

**Práctica del análisis de los datos
numéricos y textuales
con Dtm-Vic**

**Ludovic Lebart
Marie Piron**

Versión español por :

INDEC - Instituto Nacional de Estadística y Censos

República Argentina

Indice

Introducción	5
I. Presentación general de Dtm-Vic	8
1. Preparación de los archivos de datos	
2. Técnicas de análisis de datos	
3. Visualización de los resultados	
4. La caja de herramientas	
5. Formato interno de los archivos de datos	
II. Datos numéricos:	
El uso del Dtm-Vic a partir de tres ejemplos	21
1. Análisis en Componentes Principales: "presupuesto-tiempo"	
2. Análisis de Correspondencias: "encuesta-medios"	
3. Análisis de Correspondencias Múltiples: "encuesta aspiraciones"	
III. Datos textuales y mixtos:	
El uso del Dtm-Vic a partir de tres ejemplos	62
1. Análisis Textual de textos: poemas	
2. Análisis Textual de preguntas abiertas: encuesta "Life"	
3. Análisis directo de respuestas libres, con clasificación.	
IV. Importación, creación y exportación de archivos	102
1. Archivos numéricos y textuales a partir de Excel [®]	
2. Captura manual de datos numéricos	
3. Exportación de archivos de datos en formato "Excel" [®]	
V. Recodificación, almacenamiento, herramientas diversas	120
1. Recodificación, almacenamiento	
2. Intervención elemental sobre una base de datos	
3. Herramientas específicas del procesamiento previo	
VI. Otros análisis con Dtm-Vic	136
1. Datos numéricos: Semiometría	
2. Datos numéricos: Contigüidad (Iris de Fisher / Anderson)	
3. Descripción de grafos	
4. Reconstitución de imágenes	
Índice de referencias bibliográficas	180

Dtm-Vic

Data and text mining

Visualization, inference, classification

**Software de análisis exploratorio multidimensional
de datos numéricos y textuales**

Libremente descargable sobre: www.dtm-vic.com

Introducción

Dtm-Vic es un software dedicado **al análisis exploratorio multidimensional de datos numéricos y textuales**.

El análisis exploratorio, como su nombre lo indica, es un paso preliminar para abordar un conjunto de datos que prosigue con investigaciones y una descripción, sin limitarse por esto, a un protocolo predeterminado. La exploración supone que los datos son complejos y que, a priori, los conocimientos sobre los mismos son limitados.

El análisis **multidimensional**, se aplica al caso en que las dimensiones (en la mayoría de los casos: variables) son numerosas lo cual ya es un factor de complejidad y, en consecuencia, un incentivo para comenzar por una vía exploratoria. Otro incentivo, más técnico, para utilizar este enfoque se relaciona con el carácter poco realista de las hipótesis sobre las distribuciones estadísticas en el caso multidimensional, lo que torna difícil la utilización codificada de pruebas de hipótesis.

El análisis exploratorio multidimensional de datos numéricos es un componente importante del software Dtm-Vic. Las herramientas básicas son, por una parte, los métodos factoriales (o análisis en ejes principales), tal como el análisis en componentes principales, los análisis de correspondencias simples y múltiples y, por otro lado, los métodos de clasificación (clasificación jerárquica, los métodos de partición, mapas auto-organizados). Estas técnicas no se excluyen mutuamente sino que, al contrario, se utilizan sistemáticamente como técnicas complementarias que ofrecen, cada una, los puntos de vista indispensables sobre la realidad estadística. El libro básico que acompaña a los métodos que se utilizan en esta parte del software Dtm-Vic se titula: "*Statistique Exploratoire Multidimensionnelle*"¹

En particular, los **datos textuales** son a la vez multidimensionales y complejos. Por consiguiente, son posibles candidatos para el tratamiento propuesto para el análisis exploratorio. Frecuentemente se los asocia con los datos numéricos. Es el caso emblemático de las encuestas por muestreo con preguntas cerradas (datos numéricos continuos y variables nominales) y preguntas abiertas (datos textuales). Estos datos de encuestas constituyen el ejemplo-tipo alrededor del cual se desarrolló el Dtm-Vic. Una parte importante de los métodos utilizados en la parte textual del software Dtm-Vic son presentados y comentados en la obra: "*Statistique textuelle*"².

El análisis exploratorio multidimensional de datos numéricos y textuales aparece como una etapa esencial del tratamiento en estas recolecciones complejas.

Se sabe, y los ejemplos son conocidos, que habitualmente los investigadores descubren algo distinto de lo que buscan. Los usuarios de Dtm -Vic tienen frecuentemente la oportunidad de comprobarlo: los análisis realizados son prueba irrefutable de la coherencia y de la calidad de

¹ *Statistique Exploratoire Multidimensionnelle. Visualisation et Inférence en Fouille de Données*. Ludovic Lebart, Marie Piron, Alain Morineau (2006). 4^a Ed. Dunod, París.

² *Statistique textuelle*. Ludovic Lebart, André Salem (1994), Ed Dunod, París. La versión en español: *Análisis Estadístico de Textos* (L.Lebart, A. Salem, M. Becue-Bertaut, 2000, Ed. Milenio, Lleida). La versión en inglés: *Exploring Textual Data* (L.Lebart, A. Salem, E. Berry, 1998, Ed. Kluwer-Springer, Dordrecht) incluye ejemplos utilizados en este manual.

la información básica, resultado que no siempre es apreciado por aquellos que recolectan esa información, ni por aquellos que se apresuran a utilizarla muy rápidamente.

Pero para los usuarios experimentados, especialmente en el campo de las ciencias sociales, estas pruebas globales de coherencia no son consecuencias accidentales de las exploraciones, sino uno de sus objetivos fundamentales, de manera explícita inserta en un enfoque crítico que ve a la recolección como una construcción de información.

En la versión 5 del Dtm-Vic, a la que este manual de uso se refiere principalmente, la interfaz del software está en inglés (palabras clave, temas de ayuda, nombres de los análisis).

Sin embargo, a los investigadores y a los encargados de estudios de habla española no les será difícil manejar el software en estas condiciones. Es difícil para un equipo pequeño, y para un software de acceso libre, no subvencionado, el mantener varias versiones en distintos idiomas.

Las limitaciones actuales (revisables) del software con relación al tamaño de los datos de entrada son: 30.000 filas (las filas son individuos u observaciones); 1.200 columnas (variables numéricas continuas, variables numéricas de codificación de variables nominales - una variable nominal = una columna); 100.000 caracteres para "las respuestas textuales" de un individuo /observación, pero no hay límites para un texto que no esté asociado a un archivo numérico. Este formato corresponde a la gran mayoría de las aplicaciones en las encuestas socioeconómicas, a los archivos provenientes de las encuestas de gestión o de satisfacción, a los datos ecológicos, y a los análisis sensoriales, etc.

Después de una breve presentación del software (Capítulo I) en este manual hemos elegido mostrar seis ejemplos de procesamientos de datos ya preparados, es decir, presentados en un formato adecuado. Estos ejemplos deben ser descargados con el software (Capítulo II, III y VI) y se relacionan con usos frecuentes de Dtm-Vic. El usuario aprenderá a crear por sí mismo un archivo de comando desde la interfaz propuesta. Se puede encontrar sucesivamente un análisis de componentes principales (encadenados con una clasificación y para las clases, un posicionamiento mediante análisis factorial y una descripción automática) un análisis de correspondencias, análisis de correspondencias múltiples (completado también con una clasificación), un análisis factorial lexical de una serie de texto, y luego, en el marco de una encuesta, un análisis de correspondencias de una tabla léxica construida a partir de una pregunta abierta y de una pregunta cerrada y, finalmente, un análisis y una clasificación directa de respuestas a una pregunta abierta. Estas aplicaciones dan lugar a visualizaciones validadas por la técnica de *bootstrap*.

Con la esperanza de haber motivado al lector con esta primera presentación de las características del software, se abordan en el capítulo IV los procedimientos para la importación de datos. Concedamos que puede ser complicado, a veces, tratar unidades estadísticas tan diferentes como un número, una categoría, una respuesta concisa a una pregunta abierta, o una novela de Borges. Sin embargo, la total transparencia de los archivos de entrada o producidos por Dtm-Vic (todos los archivos están en formato de texto no propietario) debería tranquilizar a los usuarios y reducir la complejidad del proceso.

Al llegar al capítulo IV, la lectora o el lector dispondrán de una cierta autonomía. Para ayudar a perfeccionar y profundizar los análisis anteriores, algunos de los procedimientos básicos para el almacenamiento o la recodificación están contenidos en el Capítulo V.

Finalmente, en el sexto y último capítulo se presentan aplicaciones más detalladas, incluyendo la implementación de nuevas opciones de procedimientos de visualización. Este capítulo VI también contiene los análisis de contigüidad, las descripciones de grafos, y en él se ilustran las capacidades de compresión de las técnicas factoriales.

Todas estas fases del aprendizaje suponen que el software y el conjunto de ejemplos han sido copiados o descargados. Esto es posible desde el sitio³: <http://www.Dtm-Vic.com>

³ También podremos descargar sobre este sitio la obra precitada "*Statistique textuelle*" (L. Lebart y A. Salem) y la obra "*La sémiométrie, Essai de Statistique structurale*" (L. Lebart, M. Piron, J.-F. Steiner. 2003, Ed.Dunod, París), de donde son extraídos ciertos conjuntos de datos utilizados aquí. Otras obras citadas no son libres de derecho a esta fecha y deben ser consultadas en biblioteca o adquiridas en la red de librerías.

I. Presentación general de Dtm-Vic



Para ejecutar el Dtm-Vic, basta con hacer Clic en el icono del acceso directo en el escritorio de *Windows* o a través del programa ya instalado. Obtenemos la pantalla de inicio siguiente:



Dtm-Vic está estructurado en dos etapas:

I - La primera etapa **Dtm – Data and Text mining** incluye los procedimientos para la preparación de los datos (importación, captura, exportación) y los procedimientos de análisis de datos (creación y ejecución del archivo de comandos).

II - La segunda etapa **Vic – Visualization, Inference, Classification** proporciona las herramientas de validación e interpretación de los resultados.

También se puede ver en la pantalla de inicio dos elementos opcionales: la "caja de herramientas", **DtmVic Tools** que ofrece diferentes tipos de recodificación, almacenaje de datos, y la sección **DtmVic Images** dedicada a algunos análisis de imágenes.

Este manual debe permitir proceder a una implementación de estas etapas de cálculo y de visualización. Algunas de ellas, las más específicas del software, se mencionan a continuación y serán detalladas en diferentes partes del manual, sabiendo que todos los análisis se realizarán dentro de una secuencia de etapas:

1. Selección de un análisis
2. Apertura de los diferentes archivos de datos en el formato Dtm-Vic
 - Selección de variables
 - Selección de los diferentes parámetros específicos para el análisis.
3. Creación de un archivo de comandos
4. Ejecución del archivo de comandos
5. Visualización de los resultados.

Para obtener ayudas sobre los parámetros o los archivos, haga Clic en los menús **Help** en la barra superior, que se muestran en rojo. Para suprimir la presentación de un tema de ayuda, haga Clic, una vez más en el botón apropiado. El instructivo (en inglés) está disponible en esta barra.

I.1 Preparación de los archivos de datos:

- Haga Clic en **Data Importation, Preprocessing, Data Capture, Exportation** en **Data File**
 - Aparece una ventana que sugiere diferentes procedimientos. Aquí se visualizan los componentes de esta ventana:

<p>Importation of variables, observations and texts</p> <p>Importing Dictionary, Data, and Texts</p> <p>Specific preprocessing and importation tools</p> <p>- Importación de archivos de datos numéricos o archivos textuales y constitución de archivos diccionario, datos y textos en el formato Dtm-Vic. Ver el Capítulo IV</p> <p>- Algunas de las herramientas de procesamiento previo</p>
<p>Building the dictionary of variables and creating the data file</p> <p>Building the dictionary (manually)</p> <p>Creating the data file (manually)</p> <p>Módulos de captura de datos: construcción del diccionario de variables y creación del archivo de datos. Ver el capítulo IV</p>
<p>Exporting a DTM file to R or to Excel(r)</p> <p>Exporting dtm data (and dictionary) to R or Excel (r)</p> <p>Exporting dtm data, dictionary, and texts into a unique XML file</p> <p>Exportación de archivos de datos en Excel, R o XML. Ver el Capítulo IV</p>
<p>Dtm_tools: Amending or updating data and dictionary</p> <p>Dtm_tools</p> <p>Creación de nuevas variables, selección de una submuestra o la concatenación de varios archivos. Ver el acceso directo a la caja de herramientas <i>DtmVic Tools</i> y el capítulo V</p>

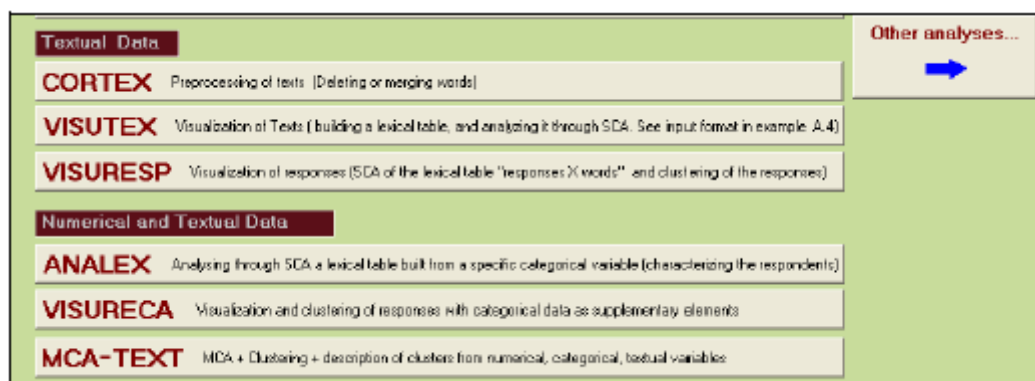
I.2 Técnicas de Análisis de Datos

- Haga Clic en **Create a command file** dentro de la sección **Command File** de **Dtm – Data and Text mining**
 - Aparece una ventana que muestra las diferentes posibles técnicas de análisis, en función de la naturaleza numérica o textual de los datos.

La parte superior de esta ventana se refiere a datos numéricos:



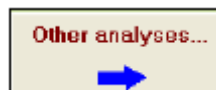
La parte inferior de la misma ventana se refiere a datos textuales:



Explicaciones resumidas de los tratamientos:

<p>Numerical Data (basics)</p> <p>BAS Basic Statistics about numerical and categorical data</p> <p>TAB CrossTabulating a series of categorical variables</p> <p>DECAT Automatic description of a series of categorical variables</p> <p>IPFIT Re-Weighting the observations/individuals of a series of categorical variables</p>	<p><i>Análisis descriptivo univariado, BAS ; Demanda de cuadros cruzados TAB de las variables continuas o nominales;</i></p> <p><i>Descripción automática de una variable por una serie de variables nominales DECAT. Reponderación de la muestra, IPFIT (Iterative Proportional Fitting).</i></p>
<p>Numerical Data (principal axes techniques)</p> <p>PCA Principal Components Analysis (complete)</p> <p>SCA Simple Correspondence Analysis (to be used with numerical variables)</p> <p>MCA Multiple Correspondence Analysis (complete)</p>	<p><i>Análisis estadístico exploratorio de datos numéricos: Encadenamiento de un análisis factorial (Análisis en Componentes Principales PCA, Análisis de Correspondencias Simples SCA, Análisis de Correspondencias Múltiples MCA) y de una clasificación (k-means y clasificación jerárquica ascendente).</i></p> <p>Ver capítulo II</p>
<p>Textual Data</p> <p>CORTEX Preprocessing of texts (Deleting stopwords, stemming, lemmatization)</p> <p>VISUTEX Visualization of Texts (building a word cloud, a network graph)</p> <p>VISURESP Visualization of responses (SCA)</p>	<p><i>Análisis estadístico exploratorio de un corpus de textos: CORTEX suprime o reagrupa palabras (lematización resumen empírico); VISUTEX realiza un análisis de correspondencias simples de una tabla léxica (ver capítulo III); VISURESP realiza un análisis directo de respuestas abiertas.</i></p>
<p>Numerical and Textual Data</p> <p>ANALEX Analysing through SCA a lexical table</p> <p>VISURECA Visualization and clustering of responses</p> <p>MCA-TEXT MCA + Clustering + descriptive statistics</p>	<p><i>Análisis estadístico exploratorio de respuestas abiertas (ver capítulo III): ANALEX realiza un análisis de correspondencias simples de una tabla léxica agregada; VISURECA realiza un análisis análogo a VISURESP, pero lo ilustra con variables nominales; MCA-TEXT; realiza un análisis de correspondencias múltiples (variables nominales) y una clasificación ilustrada por las variables léxicas.</i></p>

Otras técnicas de análisis textual son propuestas en el menú



- Si se hace Clic en este botón, una nueva ventana aparece.

Los análisis **CORDA** y **SEGME** suministran las concordancias y los segmentos repetidos, mientras que los análisis siguientes incluyen directamente la fase **CORTEX** (correcciones de textos) dentro de los análisis **VISUTEX**, **VISURESP**, **VISURECA**, **ANALEX**.

Textual Data	
CORDA Concordances of a series of words	CORDA proporciona las concordancias de una lista de palabras.
SEGME Lists of repeated segments	SEGME Da las listas de segmentos repetidos.
VISUTEX-CORTEXT Visual text analysis	VISUTEX-CORTEXT Realiza el análisis VISUTEX precedente, después de la corrección de textos similar a CORTEXT.
VISURESP-CORTEXT Visual text analysis	VISURESP-CORTEXT Realiza el análisis VISURESP después de CORTEXT.
Numerical and Textual Data	
ANALEX-CORTEXT Analyses	ANALEX-CORTEXT Realiza simultáneamente los procedimientos CORTEXT y ANALEX
VISURECA-CORTEXT Visual text analysis	VISURECA-CORTEXT Realiza simultáneamente los procedimientos CORTEXT y VISURECA

Podríamos realizar, al principio, la fase CORTEXT, luego los análisis precitados. Pero CORTEXT se refiere al conjunto del archivo de texto, mientras que se puede desear corregir individualmente cada pregunta abierta. Además, las respuestas más frecuentes, respuestas características de cada texto, serán las respuestas originales, y no las respuestas con palabras corregidas. Pero la selección estadística de las respuestas características se hace sobre los textos corregidos.


Una vez creado el archivo de comando durante el procedimiento **Create**, es posible, siempre en la sección: **Command File**, abrir directamente este archivo (botón: **Open an existing command file**) para modificar directamente algunos parámetros, luego de ejecutarlo (botón: **Execute**)

Los procedimientos de análisis exploratorios de datos numéricos o textuales implican la secuencia de varias técnicas, Análisis factorial, Clasificación, Mapas de Kohonen, Validación Bootstrap. Los resultados de los Análisis de base pueden ser sea consultado en la sección: **Result Files** (**Basic numerical results**) navegando sobre un archivo Html o en formato texto (**text format**) o visualizados a través de las diferentes herramientas ofrecidas en la sección: **VIC - Visualization, Inference, Classification**


I.3 Visualización de resultados

En la etapa, **VIC - Visualization, Inference, Classification**, una serie de herramientas de visualización permiten la validación de los resultados y facilitan su interpretación (Cf. capítulos II y III).

Para utilizar una de estas herramientas, haga Clic en el menú correspondiente:

 **AxesView** : ejes factoriales.

Clasificaciones, para cada eje, de las coordenadas de los individuos, de las variables activas, las variables ilustrativas (o suplementarias), etc. para una rápida evaluación de los resultados del análisis factorial.

 **PlaneView** : planos factoriales.

Descripción de los planos factoriales para todos tipos de elementos que intervienen en los análisis


 **Bootstrap** : Bootstrap (BootstrapView).


Zonas de confianza (elipses o envolventes convexas (*Convex Hull*) en los planos factoriales para los elementos seleccionados.

 **Seriation** : reordenar filas y columnas

Las filas y las columnas de la tabla de contingencia se reordenan de acuerdo con el primer eje del análisis de correspondencia de la tabla.

Las técnicas de seriación se basan en las permutaciones simples de filas y columnas de la tabla estudiada. Tienen la ventaja práctica y cognitiva de mostrar los datos brutos al usuario y en consecuencia, de evitarle la utilización de reglas complejas de lectura. Estas permutaciones pueden mostrar bloques homogéneos de valores altos o, por el contrario, valores pequeños o nulos. También pueden igualmente señalar una evolución continua y progresiva de los perfiles. Una propiedad óptima del análisis de correspondencia es la siguiente: el primer eje de un análisis de correspondencia proporciona un orden óptimo de los puntos-fila Y de los puntos-columna.

 **ClusterView** : proyección de las clases de la clasificación sobre los planos factoriales. Representación de las posiciones de los centros de clase en el plano factorial. Descripción de los elementos característicos de la clase correspondiente (variables numéricas, categorías, y también palabras o respuestas en el caso de preguntas abiertas).

 **Kohonen Map** : mapas de Kohonen.

Mapas autoorganizados de individuos, de variables y simultáneos de individuos y de variables a partir de las coordenadas factoriales (mapas cuadrados de tamaño 3 x 3 a 20 x 20).

 **Visualization** : herramientas complementarias de visualización.

Visualización complementaria de los planos factoriales y de la clasificación. Elipse de densidad o envolventes convexas de clases. Trazado del árbol de longitud mínimo, vecinos más cercano en los planos factoriales. Visualización pedagógica de la construcción progresiva de clases (caso del procedimiento k-means / nubes dinámicas). Visualización en los planos factoriales de las grillas de Kohonen y de ciertos grafos



Contiguity : el análisis de la contigüidad.

El análisis local, la estructura de grafo.

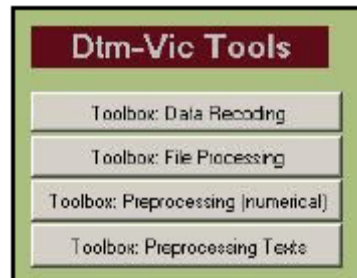
El análisis de contigüidad se refiere a las técnicas de análisis por proximidad (locales) que se presentan en el capítulo 8 del libro antes mencionado "*Statistique exploratoire multidimensionnelle*". Se considera el caso de que las observaciones tienen una estructura de grafo a priori, así como también cuando el grafo es intrínseco (grafo de los vecinos más cercanos, por ejemplo). Se generaliza el análisis discriminante de Fisher (que corresponde al caso particular del grafo asociado con una partición).

El análisis de contigüidad se trata en este manual de uso en la Sección VI.2 Capítulo VI.

Las explicaciones sobre la metodología se pueden obtener con los botones de la barra vertical del menú principal.

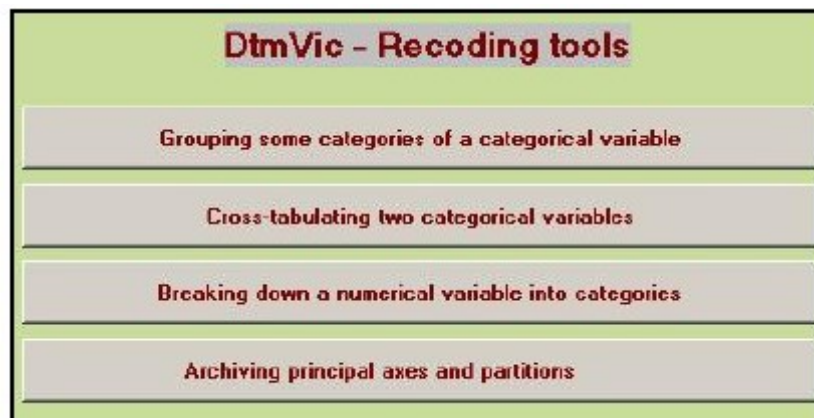
I.4. La caja de herramientas

La caja de herramientas **DtmVic Tools** ofrece diferentes tipos de recodificación, de almacenamiento y de transformación de los datos (ver el capítulo V).



- Haga Clic en **Toolbox Data Recoding**

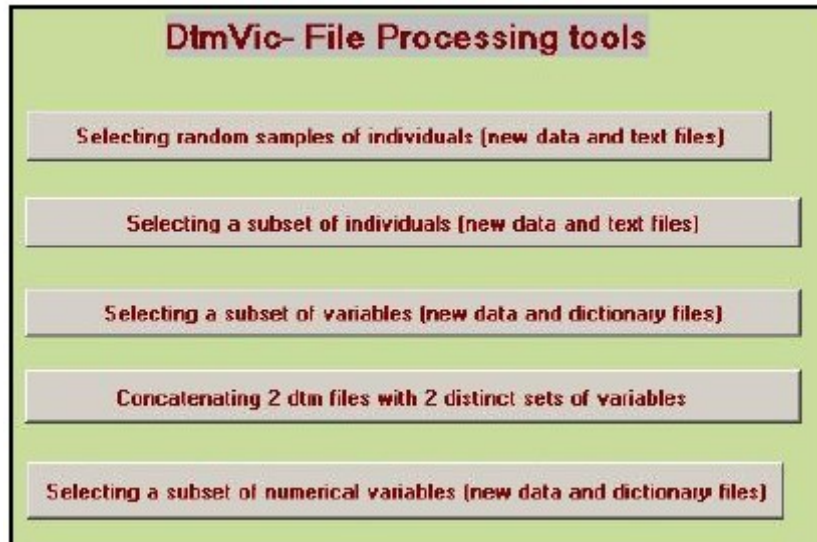
El primer menú que aparece se refiere a la recodificación de los datos y del almacenamiento de ciertos resultados.



Creación o recodificación de variables nominales:

- i) Reagrupación de modalidades;
- ii) Creación de variables nominales por cruce de variables nominales;
- iii) Transformación de una variable continua en variable nominal;
- iv) Almacenamiento de los ejes factoriales y de las particiones.

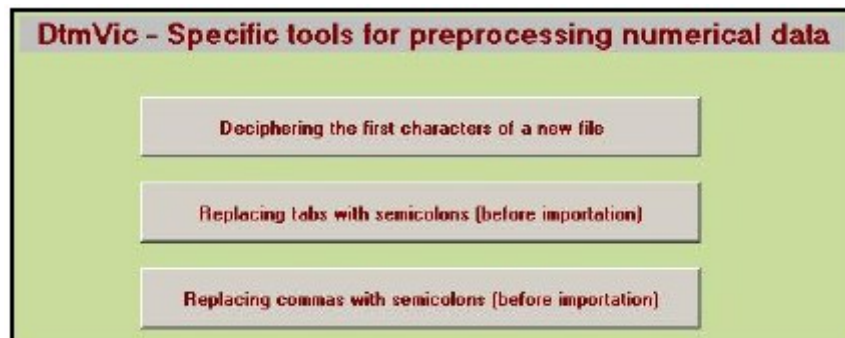
El segundo grupo de acciones concierne el menú:



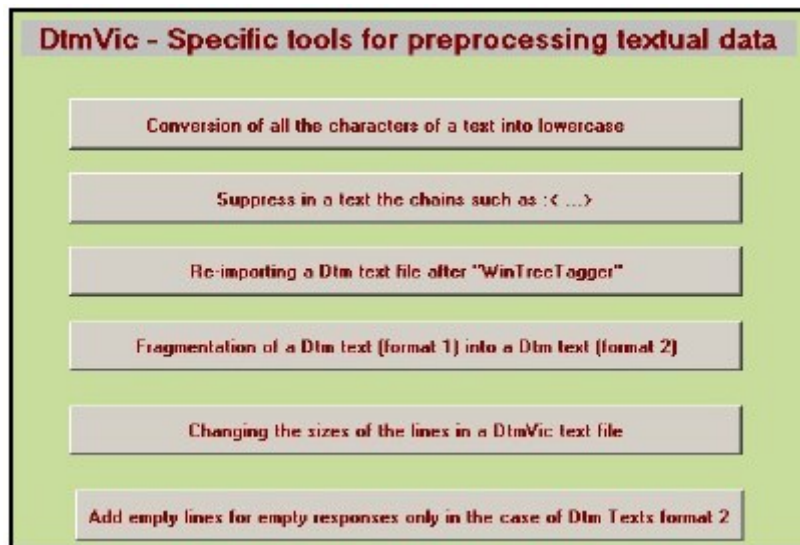
Propone modificación de la base de datos por: (ver Capítulo V)

- i) Selección de un subconjunto aleatorio de individuos (filas);
- ii) Selección de un subconjunto de individuos (filas) a partir de un filtro;
- ii) Selección de un subconjunto de variables (columnas);
- iii) Concatenación de dos bases de datos (variables diferentes);
- iv) Selección de un subconjunto de variables que tiene un peso máximo.

El menú siguiente propone ciertas herramientas básicas para conectarse con los datos y pretratamientos previos para la importación o la utilización de datos numéricos y textuales.



El último menú propone ciertos procedimientos previos para la importación o la utilización directa de los textos.



i) Conversión de los textos en minúscula

ii) Supresión de las marcas "<" y ">" y del texto que pueden contener.

iii) Reimportación en DtmVic de un archivo de texto DtmVic (tipo 1 o 2) que ha estado sujeto al software (gratuito) **TreeTagger**. Esto permite de lematizar un texto suprimiendo ciertas categorías gramaticales (preposiciones, artículos,...).

Disponible para los textos ingleses, franceses, españoles e italianos.

iv) Fragmentación de una serie de textos en formato 1 (textos separados por ***) en textos de formato 2, formados por una fila, dos filas... de los textos iniciales (se trata aproximadamente de una fragmentación en unidades de contexto). Una variable nominal es creada para conservar la información que relaciona las unidades con los textos iniciales.

v) Cambio de longitud de las filas de un archivo de texto. Al principio, formato DtmVic (1 o 2) sin limitación para la longitud de las filas. Al final: textos que tienen filas de una longitud elegida por el usuario (pero <200 caracteres). Este procedimiento permite importar textos con filas muy largas, pero también formatear las unidades de contexto (ver el punto iv más arriba).

vi) Este último procedimiento restringido y especializado permite hacer respetar la limitación "una fila vacía por respuesta abierta vacía" para archivos que utilizarían dos separadores consecutivos.

La sección **DtmVic Images**, es esencialmente pedagógica y muestra las posibilidades de compresión de imágenes ofrecidas por el análisis de correspondencias o simplemente por la descomposición a los valores simples (sección VI.4 del capítulo VI).

I.5. Formato interno de archivos de datos Dtm-Vic

[Versión inglesa de esta sección obtenida por el botón **Data Format** del menú principal].

En esta etapa, es útil conocer el formato interno de los archivos de entrada Dtm-Vic. Estos formatos serán generados por los procedimientos de importación. Tres archivos, en formato de texto, constituyen el formato de Dtm-Vic:

Nota: Los nombres de los archivos son libres, pero la extensión .txt es conveniente para una consulta rápida del contenido de los archivos⁴.

- **Exemple_dic.txt:** el archivo diccionario proporciona los nombres de las variables numéricas y nominales. Incluye las etiquetas de categoría de cada variable nominal (ver cuadro 1).

Nota: los identificadores de las variables y las etiquetas de categoría no deben contener espacios vacíos (blancos). Son a veces truncados por otro lado a 8 caracteres en las representaciones visuales.

- **Exemple_dat.txt:** el archivo de datos contiene los valores de estas variables para un conjunto de individuos (observaciones), así como también a los identificadores de individuos (cuadro 2).

- **Exemple_tex.txt:** Se consideran dos tipos de archivos de texto. Un formato de archivo de los textos simples (tipo 1) puede ser empleado cuando se trata una serie de textos (ver cuadro 3), sin archivo diccionario ni archivo de datos asociados. Cuando los textos son numerosos y calificados, en el caso de respuestas a preguntas abiertas, introducimos dos niveles de separadores (archivo tipo 2, ver cuadro 4).

Un caso de aplicación que muestra todas las posibilidades del software es una recopilación de datos de encuesta por muestreo, que contiene respuestas a las preguntas cerradas y respuestas a las preguntas abiertas. Las preguntas cerradas pueden dar lugar a variables continuas (o aún cuantitativas) o a variables nominales (o cualitativas).

El cuadro 1 muestra un ejemplo de archivo diccionario con formato Dtm-Vic que presente cuatro variables (tres nominales y una continua).

```
2 GENDER      (numero de categorías [2] en col. 1-4; blanco; titulo)
MALE MALE    (identif. cortos [col. 1-4]; blanco; identificador
FEMA FEMALE  (identif. cortos [col. 1-4]; blanco; identificador
0 AGE        (numero de categ. [0] en col. 1-4; blanco; var numer.)
4 AGE_CODE   (numero de categorías [2] en col. 1-4; blanco; titulo)
AGE1 18_24   (identif. cortos [col. 1-4]; blanco; identif. [< 20 car.]
AGE2 25_39   (identif. cortos [col. 1-4]; blanco; identif. [< 20 car.]
AGE3 40_59   (identif. cortos [col. 1-4]; blanco; identif. [< 20 car.]
AGE4 >60     (identif. cortos [col. 1-4]; blanco; identif. [< 20 car.]
3 EDUCATION  (num de categorías [3] en col. 1-4; blanco; titulo)
EDUL LOW     (identif. cortos [col. 1-4]; blanco; identificador
EDUM MEDIUM (identif. cortos [col. 1-4]; blanco; identificador
EDUH HIGH    (identif. cortos [col. 1-4]; blanco; identificador

[Los identificadores tienen menos de 20 caracteres. Nunca blanco al interior de un
identificador]
```

Cuadro 1: Archivo diccionario en formato interno a Dtm-Vic para 4 variables

Sexo (2 categorías), edad (categoría 0 = variable continua), grupo de edades
(4 categorías), nivel educativo (3 categorías).

(Los comentarios en *itálica* dan las explicaciones del formato fijo del archivo diccionario)

⁴ Estos archivos, en formato de texto (extensión ".txt"), pueden ser leídos por el "bloc-notes" o un editor de texto (TotalEdit, notepad, notepad ++, UltraEdit, etc.), o el editor de texto Dtm-Vic mediante el botón "Open" del menú principal.

El cuadro 2 muestra el ejemplo de un archivo de datos de Dtm-Vic correspondiente a las 4 variables del precedente archivo diccionario para 5 individuos (sujetos, observaciones o respondientes).

'n1006'	1	76	4	1	(Identificador de la observación : entre
'n1007'	2	20	1	2	comillas, sin blanco, < 20 caracteres.
'n1008'	2	29	2	3	Separadores entre valores: al menos un
'n950'	1	57	3	1	espacio blanco)
'n2007'	1	21	1	2	

Cuadro 2: Archivo de datos en formato interno Dtm-Vic.

