

# Análisis de Homogeneidad en ViSta “The Visual Statistics System”\*

*Rubén D. Ledesma  
Pedro M. Valero Mora  
Forrest W. Young*

Este documento describe el funcionamiento de un módulo de Análisis de Homogeneidad por Mínimos Cuadrados Alternantes integrable como plug-in al programa ViSta “The Visual Statistics System”. El programa permite analizar y visualizar, en un espacio de pocas dimensiones, la estructura de asociaciones entre variables categóricas así como las similitudes entre los objetos a los cuales esas categorías se aplican. La característica más distintiva del programa es la utilización de métodos gráficos dinámicos que facilitan la exploración y comprensión de los resultados de HOMALS. El programa también permite salvar el modelo y/o crear datos a partir de los resultados del mismo.

## 1. Introducción

En este trabajo se describe un programa informático original que permite realizar Análisis de Homogeneidad por Mínimos Cuadrados Alternantes dentro del programa ViSta “The Visual Statistics System” (Young, 1991-2004). El Análisis de Homogeneidad es uno de los modelos básicos de la familia del Escalamiento Óptimo del sistema Gifi (Gifi, 1990), el cual comprende una serie de técnicas exploratorias de análisis multivariado no lineal, extensiones del Análisis en Componentes Principales y de Correlación Canónica al caso de variables nominales o con niveles mixtos de medida. Dentro de esa familia, el Análisis de Homogeneidad se presenta como el modelo menos restrictivo, en el sentido de que trabaja solo con un grupo de variables y todas ellas son analizadas por su información nominal.

Conceptualmente, el método es similar al Análisis Factorial de Correspondencias Múltiples (AFCM) de la ‘Escuela Francesa’ (Benzécri, 1973), cuyo objetivo es identificar un espacio de pocas dimensiones o factores que permitan resumir y representar la estructura de asociaciones en un determinado grupo de variables categóricas, así como las similitudes entre los objetos a los cuales esas categorías se aplican. La diferencia esencial entre ambos métodos radica en el procedimiento de cálculo, aunque la solución es esencialmente la misma. El AFCM utiliza la función de Descomposición de Valores Propios y el Análisis de Homogeneidad se computa mediante Mínimos Cuadrados Alternantes (*Alternating Least Squares*, ALS), un algoritmo iterativo común a todos los modelos del sistema Gifi. Esta solución es conocida en la literatura con el acrónimo de HOMALS (*Homogeneity Analysis by means of ALS*). Una formulación precisa del método se puede consultar en Gifi (1990).

Lo que aquí presentamos es una implementación informática original de HOMALS, desarrollada sobre la base de una adaptación del código en LispStat de Bond y Michailidis (1996) al programa Vista “The Visual Statistics System”. A continuación, se describe brevemente el funcionamiento y las características generales del módulo HOMALS en ViSta.

## 2 Utilizando HOMALS en ViSta

### 2.1 Datos y opciones de análisis

Los datos para aplicar HOMALS en ViSta deben tener el formato de una matriz multivariada con variables categóricas, donde las filas (observaciones) son los *objetos* del análisis y los valores de las variables categóricas dispuestas en columnas, las *categorías*. La Figura 1 muestra una imagen parcial de una matriz de este tipo de datos en el entorno ViSta. Los datos corresponden a muestra de 540 víctimas de accidentes de tráfico ocurridos en la ciudad de Mar del Plata (Argentina) (Ungaro y Ledesma, 1999). Las víctimas (objetos del análisis) se han caracterizado según tres variables nominales, codificadas como se detalla en la Tabla 1.

---

\* Una versión más técnica de este trabajo puede encontrarse en: Ledesma, R, Valero Mora, P. y Young, F. (2002) Análisis de Homogeneidad en ViSta "The Visual Statistics System". *Metodología de las Ciencias del Comportamiento*. 4 (1), 139-149.

