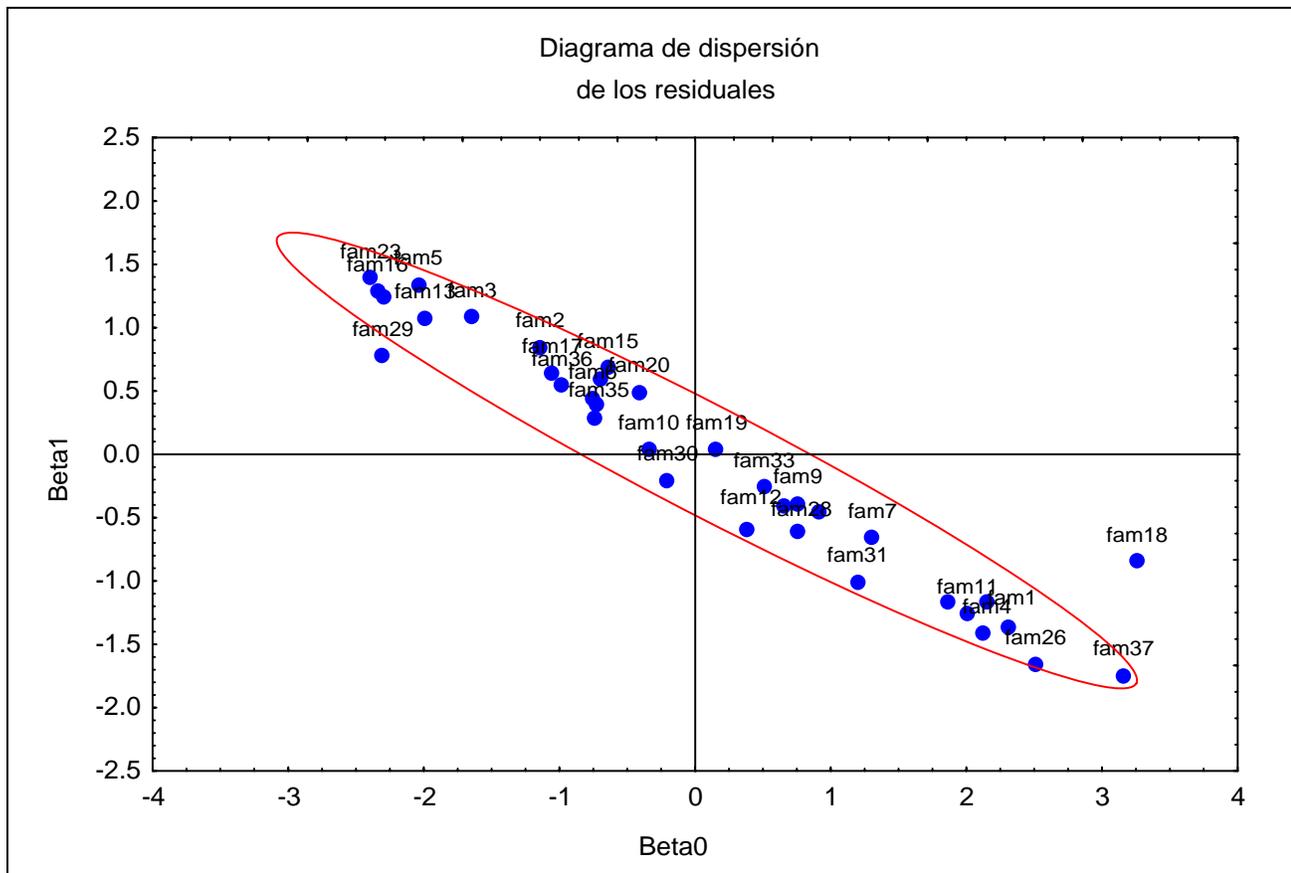


Figura 5.