Para 5 individuos (sujetos u observaciones) correspondiente a las 4 variables del diccionario precedente: sexo, edad, edad separada en 4 clases, nivel educativo (ver. cuadro 1). Longitud máxima de una fila: 5.000 caracteres. *(Comentario del formato en itálico)*

El cuadro 3 muestra el ejemplo de un archivo de texto en formato interna de Dtm-Vic para una serie de tres textos (véase el ejemplo III.1 - poemas).

```

****      LAMARTINE
Voilà les feuilles sans sève,
Qui tombent sur le gazon
Voilà le vent qui s'élève,
Et gémit dans le vallon
Voilà l'errante hirondelle,
Qui rase du bout de l'aile,
L'eau dormante des marais...
****      GAUTIER
L'automne va finir, au milieu du ciel terne,
Dans un cercle blafard et livide que cerne
Un nuage plombe, le soleil dort. Du fond
Des étangs remplis d'eau monte un brouillard qui Fond
Collines, champs, hameaux dans une même teinte.
.
****      VERLAINE
Les sanglots longs
Des violons
De l'automne
Blessent mon coeur
D'une langueur
Monotone.
=====

```

Cuadro 3: Archivo de texto en formato interno (tipo 1) Dtm-Vic

Los tres textos son de formato libre de menos de 200 columnas; los separadores de texto están separados por "****", seguido de 4 espacios y, a continuación el identificador del texto con menos de 20 caracteres; El fin del archivo está indicado por "====". Todos los separadores ocupan las 4 primeras columnas. Para algunas ediciones de cuadros, es útil que los 4 primeros caracteres del identificador de texto caractericen el texto.

Si las filas tienen más de 200 caracteres, un procedimiento de Dtm-Vic-Tools permite reformatearlas.

El cuadro 4 (a continuación) presenta un archivo de textos que se refiere a tres preguntas abiertas para tres respondientes (ver el ejemplo III.2).

¿Por qué dos formatos distintos para los datos textuales? A diferencia de los datos numéricos, los textos pueden plantear problemas de escala, de dimensiones, y por lo tanto de límites.

- El formato de tipo 1 (separadores ****) permite guardar textos muy largos, por ejemplo las novelas de la Comedia humana de Balzac. Cada texto puede ser largo, pero el número de textos está limitado aquí a 1.200.

- El formato de tipo 2 (Separadores---- [para las observaciones] luego ++++ [para las preguntas abiertas, cuyo número es limitado aquí a 12]) corresponde al archivo de encuesta (el número

de textos debe entonces ser inferior a 30.000, limita del número de observaciones de Dtm-Vic en la versión presente). El texto total de un individuo está limitado a 100.000 caracteres.

```

---- 1006
my sons, my kids are very important to me,
being on my own I am responsible for their education
++++
education and moral standard of the youngsters, law and
order
++++
basically, British culture is traditional,
people tend to keep themselves to themselves
---- 1007
job, being a teacher I love my job, for the well being
of the children
++++
law and order, drug abuse, child abuse
++++
accommodating, of course people from different races
and culture have settled in here, (i.e., Irish, Jewish,
Asians) and the British culture is working alright
---- 1008
job, sometimes it is very hard to find a job
++++

++++
=====

```

Cuadro 4: Archivo de texto de preguntas abiertas en formato interno a Dtm-Vic (tipo 2)

Tres individuos respondieron a tres preguntas abiertas. El formato es libre, dentro de las 200 columnas.

El separador entre los individuos es " ---- " seguido por el identificador del individuo (menos de 20 caracteres); las preguntas están separadas por "++++"; el fin del archivo se señala por "=====".

Todos los separadores ocupan las 4 primeras columnas. Nota: las filas vacías corresponden a las no respuestas (el último respondiente no dio respuesta a las dos últimas preguntas abiertas: por lo menos una fila vacía entera (virgen) es necesaria en este caso).

Atención: el orden de los individuos debe ser aquel del archivo de datos numéricos.

Notemos que para la importación de un archivo Excel que contiene a la vez variables numéricas y textuales, cada respuesta a una pregunta abierta está limitada a 8.000 caracteres. En los ejemplos proporcionados en Dtm-Vic, los archivos deben ser descargados a partir de los documentos preexistentes. La puesta a punto en el formato de Dtm-Vic es entonces transparente para el usuario.

Recordatorio: Dtm-Vic produce numerosos archivos de resultados intermedios vinculados a la aplicación (todos en formato .txt). Por consiguiente, es recomendado crear un repertorio por aplicación. Al principio, tal repertorio debe contener los archivos de datos, el diccionario o el texto al formato Dtm-Vic.

II. Datos numéricos:

El uso del Dtm-Vic a partir de tres ejemplos

Los ejemplos siguientes tienen por objeto presentarle el Dtm-Vic al usuario de modo pragmático. Corresponden a una carpeta incluido en una carpeta **DtmVic-Examples_A_Start** que ha sido incorporado con el software Dtm-Vic. Cada ejemplo considera un juego de datos adaptado a uno de los análisis factoriales de base (Análisis en Componentes Principales, Análisis simple de las Correspondencias, Analiza de Correspondencias Múltiples) enriquecido por herramientas complementarias (bootstrap, clasificación, mapas de Kohonen, seriación):

1. El ejemplo 1, contenido en la carpeta **EX_A01.PrinCompAnalysis**, es un análisis en componentes principales aplicado sobre un conjunto de variables continuas: variables activas y ilustrativas (o suplementarias); validación Bootstrap; clasificación de los individuos y descripción de las clases.
2. El Ejemplo 2, contenido en la carpeta **EX_A02. SimpleCorAnalysis**, presenta un análisis de las correspondencias simples adaptado al análisis de una tabla de contingencia: variables activas y ilustrativas (o suplementarias); validación Bootstrap.
3. El Ejemplo 3, contenido en la carpeta **EX_A03. MultCorAnalysis**, se refiere al análisis de las correspondencias múltiples aplicado sobre un conjunto de variables nominales surgidas de datos de encuestas: variables nominales activas, ilustrativas (o suplementarias), variables continuas; validación Bootstrap; clasificación de los individuos y descripción de las clases obtenidas.

Recordatorio: Dtm-Vic produce numerosos archivos de los resultados intermedios vinculados a la aplicación (todos en formato .txt). Por consiguiente, se recomienda especialmente **crear un directorio por aplicación**. Al principio, tal directorio debe contener los archivos de datos necesarios para el análisis al formato Dtm-Vic, descritos en el párrafo I.5.

II.1. Análisis en Componentes Principales (ACP o PCA)

Este primer ejemplo (ver directorio [DtmVic Examples_A_Start/EX_A01. PrinCompAnalysis](#)) describe un conjunto de variables continuas por el Análisis en Componentes Principales

II.1.1. Los datos y el archivo Dtm-Vic: Encuesta "presupuesto-tiempo"

Los datos son extraídos de una *Encuesta Presupuesto-tiempo Multimedia* efectuada por el Centre d'Etude des Supports de Publicité (www.cesp.org) en 1992 que abarcó cerca de 18.000 personas. Se relevó el tiempo dedicado a actividades diversas y diarias (trabajo, ocio, desplazamientos, comida, descanso) o sea 39 actividades (de V6 a V44) así como el tiempo destinado a diversos medios de comunicación (radio, televisión, prensa) o sea 5 medios de comunicación (de V45 a V49). El tiempo está expresado en minutos por día, medido el día que precede la entrevista. También han sido relevadas las características socioeconómicas del encuestado tales como la edad, el sexo, la actividad, el nivel educativo y el lugar de residencia correspondiente a 5 variables nominales (de V1 a V5). Los 18.000 respondentes originales son agrupados según las combinaciones de cinco características socioeconómicas que producen 96 grupos que constituyen en cierto modo "respondentes artificiales".

El cuadro de datos de este ejemplo presenta en fila las 96 categorías de respondentes y en columna las 5 características de base, el género, la edad, la educación y el aglomerado de residencia (o sea, 5 variables nominales), las 39 "actividades" diarias y 5 "frecuencia de uso de medios" (o sea, 43 variables continuas). En el cruce de la fila *i* y de la columna *j* se observa, después del identificador del individuo, el acumulado del tiempo dedicado (en minutos por día) para la actividad *j* por los individuos de la categoría *i*.

El objetivo es definir las asociaciones entre las diferentes actividades consideradas como variables activas y estudiar el lazo entre estas asociaciones y la frecuencia de uso de los medios de comunicación y también las características socioeconómicas [consideradas como variables ilustrativas (o suplementarias)].

A partir de un archivo de tipo Excel, se importa dos archivos en formato Dtm-Vic. Están contenidos en la carpeta [EX_A01. PrinCompAnalysis](#). Pueden ser abiertos con un editor de texto (bloque nota, notepad, Ultraedit, TotalEdit, notepad ++, o el editor de texto interno de Dtm-Vic).

Ident	Caract. socio-éco				Activités								Médias	
	Sexe	Age	Activ	Educ	Sommeil	Repos	Travail	Enfants	Ménage	Relation	Loisirs	Presse	Quotid_Nat	
1111	H	Jeun	Actif	Prim	463,8	23,8	306,5	27,9	21,3	70,2	100,6	20,9	0,8	
1115	H	Jeun	Actif	Prim	515,6	58,5	208,8	11,3	41,9	58,3	53,1	23,7	7,2	
1121	H	Jeun	Actif	Sec	463,3	34,2	317,0	22,3	18,1	66,8	94,3	24,7	1,6	
1122	H	Jeun	Actif	Sec	456,4	43,1	250,3	19,9	26,0	82,1	105,8	31,8	3,6	
1123	H	Jeun	Actif	Sec	478,0	44,2	217,9	29,6	22,3	80,4	81,1	29,3	1,9	
1124	H	Jeun	Actif	Sec	465,1	41,6	248,5	25,9	37,0	85,8	56,3	35,3	10,2	
1135	H	Jeun	Actif	Sup	458,4	47,4	328,2	24,4	25,3	72,5	65,0	45,8	10,9	
1133	H	Jeun	Actif	Sup	457,2	30,7	274,9	20,7	52,1	86,8	79,7	36,8	5,4	
1134	H	Jeun	Actif	Sup	465,2	40,2	280,0	16,5	36,3	97,5	64,1	51,8	14,9	
2111	H	Moy	Actif	Prim	449,0	42,1	316,6	5,7	15,1	46,7	133,8	28,0	1,2	
2112	H	Moy	Actif	Prim	450,2	63,1	249,6	18,1	40,4	78,0	99,1	23,5	1,2	
2115	H	Moy	Actif	Prim	455,2	47,4	251,6	15,7	30,4	53,7	82,1	31,9	4,9	
2121	H	Moy	Actif	Sec	461,9	39,3	337,1	15,1	14,9	49,6	105,3	33,3	2,0	
2122	H	Moy	Actif	Sec	453,7	44,7	274,9	23,5	23,1	72,1	106,9	37,2	3,3	
2123	H	Moy	Actif	Sec	433,1	49,8	299,7	22,6	22,4	51,4	98,9	49,4	4,1	

Cuadro de datos "presupuesto-tiempo" (primeras filas)

1. El archivo diccionario: **PCA_dic.txt**

Este archivo está accesible en la carpeta en francés (**PCA_dic_Fr.txt**) y en inglés (**PCA_dic_Eng.txt**). Contiene los identificadores de las 44 variables y de las categorías (o modalidades) de las variables nominales.

...2.Genre_V1	...0.Sommeil_V6	...0.Déma_Cours_V26
Fem Sex_Fem_1	0 Repos_V7	0 Proménad_V27
Hom Sex_Hom_2	0 Toilette_V8	0 Courses_V28
3 Age_V2	0 Repas_V9	0 Déplacem_V29
AMoy Age_Moy_1	0 Petit_Déj_V10	0 A_pied_V30
Ages Age_Ages_2	0 Repas_home_V11	0 En_Voitu_V31
Jeun Age_Jeun_3	0 Repas_rest_V12	0 Fréquent_V32
2 Activité_V3	0 Travail_V13	0 Autres_a_V33
acti Act_acti_1	0 TravailR_V14	0 Total_Do_V34
inac Act_inac_2	0 Enfants_V15	0 Total_De_V35
3 Education_V4	0 Ménage_V16	0 Total_ho_V36
prim Educ_prim_1	0 Relation_V17	0 Total_Me_V37
sec Educ_sec_2	0 Visite_amis_V18	0 Radio_V38
sup Educ_sup_3	0 Loisirs_V19	0 TV_V39
5 agglome_V5	0 Jeux_Jar_V20	0 Presse_V40
VImp aggl_Imp_1	0 Jardinag_V21	0 Quotid_N_V41
VMoy aggl_Moy_2	0 Loisirs_ext_V22	0 Quotid_R_V42
CRur aggl_Rur_3	0 Disque_V23	0 Magazine_V43
Mixt aggl_Mixte_4	0 Lecture_V24	0 Mag_TV_V44
APar aggl_Paris_5	0 Lect_livr_V25	

El identificador de una variable nominal es precedido por el número N de sus categorías (columna 5). Las N filas siguientes identifican las N categorías de respuestas: un "identificador corto" de 4 caracteres ocupa las columnas 1 - 5 y un "identificador largo" (de un máximo de 20 caracteres) comienza en la Columna 6. Convencionalmente, una variable numérica tiene cero categoría. Los espacios vacíos están prohibidos en los identificadores.

2. Extractos del archivo de dato **PCA_dat.txt**

```
'1111' 1. 1. 1. 1. 1. 463.80 23.80 26.30 139.00 16.00
'1115' 1. 1. 1. 1. 5. 515.60 58.50 19.20 138.30 13.50
'1121' 1. 1. 1. 2. 1. 463.30 34.20 28.40 126.30 16.20
'1122' 1. 1. 1. 2. 2. 456.40 43.10 29.30 118.40 15.10
'1123' 1. 1. 1. 2. 3. 478.00 44.20 28.80 115.40 15.00
'1124' 1. 1. 1. 2. 4. 465.10 41.60 30.30 135.70 17.40
'1136' 1. 1. 1. 3. 5. 458.40 47.40 28.10 133.30 15.50
'1133' 1. 1. 1. 3. 3. 457.20 30.70 25.80 137.00 17.80
'1134' 1. 1. 1. 3. 4. 465.20 40.20 28.80 136.30 16.80
'1221' 1. 1. 2. 2. 1. 523.90 41.80 26.10 112.20 15.20
```

Este archivo de datos comprende 96 filas y 45 valores. Para una fila i , el primer valor (entre comillas) corresponde al identificador del individuo i , es decir aquí el grupo i de respondentes, y los 44 otros valores corresponden a las respuestas de las 44 variables separadas por espacios blancos: los 5 primeros valores son los ítems de las 5 variables nominales (género, edad, actividad, educación, aglomerado de residencia que están en la base de la formación de los grupos), los otros 32 valores corresponden a las acumulados del tiempo dedicado (minutos por día) a las actividades por todos los individuos que constituyen el grupo i , y los 7 últimos valores corresponden a las acumulados del tiempo dedicado al contacto con un medio de comunicación.

II.1.2. Implementación del análisis (PCA)

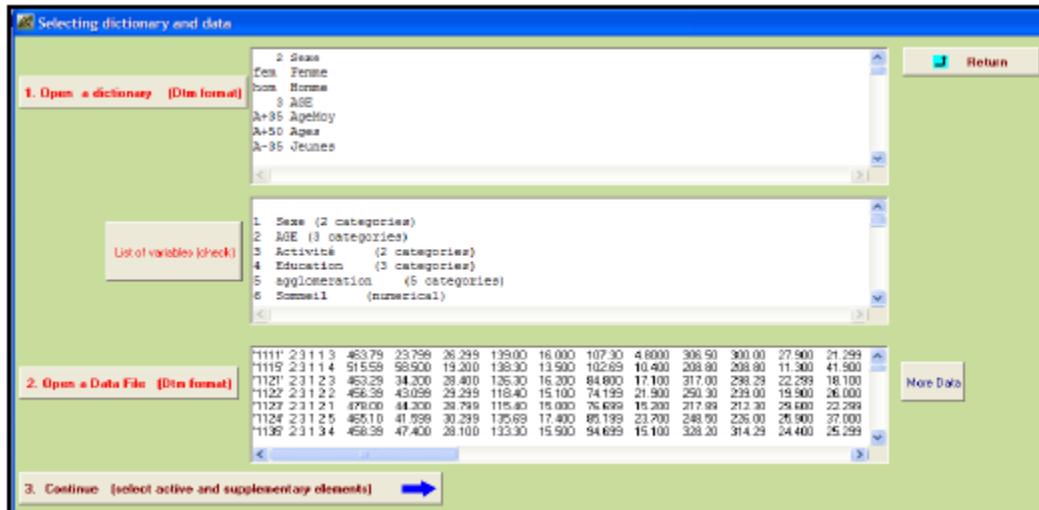
El archivo de los parámetros se crea en 5 etapas:

Etapas 1: selección del análisis

- Haga Clic en el botón **Create a Command file** de **Command File**
 - Aparece una ventana “*Choosing among some basic analyses*”.
- Seleccionar el análisis: **PCA – Principal Components Analysis** en la sección **numerical data (principal axes techniques)**
 - Una ventana “*Selecting dictionary and data*” aparece.

Etapas 2: Selección de los archivos diccionario y datas

- Haga Clic en el botón **Open a dictionary** En el directorio **EX_A01. PrinCompAnalysis**, abrir el archivo **PCA_dic.txt**. Se muestra en una primera ventana. El tipo (nominal [*categorical*] o numérico) de las variables se indica en la segunda ventana.



- Haga Clic en el botón **Open a Data File** En el directorio `DtmVic_Examples_A_Start \EX_A01.PrinCompAnalysis`, abrir el archivo `PCA_dat.txt` que se exhibe en la tercera ventana.
- Haga Clic sobre: **3. Continue** ➔ Aparece entonces una ventana "Selection of active y Supplementary elements"

Etapa 3: Selección de las variables activas e ilustrativas (o suplementarias)

Dentro de la ventana "Selection of active y Supplementary elements" se exhiben otras tres ventanas:

4. "Variables to be selected" donde figura el conjunto de las variables
5. "Active Variables" que recibe las variables activas seleccionadas
6. "Supplementary Variables" que recibe las variables ilustrativas (o suplementarias) seleccionadas.

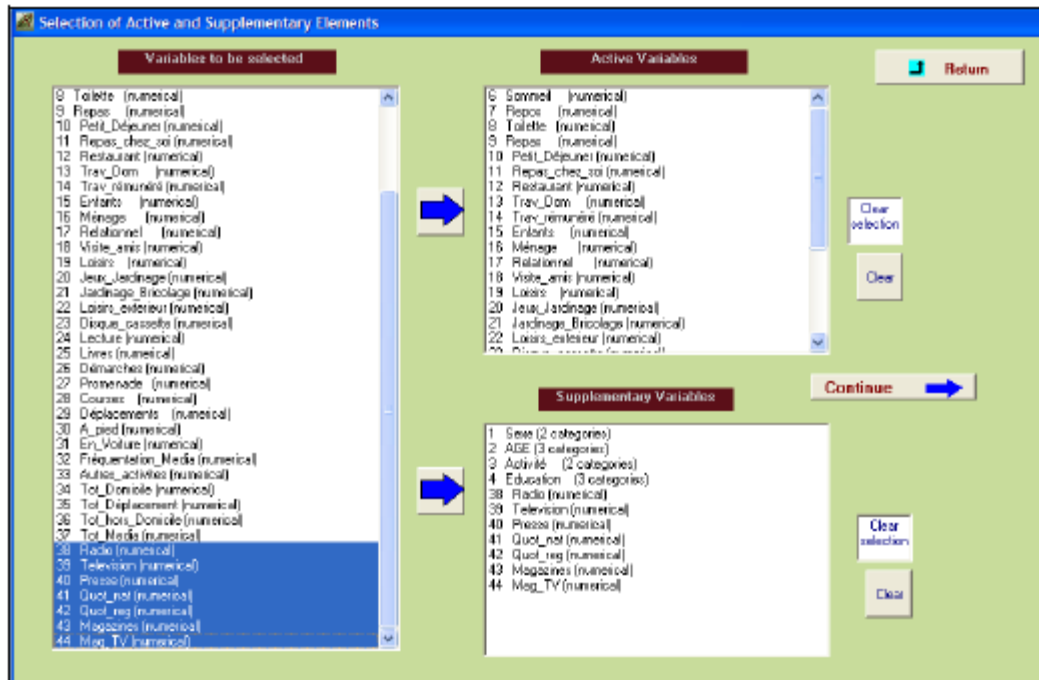
Para el ACP, las variables activas deben ser continuas (*numerical*). Las variables ilustrativas pueden ser continuas o nominales. Proponemos seleccionar las variables siguientes:

Selección de las variables continuas activas: V6 a V32 a trasladar en la ventana titulada "Active Variables":

6. Sommeil_V6	15. Enfants_V15	24. Lecture_V24
7. Repos_V7	16. Ménage_V16	25. Lect_livr_V25
8. Toilette_V8	17. Relation_V17	26. Démarche_Course_V26
9. Repas_V9	18. Visite_amis_V18	27. Promenad_V27
10. Petit_Déj_V10	19. Loisirs_V19	28. Courses_V28
11. Repas_home_V11	20. Jeux_Jar_V20	29. Déplacem_V29
12. Repas_rest_V12	21. Jardinag_V21	30. A_pied_V30
13. Travail_V13	22. Loisirs_ext_V22	31. En_Voitu_V31
14. TravailR_V14	23. Disque_V23	32. Fréquent_V32

Selección de las variables ilustrativas a trasladar a la ventana "Supplementary Variables".

variables continuas ilustrativas : V38 à V44	38. Radio 39. TV 40. Presse 41. Quotid_N	42. Quotid_R 43. Magazine 44. Mag_TV
variables nominales ilustrativas : V1 à V4	1. Sexe 2. Age	3. Activité 4. Education

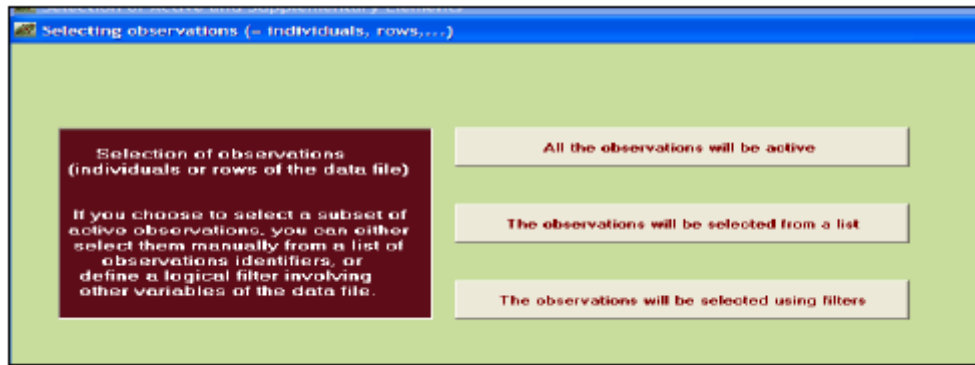


- Haga Clic sobre : **3. Continue** →
 - Aparece una ventana "Selecting observations".

Etapa 4: Selección de las observaciones (individuos)

Tres escenarios son posibles:

- Guardar todas las observaciones
- Seleccionar las observaciones sobre una lista
- Seleccionar las observaciones por un filtro



Tomamos en consideración aquí el conjunto de las observaciones.

- Haga Clic sobre: **All the observations will be active**
 - Aparece una ventana "Create a starting parameter file"

Etapa 5: creación del archivo de comando (archivo de parámetros)



A esta etapa, es posible seleccionar, como opción, los procedimientos de bootstrap y/o de clasificación. En efecto, en Dtm-Vic, los análisis factoriales pueden completarse por:

- Un procedimiento de bootstrap que permite validar la posición de las variables sobre el plano factorial
- Y/o una clasificación con una descripción automática de las clases.

a. Selección de una opción

- Haga Clic sobre **1-Select some options**
 - Aparece una ventana "Options: Bootstrap and/or Clustering of observations".
- Haga Clic sobre: "yes" para el procedimiento "bootstrap"; indicar el número de replicaciones (por defecto 25) luego **enter** Es el bootstrap parcial que se aplica por defecto. Si el bootstrap no es adoptado, haga Clic sobre: "no".
- Seleccione el número deseado de clases (sugerimos 7 clases). Luego haga Clic en **enter**

Nota técnica: los diferentes tipos de bootstrap para variables no textuales en Dtm-Vic son:

a _ Bootstrap parcial para las variables activas

Con este tipo de bootstrap, el plano inicial sirve de espacio de referencia para recibir las replicaciones, que son proyectadas como variables ilustrativas. El bootstrap parcial no tiene por objeto validar la estabilidad del espacio de salida que no está en discusión. Da una idea de la variabilidad imputable a las replicaciones para cada punto de modalidad tomado aisladamente.

b _ Bootstrap parcial para las variables ilustrativas

Para las variables ilustrativas, el bootstrap puede sólo ser parcial. Se trata de una validación externa, y por lo tanto una prueba estadística perfectamente legítima, no habiendo estas variables participado en la construcción del subespacio de referencia.

c _ Bootstrap total para las variables activas

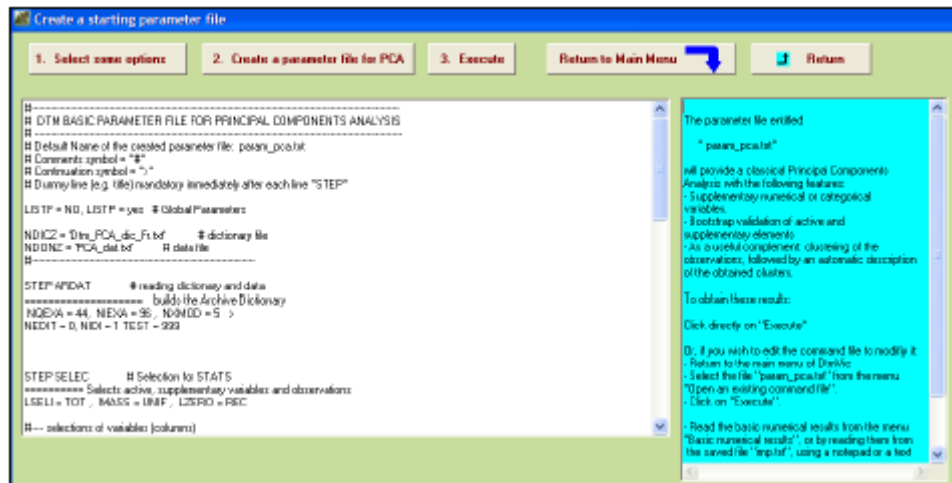
Recordemos que en este caso, cada replicación da lugar a un análisis en componentes principales específicos. Existen tres implementaciones del bootstrap total en Dtm-Vic.

- El bootstrap de tipo 1 (correcciones simples del signo de los ejes para los análisis de las replicaciones).
- El bootstrap de tipo 2 (correcciones de las inversiones de ejes) es más elaborado.
- El bootstrap de tipo 3 (“Rotaciones procrusteanas” de los ejes replicados de forma de traerlos en correspondencia con los ejes iniciales). Reunimos tan a menudo los resultados del bootstrap parcial. Las opciones de bootstrap total pueden ser puestas de ejecución por los usuarios avanzados, pero no son utilizadas en este manual.

- Haga Clic sobre: **3. Continue** ➔
 - Reaparece la ventana : “create a starting parameter file”.

b. Creación del archivo de parámetros

- Haga Clic sobre: **2-Create a parameter file for PCA.**

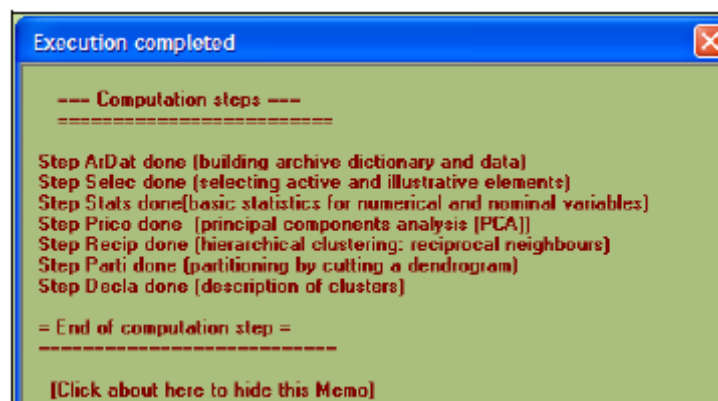


Se crea un archivo de parámetros bajo el nombre **param_PCA.txt** en la carpeta **EX_A01.PrinCompAnalysis** (de la carpeta **DtmVic_Examples_A_Start**). Para conservarlo con vistas a análisis posteriores, será prudente de renombrarlo, después de haber dejado a Dtm-Vic.

c. Ejecución

➤ Haga Clic sobre: **3-Execute**

La secuencia de los procedimientos se exhibe en bloque después de la ejecución:



Comentarios:

Ardat, (Almacenamiento de los datos), **Selec** (Selección de los elementos activos e ilustrativos), **Stats** (estadísticas básicas), **Prico** (Analiza en Componentes Principales), **Recip** (Clasificación mixta que utiliza la clasificación ascendente jerárquica - método de los vecinos recíprocos), **Parti** (Corte del dendrogramme y la optimización de la partición por el método de los centros móviles [k-means]), **Decla** (Descripción automática de las clases de la partición).

Nota: en el momento de una utilización posterior de Dtm-Vic, es posible abrir el archivo parámetro **param_PCA.txt** en el menú principal **Command File** con el procedimiento **Open an existing command file** y luego ejecutar directamente este archivo: **Execute**

Los usuarios experimentados pueden modificar parámetros directamente bajo el editor interno o fuera de Dtm-Vic con un editor de texto (ver "Help about parameters" disponible a partir del editor).

II.1.3 Archivo de los resultados

Los resultados pueden ser consultados a partir de la sección: **Result Files**

- Haga Clic sobre: **Basic numerical results** para navegar en el archivo de los resultados, luego sobre: **Return** para volver al menú principal.



- haga Clic sobre: **Basic numerical results (text format)** para abrir el archivo de resultados en formato texto.

El archivo resultado denominado **imp.txt** es contenido en el directorio **EX_A01.PrinCompAnalysis**. También se guarda bajo el nombre "imp" seguido por la fecha y la hora del análisis: "imp_08.07.11_14.45.txt" significa el 8 de julio de 2011, a las 14h 45. Este archivo de copia de seguridad conserva los resultados numéricos principales, mientras que el archivo **imp.txt** se sobrescribe para cada nuevo análisis realizado en el mismo directorio.

Después de haber consultado los resultados numéricos, vuelva al menú principal. Estos resultados serán visualizados entonces en la etapa VIC de Dtm-Vic que facilita considerablemente la interpretación (el histograma de los valores propios, el de los índices de nivel y el dendograma, sin embargo deben ser consultados en uno de los archivos **imp.txt** o **imp.html**).

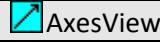
II.1.4 Visualización de los resultados

Esta segunda fase fundamental de Dtm-Vic proporciona los instrumentos de visualización necesarios para la interpretación y la validación de los resultados.



1-Ejes factoriales

Esta herramienta proporciona y clasifica las coordenadas sobre los ejes factoriales de las variables activas y ilustrativas, o de las observaciones.

- Haga Clic sobre:  **AxesView**
 - Aparece una ventana que propone visualizar las coordenadas de las variables activas, ilustrativas y de las observaciones sobre los primeros ejes (estos resultados son también los de la etapa DEFAC del archivo resultado).

Coordenadas de las variables continuas activas e ilustrativas: (ordenadas en el eje 1)

Identifia	axis 1	axis 2	axis 3	axis 4	axis 5
Repos_chet_s	731	48	-559	-189	116
Démarches	708	158	157	486	194
Repas	689	43	-492	-123	188
Courses	683	149	-72	483	150
Petit_Déjeun	666	-268	-8	347	104
Television	636	-291	28	-497	17
Ménage	620	541	-277	-204	-29
Fréquentatio	570	-439	412	-254	-28
Repos	568	-541	93	-95	-164
Mag_TV	467	126	-52	-74	37
Promenade	432	-28	492	19	155
Lecture	386	252	446	573	-294
Télévis	281	186	60	491	82

Coordenadas de las variables nominales ilustrativas: (Suplt categorías)

Identifia	axis 1	axis 2	axis 3	axis 4
actifs	-1667	97	-1024	293
AgeMoy	-495	166	-1434	441
Agee	1866	-1475	246	377
Femme	1312	1197	-655	-90
Homme	-1486	-1366	968	103
inactifs	1970	115	1212	-469
Jeunes	-1486	1373	1086	-776
primaire	939	-1195	-1070	-1215
secondaire	-119	68	239	-277
superieur	555	802	503	1258

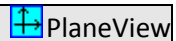
Observación: haciendo clic en la parte superior del eje 1, identificamos rápidamente las oposiciones visibles sobre este eje: oposición entre las actividades exteriores (relación, comida en el restaurante, traslado) sobre la parte positiva y las actividades del hogar (jardinería, comida en su casa) sobre la parte negativa; sobre el eje 2, el trabajo remunerado (parte positiva) se opone al descanso (parte negativa)

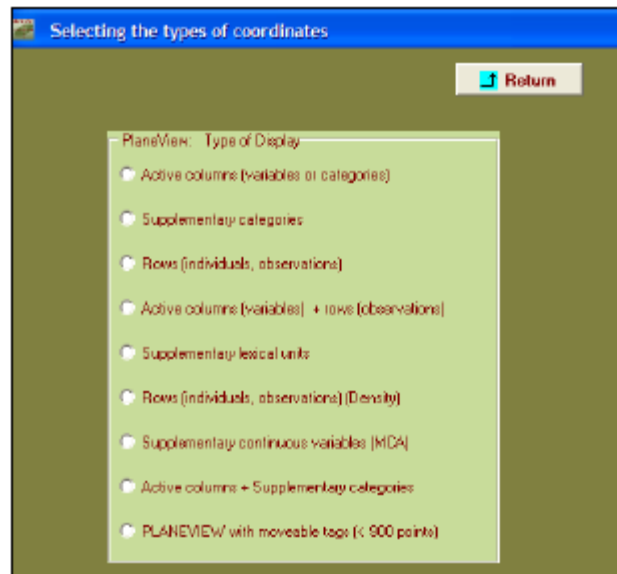
En el marco del análisis en componentes principales, tres elementos pueden ser examinados, *las variables continuas activas e ilustrativas, las variables nominales e ilustrativas y las observaciones.*

- Haga Clic en la pestaña de los elementos que hay que examinar, **Active variables** por ejemplo, luego sobre **View**. Es posible ordenar las coordenadas de un eje dado, haciendo Clic en la etiqueta "axis **x**" en la parte superior del eje **x**.
- Haga Clic en: **Exit** para salir de esta herramienta.