**Tabla 1 - Codificación de las variables y categorías del ejemplo**

Variable	Etiqueta de la Variable	Categorías	Etiqueta de las categorías
Edad de las víctimas	“Edad”	0 a 17 años; 18 a 30 años; 31 a 50 años; 51 a 60 años y más de 61 años.	0-17; 18-30; 31-50; 51-60; Mas61.
Gravedad de las lesiones sufridas	“Lesión”	Leve, Grave, Fatal.	Leve, Grave, Fatal.
Tipo de accidente protagonizado	“Accidente”	Atropello a peatón; Atropello a ciclista; Accidente con moto y Colisiones entre vehículos.	Peatón; Ciclista; Moto; Auto.

Aplicando HOMALS a estos datos se busca lograr una representación de los objetos (víctimas de accidentes) y de las categorías de las distintas variables en un mismo espacio, de modo que sus proximidades y alejamientos permitan describir, al mismo tiempo, las similitudes y diferencias entre los objetos y la estructura de relaciones entre variables. Al seleccionar HOMALS en ViSta aparecerá un cuadro de diálogo donde el usuario podrá definir ciertas opciones de análisis (Figura 1).

La primera opción (*Dimensions*) permite determinar el número de dimensiones en la solución, el resto de las opciones son relativas al computo. La opción *Iterations* permite fijar el número máximo de iteraciones del algoritmo ALS; las opciones *Solution Change* y *Function Change* permiten modificar los criterios de cambio entre un paso y otro de la iteración para la convergencia del cómputo. *Solution Change* mide el cambio en los puntajes de los objetos en cada paso de la iteración y *Function Change* en la función de pérdida o ajuste global de la solución. Finalmente, la opción *Label object by* permite definir un identificador para los objetos en los gráficos, pudiendo utilizarse las etiquetas de una variable categórica activa o no activa. En el ejemplo se han dejado las opciones tal como aparecen por defecto.

**Figura 1 - Imagen de una matriz de datos categóricos en ViSta y cuadro de diálogo para definir las opciones de análisis de HOMALS**

The screenshot displays the ViSta software interface. The main window shows a data matrix with columns for 'Type: Categ', 'Edad', 'Lesion', and 'Accidente'. The data rows are labeled 'Obs1' through 'Obs26'. A dialog box titled 'Dialog' is open in the foreground, showing the following options:

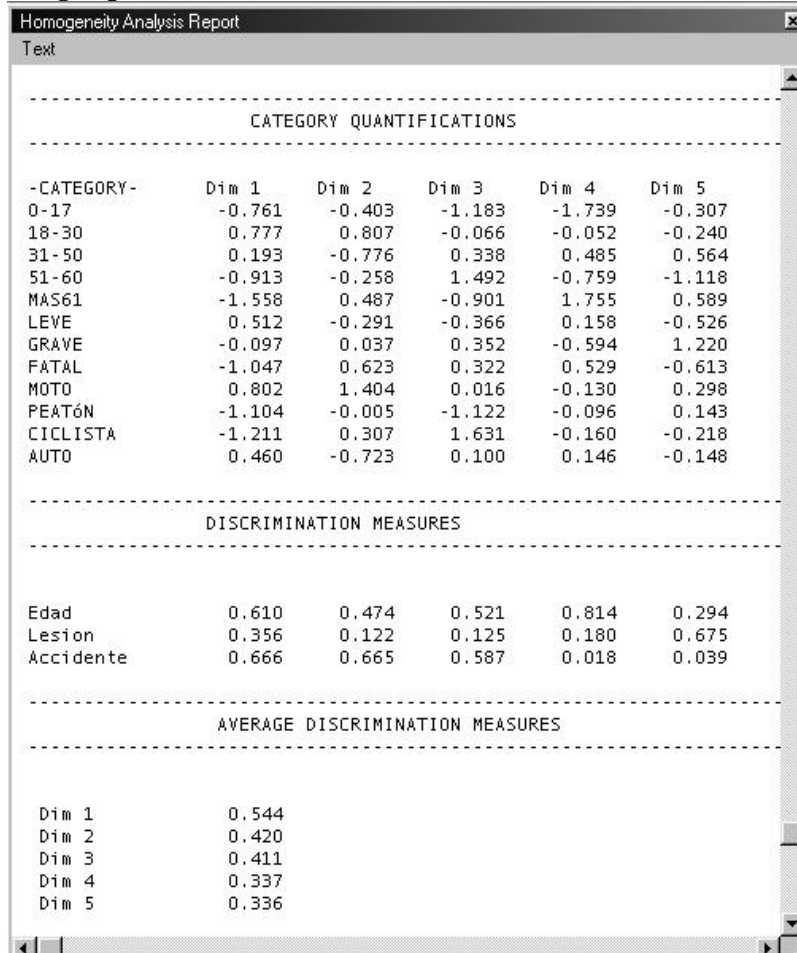
- Dimensions: 5
- Iterations: 100
- Solution change: 0.0001
- Function change: 0.0001
- Label object by: Obs labels