Diagrama de dispersión
de los residuales

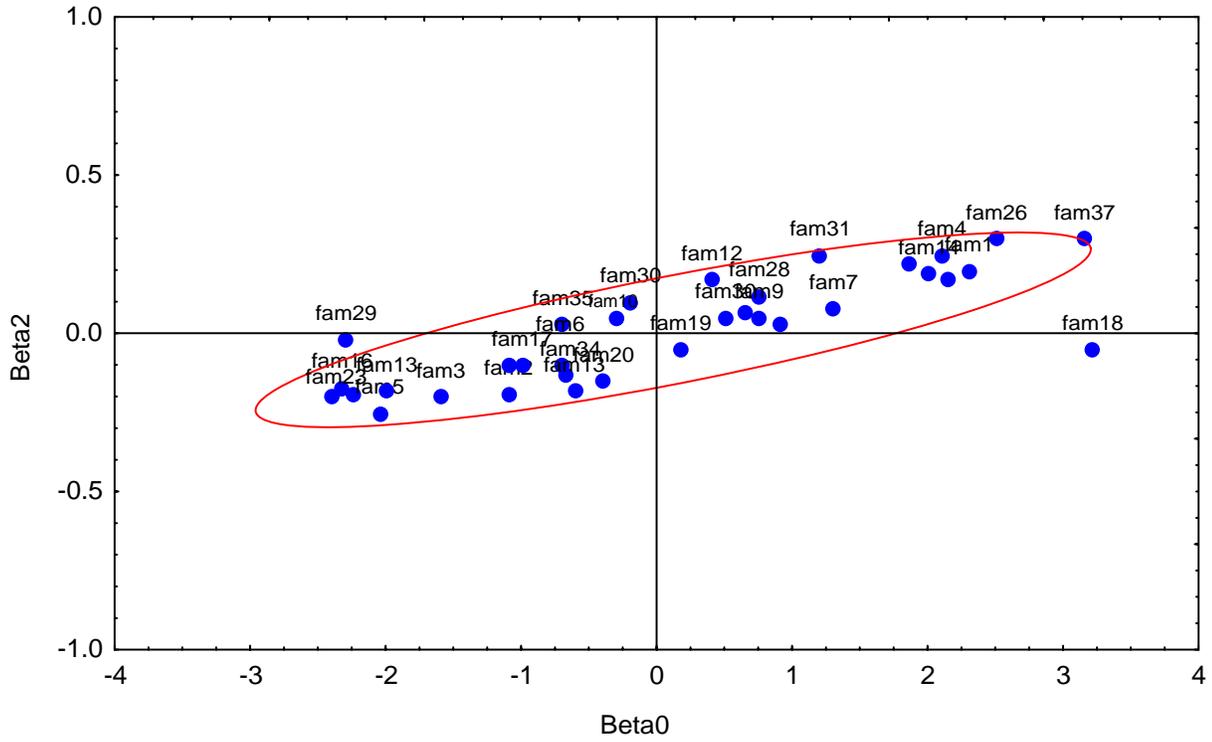


Figura 6.