2-Planos factoriales

Esta herramienta proporciona los planos factoriales separados o superpuestos de las variables activas, ilustrativas o de las observaciones.

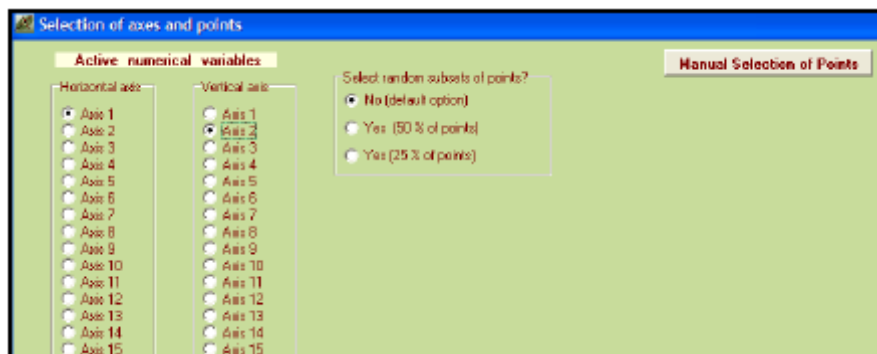
- Haga Clic sobre:  **PlaneView**
 - Una ventana propone diferentes visualizaciones de planos factoriales.



En este ejemplo de análisis, son posibles seis secciones del menú: "las columnas activas (variables o categorías)", "categorías ilustrativas", "filas activas (individuos, observaciones)", "columnas activas + filas activas", "individuos activos (densidad)" y "columnas activas + categorías ilustrativas". "PLANEVIEW with moveable tags" retoma algunas de las secciones precedentes y permite desplazar los puntos sobrepuestos para hacer más legible el gráfico.

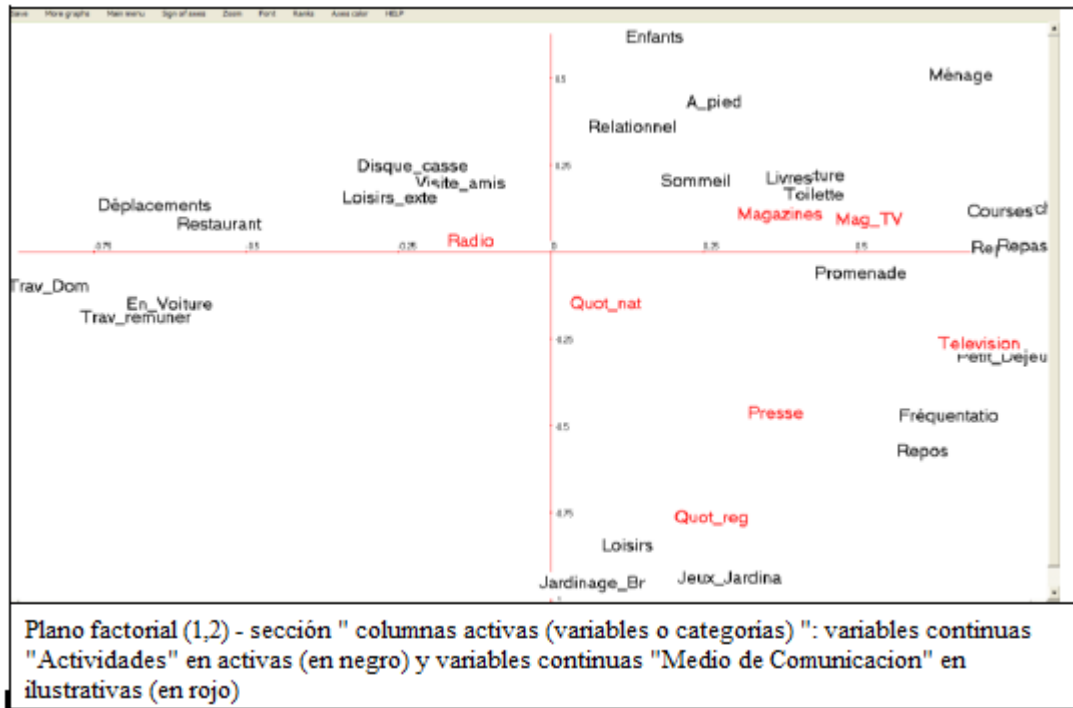
6. Seleccione la sección "Active columns (variables or categories)".

- Aparece una ventana para seleccionar el plano factorial siguiente y la dupla de ejes deseada.



- Elegir los ejes 1 y 2 luego haga Clic sobre: **Display** Es posible que sólo puedan ser mostradas en los planos algunas variables. Haga Clic entonces sobre: **Manual Selection of points** Seleccionar las variables y trasladarlas en la segunda ventana haciendo Clic en: **Select**

- Aparece la ventana del plano factorial.




En el caso de este ejemplo, la primera sección de menú "columnas activas (variables o categorías)" contiene de hecho las variables numéricas activas (en negro) y variables numéricas ilustrativas (en rojo).

Nota: Para cada gráfico, la banda superior contiene opciones:

- "Save" guarda el gráfico en formato bmp;
- "Font" ofrece la posibilidad de cambiar la fuente y el color de los caracteres;
- "More graph" permite cambiar de plano factorial;
- "Sign of axes" permite cambiar la orientación de los ejes;
- "Rank", es útil solamente en el caso de pantallas muy complejas, (lo que no es el caso aquí): este botón convierte las dos coordenadas de la pantalla actual en rangos. Por ejemplo, los n valores de la abscisa son convertidos en números enteros de 1 a n , teniendo el mismo orden que los valores originales. Así las dos distribuciones son uniformes, y los identificadores se revelan ser mucho más legibles (al precio de una distorsión sustancial de la forma de la nube de puntos).

➤ Para cerrar el gráfico, haga Clic en la cruz arriba a la derecha luego sobre:

 **Return** o directamente sobre la sección de la banda "Main menu".

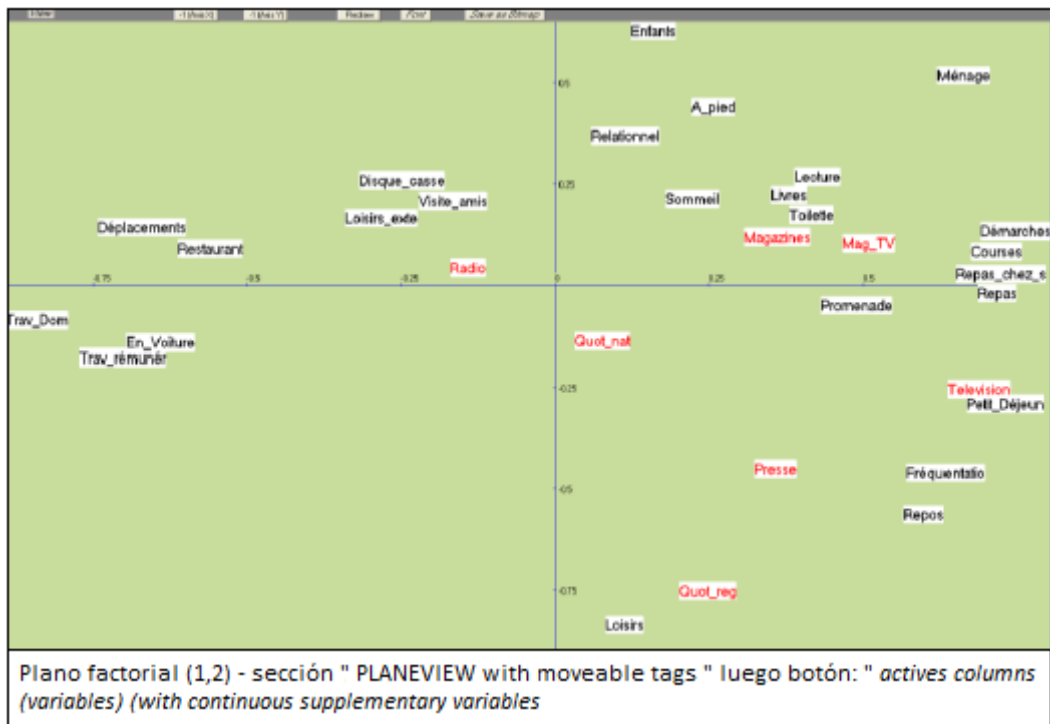
➤ Regrese luego sobre:  **PlaneView** para seleccionar otra representación factorial.

Otros gráficos:

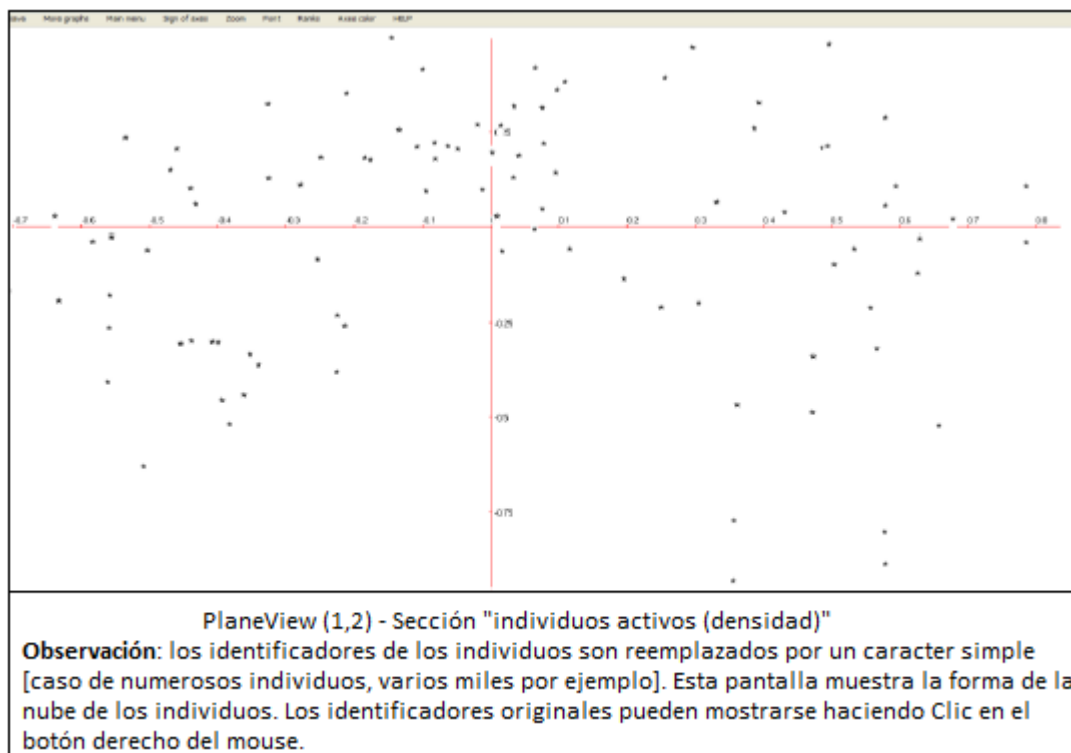
La sección: "PLANEVIEW with moveable tags" permite desplazar las etiquetas de los puntos del gráfico.

- Hacer un Clic sobre "PLANEVIEW with moveable tags" luego en **Continue**
 - Aparece una ventana.

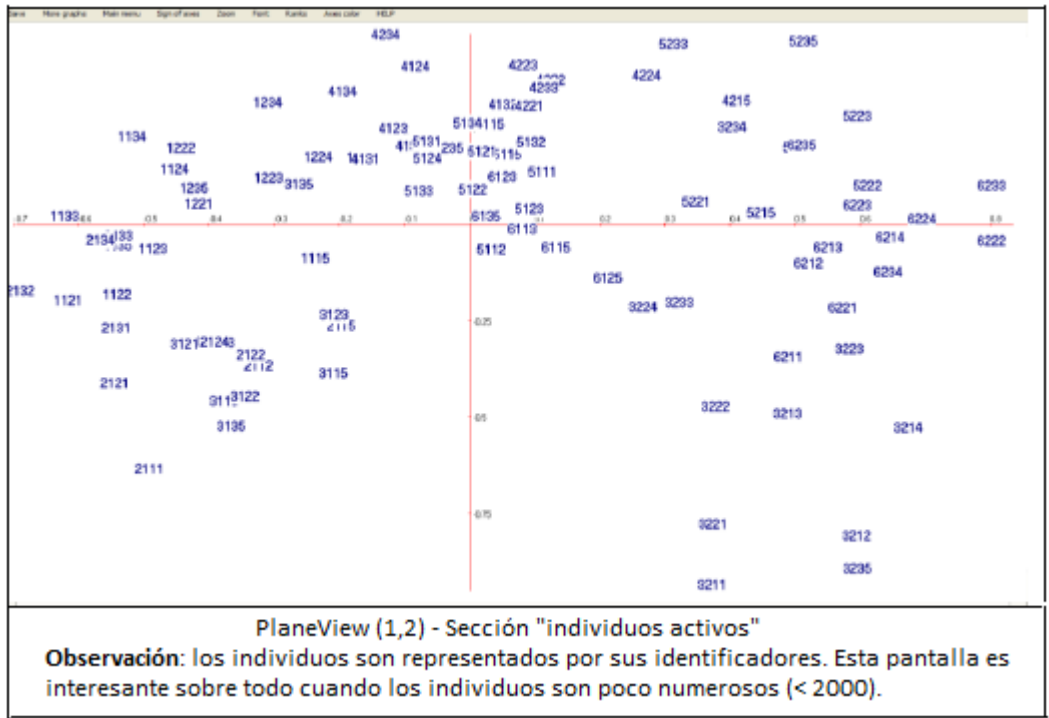
- Elegir por ejemplo "actives columns (variables) (with continuous supplementary variables)", hacer Clic en **Continue** y seleccionar el plano factorial.



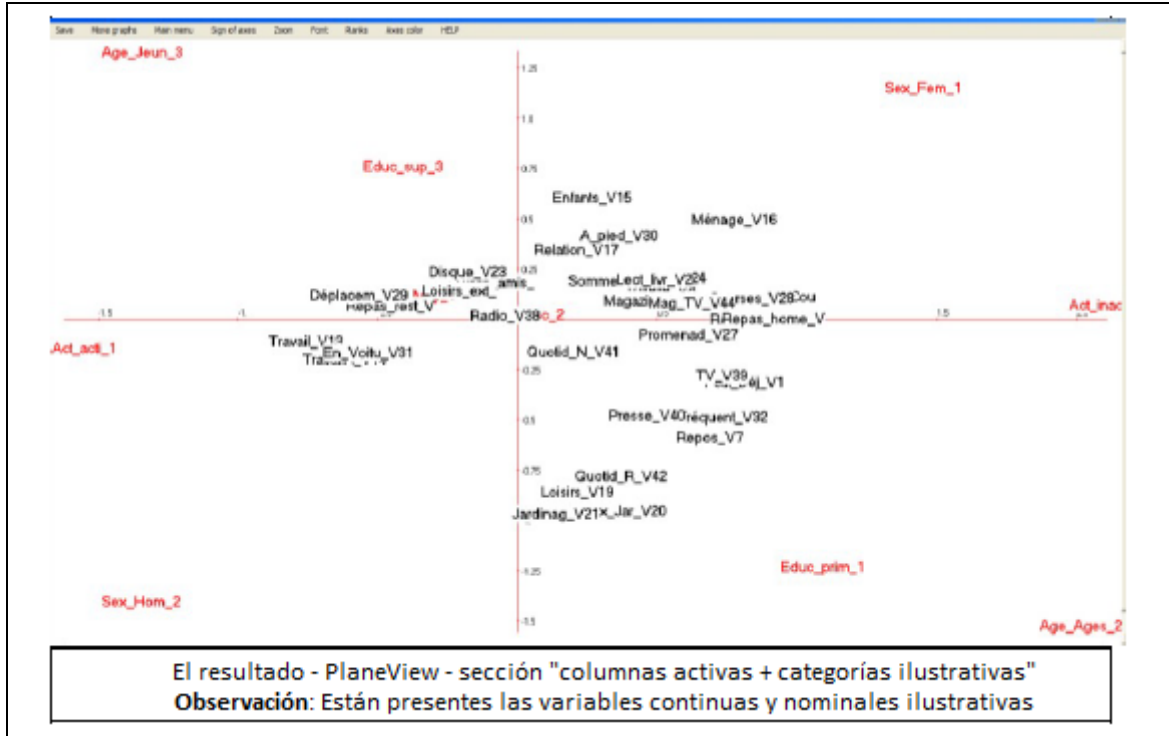
1. Sección " Individuos activos (densidad) ":



2. Sección " Individuos activos “:



3. Sección "columnas activas + categorías ilustrativas":



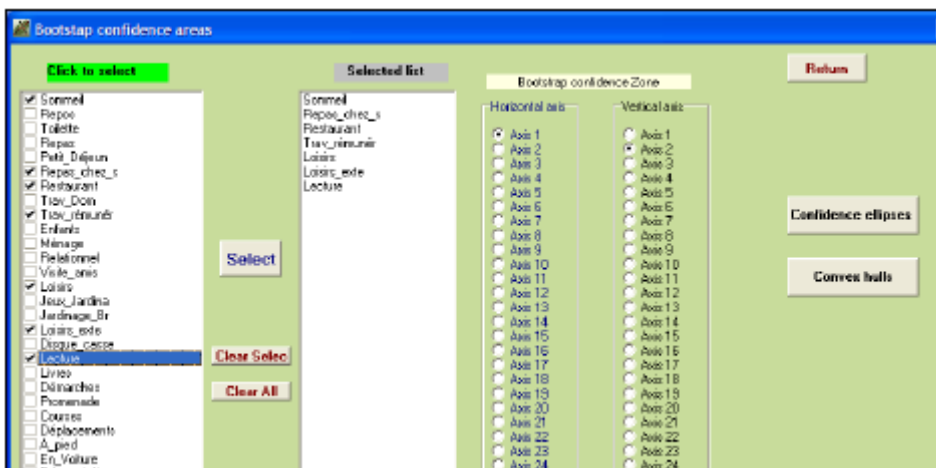
3-Validación Bootstrap

Esta herramienta permite validar la posición de las variables sobre el plano factorial.

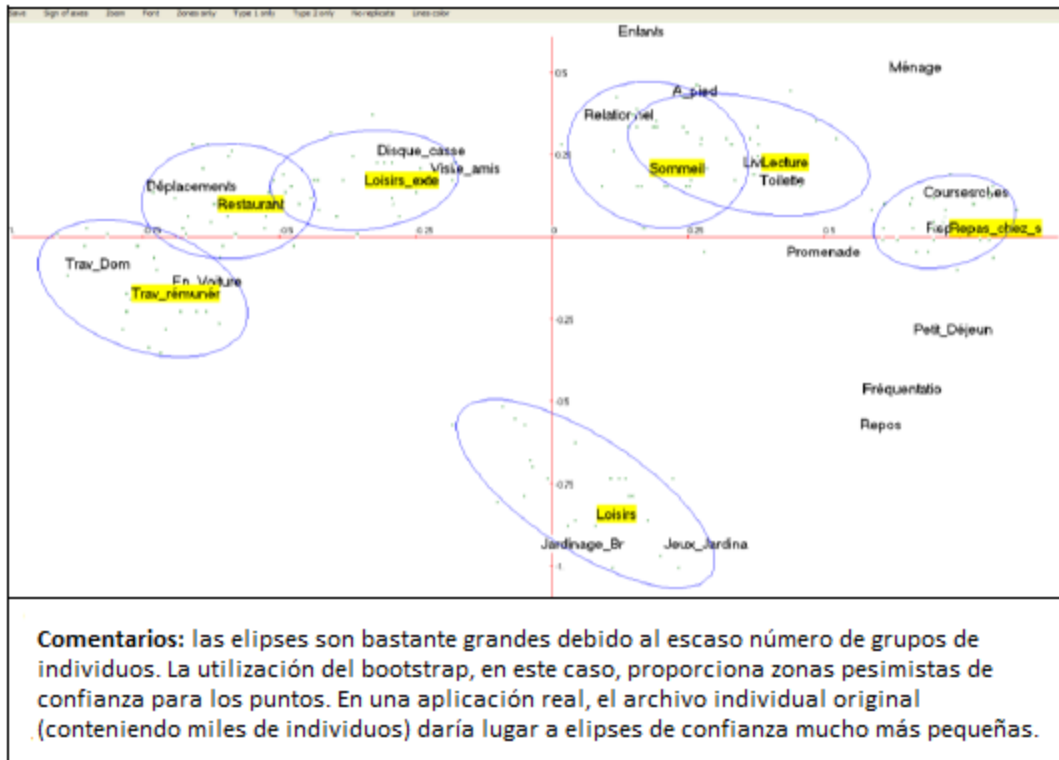
- Haga Clic en : **E** Bootstrap
 - Aparece una ventana "DtmVic - Bootstrap - Validation - Stability - Inférence".
- Hacer Clic en **Load Data** luego abrir en el directorio el archivo de las replicaciones según el bootstrap elegido. Seleccione el archivo **ngus_var_boot.txt** para un bootstrap parcial. Responder **OK** a la ventana "Set of principal coordinates loaded" que se exhibe.



- Luego haga Clic en **Confidence Areas**
 - Se muestra una ventana "Bootstrap confidence areas".
- Seleccione en la sección "Click to Select" las variables cuyas elipses se quieren visualizar. Trasladarlas con **Select**, a la ventana "selected list".



- Elegir luego el plano factorial y haga Clic en **Confidence ellipses** para obtener pantalla con el gráfico de las variables activas (si el archivo **ngus_var_boot.txt** fue cargado), o de las categorías ilustrativas (si el archivo **ngus_sup_cat_boot.txt** fue cargado).
 - Se exhibe una ventana de zonas de confianza bootstrap.



- Cierre la ventana y seleccione ahora el botón: **convex Hulls**

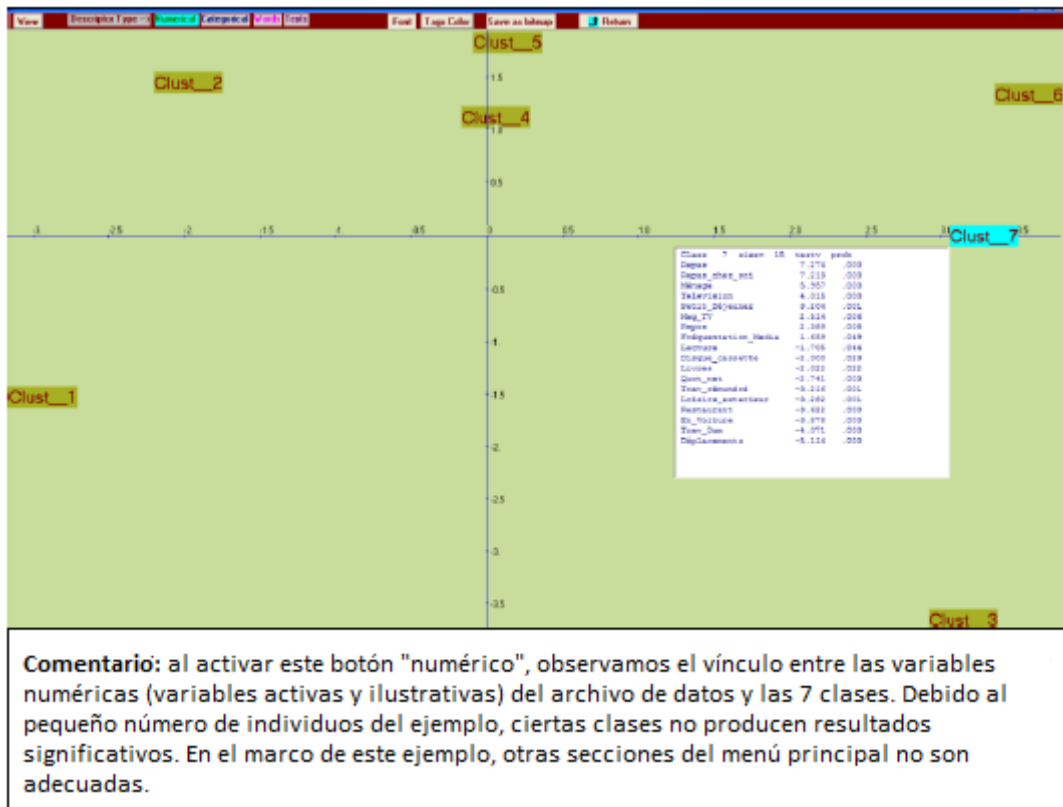
Las elipses son reemplazadas por las envolventes convexas de las replicaciones bootstrap para cada punto. Las envolventes convexas toman en consideración los puntos periféricos, mientras que las elipses son dibujadas utilizando la densidad de las nubes de las replicaciones. Las dos informaciones son complementarias.

- Para volver al menú principal de Dtm-Vic, haga Clic, sea sobre la cruz arriba a la derecha, o sea sobre **return**

4-Clasificación

Esta opción permite visualizar los centros de las clases, que son proyectados sobre el plano factorial.

- Haga Clic sobre **ClusterView**. Elegir los ejes (1 y 2 para comenzar), y **Continue**.
 - Aparece la ventana "DTM-Display of clusters".



- Haga Clic en **View**. Los centros de las 7 clases aparecen en el plano factorial. Haga Clic luego en la sección **Numerical** de la banda. Esta sección está ahora habilitada. Luego haciendo Clic (botón derecho del mouse) en una clase, aparecen las variables descriptivas de la clase.

El conjunto de los resultados figura en el procedimiento DECLA del archivo obtenido ("BASIC numerical results"). ClusterView nos permite apreciar la forma de la nube de los centros de clases e investigar interactivamente sus características.

Podemos fácilmente imaginar el interés de la herramienta para una visualización relativa a centenares de variables, los miles de individuos reagrupados, por ejemplo, en una veintena de clases.

II.2. Análisis de correspondencias (AC o SCA)

Este segundo ejemplo describe una pequeña tabla de contingencia por el análisis de correspondencias (los datos están en el directorio:

DtmVic-Examples_A_Start/ EX_A02. SimpleCorAnalysis).

II.2.1. Los datos y el archivo Dtm-Vic: *Contactos con los medios de comunicación*

Los datos provienen de una encuesta multimedia por muestreo (efectuado por el CESP en 1992) para la cual se observan aquí dos variables nominales: una variable: "medio de comunicación" con 6 clases (radio, televisión, prensas nacionales y regionales, revistas, revistas de TV) y una variable: "categoría ocupacional" con 8 clases (agricultor, pequeño empresario o patrón, alto ejecutivo, profesión intermedia, empleado, obrero calificado, obrero no calificado, inactivo). La tabla de contingencia que se considera se obtiene por cruce de estas dos variables.

Las 6 clases "medios de comunicación" están representadas en columna y las 8 clases "categoría ocupacional" son las filas de la tabla de contingencia. La celda (i, j) de la tabla contiene el número de contactos (el día precedente a la encuesta) correspondiente a los respondientes que pertenecen a la categoría i con el medio j. Recordemos que las filas y las columnas representan dos variables y desempeñan un papel idéntico (contrariamente al caso del análisis en componentes principales que distingue variables y observaciones).

Identifiers	Radio	TV	Quot_Nat	Quot_Reg	Magazine	Mag_TV
Agriculteur	96	118	2	71	50	17
Petit_patron	122	136	11	76	49	41
Aff_Cadre_sup	193	184	74	63	103	79
Prof_interm	360	365	63	145	141	184
Employé	511	593	57	217	172	306
Ouvrier_qualif	385	457	42	174	104	220
Ouvr_non_qualif	156	185	8	69	42	85
Inactif	1474	1931	181	852	642	782

Tabla de contingencia que cruza los medios y los estatutos de actividad

El objetivo es describir las relaciones entre los diferentes medios de comunicación y categorías ocupacionales para la población considerada.

También consideramos, en fila, tres otras características socio-económicas, el sexo, la edad y el nivel de estudio como variables ilustrativas. Las tablas de contingencia que cruzan estas variables con la variable "medio de comunicación" se yuxtaponen así a la tabla precedente.

La carpeta **EX_A02. SimpleCorAnalysis** contiene el archivo de datos y el archivo diccionario (que pueden ser importados a partir de un archivo de datos de tipo Excel).

- Archivo de datos: **SCA_dat.txt**

'Agriculteur'	96	118	2	71	50	17
'Petit_patron'	122	136	11	76	49	41
'Aff. Cadre sup'	193	184	74	63	103	79
'Prof. interm'	360	365	63	145	141	184
'Employ, '	511	593	57	217	172	306
'Ouvrier qualif'	385	457	42	174	104	220
'Ouvrier non qual'	156	185	8	69	42	85
'Inactif'	1474	1931	181	852	642	782
'Homme'	1630	1900	285	854	621	776
'Femme'	1667	2069	152	815	683	938
'15-24 ans'	660	713	69	216	234	360
'25-34 ans'	640	719	84	230	212	380
'35-49 ans'	888	1000	130	429	345	466
'50-64 ans'	617	774	84	391	262	263
'65 ans ou +'	491	761	70	402	251	245
'Primaire'	908	1307	73	642	360	435
'Secondaire'	869	1008	107	408	336	494
'Techn. prof.'	901	1035	80	140	311	504
'Superieur'	619	612	177	209	298	281

Este archivo de datos contiene 20 filas (entre las que 8 serán activas) y 7 columnas. Cada fila contiene el identificador de las categorías socio-económicas (entre comillas) seguido por los 6 valores que corresponden a las frecuencias absolutas de 6 medios, separadas por lo menos por un espacio vacío.

- Archivo diccionario: **SCA_dic.txt**

Radio
Television
Quot_Nat
Quot_Reg
Magazine
Mag_TV

En este formato interno de Dtm-Vic, las denominaciones de las categorías comienzan en la columna 6, [puede emplearse una fuente a intervalo tal como el "courier" para facilitar la utilización de este tipo de formato].

Nota: los espacios vacíos en los identificadores (individuos y variables) no son permitidos.

II.2.2. Implementación del análisis (SCA)

Así como en el ejemplo 1, el archivo parámetro se crea en 5 etapas:

Etapa 1: selección del análisis

En la ventana del menú principal, haga Clic sobre: **Create Command File**

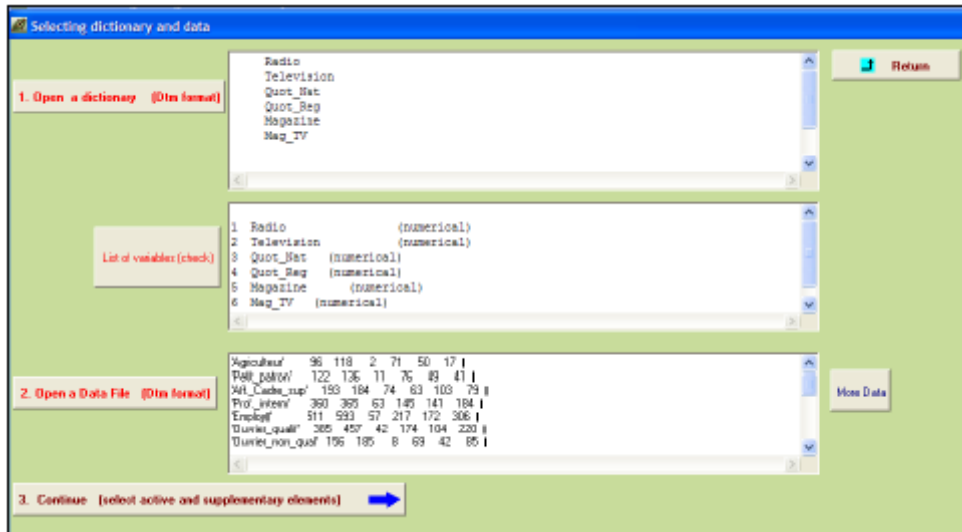
- Una ventana "*Choosing among some basic analyses*" aparece.

Seleccione el análisis: **SCA – Simple Correspondence Analysis** en la sección: **Numerical data (principal axes techniques)**.

- Aparece una ventana de apertura de los "dictionary and data file"

Etapa 2: Selección de los archivos diccionarios y de datos

- Haga Clic en el botón **Open a dictionary**. En la carpeta **EX_A02. SimpleCorAnalysis** del juego de ejemplos de Dtm-Vic, abrir el archivo **SCA_dic.txt**. Se exhibe una primera ventana. La lista y el estado (numérico por defecto en este ejemplo) de las variables se indican en una segunda ventana.



Las columnas de frecuencias, para una variable nominal dada, están consideradas aquí como variables numéricas. Veremos que para el análisis de las correspondencias múltiples (después en sección II.3), las variables nominales tienen el estado de "variable categórica", como lo vimos a propósito de ciertas variables ilustrativas en ACP.

- Haga Clic en el botón **Open a Data File** En la misma carpeta **EX_A02**.
- **SimpleCorAnalysis**, abrir el archivo **SCA_dat.txt** que se muestra en una tercer ventana.

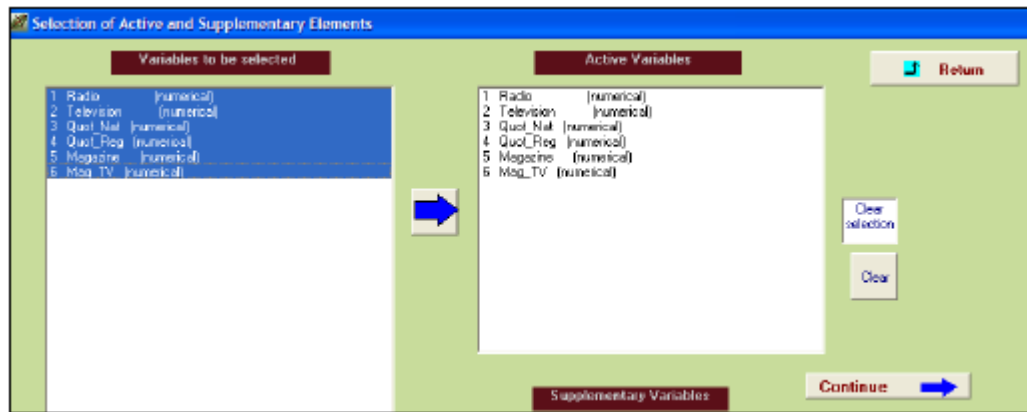
Nota: es posible que una caja de mensaje anuncie la existencia de una última fila vacía. Hacer Clic entonces sobre **OK** dos veces.