The dialog box also includes 'Ok' and 'Cancel' buttons. The background window shows the ViSta logo and a menu bar with options like File, Edit, Data, Transform, Analyze, Model, Options, Help, and Window. The status bar at the bottom indicates processing times and file information.

## 2.2 Salidas numéricas y gráficas de HOMALS en ViSta

Finalizada la fase de cómputo, puede obtenerse un reporte numérico de la solución o bien visualizarse los resultados mediante un SpreadPlot. La información de salida para el reporte numérico puede ser definida por el usuario mediante un cuadro de diálogo. La Figura 2 muestra una imagen parcial de la salida proporcionada por HOMALS en ViSta (no se incluyen las puntuaciones de los objetos).

Figura 2 - Imagen parcial de la salida numérica de HOMALS en ViSta



-----  
 CATEGORY QUANTIFICATIONS  
 -----

-CATEGORY-	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
0-17	-0.761	-0.403	-1.183	-1.739	-0.307
18-30	0.777	0.807	-0.066	-0.052	-0.240
31-50	0.193	-0.776	0.338	0.485	0.564
51-60	-0.913	-0.258	1.492	-0.759	-1.118
MAS61	-1.558	0.487	-0.901	1.755	0.589
LEVE	0.512	-0.291	-0.366	0.158	-0.526
GRAVE	-0.097	0.037	0.352	-0.594	1.220
FATAL	-1.047	0.623	0.322	0.529	-0.613
MOTO	0.802	1.404	0.016	-0.130	0.298
PEATÓN	-1.104	-0.005	-1.122	-0.096	0.143
CICLISTA	-1.211	0.307	1.631	-0.160	-0.218
AUTO	0.460	-0.723	0.100	0.146	-0.148

-----  
 DISCRIMINATION MEASURES  
 -----

Edad	0.610	0.474	0.521	0.814	0.294
Lesion	0.356	0.122	0.125	0.180	0.675
Accidente	0.666	0.665	0.587	0.018	0.039