Diagrama de dispersión
de los residuales

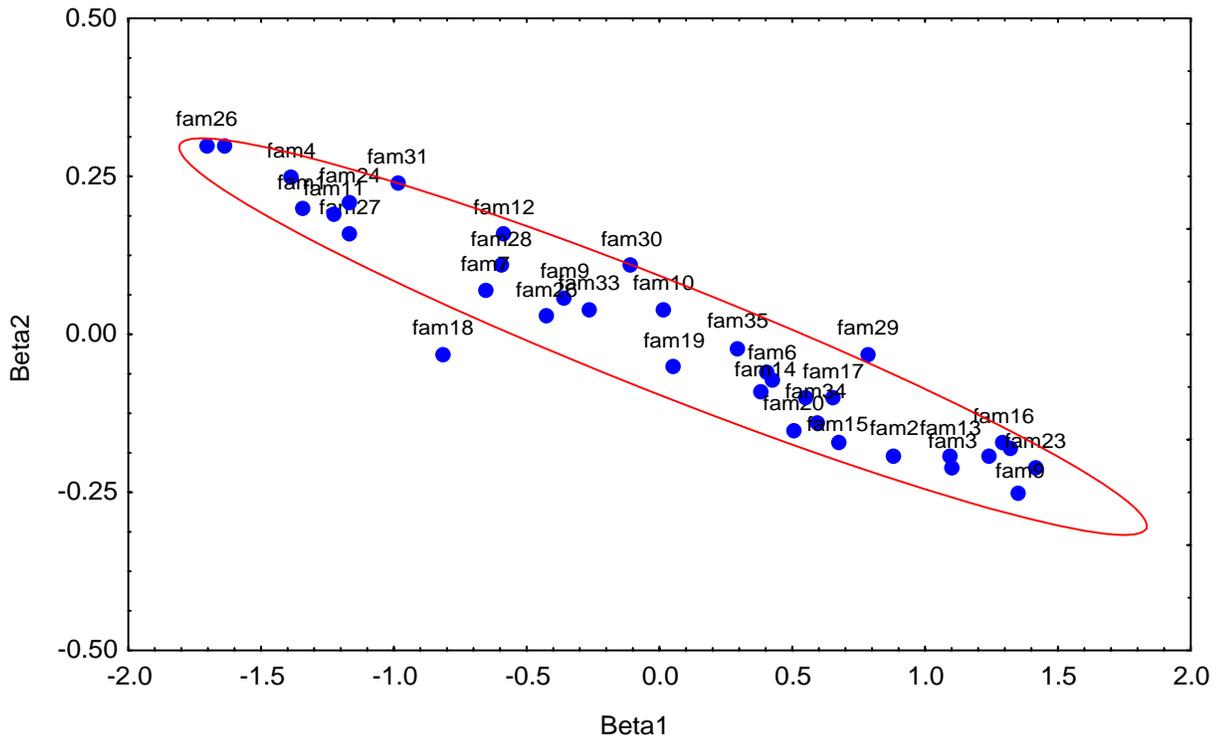


Figura 7.

AGRADECIMIENTOS

Agradezco a la Bióloga Lilia Baizabal, del Centro de Genética Forestal de la Universidad Veracruzana, el que permitiera usar los datos del ejemplo que se presenta en este trabajo, el cual forma parte del proyecto registrado ante CONACyT con el número 4066-49404. Agradezco también el apoyo de Cecilia Cruz y Sergio Juárez en la edición del trabajo.

REFERENCIAS

- BARNETT, V. and T. LEWIS (1977): "Outliers in Statistical Data", third edition, New York, Wiley.
- BRYK, A.S. and S.W. RAUDENBUSH (1992): "Hierarchical Linear Models: Applications and Data Analysis Methods", London, Sage.
- BRADU, D. and K.R. GABRIEL (1978): "The biplot as a diagnostic tool for models of two way tables", **Technometrics** 20, 47-68.
- GABRIEL, K.R. (1971): "The biplot graphic display of matrices with application to principal components analysis", **Biometrika**, 58, 453-467.
- GNANADESIKAN, R. and J.R. KETTENRING (1972): "Robust estimates, residuals and outlier detection with multiresponse data", **Biometrics** 28, 81-124.
- GOLDSTEIN, H. (1995): **Multilevel Statistical Models**, second edition. London: Edward Arnold, New York: Halsted.
- GREENACRE, M.J. (1984): **Theory and application of Correspondence Analysis**, London, Academic Press.
- OJEDA, M.M. (1992): Aspectos teóricos, metodológicos y computacionales en el análisis de datos para muestras complejas, Tesis de Doctorado, Universidad de La habana, Cuba.
- _____ (1993): Multilevel modelling for complex samples. Trabajo presentado en el 1993 European Meeting of Psychometric Society, Barcelona, España.
- PROSER, R.; J. RABASH and H. GOLDSTEIN (1991): **ML3: Software for three-level analysis: User's guide for version 2**. Institute of Education, London.
- SEARLE, S.R. (1982): **Matrix Algebra Useful for Statistics**, New York, Wiley.
- SELTZER, M.H. (1993): "Sensitivity analysis for fixed effects in the hierarchical model: A Gibbs sampling approach", **Journal of Educational Statistics** 18(3), 207-235.
- SMITH, W. and J.A. CORNELL (1993): "Biplot displays for looking at multiple response data in mixture experiments", **Technometrics** 35, 337-350.