- Haga Clic sobre: **3. Continue** ➔
 - Aparece una ventana "Selection of active and supplementary elements".

Etapa 3: selección de las variables activas e ilustrativas (o suplementarias)

En el caso de una tabla de contingencia, las variables son de hecho las modalidades de la variable considerada en columna, es decir en este caso los medios. El juego de datos presenta aquí pocas variables (tipos de medios de comunicación) que están consideradas todas como activas.

- Selección de las variables continuas activas: V1 en V6 a trasladar a la ventana "Active Variables"

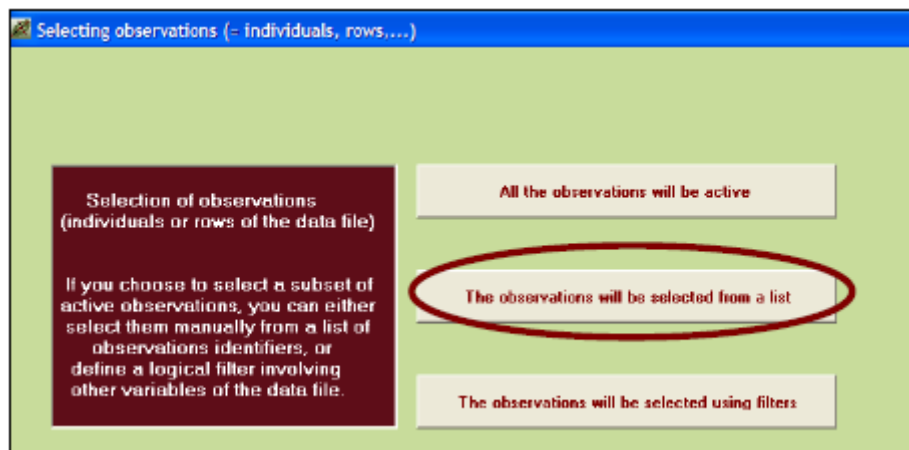


- Haga Clic en : **Continue** ➔
 - Aparece una ventana "Selecting observations".

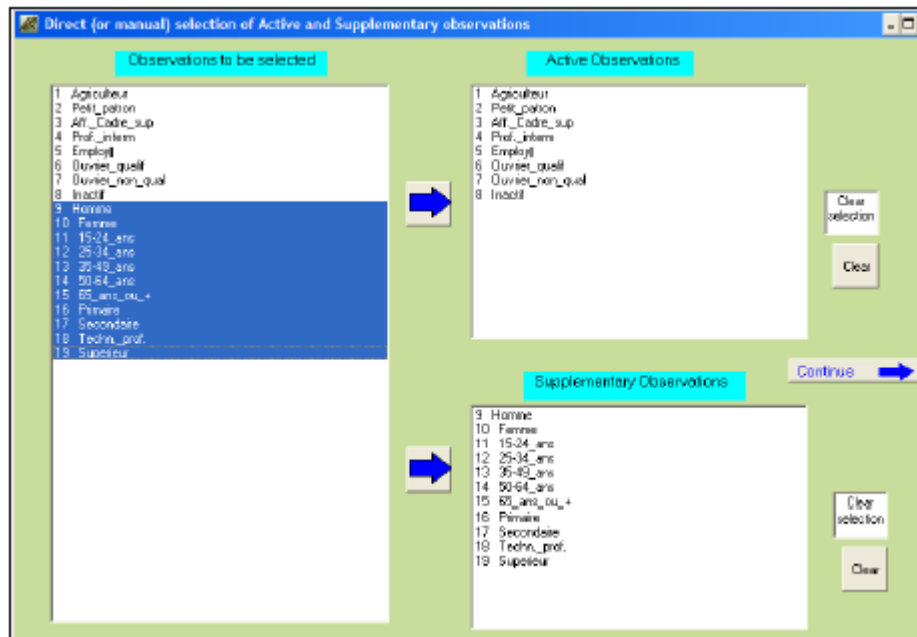
Etapa 4: selección de las observaciones (individuos)

Las filas no representan aquí observaciones o individuos como para el ACP o el Análisis de Correspondencias Múltiples (posteriormente) sino modalidades de variables. También de la misma manera que se considera variables activas y/o ilustrativas, procedemos a la selección de las modalidades activas y/o ilustrativas representadas en fila. Retenemos aquí el conjunto de las 8 categorías ocupacionales como variable activa, y el sexo, la edad y el nivel de estudio como variables ilustrativas.

- Haga Clic sobre: **The observations will be selected from a list**



- Aparece la ventana "selection of Active and Supplementary observations".
- Seleccione las modalidades de la variable "categoría ocupacional" como elementos activos. Luego seleccione las clases de las variables "sexo", "edad", "nivel de estudio" como elementos ilustrativos.



- Haga Clic sobre : **Continue →**
 - Aparece una ventana: "Create a starting parameter file".

Etapa 5: Creación del archivo parámetro

Elegimos aquí un procedimiento bootstrap. (Si no aparece, haga Clic directamente en: **2-Create a parameter file for SCA**).

- Haga Clic sobre **1-Select some options**
 - Se abre una ventana "Options: Bootstrap and / or Clustering of observations".

Como hay un pequeño número de individuos, no es necesaria ninguna clasificación: consideramos aquí sólo el procedimiento del *bootstrap*.

- Haga Clic sobre "yes" para el procedimiento *bootstrap*; indicar el número de replicaciones (por defecto 25) luego: **Enter**. Es el *bootstrap* parcial que es aplicado por defecto (ver Cuadro técnico sección II.1.2 Etapa 5 a propósito del ACP).
- Elegir la clase 0 o 1, luego haga Clic sobre: **Enter**. No queremos efectuar de clasificación.
- Haga Clic en: **Continue →**

Options: bootstrap and /or clustering of observations

Number of replicates (between 5 and 30)

Suggested value = 25

25 Enter

Bootstrap options

Partial (default) Total

1. Do you want a bootstrap validation?

Bootstrap

yes

no

2. How many clusters? (to begin with...)

(0 or 1 means no clustering at all)

0 Enter

Continue →

- Reaparece la ventana: "Create a starting parameter file".
- Haga Clic en: **2-Create a parameter file for SCA**. Un archivo parámetro acaba de ser creado bajo el nombre **param_SCA.txt** y almacenado en la carpeta **EX_A02.SimpleCorAnalysis** del directorio **DtmVic_Examples_A_Start**. (Para conservarlo con vistas a reiterar más tarde directamente el mismo análisis, habrá que renombrarlo después del análisis).

Create a starting parameter file

1 - Select some options 2 - Create a parameter file for SCA 3 - Execute Return to Main Menu

```
#-----
# DTM BASIC PARAMETER FILE FOR SIMPLE CORRESPONDENCE ANALYSIS
#-----
# Default Name of the created parameter file: param_sca.txt
# Comments symbol = "#"
# Continuation symbol = "\n"
# Dummy line (e.g. file) mandatory immediately after each line "STEP"
#-----
# Global Parameters
LUSTF = NO, LUSTP = yes # Global Parameters
NDICZ = Dtm_SCA_dic_Fr.txt # dictionary file
NDONZ = SCA_da_Fr.txt # data file
#-----
#----- Comments about step ARDAT -----
# NDEVA = ... number of questions (or variables) in both the dictionary
# and the data file
# NIEVA = ... number of "individuals" (or rows) in the data file
# NIDI = ... indicate the presence of an identifier (recommended)
#-----
STEP ARDAT # reading dictionary and data
#----- builds the Archive Dictionary
NDEVA = 6, NIEVA = 19, NIMOD = 1 >
MEDIT = 0, NIDI = 1 TEST = 999
```

The parameter file entitled "param_sca.txt" will provide a classical Simple Correspondence Analysis with the following features:

- Supplementary variables.
- Bootstrap validation of elements
- Clustering of rows.

To obtain these results:

- Click directly on "Execute",
- or, if you wish to study or edit the parameter file,
- Return to the main menu of DtmVic
- Select the file "param_sca.txt" from the button "Open an existing command file".
- Click on "Execute".

- Read the basic numerical results from the button "Basic numerical results", or by loading them from the "exp.txt", using a notepad or a text editor.

Please, have a look at the tutorial to visualize the results.

- Haga Clic en : **3-Execute**

Execution completed

=== Computation steps ===

Step ArDat done (building archive dictionary and data)

Step Selec done (selecting active and illustrative elements)

Step Afor done (correspondence analysis [CA])

Step Defoc done (description of principal axes)

- End of computation step -

[Click about here to hide this Memo]

Los procedimientos se muestran en bloque al final de la ejecución: **ArdaT** (Almacenamiento de los datos), **Selec** (Selección de los elementos activos e ilustrativos), **Afcor** (Análisis de correspondencias) y **Defac** (Descripción de los ejes factoriales).

Nota: Para una utilización posterior de Dtm-Vic, es posible abrir el archivo parámetro **param_SCA.txt** en el menú principal **Command File** con el procedimiento **Open an existing command file** y ejecutar este archivo: **Execute**

II.2.3 Archivo de los resultados

Los resultados pueden consultarse en la etapa **Result Files**


- Haga Clic sobre: **Basic numerical results** para abrir el archivo en formato html luego o en: **Basic numerical results (text format)** para abrir el archivo resultado en formato texto luego hacer Clic en **Return** para salir de allí y volver al menú principal.

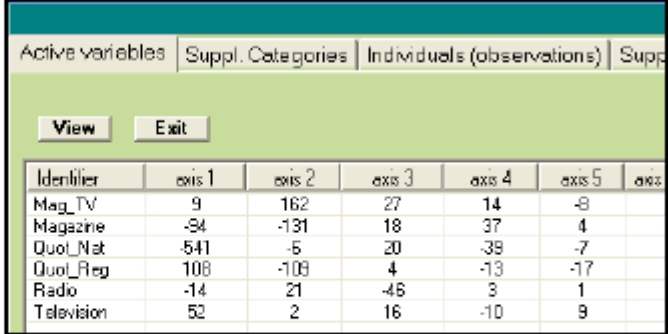
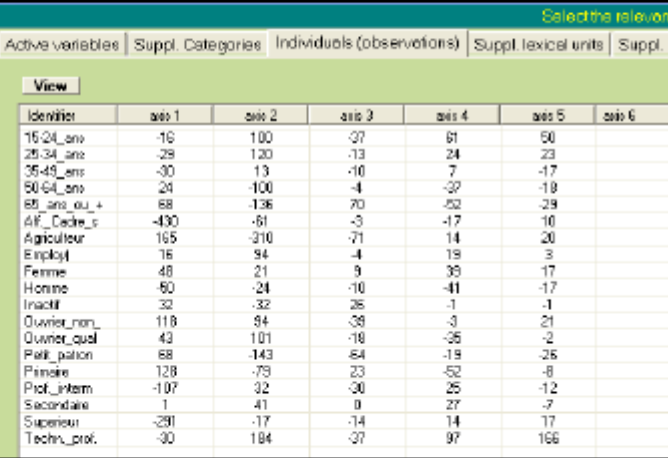
La denominación del archivo resultado se hace según los mismos principios que para el análisis en componentes principales.

II.2.4 Visualización de los resultados

Reenviamos al lector el párrafo II.1.4 para la presentación de la segunda fase de Dtm-Vic y el detalle de las diferentes herramientas de visualización. Consideramos aquí como herramientas: AxesView, PlaneView y Bootstrap.

1-Ejes factoriales

- Haga Clic sobre:  **AxesView** Una ventana propone visualizar las coordenadas de las variables activas, ilustrativas y de las observaciones, sobre los primeros ejes (resultados que corresponden a la etapa DEFAC del archivo resultado).
- Haga Clic sobre: **Active variables** luego sobre: **View** para obtener las coordenadas de las modalidades "medio de comunicación". Haga Clic luego sobre: **Individuals (observations)** luego sobre: **View** para conseguir las coordenadas de las modalidades activas "categoría ocupacional" y modalidades ilustrativas

<p>Coordenadas de las modalidades de la variable « Medio de comunicación »</p>	 <table border="1"> <thead> <tr> <th>Identifier</th> <th>axis 1</th> <th>axis 2</th> <th>axis 3</th> <th>axis 4</th> <th>axis 5</th> <th>axis 6</th> </tr> </thead> <tbody> <tr> <td>Mag_TV</td> <td>9</td> <td>162</td> <td>27</td> <td>14</td> <td>-8</td> <td></td> </tr> <tr> <td>Magazine</td> <td>-94</td> <td>-131</td> <td>18</td> <td>37</td> <td>4</td> <td></td> </tr> <tr> <td>Quot_Nat</td> <td>541</td> <td>-5</td> <td>20</td> <td>-39</td> <td>-7</td> <td></td> </tr> <tr> <td>Quot_Reg</td> <td>108</td> <td>-109</td> <td>4</td> <td>-13</td> <td>-17</td> <td></td> </tr> <tr> <td>Radio</td> <td>-14</td> <td>21</td> <td>-45</td> <td>3</td> <td>1</td> <td></td> </tr> <tr> <td>Television</td> <td>52</td> <td>2</td> <td>16</td> <td>-10</td> <td>9</td> <td></td> </tr> </tbody> </table>	Identifier	axis 1	axis 2	axis 3	axis 4	axis 5	axis 6	Mag_TV	9	162	27	14	-8		Magazine	-94	-131	18	37	4		Quot_Nat	541	-5	20	-39	-7		Quot_Reg	108	-109	4	-13	-17		Radio	-14	21	-45	3	1		Television	52	2	16	-10	9																																																																																												
Identifier	axis 1	axis 2	axis 3	axis 4	axis 5	axis 6																																																																																																																																							
Mag_TV	9	162	27	14	-8																																																																																																																																								
Magazine	-94	-131	18	37	4																																																																																																																																								
Quot_Nat	541	-5	20	-39	-7																																																																																																																																								
Quot_Reg	108	-109	4	-13	-17																																																																																																																																								
Radio	-14	21	-45	3	1																																																																																																																																								
Television	52	2	16	-10	9																																																																																																																																								
<p>Coordenadas de las modalidades de la variable "categoría ocupacional" (Esta variable es situada en fila y considerada aquí como individuos)</p>	 <table border="1"> <thead> <tr> <th>Identifier</th> <th>axis 1</th> <th>axis 2</th> <th>axis 3</th> <th>axis 4</th> <th>axis 5</th> <th>axis 6</th> </tr> </thead> <tbody> <tr> <td>15-24_ans</td> <td>-16</td> <td>100</td> <td>-37</td> <td>61</td> <td>50</td> <td></td> </tr> <tr> <td>25-34_ans</td> <td>-28</td> <td>120</td> <td>-13</td> <td>24</td> <td>23</td> <td></td> </tr> <tr> <td>35-44_ans</td> <td>-30</td> <td>13</td> <td>-10</td> <td>7</td> <td>-17</td> <td></td> </tr> <tr> <td>50-64_ans</td> <td>24</td> <td>-100</td> <td>-4</td> <td>-37</td> <td>-18</td> <td></td> </tr> <tr> <td>65_ans_ou_+</td> <td>68</td> <td>-136</td> <td>70</td> <td>-52</td> <td>-28</td> <td></td> </tr> <tr> <td>AF_Cadre_s</td> <td>-430</td> <td>-61</td> <td>-3</td> <td>-17</td> <td>10</td> <td></td> </tr> <tr> <td>Agriculteur</td> <td>165</td> <td>-310</td> <td>-71</td> <td>14</td> <td>20</td> <td></td> </tr> <tr> <td>Employé</td> <td>16</td> <td>94</td> <td>-4</td> <td>19</td> <td>3</td> <td></td> </tr> <tr> <td>Femme</td> <td>48</td> <td>21</td> <td>3</td> <td>35</td> <td>17</td> <td></td> </tr> <tr> <td>Homme</td> <td>-60</td> <td>-24</td> <td>-10</td> <td>-41</td> <td>-17</td> <td></td> </tr> <tr> <td>Inactif</td> <td>-32</td> <td>-32</td> <td>26</td> <td>-1</td> <td>-1</td> <td></td> </tr> <tr> <td>Ouvrier_non_</td> <td>118</td> <td>94</td> <td>-38</td> <td>-3</td> <td>21</td> <td></td> </tr> <tr> <td>Ouvrier_qual</td> <td>43</td> <td>101</td> <td>-18</td> <td>-35</td> <td>-2</td> <td></td> </tr> <tr> <td>Patron_pation</td> <td>68</td> <td>-143</td> <td>64</td> <td>-19</td> <td>-25</td> <td></td> </tr> <tr> <td>Primeire</td> <td>128</td> <td>-73</td> <td>23</td> <td>-52</td> <td>-8</td> <td></td> </tr> <tr> <td>Prof_interm</td> <td>-107</td> <td>32</td> <td>-30</td> <td>25</td> <td>-12</td> <td></td> </tr> <tr> <td>Secondaire</td> <td>1</td> <td>41</td> <td>0</td> <td>27</td> <td>-7</td> <td></td> </tr> <tr> <td>Supérieur</td> <td>-231</td> <td>-17</td> <td>-14</td> <td>14</td> <td>17</td> <td></td> </tr> <tr> <td>Techn_prof.</td> <td>-30</td> <td>184</td> <td>-37</td> <td>97</td> <td>166</td> <td></td> </tr> </tbody> </table>	Identifier	axis 1	axis 2	axis 3	axis 4	axis 5	axis 6	15-24_ans	-16	100	-37	61	50		25-34_ans	-28	120	-13	24	23		35-44_ans	-30	13	-10	7	-17		50-64_ans	24	-100	-4	-37	-18		65_ans_ou_+	68	-136	70	-52	-28		AF_Cadre_s	-430	-61	-3	-17	10		Agriculteur	165	-310	-71	14	20		Employé	16	94	-4	19	3		Femme	48	21	3	35	17		Homme	-60	-24	-10	-41	-17		Inactif	-32	-32	26	-1	-1		Ouvrier_non_	118	94	-38	-3	21		Ouvrier_qual	43	101	-18	-35	-2		Patron_pation	68	-143	64	-19	-25		Primeire	128	-73	23	-52	-8		Prof_interm	-107	32	-30	25	-12		Secondaire	1	41	0	27	-7		Supérieur	-231	-17	-14	14	17		Techn_prof.	-30	184	-37	97	166	
Identifier	axis 1	axis 2	axis 3	axis 4	axis 5	axis 6																																																																																																																																							
15-24_ans	-16	100	-37	61	50																																																																																																																																								
25-34_ans	-28	120	-13	24	23																																																																																																																																								
35-44_ans	-30	13	-10	7	-17																																																																																																																																								
50-64_ans	24	-100	-4	-37	-18																																																																																																																																								
65_ans_ou_+	68	-136	70	-52	-28																																																																																																																																								
AF_Cadre_s	-430	-61	-3	-17	10																																																																																																																																								
Agriculteur	165	-310	-71	14	20																																																																																																																																								
Employé	16	94	-4	19	3																																																																																																																																								
Femme	48	21	3	35	17																																																																																																																																								
Homme	-60	-24	-10	-41	-17																																																																																																																																								
Inactif	-32	-32	26	-1	-1																																																																																																																																								
Ouvrier_non_	118	94	-38	-3	21																																																																																																																																								
Ouvrier_qual	43	101	-18	-35	-2																																																																																																																																								
Patron_pation	68	-143	64	-19	-25																																																																																																																																								
Primeire	128	-73	23	-52	-8																																																																																																																																								
Prof_interm	-107	32	-30	25	-12																																																																																																																																								
Secondaire	1	41	0	27	-7																																																																																																																																								
Supérieur	-231	-17	-14	14	17																																																																																																																																								
Techn_prof.	-30	184	-37	97	166																																																																																																																																								
<p><i>El factor 1 opone la prensa diaria nacional a otros medios y los ejecutivos a otras categorías. El factor 2 opone la prensa regional y la revista a la prensa TV, y los agricultores e independientes a los empleados y obreros.</i></p>																																																																																																																																													

- Haga Clic sobre: **exit** para salir de esta herramienta

2-Planos factoriales

- Haga Clic sobre: **PlaneView**.

Se muestra una ventana en la que se proponen diferentes visualizaciones de planos factoriales.

Esta opción proporciona los planos factoriales separados o superpuestos de las variables activas e ilustrativas, o las observaciones. Acá todavía, variables y observaciones representan las modalidades de las dos variables de la tabla de contingencia. En este caso, el submenú "Actives columns + Active rows" es adaptado para la tabla de contingencia.

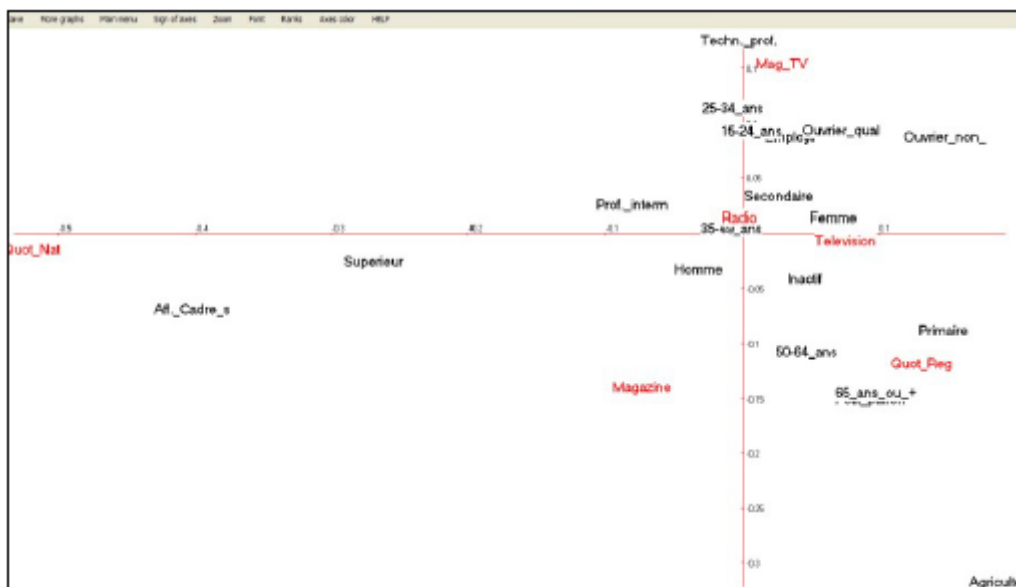
- Haga Clic en la sección: "**Actives columns + Active rows**" luego seleccione los ejes principales deseados (aquí los ejes 1 y 2). Haga Clic luego sobre: **display**
 - Aparece una ventana para seleccionar el plano factorial según el par de ejes deseado.
- Elegir los ejes 1 y 2 (elección por defecto) luego haga Clic sobre: **display**. Es posible mostrar en los planos sólo algunas variables. Haga Clic entonces sobre: **Manual**

Selection of points Seleccionar las variables y trasladarlas en la segunda ventana haciendo Clic sobre: **select**

Recordar que: para cada gráfico, la banda contiene opciones:

- "Save" guarda el gráfico en formato bmp;
- "Font" ofrece la posibilidad de modificar la fuente y el color de los caracteres;
- "More graph" permite cambiar de plano factorial;
- "Sign of axes" permite invertir los ejes;
- "Rank", es útil solamente en el caso de pantallas muy complejas, (lo que no es el caso aquí): este botón convierte las dos coordenadas de la pantalla actual en rangos (ver nota de la sección precedente).

- Aparece la ventana del plano factorial. Elegir una opción luego haga Clic sobre: **View**
- Regrese luego sobre: "PlaneView" para seleccionar otra representación factorial. Para cerrar el gráfico, haga Clic sobre: **Return** o sobre la cruz arriba a la derecha, luego sobre: **Return** en la ventana de selección de los ejes principales.

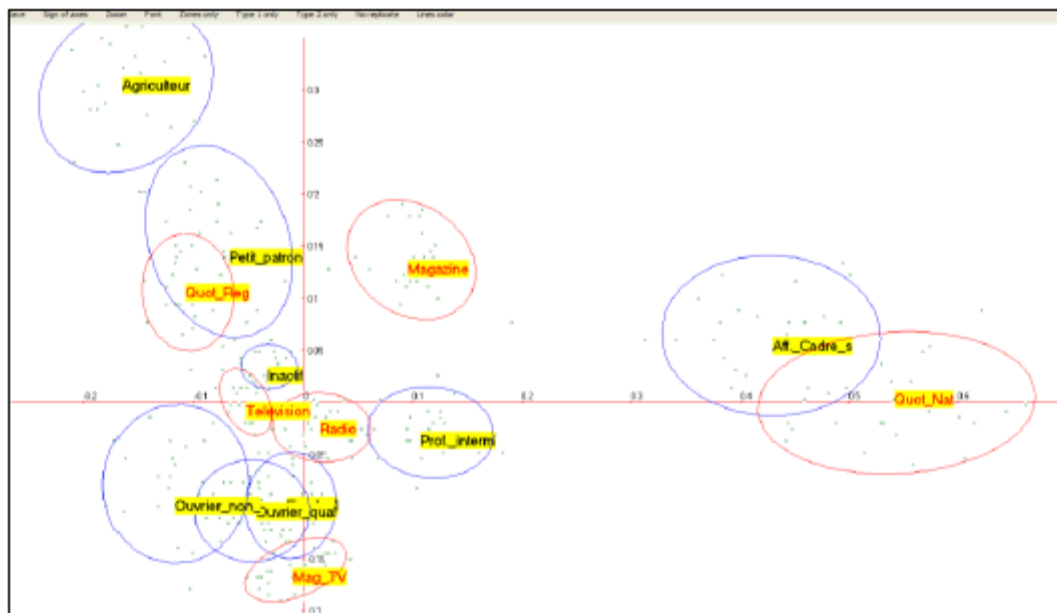
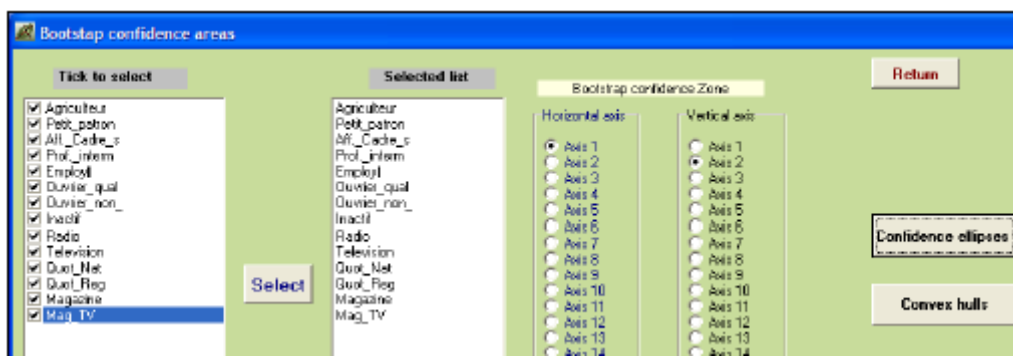


Comentario: también hay en el plano factorial principal, la oposición entre Prensa diaria Nacional y Regional, y también entre Ejecutivos y otras Categorías. Luego, sobre el segundo eje, la oposición entre las revistas TV y otros medios de prensa.

- Regresar luego sobre: "PlaneView" para seleccionar otra representación factorial. Para cerrar el gráfico, haga Clic sobre: **Return** o sobre la cruz arriba a la derecha, luego sobre: **return** en la ventana de selección de los ejes principales.
- Para volver al menú principal de Dtm-Vic, haga Clic o sea en la cruz arriba a la derecha, ya sea sobre "Main menu" o sobre: **return**

3-Validación Bootstrap

- Haga Clic sobre: **B** Bootstrap para validar la posición de las variables en los planos factoriales.
 - Aparece una ventana: "DtmVic - Bootstrap - Validation - Stability - Inférence"
- Haga Clic luego sobre: **Load Data** luego abrir en el directorio el archivo de las replicaciones según el bootstrap escogido. Seleccionamos aquí el archivo **ngus_var_boot.txt** para un bootstrap parcial. Responder: **OK** a la caja de mensaje: "set of principal coordinates loaded" que se exhibe.
- Seleccione ("Tick to select") las variables cuyas elipses se quieren visualizar. Traslados con **Select**, en la ventana "selected list". Elegir luego el plano factorial luego haga Clic en **Confidence ellipses** para la pantalla gráfica de las variables activas (archivo **ngus_var_boot.txt**).



Comentario: las zonas de confianza de los puntos-medios de comunicación (en rojo) muestran que éstos tienen posiciones distintas, es decir perfiles bastante caracterizados de categorías ocupacionales. Por el contrario, ciertos puntos - categorías ocupacionales (en negro) tienen zonas de confianza que se superponen. De esta forma, no se puede concluir que los "ouvriers non qualifiés" (obreros no calificados), los "ouvriers qualifiés" (obreros calificados) y los "employés" (empleados) ocupan posiciones distintas, y entonces que estas categorías tienen perfiles de frecuencias de medios de comunicación distintos.

Para cerrar el gráfico, haga Clic sobre: **return**

II.3. Análisis de Correspondencias Múltiples (ACM o MCA)

El tercer ejemplo (ver directorio: [DtmVic-Examples_A_Start/ EX_A03.MultCorAnalysis](#)) describe un conjunto de variables nominales por el Análisis de Correspondencias Múltiples.

II.3.1. Los datos - Extractos de la encuesta:

"Condiciones de vida y Aspiraciones de los franceses"

Los datos son extraídos de una encuesta por muestreo efectuada por el CREDOC en 1986 sobre "las condiciones y las aspiraciones de los franceses"⁵. Tratan respuestas de una pequeña submuestra de 315 individuos y 49 preguntas. Una primera serie de preguntas se refiere a las características objetivas del encuestado o de su hogar (edad, categoría ocupacional, género, equipamiento...). Otras series de preguntas se refieren a la actitud o a las opiniones de los respondentes sobre la percepción del nivel de vida, la familia, el medio ambiente físico y tecnológico, la salud, la justicia, la sociedad.

En la carpeta [EX_A03. MultCorAnalysis](#) del directorio [DtmVic-Examples_A_Start](#), son contenidos los archivos diccionario y de datos en formato Dtm-Vic:

1. El Archivo diccionario [MCA_dic.txt](#)

8 region	BC01 satisf.log:tres
AA01 region_paris	BC02 satisf.log:assez
AA02 bassin_parisien	BC03 satisf.log:peu
AA03 nord	BC04 satisf.log:pas du tout
AA04 est	4 satisfaction cadre de vie
AA05 ouest	BD01 cdv:tres
AA06 sud-ouest	BD02 cdv:assez
AA07 centre-est	BD03 cdv:peu
AA08 mediterranee	BD04 cdv:pas du tout
9 taille d'agglomeration	5 statut d'occupat logement
AB01 <2000	BE01 accedant
AB02 2001-5000	BE02 proprietaire
AB03 5001-10000	BE03 locataire
AB04 10001-20000	BE04 loge gratuit
AB05 20001-50000	BE05 autre
AB06 50001-100000	6 depenses de logement

⁵ Cf. Lebart L. (1987) - Conditions de vie et aspirations des Français. Evolution et structure des opinions de 1978 à 1984. *Futuribles*, 1, p 25-56. Cf. también: Lebart L. (1986) - Qui pense quoi ? Evolution et structure des opinions en France de 1978 à 1984. *Consommation Revue de Socio-Economie*, Ed. Dunod, 4, p 3-22.

AB07 100001-200000	BF01 negligeeable
AB08 >200000	BF02 sans gros probleme
AB09 paris.agglo.paris	BF03 une lourde charge
2 sexe	BF04 tres lourde charge
AC01 masculin	BF05 ne fait pas face
AC02 feminin	BF06 ne sait pas
0 age
7 situation	4 activite professionnelle
AD01 actif	GB01 plein temps
AD02 etudiant	GB02 temps partiel
AD03 menagere s.prof.	GB03 non activite
AD04 malade invalide	GB04 n'a jamais travail.
AD05 retraite	2 conflits travail-vie person
AD06 militaire	GC01 conflits oui
AD07 chomeur	GC02 conflits non
5 Al-statut matrimonial	2 chomage douze derniers mois
AG01 celibataire	GD01 chomage oui
AG02 marie(e)	GD02 chomage non
AG03 concubinage	2 maux de tete
AG04 separe(e) divorce	HA01 maux de tete oui
AG05 veuf(ve)	HA02 maux de tete non
3 la famille est le seul end	2 mal au dos
AI01 famille:-oui-	HB01 mal au dos oui
AI02 famille:-non-	HB02 mal au dos non
AI03 famille:nsp-nr	2 nervosite
4 opinion sur le mariage	HC01 nervosite oui
AJ01 mariage:indissoluble	HC02 nervosite non
AJ02 mariage:dissout si pb grave	2 etat depressif
AJ03 mariage:dissout si accord	HD01 etat depressif oui
AJ04 mariage:ne sait pas	HD02 etat depressif non
4 travaux/menage/enfants	4 satisfaction sante
AK01 la femme seule	HG01 satisfaction sante:tres
AK02 plutot la femme	HG02 satisfaction sante:satisf
AK03 homme et femme	HG03 satisfaction sante:peu
AK04 tr.femmes:ne sait pas	HG04 satisfaction sante:pas du t
4 satisfaction_logement	0 nombre de personnes logt

El diccionario MCA_dic.txt contiene los identificadores de 49 variables (39 nominales y 10 continuas).

Recordatorio: El identificador de una variable nominal es precedido por el número N de sus categorías (en columna 5). Las N filas siguientes identifican las N categorías de las respuestas: un identificador en 4 caracteres ocupa las columnas 1 a 4 y un identificador largo (de un máximo de 20 caracteres) comienza a la columna 6 [utilizar una fuente a intervalo fijo]. Una variable numérica tal como la edad tiene convencionalmente 0 categoría.
Los espacios vacíos en los identificadores no están permitidos.