-----  
 AVERAGE DISCRIMINATION MEASURES  
 -----

Dim 1	0.544
Dim 2	0.420
Dim 3	0.411
Dim 4	0.337
Dim 5	0.336

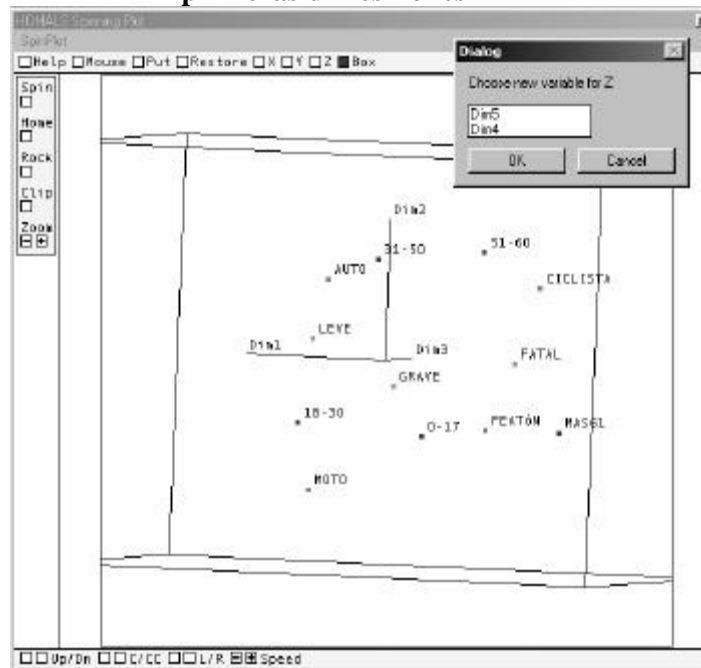
Esa salida incluye las cuantificaciones de las categorías (*Category Quantifications*); las medidas de discriminación de las variables (*Discrimination Measures*) y las medias de dichas medidas (*Average Discrimination Measures*). Las cuantificaciones de las categorías se presentan en una matriz donde las filas son las categorías de las variables y las columnas son las cuantificaciones obtenidas en cada dimensión definida para la solución. Las medidas de *discriminación* son las varianzas de las variables cuantificadas, una medida de la contribución de cada variable a la conformación de cada dimensión. Esas medidas tiene un valor máximo de 1 que se logra si los puntajes de los objetos están en grupos excluyentes y todos los individuos (objetos) dentro de una categoría son idénticos. En el ejemplo, encontramos que la variable *Accidente* obtiene las mejores medidas de discriminación en las dos primeras dimensiones. En términos de representación eso significa que las categorías de esa variable poseen la máxima dispersión respecto al origen del referencial, discriminando bien entre los objetos del análisis. La variable *Lesion* presenta el caso contrario, pues no parece contribuir sustancialmente a la conformación de las primeras dimensiones, obteniendo un valor alto en la cuarta.

Finalmente, el reporte muestra las medias de las medidas de discriminación de las variables en cada dimensión. En HOMALS estos son los *autovalores* de la solución, una medida de la varianza explicada por cada dimensión. Esos valores indican la cantidad de información recogida o contenida en cada dimensión y son una medida del *ajuste* de la solución.

Como se ha indicado previamente, puede resultar difícil interpretar de forma directa los resultados de HOMALS, sobre todo cuando el número de categorías, de variables o de dimensiones en la solución es grande. Teniendo presente además que las interpretaciones más interesantes surgen de considerar más de una dimensión por vez, el soporte gráfico resulta un elemento clave. Por ello, los programas informáticos incluyen gráficos como parte de la salida del análisis, de los cuales el más popular es el diagrama de dispersión de las puntuaciones y las cuantificaciones en las dimensiones tomadas de a par (denominado ‘plano factorial’ en la tradición francesa). ViSta proporciona gráficos de ese tipo y también dos extensiones del mismo al caso multivariado. Esos gráficos son: el diagrama de dispersión en tres dimensiones (Spin plot) y la matriz de diagramas de dispersión (Scatter matrix plot).