2. Archivo de datos (extraídos): MCA_dat.txt

'0005'	8. 1. 2. 27. 3. 2. 7. 1. 2. 3. 1. 1. 2. 2. 2. 2. 2. 3. 0. 0. 1. 1.....	4. 7. 7. 6. 6. 6. 3. 3. 2. 4. 1 3
'0011'	8. 1. 2. 32. 3. 2. 2. 1. 3. 3. 1. 2. 3. 3. 2. 2. 2. 4. 0. 0. 2. 1.....	1. 7. 5. 4. 7. 7. 1. 5. 3. 4. 2 1
'0018'	8. 8. 1. 21. 2. 1. 8. 2. 1. 3. 2. 3. 1. 4. 2. 2. 1. 4. 0. 0. 2. 1.....	4. 7. 7. 7. 5. 7. 3. 7. 2. 4. 1 3
'0024'	5. 1. 2. 42. 1. 2. 3. 1. 2. 3. 1. 2. 1. 3. 2. 2. 2. 2. 1. 2. 2. 1.....	1. 7. 6. 7. 5. 5. 7. 5. 2. 4. 3 1
'0030'	5. 1. 1. 29. 1. 2. 2. 1. 2. 3. 1. 2. 1. 2. 2. 2. 2. 2. 2. 1. 1. 2.....	3. 7. 7. 4. 4. 7. 4. 3. 4. 4. 1 1
'0036'	2. 4. 2. 35. 1. 2. 7. 1. 2. 2. 1. 1. 2. 2. 1. 1. 2. 1. 1. 2. 1. 1.....	4. 7. 7. 5. 6. 7. 5. 5. 2. 4. 2 3
'0042'	2. 4. 1. 71. 5. 2. 8. 1. 3. 3. 4. 2. 3. 2. 2. 2. 1. 3. 0. 0. 2. 2.....	2. 5. 7. 7. 5. 5. 1. 3. 4. 4. 4 3
'0048'	5. 1. 1. 62. 1. 2. 1. 1. 3. 2. 2. 2. 2. 3. 2. 2. 2. 1. 1. 2. 1. 1.....	3. 6. 6. 6. 6. 6. 3. 3. 3. 1. 4 1
'0054'	5. 5. 1. 24. 1. 3. 3. 1. 3. 2. 2. 2. 3. 2. 2. 2. 2. 1. 2. 2. 2. 2.....	4. 7. 4. 7. 5. 7. 4. 3. 3. 3. 1 1
'0060'	4. 1. 1. 52. 1. 2. 3. 1. 2. 3. 2. 2. 2. 2. 2. 2. 2. 1. 2. 2. 1. 1.....	2. 7. 7. 5. 4. 5. 7. 3. 3. 2. 3 1

El archivo de datos contiene 315 filas que corresponden a los individuos respondentes y 50 valores. Para una fila i , el primer valor (entre comilla) corresponde al identificador del individuo i , y los 49 otros valores corresponden a las respuestas de las 49 variables numéricas o a los valores que codifican los ítems de respuesta a las variables nominales, separadas por espacios blancos.

II.3.2. Implementación del ACM

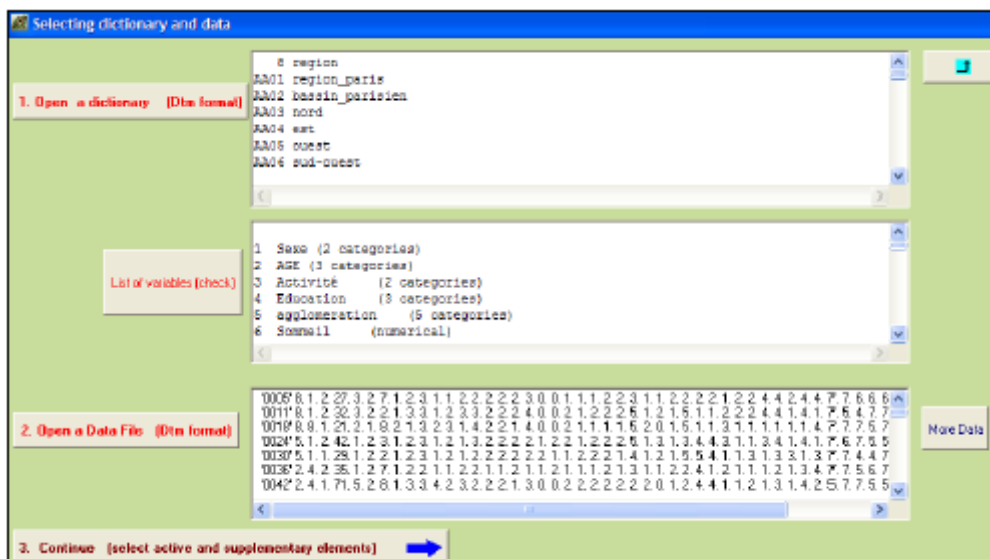
Según el mismo principio de implementación del análisis en componentes principales (ver § II.1.2), el archivo parámetro se crea en 5 etapas:

Etapa 1: selección del análisis

- Haga Clic en el botón: **Create a command file** de la fila: **Command File**
 - Aparece una ventana: "Choosing among some basic analyses".
- Seleccione el análisis: **MCA – Multiple Correspondances Analysis** en la sección **Numerical Data (principal axes techniques)**
 - Aparece una ventana de apertura "Selecting dictionary and data"

Etapa 2: selección de los archivos diccionarios y de datos

- Haga Clic en el botón: **Open a dictionary** en el directorio: **DtmVic-Examples_A_Start / EX_A03. MultCorAnalysis**. Abrir: **MCA_dic.txt**. Este archivo se muestra en una primera ventana. El estado (*categorical* o *numerical*) de las variables es indicado en la segunda ventana.
- Haga Clic en el botón: **Open a Data File** en el directorio **DtmVic-Examples_A_Start / EX_A03. MultCorAnalysis**. Abrir el archivo **MCA_dat.txt** que se exhibe en la tercera ventana.



- Haga Clic en **3. Continue**
 - Aparece una ventana "Selection of active and supplementary elements".

Etapa 3: selección de las variables activas e ilustrativas (o suplementarias)

Dentro de la ventana "*Selection of active and supplementary elements*" se exhiben tres otras ventanas:

- "*Variables to be selected*" donde figura el conjunto de las variables
- "*Active Variables*" que incluye las variables activas seleccionadas
- "*Supplementary Variables*" que incluye las variables ilustrativas seleccionadas

En el marco del análisis de las correspondencias múltiples, las variables activas deben ser nominales (categóricas). Las variables ilustrativas pueden ser continuas o nominales.

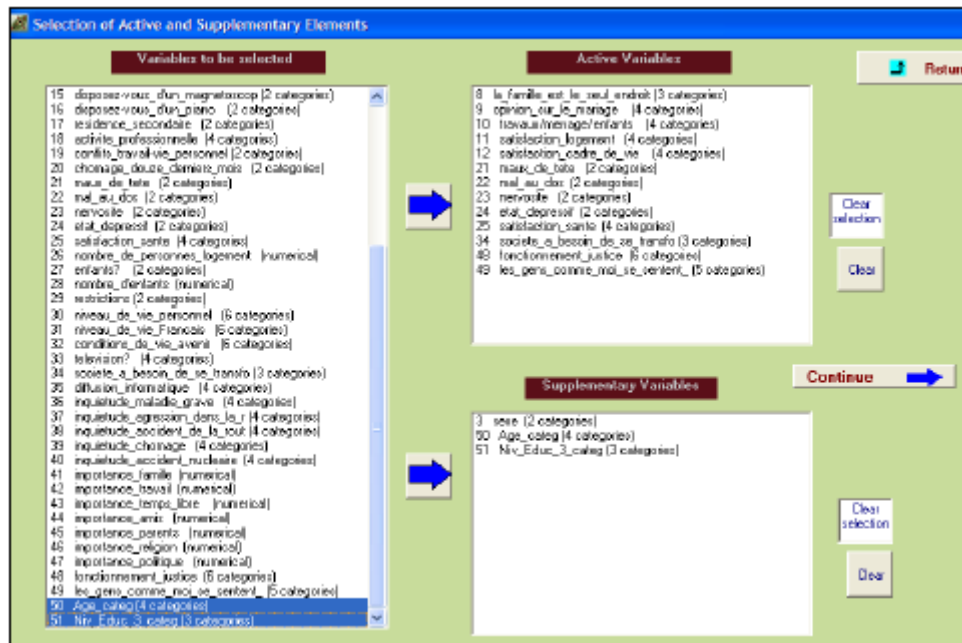
Sugerimos seleccionar las variables siguientes como variables activas e ilustrativas:

- Selección de las variables nominales activas que hay que trasladar en la ventana "*Active Variables*"

8 . la_famille_est_le_seul_endroit_où ...	23 . nervosite
9 . opinion_sur_le_mariage	24 . etat_depressif
10 . travaux/menage/enfants	25 . satisfaction_sante
11 . satisfaction_logement	34 . societe_a_besoin_de_se_transf
12 . satisfaction_cadre_de_vie	48 . fonctionnement_justice
21 . maux_de_tete	49 . les_gens_comme_moi_se_sentent_seuls
22 . mal_au_dos	

- Selección de las variables ilustrativas que hay que trasladar en la ventana "*Supplementary Variables*"

variables nominales ilustrativas	3 . sexe
:	50 . Age_categ
	51 . Niv_Educ_3_categ



- Haga Clic sobre: **Continue** ➔
 - Aparece una ventana: " *Selecting observations*".

Etapa 4: selección de las observaciones (individuos)

Tres casos de figura son posibles:

1. Guardar todas las observaciones
2. Seleccionar las observaciones de una lista
3. Seleccionar las observaciones por un filtro

Aquí tenemos en cuenta todas las observaciones.

- Haga Clic sobre: **All the observations will be active**
 - Aparece una ventana: "*Create a starting parameter file*"

Etapa 5: Creación del archivo parámetro



A esta etapa, es posible seleccionar, como opción, los procedimientos de *bootstrap* y/o de clasificación. Recordemos que en Dtm-Vic los análisis factoriales son sistemáticamente completados por:

- Un *bootstrap* que permite validar las posiciones de las variables.
- Una clasificación con una descripción automática de las clases.

- Haga Clic sobre: **1-Select some options**
 - Una ventana "*Option : Bootstrap and/or Clustering of observations*" aparece

Para recordar los diferentes tipos de bootstrap en Dtm-Vic, ver el cuadro técnico a propósito del ACP, **sección II.1.2, Etapa 5.**

- Haga Clic sobre: "yes" para el procedimiento "bootstrap"; indicar el número de repeticiones (por defecto 25) luego: **Enter** Es el bootstrap parcial que se aplica por defecto. Si no se adopta el bootstrap, haga Clic sobre "no" y directamente pase a la opción de clasificación.

- Seleccione el número de clases deseado (sugerimos 5 clases) Luego haga Clic sobre: **Enter**
- Haga Clic en **Continue** ➔
 - La ventana "Create a starting parameter file" reaparece.
- Haga Clic sobre **2-Create a parameter file for MCA** Un archivo parámetro acaba de ser creado con el nombre **param_MCA.txt** y almacenado en la carpeta **EX_A03. MultCorAnalysis** del directorio **DtmVic-Examples_A_Start**. Para conservarlo con vistas a repetir el análisis posteriormente, tendrá que renombrarlo.
- Haga Clic sobre **3-Execute**

```

Execution completed

--- Computation steps ---
-----

Step ArDat done (building archive dictionary and data)
Step Selec done (selecting active and illustrative elements)
Step Multm done (multiple correspondence analysis (MCA))
Step Recip done (hierarchical clustering: reciprocal neighbours)
Step Parti done (partitioning by cutting a dendrogram)
Step Decla done (description of clusters)

= End of computation step =
-----

[Click about here to hide this Memo]

```

Los procedimientos se muestran en bloque al final de la ejecución.

Comentarios sobre los procedimientos:

ArdaT (Almacenamiento de los datos), **Selec** (Selección de los elementos activos e ilustrativos), **Multm** (Análisis de correspondencias múltiples), **Recip** (Clasificación mixta que utiliza la clasificación ascendente jerárquica, el método de los vecinos recíprocos), **Parti** (Corte del dendrograma y la optimización de la partición por el método de los centros móviles [k-means]), **Decla** (Descripción automática de las clases).

Nota: una vez creado, es posible, en el momento de una utilización posterior de Dtm-Vic abrir el archivo parámetro **param_MCA.txt** en el menú principal con procedimiento **Open an existing command file** luego, ejecutar de nuevo este archivo **Execute**. Los usuarios experimentados pueden modificar parámetros directamente, o con cualquier otro editor de textos después de haber abandonado a Dtm-Vic.

II.3.3 Archivo de los resultados

Los resultados pueden ser consultados en la sección: **Result Files**

- Haga Clic en **Basic numerical results** para navegar en el archivo en formato html luego por **Return** para salir de allí y volver al menú principal.



- haga Clic en **Basic numerical results (.txt format)** para abrir el archivo de los resultados en formato texto.

Ambos archivos "imp.txt" e "imp.html" están incluidos en el directorio **EX_A03. MultCorAnalysis**. Están también guardados bajo el nombre "imp" con la fecha y la hora del análisis. Estos archivos de copia de seguridad guardan los resultados numéricos principales mientras que los archivos "imp.txt / html" sobrescribirán para cada nuevo análisis ejecutado en el mismo directorio.

Después de haber recorrido los resultados numéricos, vuelva al menú principal. Estos resultados son visualizados entonces en la etapa VIC de Dtm-Vic. Esta visualización va a facilitar las interpretaciones.

II.3.4 Visualización de los resultados

Esta segunda fase de Dtm-Vic proporciona las herramientas de visualización necesarias para la interpretación y la validación de los resultados.



1-Ejes factoriales

- Haga Clic en **AxesView**. Una ventana propone visualizar las coordenadas de las variables activas, ilustrativas y de las observaciones, sobre los primeros ejes [ver también la etapa DEFAC del archivo resultados].

En el marco de un ACM, tres elementos pueden ser examinados, *las variables nominales activas e ilustrativas, las variables continuas ilustrativas y las observaciones.*

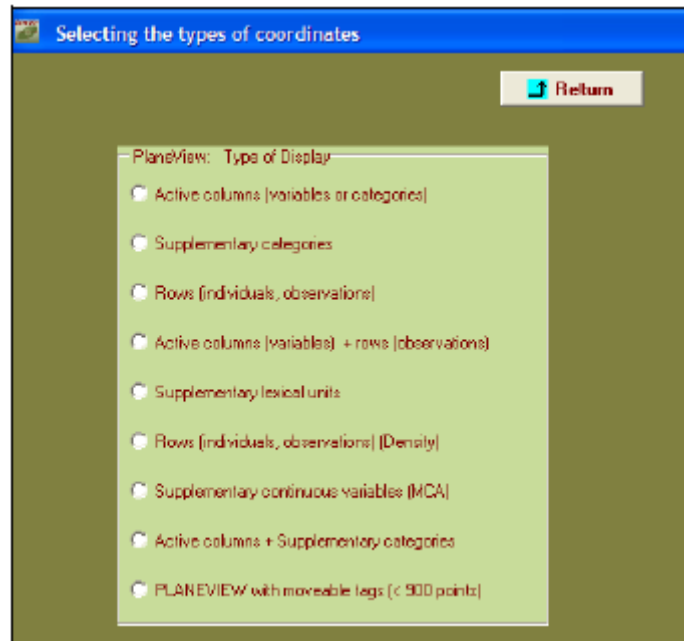
- Haga Clic en la pestaña de los elementos que hay que examinar, por ejemplo **Active variables** luego sobre: **View**. Es posible ordenar las coordenadas de un eje dado, por ejemplo el eje 2, haciendo Clic en "Axis 2".

Active variables	Suppl. Categories	Individuals (observ)	Active variables	Suppl. Categories	Individuals (observ)																																																																																																																								
<p>View Exit</p> <table border="1"> <thead> <tr> <th>Identif</th> <th>axis 1</th> <th>axis 2</th> <th>axis 3</th> </tr> </thead> <tbody> <tr><td>satisfaction_sante_p</td><td>-2256</td><td>-300</td><td>-1250</td></tr> <tr><td>satisfaction_sante_p</td><td>-1370</td><td>122</td><td>898</td></tr> <tr><td>etat_depressif_oui</td><td>-1350</td><td>-317</td><td>-569</td></tr> <tr><td>justice.ne_sait_pas</td><td>-1001</td><td>535</td><td>-716</td></tr> <tr><td>marriage.ne_sait_pas</td><td>-906</td><td>1282</td><td>-698</td></tr> <tr><td>la_femme_seule</td><td>-879</td><td>1442</td><td>626</td></tr> <tr><td>transf_soc.ne_sait_p</td><td>-865</td><td>1383</td><td>-307</td></tr> <tr><td>maux_de_tete_oui</td><td>-785</td><td>-145</td><td>-61</td></tr> <tr><td>solitude:assez_d'acc</td><td>-694</td><td>-363</td><td>17</td></tr> <tr><td>solitude:tres_d'acc</td><td>-651</td><td>-848</td><td>995</td></tr> <tr><td>nervosite_oui</td><td>-640</td><td>-160</td><td>-160</td></tr> <tr><td>mal_au_dos_oui</td><td>-570</td><td>150</td><td>54</td></tr> <tr><td>satis.log.peu</td><td>-358</td><td>-680</td><td>1883</td></tr> <tr><td>justice.refus/repond</td><td>-144</td><td>1110</td><td>-17</td></tr> <tr><td>satis.log.assez</td><td>-128</td><td>-358</td><td>-19</td></tr> <tr><td>plutot_la_femme</td><td>-119</td><td>280</td><td>54</td></tr> <tr><td>marriage.dissout_si_p</td><td>-83</td><td>50</td><td>-165</td></tr> <tr><td>odre.assez</td><td>-70</td><td>-123</td><td>231</td></tr> <tr><td>famille_oui</td><td>-63</td><td>184</td><td>288</td></tr> </tbody> </table> <p>Coordenadas (x 1000) de las variables nominales activas</p>			Identif	axis 1	axis 2	axis 3	satisfaction_sante_p	-2256	-300	-1250	satisfaction_sante_p	-1370	122	898	etat_depressif_oui	-1350	-317	-569	justice.ne_sait_pas	-1001	535	-716	marriage.ne_sait_pas	-906	1282	-698	la_femme_seule	-879	1442	626	transf_soc.ne_sait_p	-865	1383	-307	maux_de_tete_oui	-785	-145	-61	solitude:assez_d'acc	-694	-363	17	solitude:tres_d'acc	-651	-848	995	nervosite_oui	-640	-160	-160	mal_au_dos_oui	-570	150	54	satis.log.peu	-358	-680	1883	justice.refus/repond	-144	1110	-17	satis.log.assez	-128	-358	-19	plutot_la_femme	-119	280	54	marriage.dissout_si_p	-83	50	-165	odre.assez	-70	-123	231	famille_oui	-63	184	288	<p>View Exit</p> <table border="1"> <thead> <tr> <th>Identif</th> <th>axis 1</th> <th>axis 2</th> <th>axis 3</th> </tr> </thead> <tbody> <tr><td>Age_super_60</td><td>-333</td><td>374</td><td>363</td></tr> <tr><td>leminin</td><td>-204</td><td>-54</td><td>-101</td></tr> <tr><td>Niv_Educ_bas</td><td>-203</td><td>59</td><td>142</td></tr> <tr><td>Age_inf.60</td><td>-85</td><td>104</td><td>87</td></tr> <tr><td>Niv_Educ_moyen</td><td>14</td><td>-64</td><td>-224</td></tr> <tr><td>Age_inf.40</td><td>82</td><td>-14</td><td>-254</td></tr> <tr><td>Age_inf.30</td><td>248</td><td>-347</td><td>-133</td></tr> <tr><td>masculin</td><td>251</td><td>70</td><td>129</td></tr> <tr><td>Niv_Educ_haut</td><td>335</td><td>-65</td><td>-115</td></tr> </tbody> </table> <p>Coordenadas (x1000) de las variables nominales ilustrativas</p>			Identif	axis 1	axis 2	axis 3	Age_super_60	-333	374	363	leminin	-204	-54	-101	Niv_Educ_bas	-203	59	142	Age_inf.60	-85	104	87	Niv_Educ_moyen	14	-64	-224	Age_inf.40	82	-14	-254	Age_inf.30	248	-347	-133	masculin	251	70	129	Niv_Educ_haut	335	-65	-115
Identif	axis 1	axis 2	axis 3																																																																																																																										
satisfaction_sante_p	-2256	-300	-1250																																																																																																																										
satisfaction_sante_p	-1370	122	898																																																																																																																										
etat_depressif_oui	-1350	-317	-569																																																																																																																										
justice.ne_sait_pas	-1001	535	-716																																																																																																																										
marriage.ne_sait_pas	-906	1282	-698																																																																																																																										
la_femme_seule	-879	1442	626																																																																																																																										
transf_soc.ne_sait_p	-865	1383	-307																																																																																																																										
maux_de_tete_oui	-785	-145	-61																																																																																																																										
solitude:assez_d'acc	-694	-363	17																																																																																																																										
solitude:tres_d'acc	-651	-848	995																																																																																																																										
nervosite_oui	-640	-160	-160																																																																																																																										
mal_au_dos_oui	-570	150	54																																																																																																																										
satis.log.peu	-358	-680	1883																																																																																																																										
justice.refus/repond	-144	1110	-17																																																																																																																										
satis.log.assez	-128	-358	-19																																																																																																																										
plutot_la_femme	-119	280	54																																																																																																																										
marriage.dissout_si_p	-83	50	-165																																																																																																																										
odre.assez	-70	-123	231																																																																																																																										
famille_oui	-63	184	288																																																																																																																										
Identif	axis 1	axis 2	axis 3																																																																																																																										
Age_super_60	-333	374	363																																																																																																																										
leminin	-204	-54	-101																																																																																																																										
Niv_Educ_bas	-203	59	142																																																																																																																										
Age_inf.60	-85	104	87																																																																																																																										
Niv_Educ_moyen	14	-64	-224																																																																																																																										
Age_inf.40	82	-14	-254																																																																																																																										
Age_inf.30	248	-347	-133																																																																																																																										
masculin	251	70	129																																																																																																																										
Niv_Educ_haut	335	-65	-115																																																																																																																										

2-Planos factoriales

Esta herramienta propone los planos factoriales separados o superpuestos de las variables activas, ilustrativas, u observaciones.

- Haga Clic sobre: ↕ PlaneView
 - Se muestra una ventana en la que se proponen diferentes visualizaciones.



En este ejemplo de análisis, son posibles seis secciones: "columnas activas (variables, categorías)", "categorías ilustrativas", "filas activas (individuos, observaciones)", "columnas activas + filas activas", "individuos activos (densidad)" y "columnas activas + categorías ilustrativas".

El ítem: "PLANEVIEW with moveable tags" retoma algunas de las secciones precedentes y permite desplazar los puntos sobrepuestos para hacer más legible el gráfico.

- Seleccione: "columnas activas + categorías ilustrativas".
 - Aparece una ventana para seleccionar la dupla de ejes deseados.
- Dejar los ejes 1 y 2 (opción por defecto) luego hacer Clic sobre: display. Es posible hacer permanecer en los planos sólo ciertas variables.
- Haga Clic entonces sobre: Manual Selection of points Seleccionar las variables y trasladarlas en la segunda ventana haciendo Clic sobre: select
 - Aparece la ventana del plano factorial (ver a continuación)

- Para volver al menú principal de Dtm-Vic, haga Clic, según sea la ventana, en la cruz arriba a la derecha, o en **Return**

3-Validación Bootstrap

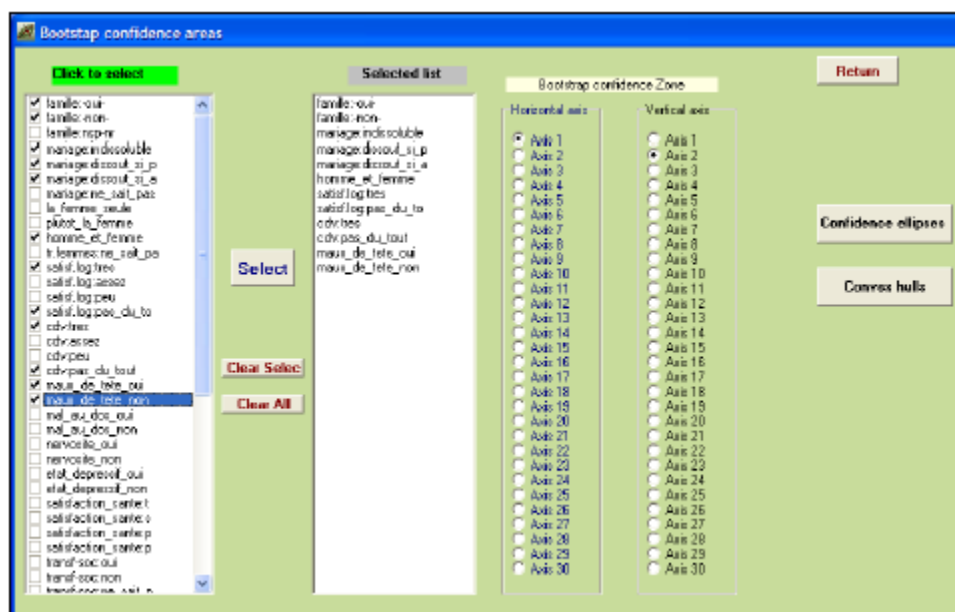
Esta herramienta permite validar la posición de las variables sobre el plano factorial.

- Haga Clic sobre **B** Bootstrap
 - Aparece una ventana "DtmVic - Bootstrap - Validation - Stability - Inférence"
- Hacer Clic en **Load Data** luego abrir en el directorio el archivo de las replicaciones según el bootstrap elegido.
- Seleccione el archivo **ngus_var_boot.txt** para un bootstrap parcial.
Responder **OK** a la ventana "Set of principal coordinates loaded" que se exhibe.

- Luego haga Clic en **Confidence Ellipse**

Se muestra una ventana "Bootstrap confidence areas"

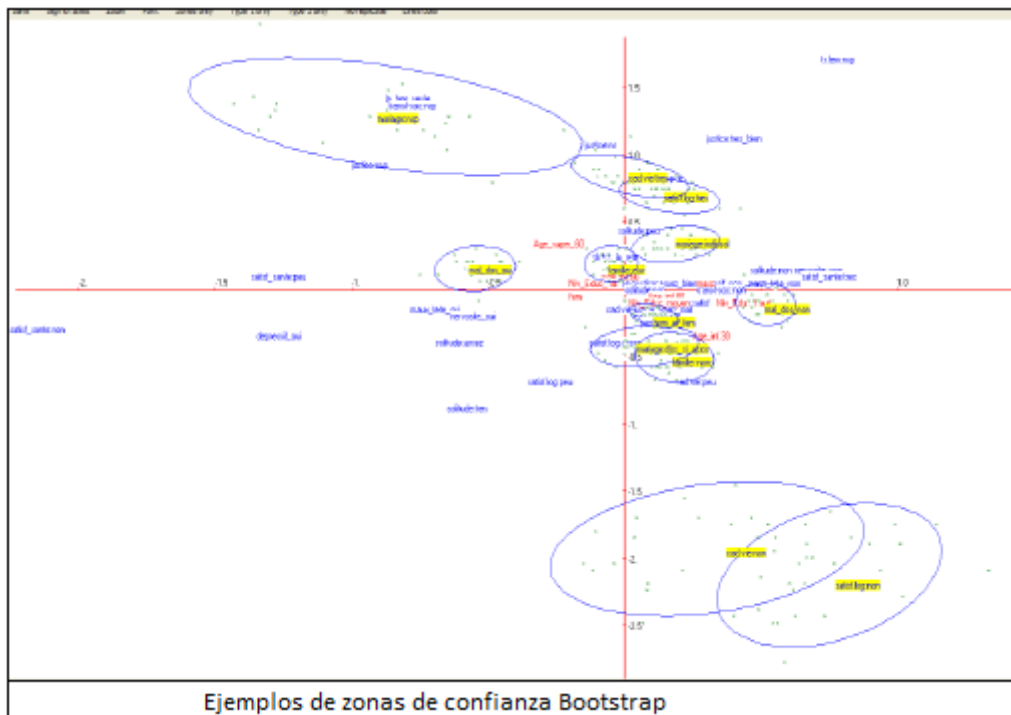
- Seleccione en la sección "Click to select" las variables cuyas elipses se quieren visualizar.
- Trasladarlos con **Select**, en la ventana "Selected list".
- Elegir luego el plano factorial, luego haga Clic en **Confidence ellipses** o sobre **Convex Hulls** para obtener pantalla gráfica de las variables activas (si fue cargado el archivo **ngus_var_boot.txt**), o de la categoría ilustrativa (si fue cargado el archivo **ngus_sup_cat_boot.txt**).



Las elipses de confianza toman en consideración la densidad de la nube de puntos-replicaciones, pero pueden dejar algunos puntos fuera. Para cada punto, la elipse de confianza es calculada a partir de un análisis en componentes principales específico del conjunto de las replications.

Las envolventes convexas (Convex hulls) circundan todas las replications, pero dan el peso a los puntos periféricos sin ninguna consideración de densidad. Ver por ejemplo el capítulo 7 de: *Multiple Correspondence Analysis and Related Techniques* (M. Greenacre and J. Blasius, eds): *Tecnical Validation in Multiple Correspondence Analysis* (L. Lebart). Chapman and Hall, 2006.

- Para volver al menú principal VIC, haga Clic, según sea la ventana, en la cruz arriba a la derecha, o en **Return**



4-Clasificación

Esta opción sitúa las clases obtenidas sobre el plano factorial.

- Haga Clic sobre **ClusterView**. Elija los ejes (1 y 2 para comenzar), y **Continue**
 - Aparece la ventana "DTM-Display of clusters".
- Haga Clic en **View**. Los centroides de las 5 clases aparecen en el plano factorial.
- Accione el botón **Categorical** de la barra. Luego haciendo Clic (derecho) en una clase, aparecen las variables descriptivas de la clase. El conjunto de los resultados figura en el procedimiento DECLA del archivo de los resultados.



Veremos a propósito de los análisis textuales que el mismo procedimiento ClusterView permite mostrar también las palabras características de las clases (para la respuesta de los individuos a una pregunta abierta) y las respuestas características (bajo la forma de texto) de las clases.

Recordatorio: Dtm-Vic produce numerosos archivos de resultados intermedios vinculados a la aplicación (todos en formato .txt) **Se recomienda, por consiguiente, crear un directorio para cada aplicación.** Al principio, tal directorio deber contener los archivos de datos, diccionario y/o texto con el formato Dtm-Vic

III Datos textuales y mixtos: El uso de Dtm-Vic a partir de tres ejemplos

Este capítulo presenta un ejemplo de análisis textual simple y dos ejemplos de análisis elaborados que utilizan a la vez datos numéricos y textuales (Carpeta:

DtmVic_Examples_A_Start de **DtmVic_Examples**).

- El Ejemplo 4, contenido en la subcarpeta **EX_A04.Text-Poems**, realiza un análisis léxico a partir de una serie de textos (poemas): codificación numérica de las respuestas; aplicación del análisis de correspondencias a la tabla léxica que cruza las palabras y los poemas; validación *Bootstrap*; descripción de los poemas por sus palabras y versos característicos; mapa de Kohonen de las palabras y los poemas; seriación.
- El Ejemplo 5, contenido en la subcarpeta **EX_A05.Text-Responses_1**, se refiere al análisis de un juego de datos numéricos y textuales correspondiente a preguntas cerradas y abiertas de una encuesta : tratamiento de las respuestas a una pregunta abierta que utiliza una variable nominal específica para reagrupar las respuestas; codificación numérica de las respuestas; análisis de correspondencias de la tabla léxica en donde se cruzan las palabras y las categorías de individuos; validación *Bootstrap*; descripción de las categorías por sus palabras y respuestas; mapa de Kohonen simultáneo de las palabras y de las categorías.
- El Ejemplo 6 utiliza los mismos datos y diccionario que el ejemplo 5. Es contenido en **EX_A06.Text-Responses_2** siempre en la carpeta **DtmVic_Examples_A_Start**. Procede a un análisis directo de las respuestas a una pregunta abierta, sin reagrupación previa, con clasificación de las respuestas y descripción de las clases a partir de las palabras, de las respuestas características y de las características de los respondentes.

Recordatorio: Dtm-Vic produce numerosos archivos de resultados intermedios vinculados a la aplicación (todos en formato .txt) **Se recomienda por consiguiente, crear un directorio para cada aplicación.** Al principio, tal directorio debe contener los archivos de datos, diccionario y/o texto con el formato Dtm-Vic

III.1 Textos Simples: serie de poemas

Este ejemplo elemental trata la forma más simple de análisis de los textos. Los datos corresponden a una serie de textos compuesta aquí de los 20 primeros sonetos de Shakespeare⁶ En este formato simple, Dtm-Vic puede tratar hasta 1.200 textos sin limitación de tamaño para cada texto. Esta parte tomada como ejemplo es así un "modelo reducido" donde se subrayan solamente las funcionalidades (pero no la potencia) de Dtm-Vic.

III.1.1 El archivo DtmVic: " serie de poemas"

En el marco de un análisis de texto, basta un solo archivo Dtm-Vic que contenga al conjunto de los textos. Nuestro ejemplo es denominado **Sonnet_LowerCase.txt** y esta contenido en el directorio **DtmVic-Examples_A_Start/EX_A04.Text-Poems**.

```

****      S_1
from fairest creatures we desire increase,
that thereby beauty's rose might never die,
but as the ripper should by time decease,
his tender heir might bear his memory:
but thou, contracted to thine own bright eyes,
feed'st thy light'st flame with self-substantial fuel,
making a famine where abundance lies,
thyself thy foe, to thy sweet self too cruel.
thou that art now the world's fresh ornament
and only herald to the gaudy spring,
within thine own bud buriest thy content
and, tender churl, makest waste in niggarding.
pity the world, or else this glutton be,
to eat the world's due, by the grave and thee.

****      S 2
when forty winters shall beseege thy brow,
and dig deep trenches in thy beauty's field,
thy youth's proud livery, so gazed on now,
will be a tatter'd weed, of small worth held:
then being ask'd where all thy beauty lies,
where all the treasure of thy lusty days,
to say, within thine own deep-sunken eyes,
were an all-eating shame and thriftless praise.
how much more praise deserved thy beauty's use,
if thou couldst answer 'this fair child of mine

****      S 20
a woman's face with nature's own hand painted
hast thou, the master-mistress of my passion;
a woman's gentle heart, but not acquainted
with shifting change, as is false women's fashion;
an eye more bright than theirs, less false in rolling,
gilding the object whereupon it gazeth;
a man in hue, all 'hues' in his controlling,
much steals men's eyes and women's souls amazeth.
and for a woman wert thou first created;
till nature, as she wrought thee, fell a-doting,
and by addition me of thee defeated,
by adding one thing to my purpose nothing.
but since she prick'd thee out for women's pleasure,
mine be thy love and thy love's use their treasure.

=====

```

⁶ Para un conjunto más importante de sonetos y los comentarios correspondientes, reenviamos al sitio: <http://www.shakespeare-online.com/sonetos/>.

Los textos pueden tener longitudes muy diferentes, una fila específica separa un soneto de otro. Es caracterizada por separadores "****" continuos de 4 espacios blancos y del nombre del texto. El símbolo "====" indica el fin del archivo. Como todos los archivos de datos en formato Dtm-Vic, éste está en formato "txt". La conversión en minúsculas permite aquí no tratar de forma diferente la primera palabra de cada verso o de cada frase.