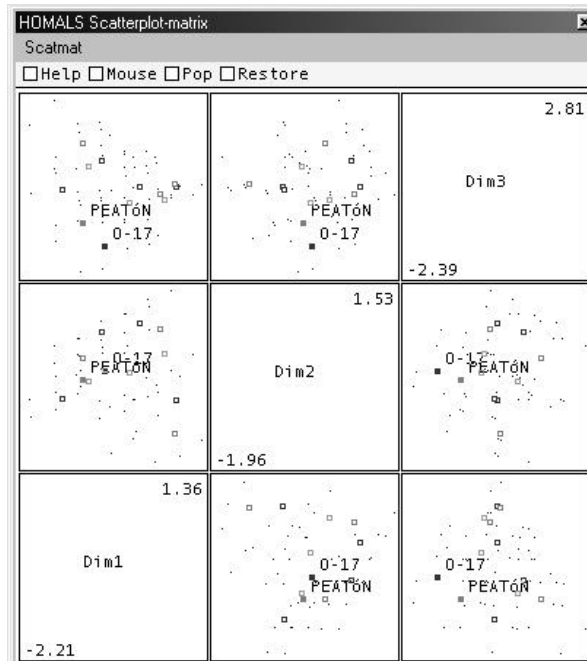
El ‘Spin plot’ es un gráfico de dispersión rotatable que permite representar objetos en tres dimensiones (Figura 3). En esta versión del método, el usuario puede controlar el sentido, la dirección y la velocidad del giro. Otras posibilidades interesantes son: control del foco (zoom) y selección de nuevas variables para definir los ejes. El Spin Plot se utiliza para representar las categorías cuantificadas y/o los puntajes de los objetos. Cada eje del gráfico es una dimensión en la solución del análisis, la posición de los puntos-objetos y de los puntos-categorías viene determinada por sus respectivos valores (puntajes y cuantificaciones) en cada dimensión. Un plano factorial clásico puede obtenerse poniendo el gráfico en la posición HOME, de este modo se ofrece un diagrama de dispersión en dos dimensiones.

**Figura 3 - Imagen de ‘Spin Plot’ que representa las categorías cuantificadas en las tres primeras dimensiones**



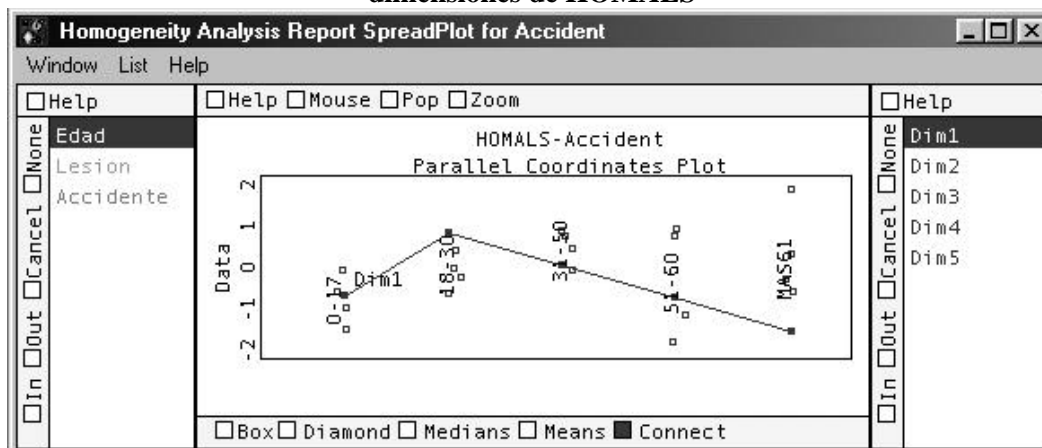
La segunda extensión del diagrama de dispersión es la matriz de diagramas de dispersión, que extiende dicho método al caso multivariado (Figura 4). En este gráfico cada celda muestra los puntajes de los objetos y/o las cuantificaciones de las categorías en el espacio de todas las dimensiones tomadas de a par. En el ejemplo, se han tomado las tres primeras dimensiones, como resultado de lo cual tenemos una matriz simétrica de tres diagramas de dispersión a cada lado de la diagonal. La diagonal lista las etiquetas de las dimensiones (Dim1, Dim2, etc.) y el valor mínimo y máximo de cada una. En esta versión dinámica del método, todas las celdas están vinculadas por los objetos y/o las categorías, eso significa que una acción del usuario sobre una celda de la matriz propaga la acción al resto de las casillas. De este modo, un objeto o categoría (o grupo de objetos/categorías) puede ser explorado/a teniendo en cuenta todas las combinaciones bivariadas al mismo tiempo. En el ejemplo, se han seleccionado sobre una de las celdas del gráfico las categorías ‘0-17 años’ (variable: ‘Edad’) y ‘Atropello a peatón’ (variable: ‘Tipo de accidente’), las que aparecen ‘iluminadas’ en todos los diagramas de dispersión.

**Figura 4 - Imagen de una Matriz de diagramas de dispersión dinámica que representa los objetos y las categorías en las tres primeras dimensiones.**



Además de los gráficos anteriores, puede obtenerse una representación complementaria para visualizar de forma desagregada la transformación de las categorías en las distintas dimensiones de la solución (Figura 5). Este procedimiento es equivalente al gráfico de transformación univariada de la versión en Lisp, solo que es más completo y dinámico. La diferencia esencial es que pueden verse todas las transformaciones de una variable en las diferentes dimensiones al mismo tiempo en el mismo gráfico, y pasar rápidamente de una variable a otra sin necesidad de cambiar de ventana. De otra manera, para este caso se dispondría de una salida de 15 gráficos diferentes (3 variables por 5 dimensiones). Este método de visualización está formado por tres ventanas: una lista con los nombres de las variables del análisis; una lista con las dimensiones de la solución y un gráfico de puntos paralelos para visualizar la transformación de cada variable. El método funciona de la siguiente manera: al seleccionar una variable en la lista, el gráfico de puntos presenta verticalmente la cuantificación de cada categoría en las distintas dimensiones. Cuando seleccionamos una o varias dimensión en la Lista, podemos ver la transformación de la variable completa en tales dimensiones.

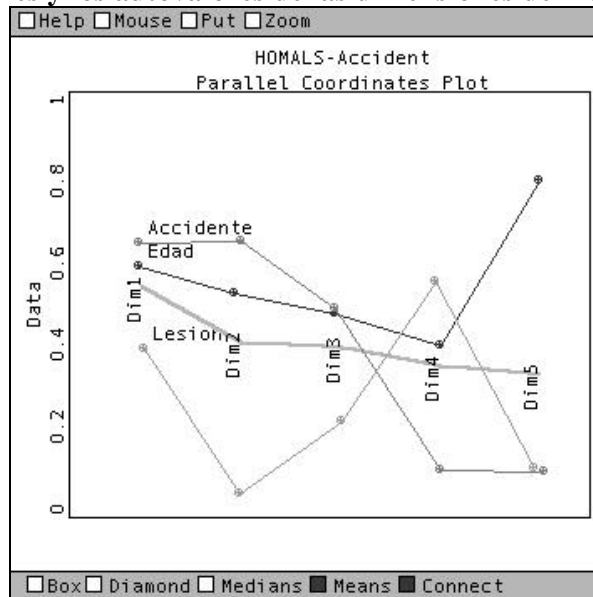
**Figura 5 - Imagen de un método dinámico para visualizar la transformación de las variables en las dimensiones de HOMALS**



Por último, las medidas de discriminación pueden ser visualizadas mediante un gráfico propio. Se trata de un diagrama de líneas paralelas donde las variables del análisis se presentan como puntos

dispuestos verticalmente, ubicados según su correspondiente valor de discriminación en cada dimensión (columnas). El eje vertical está ajustado al rango 0-1, valores posibles de las medidas de discriminación. Los puntos-variables están ligados, lo que permite explorar el perfil de una o varias variables en el espacio de todas las dimensiones al mismo tiempo. El gráfico presenta las medias de los puntos conectadas entre sí por medio de una línea verde, esa línea representa los autovalores de cada dimensión. Eso permite evaluar, al mismo tiempo, las medidas de discriminación de las variables y la importancia relativa de cada dimensión en la solución global.