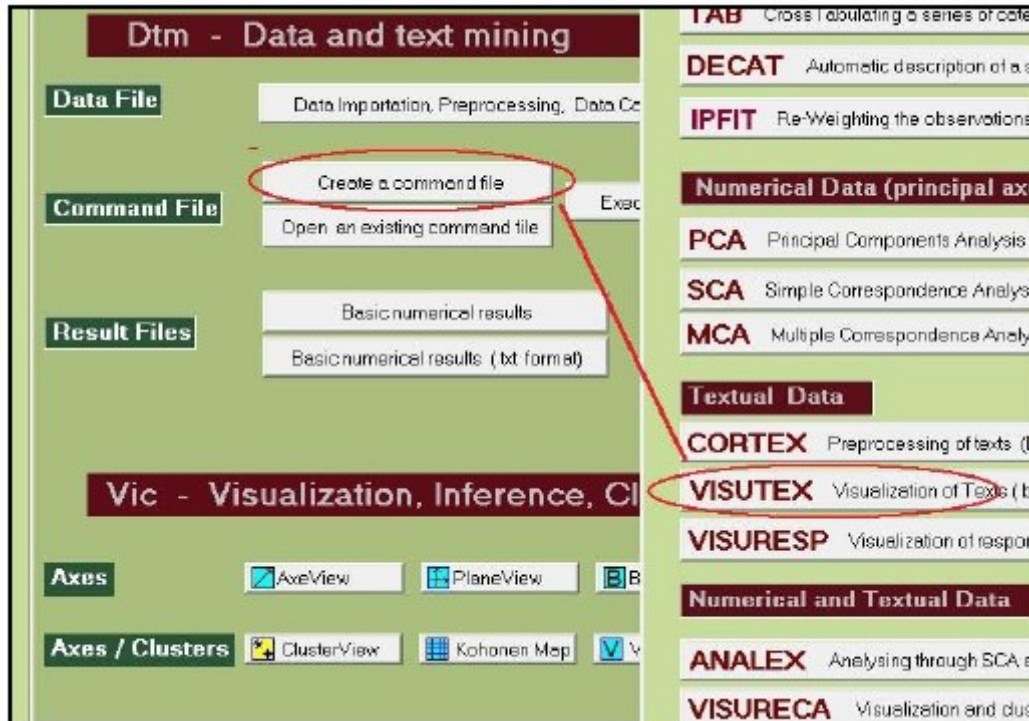
El objetivo es describir los textos a partir de la tabla de contingencia léxica cruzando los textos con las palabras más frecuentes. La metodología general que sirve de base del tratamiento esta presentada en los libros *"Statistique textuelle"* (L. Lebart, A. Salem ; ed. Dunod Paris, 1994), *"Exploring Textual Data"* (L. Lebart, A. Salem, L. Berry ; ed. Kluwer-Springer, Dordrecht, 1998) y *"Análisis estadístico de textos"* (L. Lebart, A. Salem, M. Becue-Bertaut ; ed. Milenio, lleida, 2000) La obra *"Statistique textuelle"* puede libremente ser descargada a partir del sitio: www.dtmvic.com.

III.1.2. Implementación del análisis textual: "VISUTEXT"

El archivo de comando, o archivo parámetro, se crea en 4 etapas:

Etapa 1: selección del análisis

- En la ventana del menú principal, haga Clic en : **Create a ccommand file** de **Command File**
 - Aparece una ventana *"Choosing among some basic analyses"*



- Seleccione el análisis: **VISUTEX – Visualization of texts** de la sección: **Textual Data**
 - Aparece una ventana: *"Opening a text file"*.

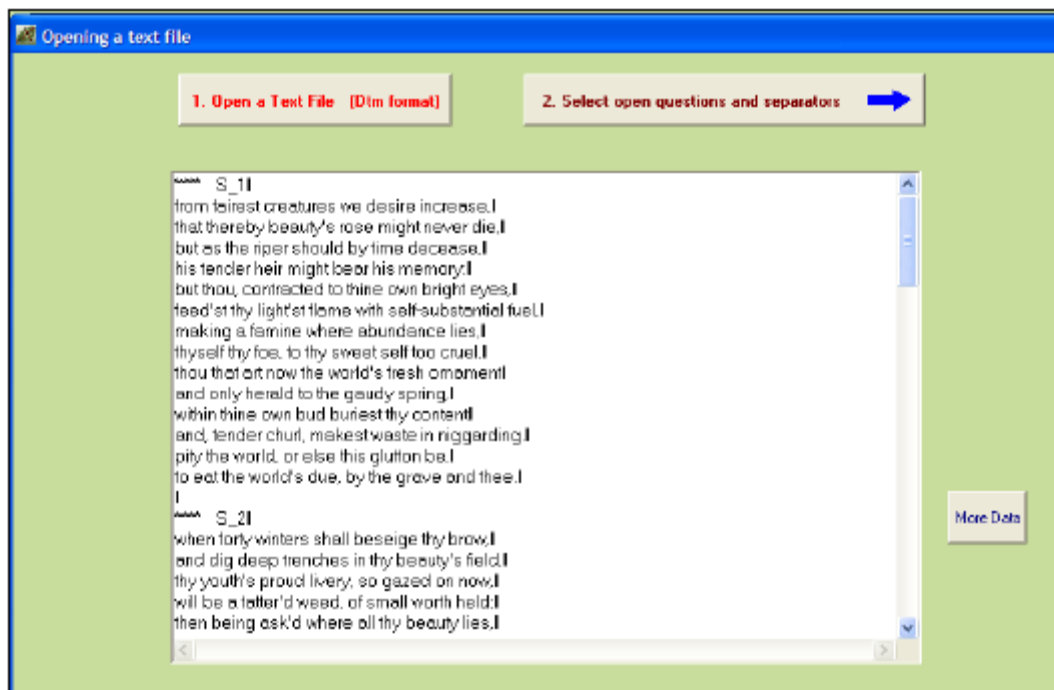
Etapa 2: selección del archivo de texto

- Haga Clic en el botón: **1. Open a text File**

En el directorio **EX_A04.Text-Poems**, abrir el archivo **Sonnet_LowerCase.txt**.

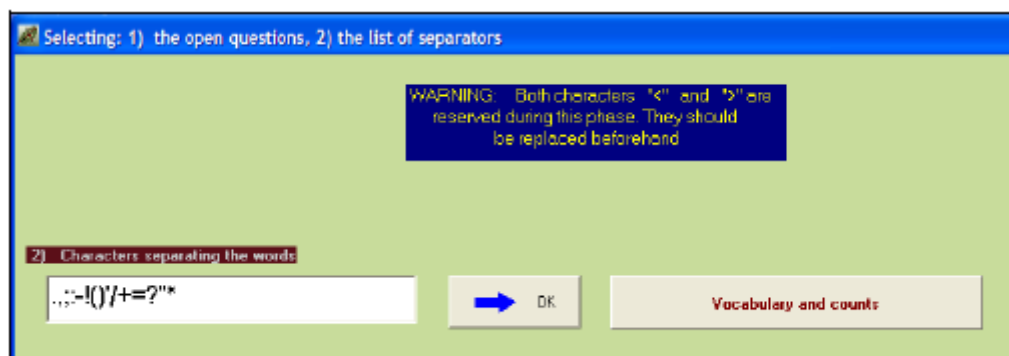
Después de haber hecho Clic sobre: **OK** en la caja de mensaje que da el número de filas y de textos, se exhibe el archivo en una primera ventana.

- Haga Clic sobre: **2. Select Open questions and separators** ➔



Etapa 3: selección de las preguntas, palabras y vocabulario

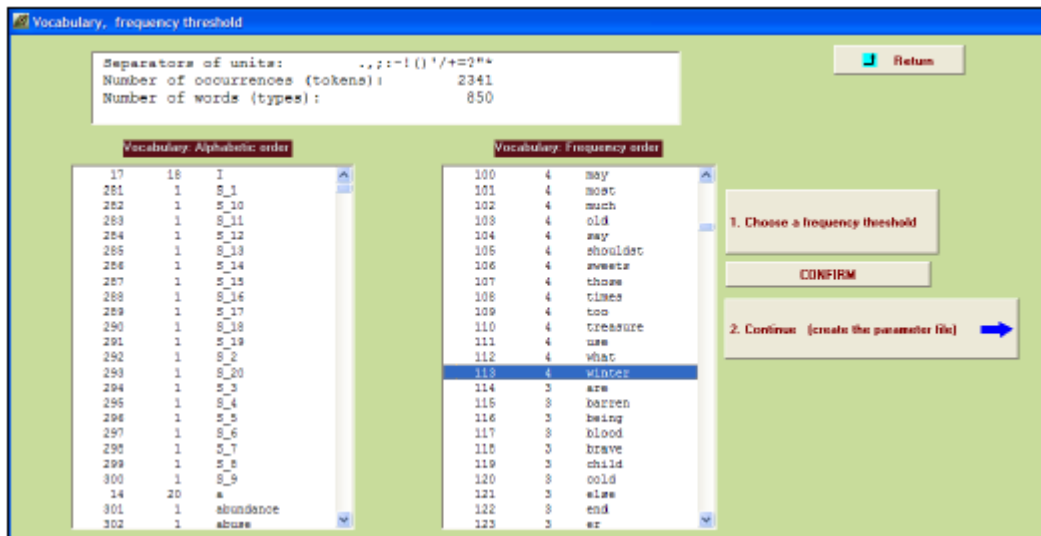
La ventana siguiente permite seleccionar tanto las preguntas abiertas (lo que no es el caso aquí), como completar la lista de los separadores de las palabras.



- Directamente haga Clic sobre: **vocabulary and counts**

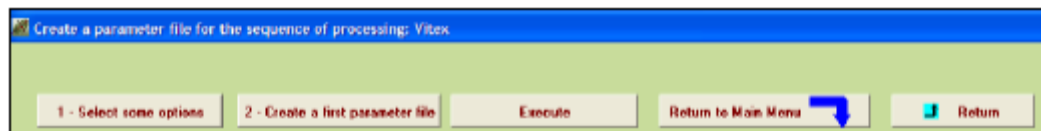
La ventana siguiente presenta el vocabulario (orden alfabético y orden de frecuencia).

Debemos elegir un rango de frecuencia seleccionando una fila en la sección: **Vocabulary: Frequency order**. La fila 113 corresponde a la frecuencia 4 (es una frecuencia pequeña, adaptada a un pequeño corpus. Simplemente se trata aquí de explorar el abanico de los comandos, sin interpretación lingüística pertinente).



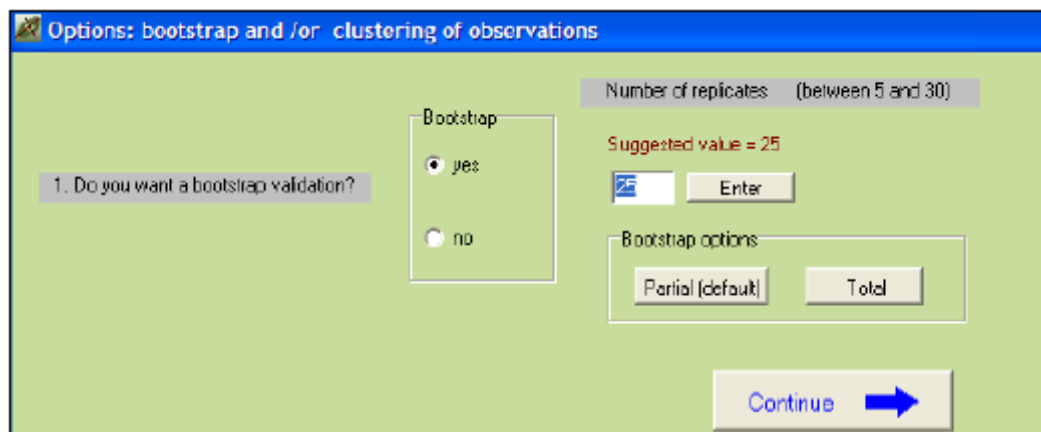
- Seleccione esta fila 113 luego haga Clic en **CONFIRM**. La frecuencia aparece. Responder **OK** a la caja de mensaje.
- Haga Clic en **2. continue (create a parameter file)**

Etapa 4: creación del archivo parámetro

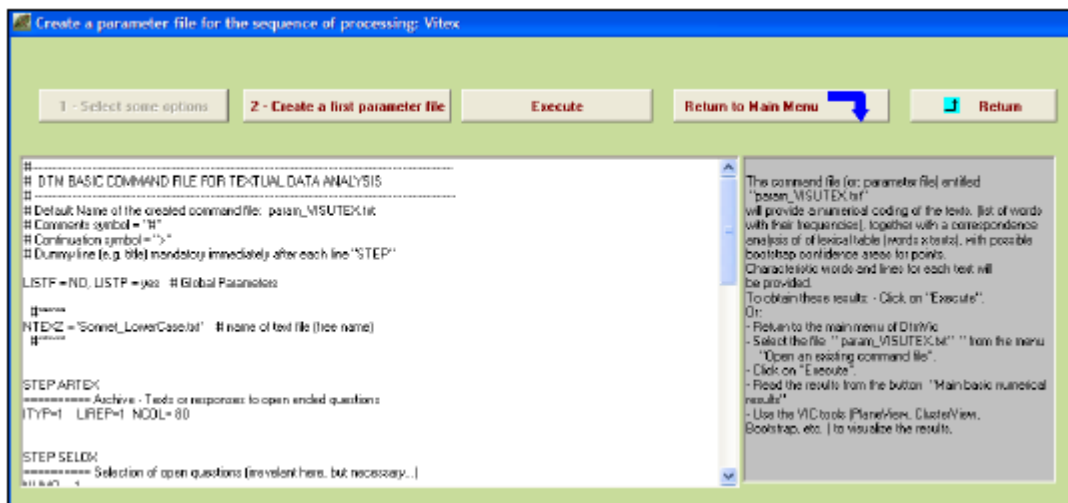


Es en esta etapa de constitución del archivo parámetros que es propuesta la opción *bootstrap* (ver los tres ejemplos precedentes).

- Haga Clic sobre **1-Select some options**
 - Aparece una ventana "Options: Bootstrap and / or Clustering of observations".

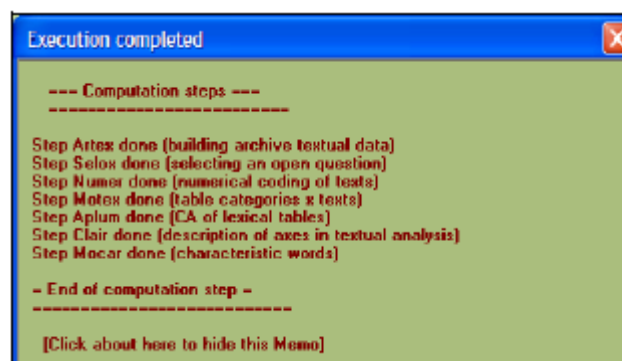


- Haga Clic sobre "yes" para el procedimiento "bootstrap"; indique el número de replicaciones (por defecto 25) luego **Enter**. Si no adopta el bootstrap, haga Clic sobre "no".
- Haga Clic en **Continue** ➔
 - Aparece de nuevo la ventana "Create a parameter file".
- Haga Clic sobre **2-Create a first parameter file**. Un archivo de comando (parameter file) acaba de ser creado bajo el nombre **param_VISUTEX.txt** y almacenado en la carpeta **EX_A04.Text-Poems** del directorio **DtmVic-Examples_A_Start**. (Para conservarlo con miras a análisis posteriores, habrá que renombrarlo).



- Haga Clic sobre **3-Execute**

Los procedimientos se muestran en bloque después de la ejecución: **Artex** (Almacenamiento de los textos), **Selox** (Selección de las preguntas abiertas), **Numer** (digitalización del texto), **Motex** (Tabla de contingencia Palabras-textos), **Aplum** (Análisis de correspondencias para este tipo de tablas), **Clair** (Descripción breve de los ejes factoriales), **Mocar** (Palabras y filas características).



Nota: Una vez que el archivo de comando fue creado (archivo parámetro): **param_VISUTEX.txt**), es posible abrirlo, en el momento de una utilización posterior de DtmVic, en el menú principal **Command File** con botón: **Open an existing command file** luego de ejecutar este archivo: **Execute**. Los usuarios experimentados pueden también modificar los parámetros directamente bajo el editor propuesto por Open (con la ayuda del botón "Help about parameters" disponible en el editor) o con otro editor de texto fuera de Dtm-Vic.

III.1.3 Archivos de los resultados

Los resultados pueden ser consultados en la sección: **Result Files**

- Haga Clic sobre: **Basic numerical results** para navegar en el archivo de resultados en formato HTML, luego sobre: **Return** para salir de allí y volver al menú principal o haga Clic sobre **Basic numerical results (text format)** para abrir el archivo de resultados en formato texto.

Los archivos de los resultados están en el directorio **EX_A04.Text-Poems**.

Recordatorio: El archivo resultado "imp.txt" (como su homólogo "imp.html") está guardado bajo el nombre "imp" seguido de la fecha y la hora del análisis: "imp_18.07.11_14.45.txt" significa el 18 de julio de 2011, a las 14h 45. Este archivo de copia de seguridad guarda como archivos los resultados numéricos principales mientras que las carpetas "imp.txt" e "imp.html" son sobrescritos a cada nuevo análisis ejecutado en el mismo directorio.



```

-----
DtmVic ----- Assignments: --> listf - no. lista - ver
-----
listing of parameters
-----
1 #
2 # DTM BASIC COMMAND FILE FOR TEXTUAL DATA ANALYSIS
3 #
4 # Default Name of the created command file: param_VICDTM.txt
5 # Comments symbol - "*"

```

La lectura de este archivo es útil para informarse de los ciertos resultados que no pueden ser visualizados. El procedimiento NUMER, nos dice, por ejemplo, que la tabla léxica se presenta bajo la forma de 280 respuestas (filas), con un número total de palabras (casos) de 2.321, lo que implica 830 palabras distintas. Al utilizar un umbral de frecuencia de 4, lo que significa que se conservan las palabras de frecuencia superior a tres el número de palabras guardadas se reduce a 1.384, mientras que el número de palabras distintas se reduce a 114.

III.1.4 Visualización de los resultados y la interpretación

Esta segunda fase fundamental de Dtm-Vic proporciona las herramientas de visualización necesarias para la validación y la interpretación de los resultados.

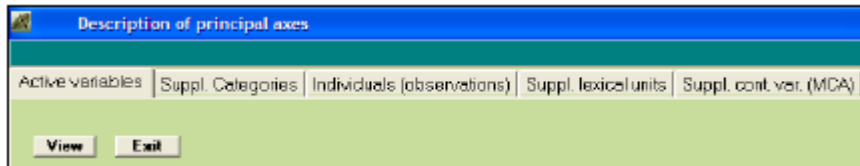


1-Ejes factoriales

Esta herramienta proporciona las coordenadas en los ejes factoriales de las variables activas e ilustrativas, o de las observaciones.

- Haga Clic sobre: AxesView

En el contexto de este análisis textual, solamente dos opciones son factibles: "variables activas" (que corresponden aquí a los poemas) y las "observaciones" (que corresponden aquí a las palabras).




- Haga Clic en la barra de los elementos que hay que examinar, **Active variables** o **Individuals (observations)** luego sobre **View**. Es posible ordenar las coordenadas de un eje dado, haciendo Clic en este eje.
- Haga Clic: **Exit** para salir de esta herramienta

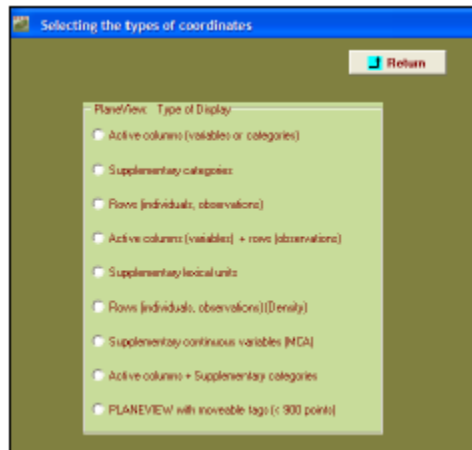
Active variables						Active variables						
Suppl. Categories						Suppl. Categories						
Individuals (observatio						Individuals (observatio						
View						View						
Identif	axis 1	axis 2	axis 3	axis 4	axis 5	Identif	axis 1	axis 2	axis 3	axis 4	axis 5	axis 6
S_1	-263	237	4	214	580	a	316	14	-504	-114	157	321
S_10	-340	-360	273	-9	634	age	83	562	-442	-776	221	-1047
S_11	-321	-158	246	-296	-136	al	-8	463	301	353	-256	64
S_12	68	744	331	370	-563	an	-17	172	-910	-75	783	172
S_13	1402	-799	50	298	-46	and	87	328	90	158	-156	-35
S_14	-61	535	442	465	-17	another	-713	-177	414	-470	212	-686
S_15	574	337	25	104	-239	art	-601	-370	221	578	736	123
S_16	1196	-236	247	-81	119	az	34	418	289	259	-39	-2
S_17	583	-98	-172	108	137	be	-648	-774	-222	143	279	239
S_18	-64	370	20	540	-59	bear	565	-505	832	-615	402	104
S_19	25	319	354	74	4	beauty	-149	68	-423	90	-266	216
S_2	-136	202	-196	381	197	but	250	104	-174	-182	43	187
S_20	-136	-10	-195	50	211	by	-61	293	365	-100	223	270
S_3	-307	34	70	208	381	can	385	740	495	933	-314	319
S_4	-741	-612	-237	-750	-693	change	-114	-203	218	86	810	634
S_5	104	9	-1052	167	-832	d	-72	-35	-486	246	-193	-188
						day	691	686	-59	-28	-488	-391
						decib	1	1201	97	1006	179	692




2-Planos factoriales

Esta opción da los planos factoriales separados o superpuestos de los sonetos (variables activas) y de las palabras (observaciones).

- Haga Clic sobre  PlaneView
 - Se muestra una ventana en la que se proponen diferentes planos factoriales.

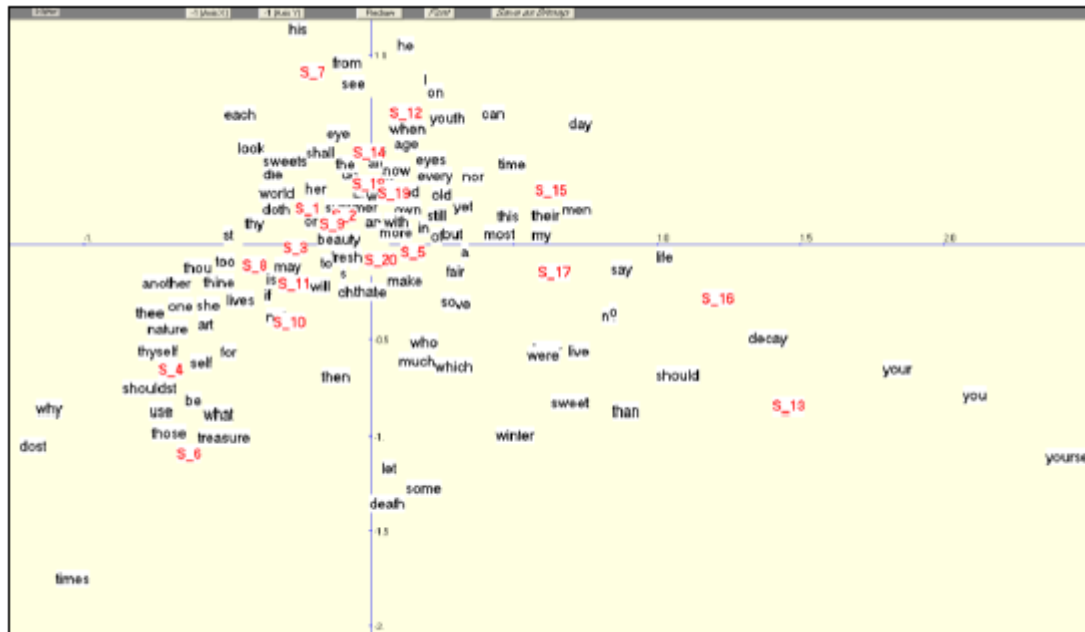
Entre las configuraciones de los planos factoriales propuestos, la opción "active columns + active rows" es apropiada a este análisis.



- Seleccione la sección " **Active columns (variables) + rows (observaciones)**".
 - Aparece una ventana para seleccionar el plano factorial y el par deseado de ejes.
- Elegir los ejes 1 y 2 luego hacer Clic en: . Es posible omitir ciertas variables en los planos. Haga Clic entonces sobre: . Seleccionar las variables y trasladarlas a la segunda ventana haciendo Clic sobre: 
 - Aparece la ventana del plano factorial.

También podemos elegir este menú a través de "PLANEVIEW with moveable tags" que retoma algunas de las secciones precedentes y permite desplazar los puntos superpuestos para hacer más legible el gráfico.

Recordatorio: Para cada gráfico, la barra superior contiene opciones: "Font" ofrece la posibilidad de modificar la fuente y el color de los caracteres; "Sign of axes" permite invertir la orientación de los ejes; "Save" guarda el gráfico en formato bmp; "Rank", es útil solamente en el caso de presentaciones muy complejas: este botón convierte ambas coordenadas de la presentación actual en números enteros (en rangos). Por ejemplo, las n valores de la abscisa son convertidos en números enteros de 1 a n, teniendo el mismo orden que los valores originales. Así ambas distribuciones son uniformes, y los identificadores se superponen menos a costa de una deformación sustancial de la presentación.



Posición de los sonetos y de las palabras en el plano factorial principal.

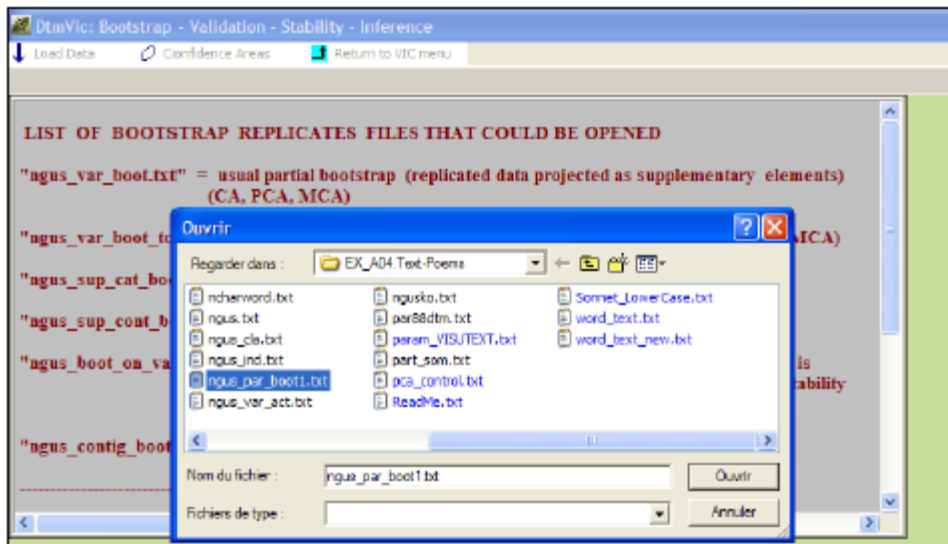
Elegir una opción luego haga Clic sobre: **View**

- Para volver al menú principal de Dtm-Vic, haga Clic sobre: **return**

3-Validación Bootstrap

(Ver el cuadro técnico sobre el bootstrap, chap. II, sección II.1.2, Etapa 5)

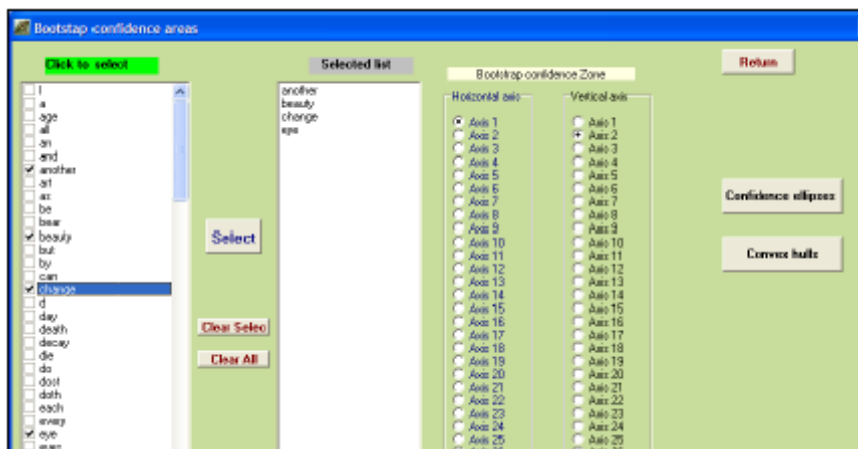
- Haga Clic sobre: **B Bootstrap** para validar la posición de las variables sobre los planos factoriales.
 - Aparece una ventana: "DtmVic - Bootstrap - Validation - Stability- Inférence".
- Haga Clic sobre: **Load Data**, luego abrir en el directorio el archivo de las replicaciones según el bootstrap elegido. Seleccione el archivo **ngus_par_boot1.txt** para un bootstrap parcial en el caso textual.
- Responder: **OK** a la ventana: "Set of principal coordinates loaded" que se exhibe.

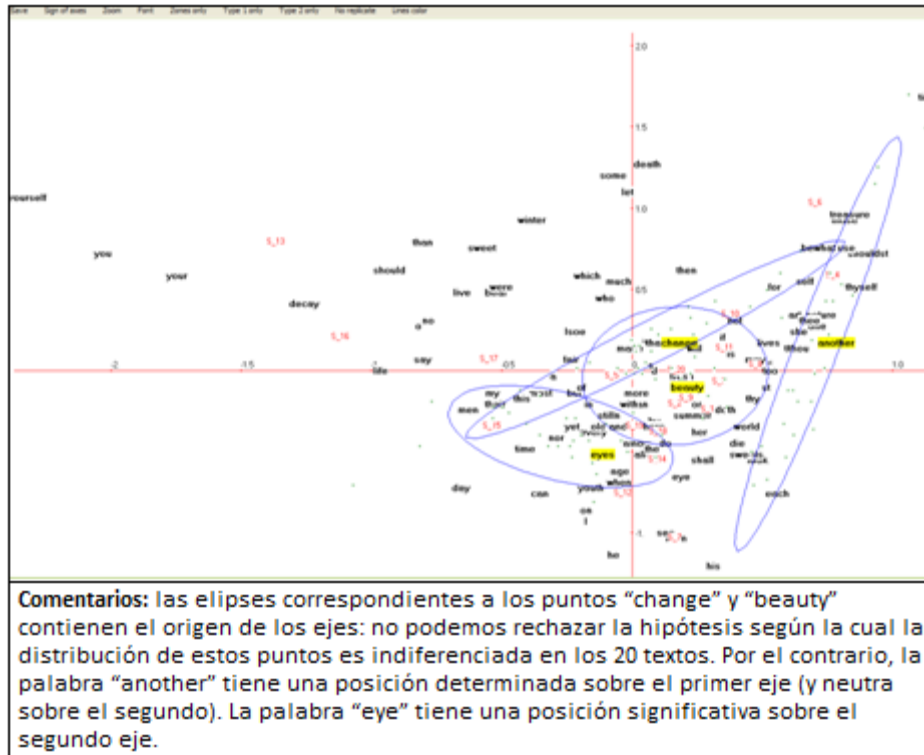


➤ Luego haga Clic sobre: **Confidence Ellipse**


Se muestra una ventana: “Bootstrap confidence areas”.

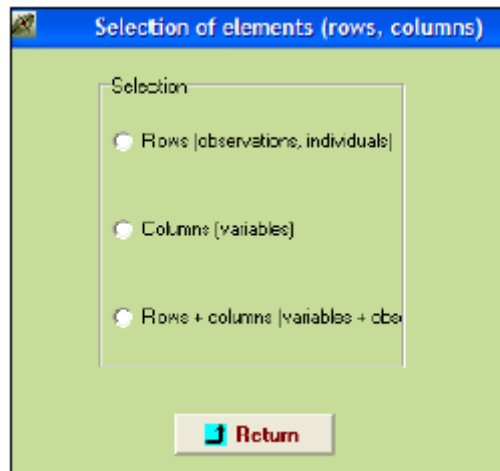
➤ Seleccione en la sección: “Click to select” las variables cuyas elipses se quieren visualizar. Trasladarlas con: **Select**, en la ventana “selected list”. Elegir luego el plano factorial, luego hacer Clic en: **Confidence Ellipse** o en : **Convex Hulls** (ver § II.1.4.3-Bootstrap) para obtener la presentación gráfica de los elementos activos (si fue cargado la carpeta `ngus_par_boot1.txt`)





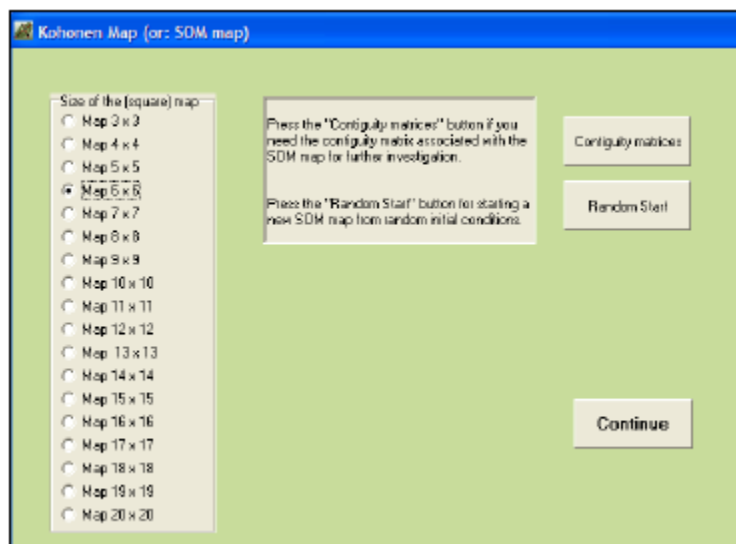
4-Mapas autoorganizados de Kohonen

- Haga Clic sobre  Kohonen Map.
 - Aparece una ventana "Selection of elements"



Las columnas, es decir las variables activas, son los poemas, y las filas es decir las observaciones, son las palabras. Deseamos representar sobre un mismo mapa las palabras y los poemas.

- Haga Clic en "Rows + Columns"
 - Aparece una ventana "Kohonen map"



- Elegir el mapa "map 5x5" luego **Continue** y responder **OK**. a la caja de mensaje: "SOM map completed".
 - Una nueva ventana se exhibe.
- Accione **Draw**. El Mapa de Kohonen aparece.

Obtuvimos una representación simultánea de las filas y de las columnas, debido a la utilización, como archivo de entrada, de las coordenadas del análisis de correspondencia de la tabla léxica. En el marco de este ejemplo, los otros artículos del menú principal no son adecuados.

Observemos que, para todo el análisis presentado, ninguna transformación previa ha sido operada en el vocabulario. El procedimiento CORTEX habría podido preceder el procedimiento VISUTEXT para fusionar palabras (formas gráficas relativas al mismo lema) o para suprimir ciertas palabras (palabras herramientas por ejemplo). No obstante, se aconseja realizar un análisis previo de los materiales brutos.

El uso del software libre "WinTreeTagger" directamente sobre el archivo de texto, seguido por la etapa propuesta en la sección 5.3.2 de este manual permite una lematización del texto.


world thy s own old her an S_9 S_20 S_2 S_1	now fresh die	youth on look his he from age S_7	when day	time see do can as and all S_15 S_12 I
thine much if for	thou she more S_3 S_11	the	yet eyes	this shall say of my men fair S_19 S_18 S_14
use treasure times ten some death be S_6	shouldst not let	so make in	most by	too of is

Extractos del mapa de Kohonen que representan simultáneamente los sonetos y las palabras.
Observación: es posible cambiar de tamaño de la fuente ("Font") y expandir el mapa de Kohonen obtenido ("Dilat") para hacer más legible el gráfico. Las palabras que aparecen en la misma celda son a menudo asociadas con las mismas respuestas (sonetos). Esta propiedad se mantiene en un menor grado, para las celdas vecinas.

5- Seriación

(Ver el recordatorio del párrafo 1.3 del capítulo 1)

La seriación se aplica aquí sobre la tabla léxica que cruza los 20 sonetos y las palabras elegidas (palabras que aparecen por lo menos 4 veces en el corpus)

- Haga Clic sobre  Seriation
 - La ventana "Reordering" aparece.
- Haga Clic en **Reordering the rows and the columns of a word-text table**
- Responder **OK** en "Seriation of rows and columns of the lexical table completed".

Queda constituida la tabla reordenada en filas y en columnas que cruzan los 20 sonetos y las palabras retenidas.

Reordering the rows and columns of a word list table

Help Original table

The rows and columns of the lexical table below have been sorted according to the coordinates on the first axis from the correspondence analysis of the table.

		r_4	r_6	r_8	r_10	r_11	r_3	r_3	r_9	r_7	r_2	r_20	r_18	r_14	r_13	r_12	r_5	r_15	r_17	r_16	r_12	
1	tom	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	does	4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	why	3	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	times	0	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	shoulder	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	thymself	3	2	0	5	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0
7	use	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
8	those	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
9	treasure	0	2	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
10	another	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	what	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
12	one	0	1	4	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
13	hatsun	2	0	0	0	1	0	0	0	0	0	2	1	0	0	2	0	0	0	0	0	0
14	he	3	5	0	0	5	0	2	1	1	0	2	1	0	0	0	0	0	0	2	0	0
15	three	3	4	2	2	1	2	1	2	0	0	3	2	1	1	1	0	0	0	0	0	0
16	best	1	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
17	who	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	thus	5	5	5	4	0	4	2	3	2	3	4	1	3	1	0	0	0	0	0	0	0
19	was	0	2	0	0	0	3	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0
20	where	0	2	0	1	2	2	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0
21	how	1	0	0	4	2	1	0	2	0	0	2	0	0	1	0	1	1	0	0	0	0
22	each	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
23	cup	0	1	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0
24	is	0	0	1	2	0	0	2	0	0	1	0	1	0	0	0	0	0	0	0	0	0
25	lives	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
26	lost	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
27	day	3	1	0	0	0	4	4	1	7	2	1	1	4	1	0	0	0	0	0	0	0
28	may	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
29	where	0	0	0	0	1	3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
30	she	0	0	0	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
31	awake	0	0	2	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
32	not	1	0	2	1	2	2	0	0	0	0	1	2	1	0	1	0	1	0	1	1	0
33	is	0	1	0	2	0	0	1	0	0	1	1	1	0	0	0	0	0	0	1	0	0
34	it	0	0	1	1	1	2	0	1	0	1	0	0	2	0	0	0	0	0	0	0	0
35	both	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
36	the	0	0	0	0	0	2	2	1	7	1	1	2	1	0	0	0	0	0	0	0	0
37	it	0	1	0	2	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
38	the	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
39	will	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
40	will	0	1	0	0	0	0	2	0	0	1	0	0	0	0	0	1	0	1	0	0	0
41	them	2	2	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	0	1	0
42	rather	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0
43	to	4	2	3	5	0	2	4	2	3	2	1	3	4	2	1	2	2	2	2	1	2
44	eye	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	1	0	0	0	0	0
45	beauty	3	1	0	1	1	0	1	1	3	4	0	0	0	0	0	0	0	0	0	0	0
46	fresh	0	0	0	0	1	3	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
47	from	0	0	0	0	2	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0

Comentario: podemos ver (o adivinar si los caracteres son demasiados pequeños) que las primeras palabras de la lista reordenada de palabras caracterizan (a veces en forma exclusiva) los primeros sonetos en la lista, ella misma reordenada de sonetos. Las últimas palabras de la misma lista ordenada están ausentes o raramente observadas entre estos sonetos. Sin embargo, son frecuentes entre los últimos sonetos (lado derecho de la tabla).

El botón: **Original table** permite inspeccionar la tabla léxica para la cual las filas y las columnas tienen su disposición inicial.

III.2. Análisis textual de preguntas abiertas

Este ejemplo describe las respuestas a una pregunta abierta en una encuesta por muestreo en relación con respuestas a preguntas cerradas. Se trata de confrontar los perfiles léxicos de las respuestas de ciertas categorías de respondentes elegidos *a priori*.

III.2.1 Los datos y archivos Dtm-Vic:

"Encuesta internacional sobre las actitudes y los valores"

La encuesta que va a servirnos de ejemplo ha sido realizada en siete países (Japón, Francia, Alemania, el Reino Unido, los Estados Unidos, los Países Bajos, Italia) hacia el fin de los años 80⁷. Presentamos aquí el componente británico de esta encuesta, que designaremos por Encuesta "Life", que trata las respuestas de 1.043 individuos a 14 preguntas cerradas y a 3 preguntas abiertas. Las preguntas cerradas se refieren tanto a las características objetivas del encuestado o de su hogar (edad, categoría ocupacional, género, equipamiento) como a preguntas de las actitudes y a los valores de las personas interrogadas, de las que la mayor parte fueron extraídas del cuestionario de la encuesta "Aspiración" (ejemplo de la sección II.3, ACM).

Se hicieron tres preguntas abiertas:

- "¿Que es lo más importante para usted en la vida?"
- "¿Cuáles son las otras cosas muy importantes para usted?" (Reactivación de la primera pregunta)
- "¿Que piensa usted de la cultura de su país?"

Nos interesamos aquí por las dos primeras preguntas que queremos, más tarde, relacionar con la edad y el nivel de instrucción del encuestado. Se crea una variable nominal con 9 categorías combinando los tres niveles de edad con los tres grados de instrucción.

Este ejemplo está disponible en la carpeta **EX_A05.Text-Responses_1** incluía en el directorio **DtmVic-Examples_A_Start**. Encontramos allí 3 archivos de entrada Dtm-Vic: Diccionario, Datos numéricos y Dato textuales.

Estos archivos en formato Dtm-Vic pueden ser generados por un procedimiento de importación a partir de un archivo Excel único (ver capítulo IV).

⁷ Cf. Hayashi C., Suzuki T., Sasaki M. (1992): *Data Analysis for Social Comparative research: International Perspective*, North-Holland, Amsterdam. El Profesor Chikio Hayashi, fue Director del Instituto de Estadística Matemática (Tokio) y director a cargo de estas encuestas, fue también uno de los primeros descubridores del análisis de correspondencias

1. Archivo de datos para las preguntas cerradas: TDA_dat.txt (extracto)

' ___1'	1	12	80	1	2	3	3	3	2	1	3	3	1	3
' ___2'	1	8	54	1	1	1	3	1	1	1	2	2	1	2
' ___3'	1	6	40	1	1	2	1	2	2	2	2	2	1	2
' ___4'	2	3	27	2	1	2	1	1	1	1	1	4	5	4
' ___5'	2	5	39	2	2	1	3	1	1	1	2	5	5	5
'1039'	1	8	54	2	2	4	2	0	0	1	2	2	2	5
'1040'	2	3	27	2	5	4	2	1	1	1	1	4	5	4
'1041'	1	2	23	3	3	2	1	2	2	1	1	1	3	7
'1042'	1	9	57	2	4	3	1	1	2	2	3	3	2	6
'1043'	2	5	38	1	5	3	5	2	2	2	2	5	4	2

Este archivo comprende 1.043 filas (los individuos) y 15 columnas separadas por espacios en blanco. La primera columna corresponde al identificador del individuo, las otras 14 son los valores de las respuestas a las preguntas cerradas representadas por variables nominales o numéricas continuas.

2. Archivo diccionario de las preguntas cerradas: TDA_dic.txt (extractos)

2 GENDER	EDUM MEDIUM
MALE MALE	EDUH HIGH
FEMA FEMALE	3 WILL_PEOLE_BE_HAPPIER?
12 AGE_CODE	HAP1 Happier
AGE1 18_19	HAP2 LESS_happy
AGE2 20_24	HAP3 About_the_same
AGE3 25_29	4 PEOLE_PEACE_OF_MIND...
AGE4 30_34	PEA1 INCREASES
AGE5 35_39	PEA2 DECREASES
AGE6 40_44	NOT_CHANGES
AGE7 45_49	PEA4 OTHER
AGE8 50_54	3 MORE_OR_LESS_FREEDOM
AGE9 55_59	FRE1 MORE_FREEDOM
AG10 60_65	FRE2 LESS_FREEDOM
AG11 65_70	FRE3 THE_SAME
AG12 71_et_+	3 Age_3_classes
0 AGE	-30 less_than_30
3 EDUCATION	3055 from_30_to_55
EDUL LOW	+ 55 over_55

El archivo diccionario contiene los identificadores de las 14 variables

Recordatorio 1: el identificador de una variable nominal está precedido por el número N de sus categorías (en columna 5). Las N filas siguientes identifican las N categorías de las respuestas: un "identificador corto" (facultativo) de 4 caracteres ocupa las columnas 1 - 5 y un "identificador largo" (de un máximo de 20 caracteres) comienza en la columna 6. Una variable numérica tal como la edad o el número de niños, tiene categoría 0.

Recordatorio 2: no se permiten los espacios vacíos en los identificadores.

3. Archivo de los textos de las preguntas abiertas: TDA_tex.txt: (extractos)

```

----' 1'
  good health
++++
  happiness
++++

----' 2'
  happiness in people around me, contented family, would make me happy
++++
  contented with life as a whole
++++
  education
----' 3'
  contentment
++++
  family
++++
  arts
-----

----1042
  to see my daughter settled in a job
++++
  health, healthy enough to keep them secure, that I get
  on well with my neighbours, a life outside my family circle,
++++
  folk music, architecture, particularly religious
  architecture,
----1043
  contentment
++++
  my children's health and happiness
++++

=====

```

Este archivo contiene las respuestas libres de 1.043 individuos a las tres preguntas abiertas citadas anteriormente. El formato del archivo de los textos es bastante específico, pero transparente para el usuario (format.txt).

Recordatorio sobre el formato interno Dtm-Vic: ya que las respuestas pueden tener longitudes muy diferentes, se utilizan separadores para distinguir las preguntas de los individuos (o respondentes). Los individuos [que deben obligatoriamente estar en el mismo orden que en el archivo de datos numéricos] están separados por la cadena de caracteres "----" (comenzando en la columna 1) seguidos eventualmente por el identificador del individuo. Luego en la fila siguiente, vienen las respuestas a las preguntas abiertas, separadas por "++++" (comenzando en la columna 1). El símbolo "====" indica el fin del archivo. Así como todos los archivos de datos Dtm-Vic, este archivo es un documento de texto bruto (.txt). Si el documento de los textos viene de una fase de procesamiento de textos, debe ser guardado en ".txt"

Después del almacenamiento de los diccionarios, de los datos y textos, la codificación numérica del texto nos permite construir una tabla léxica en la que se cruzan las palabras con una variable nominal seleccionada. Un análisis de correspondencia entonces es ejecutado sobre esta tabla léxica⁸. Zonas de confianza *bootstrap* podrán ser dibujadas alrededor de las palabras y las categorías de individuos.

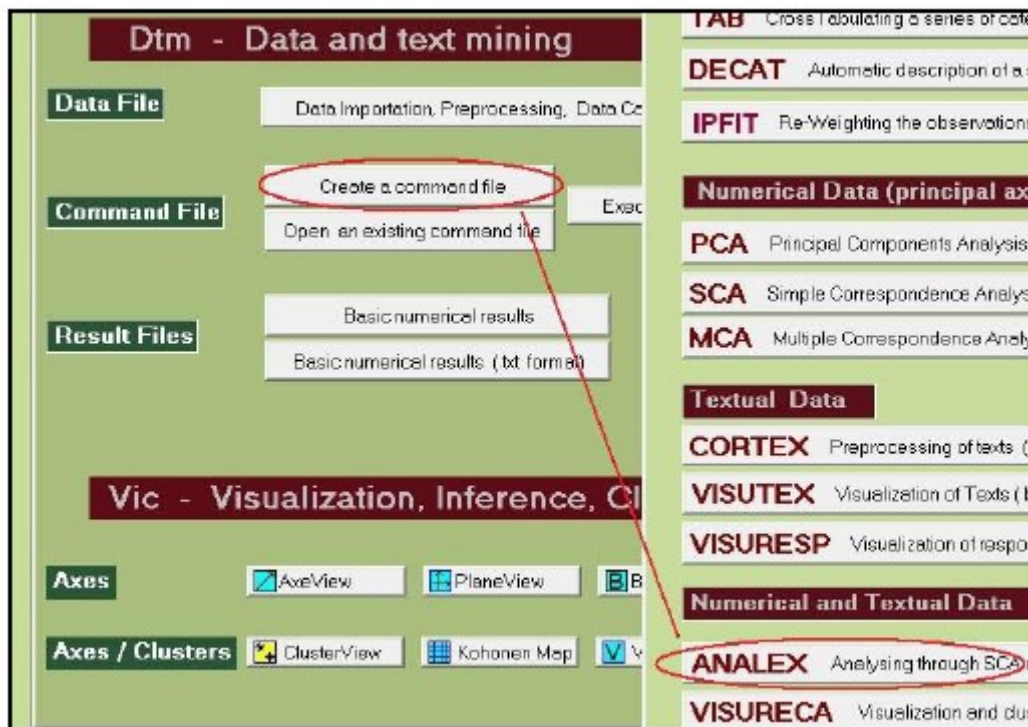
⁸ Explicaciones más amplias a propósito de este ejemplo particular y a propósito de la metodología correspondiente pueden estar encontradas en el libro: "Exploring Textual Data" (L. Lebart, A. Salem, L. Berry; Ed. Kluwer-Springer, Dordrecht, 1998).

III.2.2. Implementación del análisis textual sobre tabla léxica agregada - ANALEX

El archivo parámetro se crea en 5 etapas:

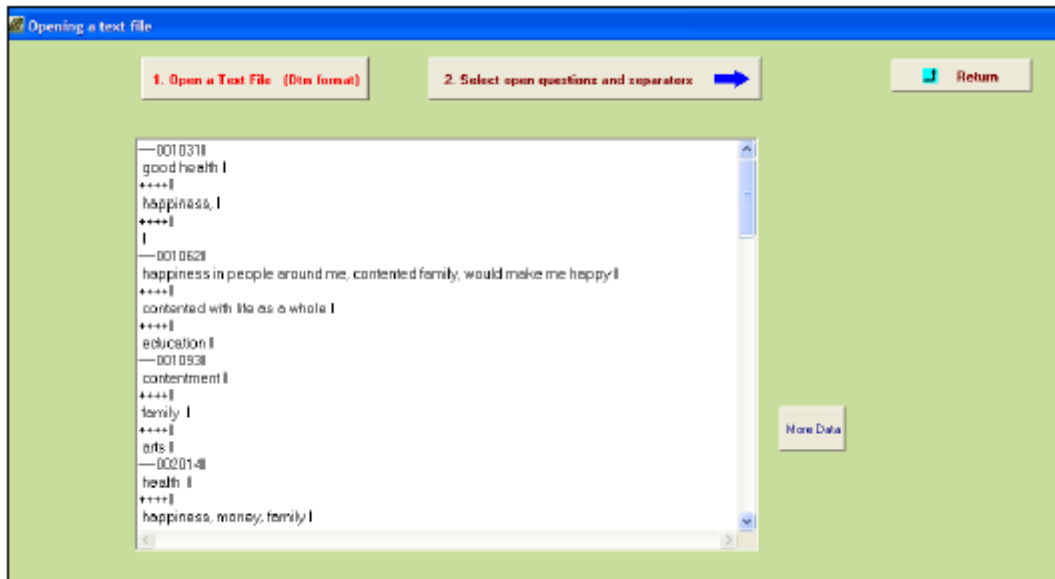
Etapa 1: selección del análisis

- En el menú principal, haga Clic sobre: **Create a command file** de **Command File**
 - Aparece una ventana: "Choosing among some basic analysis".
- Seleccione el análisis **ANALEX – Analysing through SCA of a lexical table built from a specific categorical variable** en la sección **Numerical and Textual Data**.
 - Una ventana: "Opening a text file" aparece.



Etapa 2: selección del archivo de texto

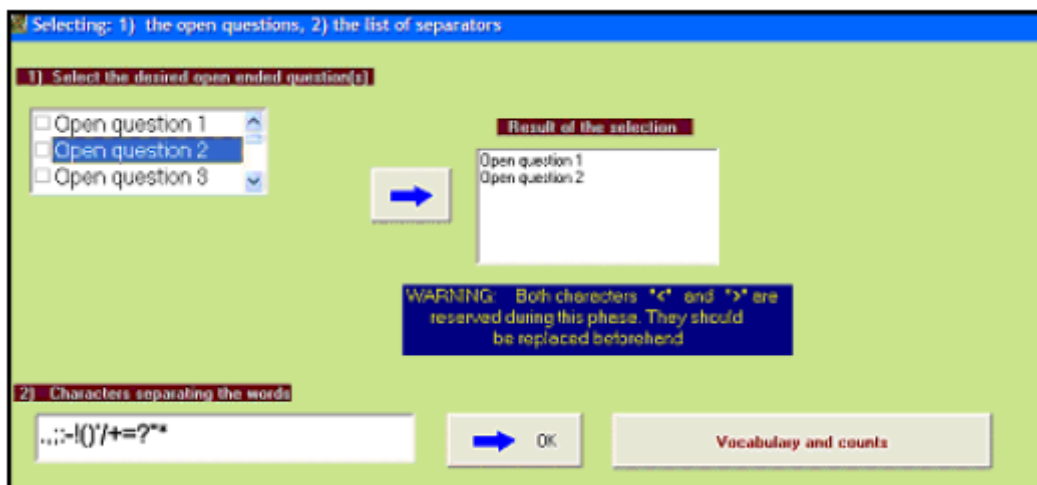
- Haga Clic en el botón: **Open a text File**
- En el directorio **EX_A05.Text-Responses**, abrir el archivo: **TDA _tex.txt**.
- Una caja de mensaje recapitula las informaciones de este archivo: 7.329 filas (correspondientes al conjunto de respuestas de las tres preguntas), 1.043 observaciones (los respondentes) y 3 preguntas abiertas
 - Haga Clic sobre: **OK**, el archivo de texto en formato Dtm-Vic se exhibe en una primera ventana.



- Haga Clic sobre: **2.Select Open questions and separators**
 - Aparece una nueva ventana que tiene para título: "Selecting: 1) the open questions, 2) the list of separators".

Etapa 3: selección de las preguntas abiertas

- Seleccione las preguntas abiertas 1 y 2 y trasladarlas en "Result of the selection". Luego elegir los separadores. Aquí, adoptamos a los valores por defecto. Haga Clic entonces en **Vocabulary and counts**



- La ventana siguiente presenta el vocabulario (alfabético y por orden de frecuencia).

Debemos elegir un umbral de la frecuencia eligiendo una fila en la sección "Vocabulary (frequency order)". La fila 135 corresponde a la frecuencia 16.

- Seleccione esta fila luego: **CONFIRM**. Aparece la frecuencia.

Responder

Vocabulary, frequency threshold

Separators of units: .,;|-|{}'+=7*+
 Number of occurrences (tokens): 13919
 Number of words (types): 1365

Vocabulary, Alphabetic order

666	1	1
667	1	100
668	1	14
669	1	18
257	4	2
470	1	3
471	1	30
472	1	6
9	286	I
473	1	If
474	1	Improving
475	1	Independence
476	1	Indoor
472	2	Ireland
478	2	It
8	300	a
296	5	ability
44	55	able
477	1	abled
70	31	about
398	3	above
474	2	abroad
478	1	absence
475	2	abuse

Vocabulary, Frequency order

132	16	long
133	16	make
134	16	own
135	16	worries
136	15	se
137	15	personal
138	15	relationship
139	15	social
140	14	an
141	14	marriage
142	14	or
143	14	sufficient
144	14	together
145	14	without
146	13	animals
147	13	got
148	13	know
149	13	making
150	13	now
151	13	old
152	13	one
153	13	order
154	13	parents
155	13	religion

1. Choose a frequency threshold

CONFIRM

2. Continue (create the parameter file) →

- Haga Clic en **2. Continue (create the parameter file)** y aparece una ventana de apertura "Selecting Dictionary and Data".

Etapa 4: selección de los archivos diccionario y de datos

- Haga Clic en el botón: **Open a dictionary**. En el directorio **EX_A05.Text-Responses**, abrir el archivo **TDA_dic.txt**. Se exhibe en una primera ventana. El estado (nominal o numérico) de las variables está indicado en la segunda ventana
- Haga Clic en el botón: **Open a Data File**. En el directorio **EX_A05.Text-Responses**, abrir el archivo **TDA_dat.txt** que se exhibe en la tercera ventana.

Selecting dictionary and data

1. Open a dictionary (Dtm format)

2 GENDER
 MALE MALE
 FEMA FEMALE
 12 AGE_CODE
 AGE1 18_19
 AGE2 20_24
 AGE3 25_29

List of variables (check)

1 GENDER (2 categories)
 2 AGE_CODE (12 categories)
 3 AGE (numerical)
 4 EDUCATION (5 categories)
 5 S1_CHANGE_IN_THE_STANDARD_OF_1 (5 categories)
 6 S2_CHANGE_IN_YOUR_STANDARD_OF_1 (5 categories)

2. Open a Data File (Dtm format)

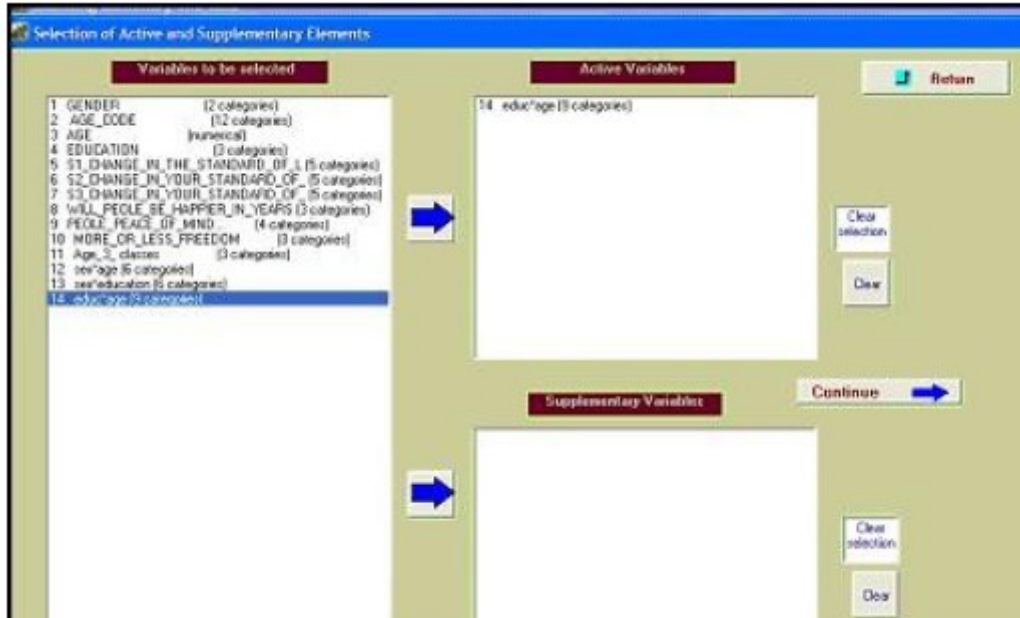
More Data

3. Continue (select active and supplementary elements) →

- Haga Clic sobre: **3. Continue →**
 - Aparece una ventana: "Selection of active et supplementary elements".

Etapa 5: Selección de las variables activas e ilustrativas (o suplementarias).

Dentro de la ventana "Selection of active and supplementary elements" se exhiben tres otras ventanas:



- "Active Variables" que recibe las variables activas seleccionadas
- "Supplementary Variables" que recibe variables ilustrativas.

Para este tipo de análisis, la variable activa y única, es aquella donde las modalidades van a servir para reagrupar las respuestas a las preguntas abiertas. Sugerimos seleccionar la variable nominal número 14 "Educ*age" como variable activa e ignorar las variables ilustrativas. En este caso, las variables ilustrativas podrían servir para describir la variable activa, para completar la etapa "ClusterView".

- Haga Clic sobre: **Continue** ➡
 - Aparece una ventana: "Selecting Observations".

Etapa 6: selección de las observaciones (individuos)

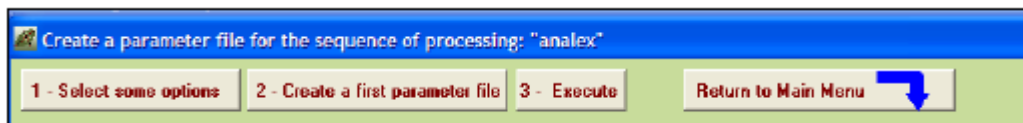
Tres casos son posibles:

1. Guardar todas las observaciones.
2. Seleccionar las observaciones sobre una lista.
3. Seleccionar las observaciones por un filtro.

Consideramos aquí el conjunto de las observaciones.

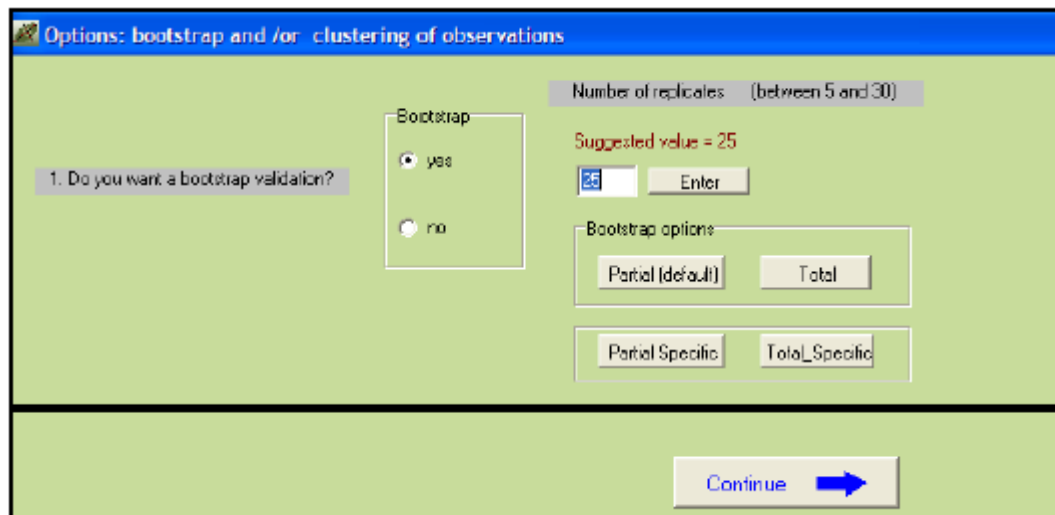
- Haga Clic sobre: **All the observations will be active**
 - Aparece una ventana: "Create a starting parameter file".

Etapa 7: creación del archivo parámetro



En esta etapa, es posible seleccionar, como opción, los procedimientos de bootstrap. Recordemos que en Dtm-Vic, los análisis factoriales pueden ser completados por un *bootstrap* que permite validar la posición de las variables sobre el plano factorial

- Haga Clic sobre **1-Select some options**
 - Se abre una ventana : "Options : *Bootstrap and / or Clustering of observations*"
- Haga Clic sobre "yes" para el procedimiento "bootstrap"; indicar el número de replicaciones (por defecto 25) luego: **Enter**. Es el bootstrap parcial que es aplicado por defecto. Si no se desea el bootstrap, haga Clic sobre "no" y continuar.



- Haga Clic sobre: **Continue** →
 - La ventana: "*Create a starting parameter file*" reaparece.
- Haga Clic sobre: **2-Create a first parameter file**

Un archivo parámetro acaba de ser creado bajo el nombre **param_ANALEX.txt** y almacenado en el directorio **EX_A05.Text-Responses**, del directorio **DtmVic-Examples_A_Start**.

- Haga Clic sobre **3-Execute**

La lista de los procedimientos se exhibe en bloque al final de la ejecución: **Ardat** (Almacenamiento de los datos), **Artex** (Almacenamiento de los textos), **Selox** (selección de las preguntas abiertas), (Selección de los elementos ilustrativos), **Numer** (digitalización del texto), **Motex** (tabla de contingencia Palabras-textos - aquí esta la reagrupación de textos de respuestas según la variable activa seleccionada), **Mocar** (palabras y respuestas características), **Aplum** (análisis de correspondencias para este tipo de tablas), **Sele**c (Selection

de las variables con vistas a la descripción de la variable activa), **Decat** (descripción automática de las modalidades de la variable activa a partir de las variables ilustrativas).

```

Execution completed

--- Computation steps ---
=====
Step ArDat done (building archive dictionary and data)
Step Artex done (building archive textual data)
Step Selox done (selecting an open question)
Step Numer done (numerical coding of texts)
Step Motex done (table categories x texts)
Step Mocar done (characteristic words)
Step Aplum done (CA of lexical tables)
Step Seleoc done (selecting active and illustrative elements)
Step Decat done (description of categories of a nominal var.)

- End of computation step -
=====

[Click about here to hide this Memo]

```

Nota: una vez creado el archivo parámetro **param_ANALEX.txt**, es posible, después de haber dejado a Dtm-Vic, abrirlo de nuevo en el menú principal **Command File** con el procedimiento **Open an existing command file** luego ejecutar este archivo **Execute**. Los usuarios experimentados pueden modificar los parámetros directamente bajo el editor propuesto por **Open an existing command file** o con otro editor de texto fuera de Dtm-Vic (ver el botón "Help about parameters", menú principal y menú del editor de texto interno).

III.2.3 Archivo de los resultados

Los resultados pueden ser consultados en la sección **Result Files** del menú principal.

- Haga Clic en **Basic numerical results** para navegar en el archivo en formato html luego en **Return** para salir de allí y volver al menú principal.

Recordatorio: el archivo resultado "imp.txt" (como su homólogo "imp.html") también se ha guardado bajo el nombre "imp" seguido de la fecha y la hora del análisis. Este archivo de copia de seguridad guarda como archivos los resultados numéricos principales mientras que las carpetas "imp.txt" e "imp.html" son destruidas en cada nuevo análisis ejecutado en el mismo directorio.

DtmVic: Main basic numerical results	
Table of content	
ArDat (building archive dictionary and data)	
Artex (building archive textual data)	
Selox (selecting an open question)	
Numer (numerical coding of texts)	
Motex (table categories x texts)	
Mocar (characteristic words)	
Aplum (CA of lexical tables)	
Seleoc (selecting active and illustrative elements)	
Decat (description of categories of a nominal var.)	
List of commands	

La lectura de este archivo es necesaria para informarse de ciertos resultados que no pueden ser visualizados. Así el procedimiento NUMER nos dice que tenemos 1.043 individuos y 13.919 palabras entre las que están 1.365 palabras distintas. Utilizando un umbral de frecuencia de 16

(lo que significa que se conservan las palabras de frecuencia superior a 16), el número de palabras conservadas se reduce a 10.738, mientras que el número de palabras distintas es disminuye a 136. El libro *"Exploring Textual Data"* (op. cit.) trata los detalles de este tratamiento previo y todos los resultados que siguen.

III.2.4 Visualización de los resultados e interpretación

Esta segunda fase fundamental de Dtm-Vic aporta las herramientas de visualización necesarias para la validación y la interpretación de los resultados.



1-Ejes factoriales


- Haga Clic en  **AxesView**.

Una ventana propone visualizar las coordenadas de las variables activas e ilustrativas y de las observaciones sobre los primeros ejes. En el contexto del análisis textual, solamente dos opciones son posibles: "variables activas" (que corresponden a las categorías) y las "observaciones" (que corresponden a las palabras).

- Haga Clic en la barra de los elementos que hay que examinar **Active variables** o **Individuals (observations)** luego sobre **View**. Es posible ordenar las coordenadas de un eje dado, haciendo Clic en este eje. Haga Clic en **Exit** para salir de esta herramienta.