**Figura 6 - Gráfico dinámico de líneas paralelas con las medidas de discriminación de las variables y los autovalores de las dimensiones de HOMALS**



### 2.3. Visualización estructurada del modelo: 'SpreadPlot' de HOMALS

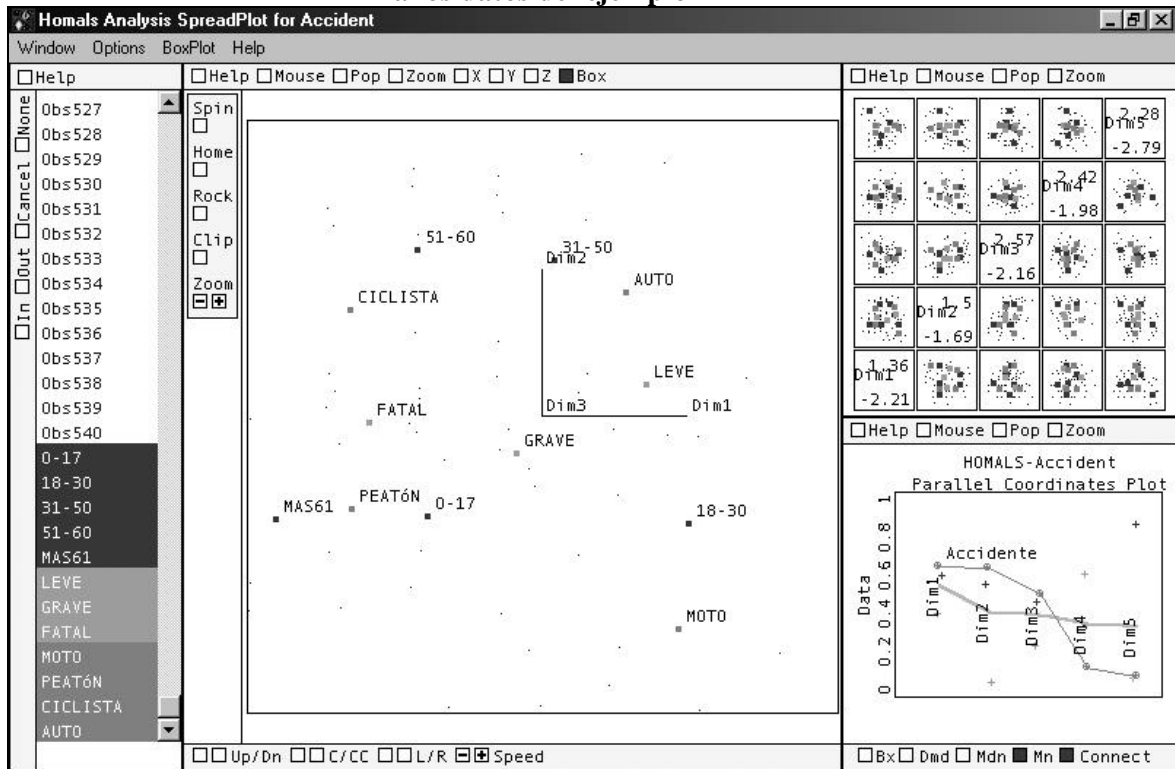
La mayor ventaja de HOMALS en ViSta es que los métodos anteriores se estructuran en una única representación, donde los gráficos, además de sus propias capacidades interactivas, están relacionados empíricamente por los objetos, las categorías y las dimensiones de la solución. Eso se logra utilizando la arquitectura de los SpreadPlots.

El SpreadPlot de HOMALS incluye cuatro gráficos básicos: una *Lista* con los nombres de los objetos y las categorías; un *Spin Plot*; una *Matriz de diagramas de dispersión* y un *Gráfico de las medidas de discriminación* (Figura 7). Asimismo, puede obtenerse desde el mismo SpreadPlot una ventana con el gráfico de transformación discutido en el apartado anterior.

Lo importante de esta estructura es que los distintos gráficos están vinculados empíricamente. Por ejemplo, puede utilizarse la Lista, a la izquierda del Spreadplot (donde se presentan los nombres de los objetos y de las categorías de las variables) como método para identificar o seleccionar observaciones y categorías en el resto de los gráficos. Esto significa que al seleccionar con el ratón un determinado objeto o categoría - o un conjunto de objetos y categorías -, los mismos se iluminarán automáticamente en el resto de los gráficos, pudiéndose identificar con facilidad su posición en el espacio. Para facilitar esa tarea se utilizan diferentes colores para los objetos y las categorías correspondientes a las distintas variables.

Otra posibilidad es utilizar la matriz de diagramas de dispersión como panel dinámico para seleccionar las dimensiones (ejes) del Spin Plot. Puede seleccionarse una o varias dimensiones pinchando con el ratón en la diagonal de la matriz, o seleccionarse un determinado par de dimensiones realizando la misma acción sobre una celda específica (diagramas de dispersión). El Spin Plot se acomodará automáticamente a la selección que el usuario realice sobre la matriz de diagramas de dispersión. De ese modo es posible explorar dinámica y secuencialmente los resultados en las distintas dimensiones.

**Figura 7 - Gráfico extendido (SpreadPlot) para visualizar los resultados de HOMALS aplicado a los datos del ejemplo**



Finalmente, el usuario puede generar sus propias representaciones o utilizar otros métodos de ViSta para analizar los resultados. Esto es posible utilizando el comando 'Crear Datos' de HOMALS para guardar las puntuaciones de los objetos, las cuantificaciones de las categorías y/o las medidas de discriminación como nuevos datos. Estos archivos pueden utilizarse para realizar nuevos análisis y visualizaciones.

### 3. Comentarios finales

Actualmente HOMALS funciona como plug-in integrable a ViSta 6.4 para Windows (95, 98, NT y 2000). El programa no corre bajo Windows 3.x y tampoco está disponible para Macintosh y Unix (aunque existen versiones previas de ViSta que operan en tales plataformas). Por otro lado, si bien existen versiones anteriores del programa en castellano, ViSta 6.4 solo se encuentran disponible en inglés.

El programa ViSta puede obtenerse gratuitamente desde la dirección URL: <http://forrest.psych.unc.edu/research/index.html> o desde el siguiente "mirror" español URL: <http://www.uv.es/~prodat>. Con respecto al plug-in de HOMALS, puede solicitarse directamente a los autores.

### 4. Referencias

- Bond, J. y Michailides, G. (1996). Homogeneity Analysis in LISP-STAT. *Journal of Statistical Software*, 1, 2, 1-31.
- Benzécri, J.P (1973). *Analyse des Données* (2 vols.). París: Dunod.
- Cleveland, W. S. (1993). *Visualizing Data*. Murray Hill, NJ: ATyT Bell Lab.
- Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Chichester: Wiley.
- SPSS Inc. (1990). *SPSS Categories User's Manual*. Chicago: SPSS Inc.
- Tierney, L. (1990). *Lisp-Stat - An Object-Oriented Environment for Statistical Computing and Dynamic Graphics*. NY: John Wiley&Sons.
- Ungaro, J., Ledesma, R. (1999). Accidentes de tránsito: un análisis tipológico mediante la aplicación de métodos estadísticos descriptivos multidimensionales. *CM-Publicación Médica*, 12, 1, 15-19.

- Young, F., Valero Mora, P., Faldowsky, R. A., Bann, C. (2000). SpreadPlots. Chapel Hill, NC: L. L. Thurstone Psychometric Lab, Univ. N. Carolina at Chapel Hill. The Visual Statistics Project, Report Number 2000-4.
- Young, F. (1991-2004). ViSta “The Visual Statistics System” (version 6.4) [software] [On-line], Disponible: URL: <http://forrest.psych.unc.edu/research/index.html>