Active variables					Individuals (observations)							
Identifier	axis 1	axis 2	axis 3	axis 4	Identifier	axis 1	axis 2	axis 3	axis 4	axis 5	axis 6	axis 7
+55/high	-86	279	279	462	e	-112	-52	12	93	-57	56	61
+55/low	305	-111	70	-14	able	-4	-127	87	-114	-27	96	10
+55/medium	114	217	8	-71	about	160	-564	68	-208	-122	126	-68
-30/high	-337	-377	219	-35	alter	541	-79	-261	100	-75	1	59
-30/low	-101	-209	-71	703	all	32	254	7	8	35	-61	-76
-30/medium	-208	-149	-199	-29	and	43	-43	41	9	-29	19	59
30-55/high	-296	104	268	-148	anything	405	-136	197	-128	226	-232	8
30-55/low	39	115	-150	-12	are	317	135	26	-115	224	-171	-14
30-55/medium	-131	177	79	23	as	423	-161	64	-4	79	-14	-45
					at	29	-54	-101	-118	57	-4	-34
					be	64	-104	-54	82	41	-103	-67
					being	-252	-248	37	-71	0	49	-4
					can	455	-259	28	83	23	18	13
					car	-182	-524	26	104	142	162	514
					children	-64	224	-156	-7	171	-114	-20
					church	-50	409	492	-470	-614	405	28
					comlatable	70	-263	81	-146	153	-180	-76

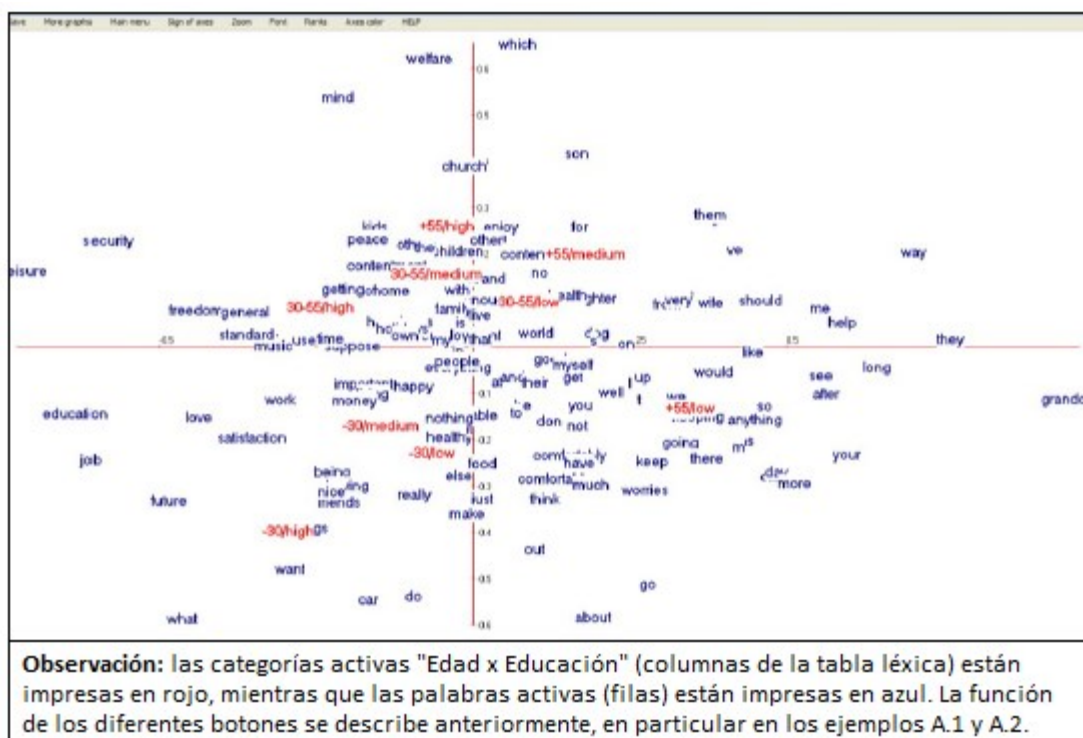
2-Planos factoriales

- Haga Clic sobre  PlaneView
 - Una ventana se exhibe proponiendo diferentes visualizaciones de planos factoriales.


Elegir la sección "**Actives columns (variables) + rows (observations)**", adaptada a este análisis. En efecto, se refiere a filas y columnas de la tabla léxica.

Aparece entonces una ventana para seleccionar el plano factorial que sigue el par deseado de ejes. Elegir los ejes 1 y 2 luego haga Clic sobre **display**. Aparece el plano factorial.

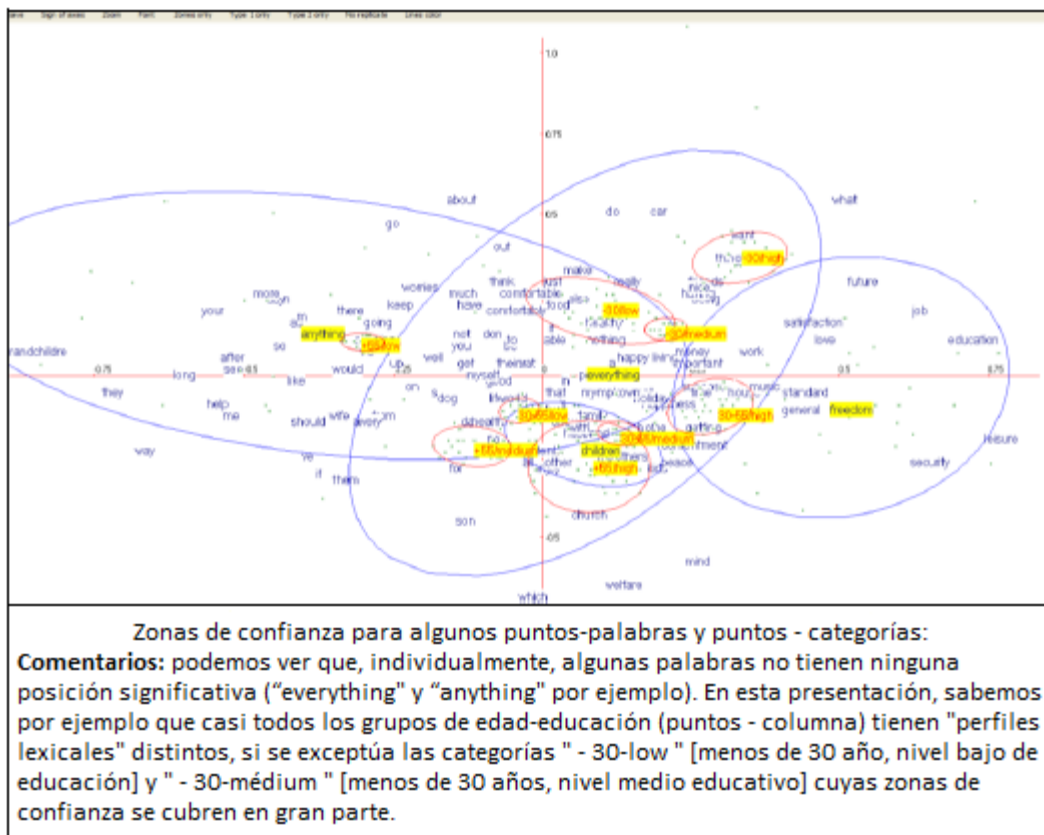
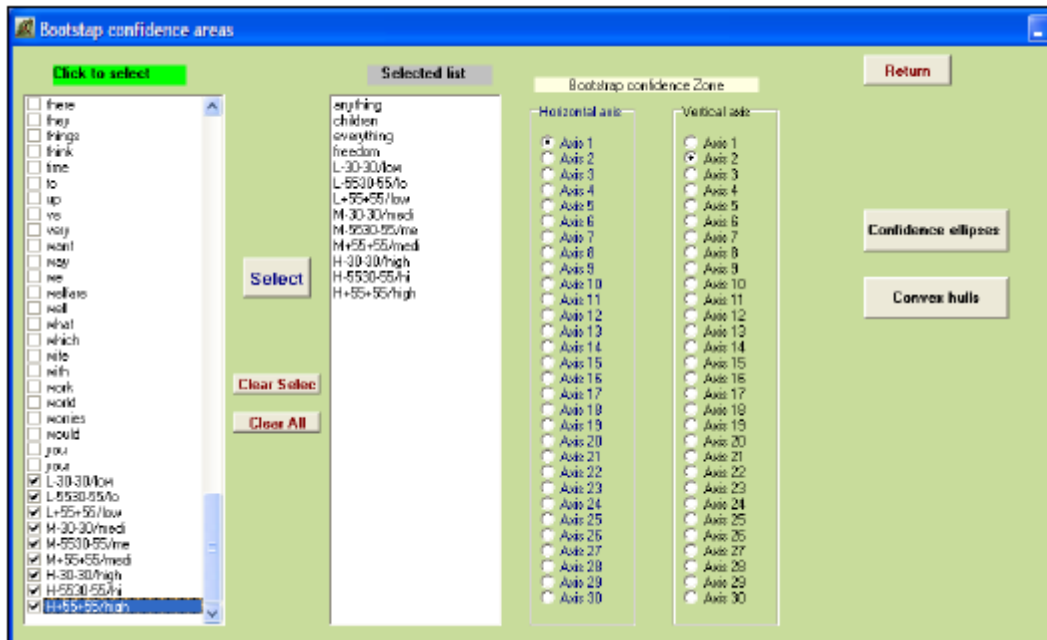
También podemos elegir este menú a través de "PLANEVIEW with moveable tags" que repite algunas de las secciones precedentes y permite desplazar los puntos sobrepuestos para hacer más legible el gráfico.



3-Validación Bootstrap


- Haga Clic sobre:  Bootstrap para validar la posición de las variables sobre el plano factorial.
 - Aparece una ventana: "*DtmVic - Bootstrap - Validation - Stability - Inférence*".
- Haga Clic sobre: **Load Data**, luego abrir en el directorio el archivo de las replicaciones según el *bootstrap* elegido. Seleccione el archivo: *ngus_par_boot1.txt* para un *bootstrap* textual parcial. Responder **OK** a la caja: "*set of principal coordinates loaded*" que se exhibe.
- Luego haga Clic sobre: **Confidence Ellipse**
 - Una ventana: "*Bootstrap confidence areas*" se exhibe

- Seleccione en la sección "Click to select" las variables de las cuales se quieren visualizar sus elipses. Trasladarlas con **Select**, a la ventana "selected list". Elegir luego el plano factorial, luego haga Clic sobre: **Confidence ellipses** o sobre: **convex Hulls** (ver § II.1.4) para obtener la presentación gráfica de las variables activas.

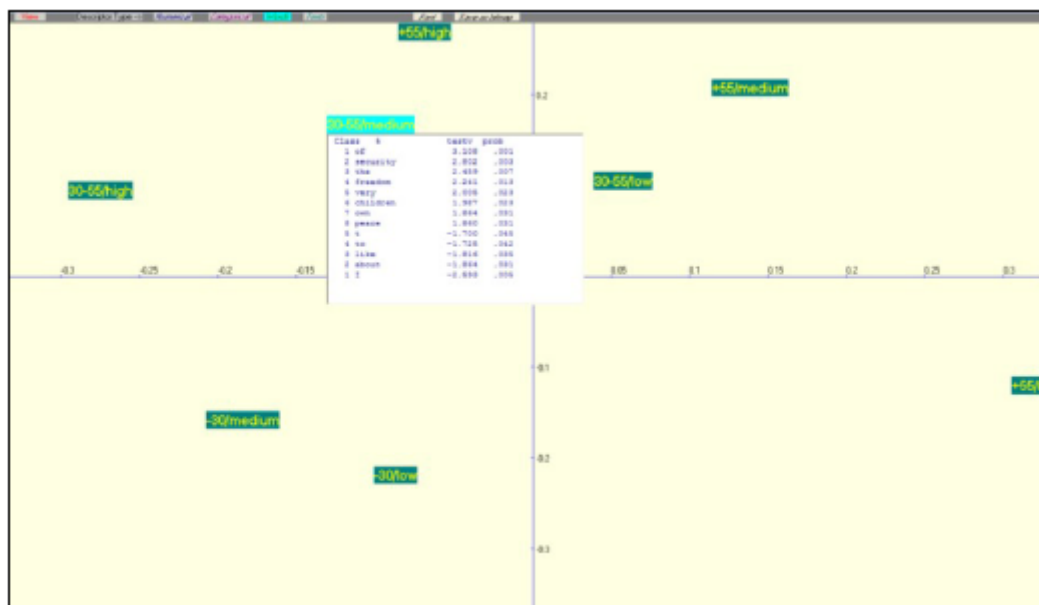


4- ClusterView

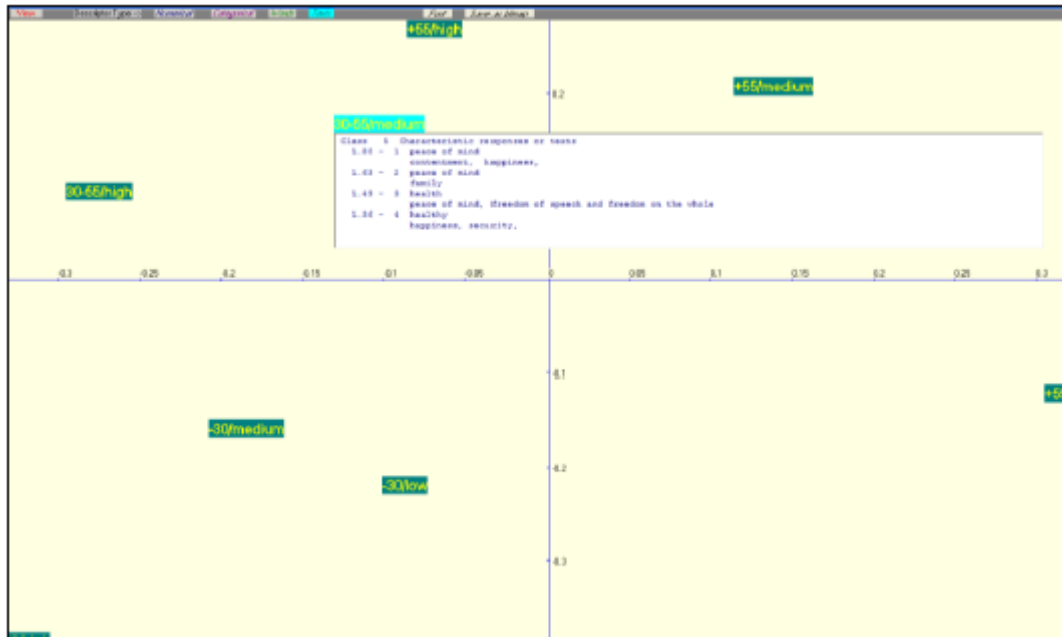
En el caso de ANALEX, no se trata de los resultados de una clasificación, sino de las categorías de la variable activa. Esta opción sitúa las 9 categorías de la variable "14_educ*age" sobre el plano factorial y proporciona las palabras y los textos característicos para cada una de estas categorías.

➤ Haga Clic sobre:  ClusterView. Elija los ejes (1 y 2 para comenzar), y: **Continue**
Se exhibe la ventana del plano factorial. Haga Clic en **View**. La ubicación de las 9 clases aparecen sobre el plano factorial.


➤ Primero active el botón **Words** de la barra. Luego haciendo Clic (derecho) en una categoría, aparecen las palabras descriptivas de la categoría.



➤ Active luego el botón **Texts** de la barra. Luego haciendo Clic (derecho) en una categoría, aparecen los textos descriptivos (respuestas más características o respuestas modales) de la categoría.



5-Mapa autoorganizado: Kohonen map

- Haga Clic sobre  Kohonen Map
 - Aparece una ventana "Selection of elements".
- Haga Clic en "Rows + columns"
 - Aparece una ventana "Kohonen map or SOM map".
- Elegir el mapa "map 5x5" luego **continue** y responder **OK** a la caja de mensaje: "SOM map completed"

Una nueva ventana "Kohonen map" se exhibe

- Active **Draw**. Aparece el Mapa de Kohonen.


Las variables activas son las palabras (en negro) y las observaciones representan las categorías de la variable (en rojo).

what want think things satisfaction nice having future trends do being about -30j/high				suppose security others music love leisure general freedom education 30-55/high
really nothing else -30j/medium		work money kids house happy happiness a	time job important	
out just go comfortable car able	to it healthy comfortably be and	their that my in family everything at 30-55/low	with the living is home holidays getting enjoy children 30-55/medium	standard of contentment
not more make m keep have employment -30j/low	worries up t s myself get don	world son no the health good dog daughter	which live husband for enough content all	welfare peace own other mind
so can	see long after	wife we keeping going are	very them on	people if from +55/high
you food church +55/medium	well way we should our	your there me like grandchildre as anything +55/low	they	would much help day
Observación: es posible cambiar el tamaño ("Fuente") y ampliar el mapa de Kohonen obtenido ("Dilat") para hacer el gráfico más legible				

6- Seriación

(Ver el recordatorio del párrafo 1.3 del capítulo 1)

La seriación es aplicada aquí sobre la tabla léxica que cruza las 9 categorías de respondentes y las palabras elegidas (palabras que aparecen por lo menos 16 veces en el corpus). En esta versión de Dtm-Vic, la seriación puede ser obtenida solamente después de los dos tipos de análisis: VISUTEX y ANALEX. Estos dos enfoques implican el análisis de correspondencia de las tablas léxicas.

- Haga Clic sobre  Seriation.
 - Aparece la ventana "reordering".
- Haga Clic en **Reordering the rows and the columns of tiene a word-text table** Y responder **OK** en "Seriation of rows and columns of the lexical table completed".

Se construye, entonces, la tabla léxica reordenada que cruza las 9 categorías de los respondentes y las palabras escogidas.

Reordering

Reordering the rows and columns of a word-text table Help Original table

The rows and columns of the lexical table below have been sorted according to the coordinates on the first axis from the correspondence analysis of the table

		H-30	H-55	M-30	M-55	L-30	H-55	L-55	M+55	L+55
1	leisure	1	3	5	5	0	0	2	1	0
2	education	4	3	4	6	2	1	4	0	1
3	job	15	17	49	21	3	2	23	3	10
4	security	4	6	6	14	0	1	6	1	2
5	future	2	1	6	3	0	0	3	0	2
6	what	4	2	7	2	0	0	3	0	4
7	freedom	3	4	9	12	0	0	3	3	4
8	love	4	5	7	7	0	1	2	1	6
9	satisfaction	3	3	5	1	0	1	5	1	3
10	standard	3	2	5	9	3	1	5	2	3
11	general	3	4	3	4	0	1	6	1	3
12	music	2	4	2	4	0	0	0	2	4
13	work	11	6	29	15	3	3	36	2	13
14	want	6	2	7	4	0	0	3	2	7
15	house	2	0	10	8	1	0	6	4	3
16	things	5	4	4	3	0	0	5	0	7
17	being	13	9	27	17	2	1	17	5	25
18	time	1	4	7	6	0	0	7	0	6
19	nice	3	3	7	3	1	1	5	0	7
20	friends	17	9	23	18	3	2	7	9	28
21	mind	0	5	6	12	1	5	8	5	5
22	getting	3	2	4	9	1	1	9	1	5
23	suppose	1	3	5	5	0	0	3	1	5
24	having	9	6	15	8	0	1	14	1	16
25	money	9	7	46	28	7	3	34	8	29
26	important	2	2	5	7	0	0	4	0	6
27	contentment	3	2	3	8	0	1	7	2	5
28	peace	5	6	10	19	1	5	10	8	13
29	living	8	4	9	13	2	2	13	3	14

Podemos ver que las primeras palabras de la lista reordenada identifican a las categorías de los jóvenes e instruidos. Las últimas palabras de la misma lista reordenada están ausentes o raramente observadas entre estas categorías. Sin embargo, ellas son frecuentes entre las últimas categorías (lado derecho del cuadro).

Recordatorio: Dtm-Vic produce numerosos archivos de resultados intermedios vinculados a la aplicación (todos en formato .txt) **Se recomienda, por consiguiente, crear un directorio para cada aplicación.** Al principio, tal directorio deber contener los archivos de datos, diccionario y/o texto con el formato Dtm-Vic

III.3. Análisis directo de respuestas libres

Este ejemplo repite el ejemplo precedente y procede a un análisis directo de las respuestas a una pregunta abierta, sin reagrupación previa.

III.3.1 Los datos y los archivos Dtm-Vic:

"Encuesta internacional sobre las actitudes y los valores".

Se trata de la encuesta "Life", componente británico de la encuesta internacional sobre las actitudes y los valores (ver sección precedente III.2.1). Nos interesamos aquí por las dos primeras preguntas que directamente queremos analizar, sin reagrupación previa de los respondientes:

- "¿Que es lo más importante para usted en la vida?"
- "¿Cuáles son las otras cosas muy importantes para usted?"

Queremos detectar cuáles son las variables nominales más vinculadas a las respuestas abiertas, para eventualmente utilizarlas para proceder al reagrupamiento de respuestas (según el procedimiento ANALEX de la sección precedente).

La sección previa III.2 proporciona todas las informaciones necesarias sobre los tres archivos Dtm-Vic de base que van a ser utilizados:

- Archivo de datos para preguntas cerradas: TDA_dat.txt
- Archivo diccionario de preguntas cerradas: TDA_dic.txt
- Archivo de los textos de preguntas abiertas: TDA_tex.txt

III.3.2. Implementación del análisis textual directo de las respuestas - "VISURECA"

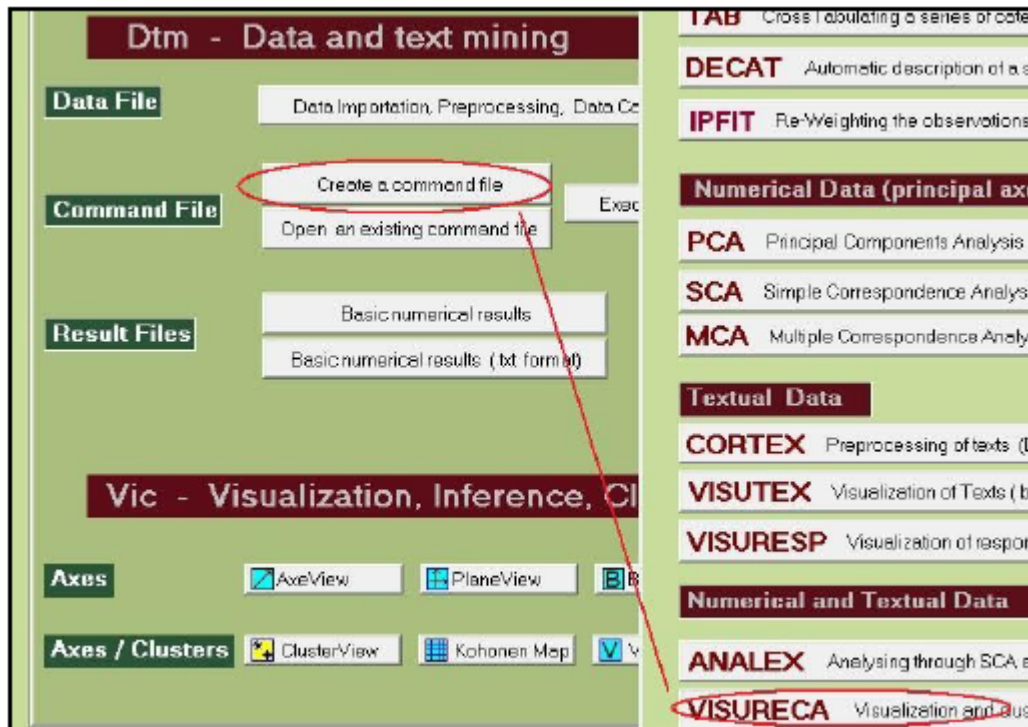
El archivo parámetro se crea en 5 etapas:

Etapas 1: Selección del análisis

- En el menú principal Haga Clic sobre: **Create a command file** de **Command File**
 - Aparece una ventana: "Choosing among some basic analysis".

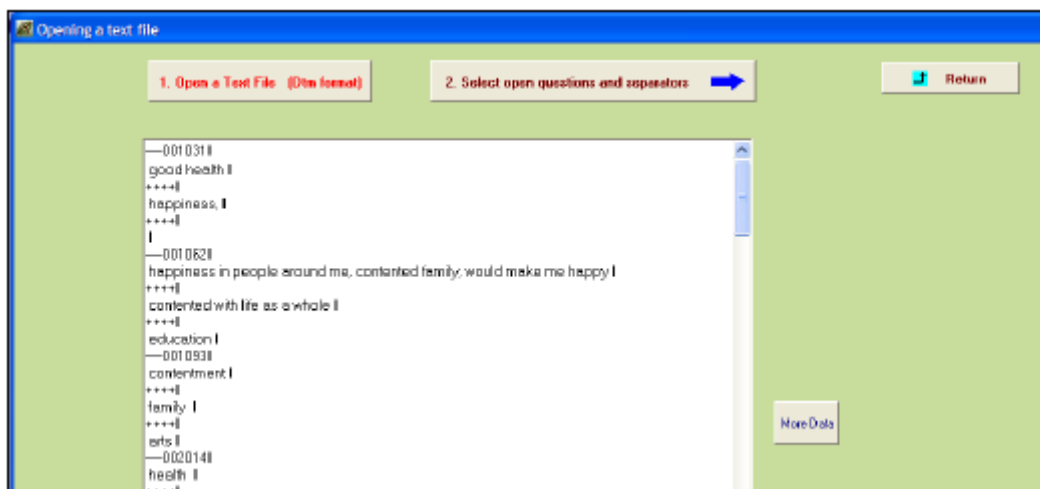
Seleccionar el análisis **VISURECA – Visualization and Clustering of responses with categorical data as supplementary elements** en la sección **Numerical and textual Date.**

- Aparece una ventana: "Opening a text file".



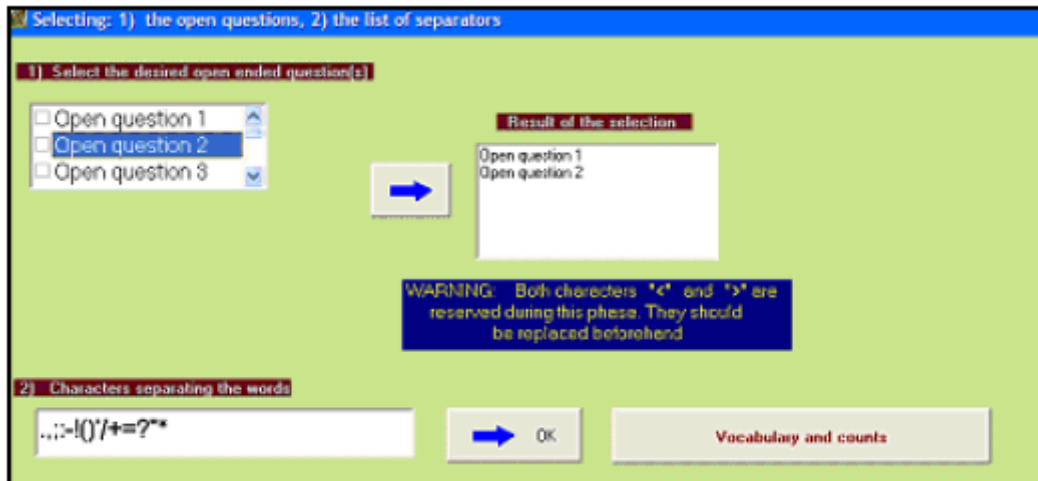
Etapa 2: Selección del archivo texto

- Haga Clic sobre el botón: **Open a text File**. En el directorio **EX_A06.Text-Responses_2**, incluido la carpeta **DtmVic_Examples_A_Start**, abrir el archivo **TDA_text.txt**.
- Una caja de mensaje resume las informaciones de ese archivo: 7.329 filas (correspondientes al conjunto de respuestas de las tres preguntas), 1.043 observaciones (los respondentes) y 3 preguntas abiertas.
- Haga Clic sobre: **OK** el archivo se exhibe en una primera ventana.



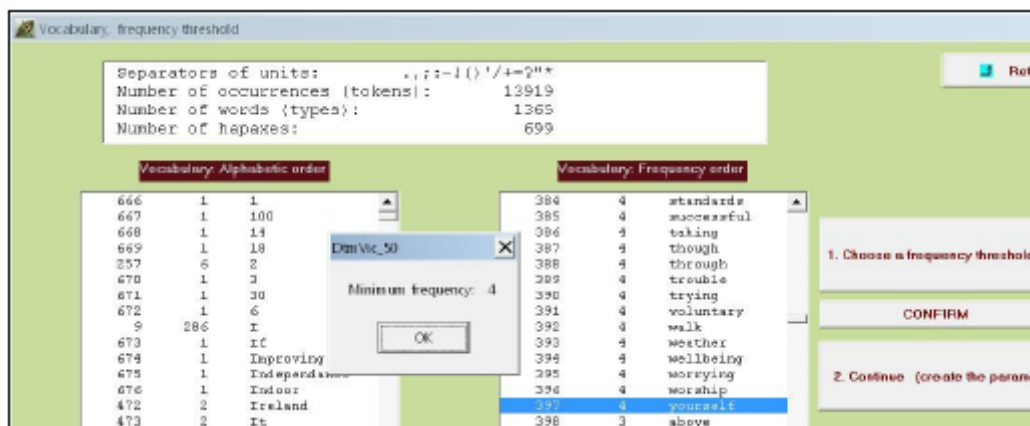
- Un segundo botón aparece: **2.Select Open questions and separators** Haga Clic sobre ese botón.

- Se presenta una nueva ventana: "Selecting: 1) the open questions, 2) the list of separators".



Etapa 3: Selección de las preguntas abiertas

- Seleccionar las preguntas abiertas 1 y 2 y transferirlas a "Result of the selection". Luego elegir los separadores. Acá adoptamos los propuestos por defecto. Haga Clic entonces sobre **Vocabulary and counts**. La siguiente ventana presenta el vocabulario (alfabético y por orden de frecuencia)



Debemos seleccionar un umbral de la frecuencia eligiendo de una fila en la sección "Vocabulary (frequency order)". La fila 397 corresponde a la frecuencia 4. (Anteriormente habíamos elegido un umbral de 16: para respuestas individuales y muy pobres léxicamente, necesitamos más palabras, para no generar demasiadas respuestas vacías después de la elección del umbral). Vamos entonces a guardar las 397 palabras más frecuentes.

- Seleccionar esta fila, luego: **CONFIRM**. Aparece la frecuencia. Responder: **OK**
- Haga Clic sobre **2. Continue (create the parameter file)**.
 - Aparece una ventana de apertura de los "archivos de diccionarios y de datos".

Etapa 4: Selección de los archivos diccionarios y datos

- Haga Clic sobre el botón: **Open a dictionary**. En el directorio **EX_A06.Text-Responses_2**, abrir el archivo **TDA_dic.txt**. Se exhibe en una primera ventana.

El estado (nominal o numérico) de las variables está indicado en una segunda ventana.

- Haga Clic sobre el botón: **Open a Data File**. En el directorio **EX_A06.Text-Responses_2**, abrir el archivo **TDA_dat.txt** que se exhibe en una tercera ventana. La imagen de la pantalla correspondiente es la misma que para el ejemplo II.2.
- Haga Clic sobre: **3. Continue →**
 - Aparece una ventana: "*Selection of active et supplementary elements*"

Etapa 5: Selección de las variables activas e ilustrativas

En el interior de la ventana "*Selection of active et supplementary elements*" se exhiben otras tres ventanas:

"*Variables to be selected*" donde figura el conjunto de las variables

"*Active Variables*": No hay variable activa, ya que aquí el que está activo es el texto de las respuestas. Habíamos elegido de hecho las variables activas seleccionando más arriba las respuestas a las preguntas abiertas 1 y 2.

"*Supplementary Variables*" admite a las variables ilustrativas seleccionadas. Podemos seleccionarlas todas: nos servirán para describir nuestros ejes y nuestras clases.

- Haga Clic sobre: **Continue →**
 - Aparece una ventana: "*Selecting observations*".

Etapa 6: Selección de las observaciones (individuos)

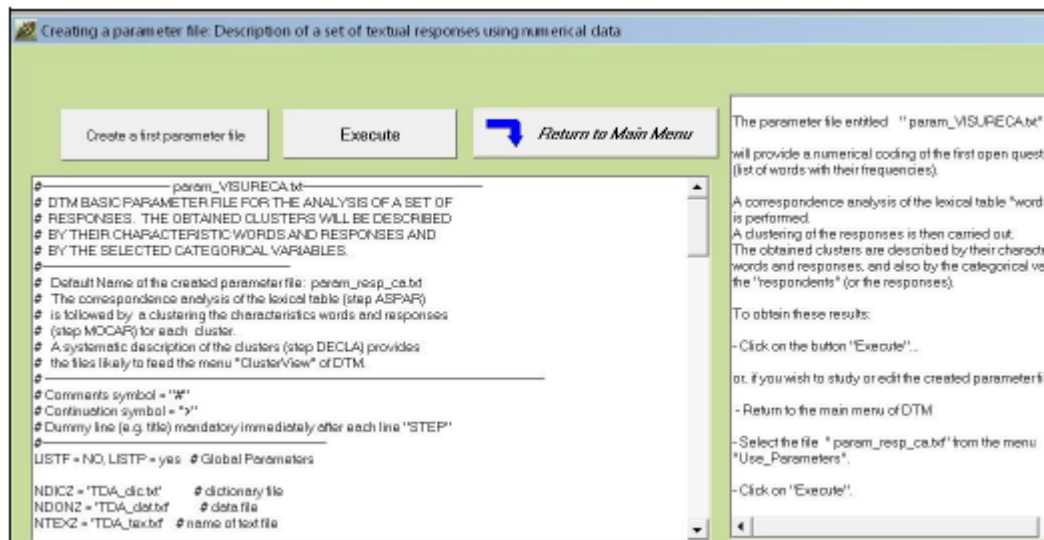
Consideramos aquí el conjunto de observaciones.

- Haga Clic sobre: **All the observations will be active**
 - Aparece una ventana "*Create a starting parameter file*".

Etapa 7: Creación del archivo parámetro

- Haga Clic sobre: **2-Create a first parameter file**

Se creó un archivo parámetro bajo el nombre **param_VISURECA.txt** y se almacenó en el directorio **EX_A06.Text-Responses_2**, del directorio **DtmVic-examples_A_Start**.



Para este tipo de análisis la validación *bootstrap*, por defecto es automática (*bootstrap* parcial). La clasificación es automática, y el número de clases es elegido (por defecto) en función del número de respuestas (aquí 30 clases). [Este número de clase puede ser modificado editando el archivo de comando (o archivo parámetro) antes de la ejecución, parámetros de etapas (STEP) "PARTI" y "DECLA"].

➤ Haga Clic en **Execute**

La lista de los procedimientos se exhibe en bloque al final de la ejecución.

Comentarios sobre las etapas de cálculos:

Ardat (Almacenamiento de datos), **Artex** (Almacenamiento de textos), **Selox** (selección de las preguntas abiertas), (selección de los elementos activos e ilustrativos), **Numer** (digitalización del texto), **Aspar** (análisis directo de las correspondencias de la tabla casi vacías (*sparse*) individuos x palabras), **Recip** (clasificación jerárquica de las respuestas por el método de los vecinos recíprocos), **Parti** (corte del árbol y optimización de la partición obtenida), **Motex** (tabla de contingencia palabras-textos - los textos son aquí reagrupaciones de respuestas según las clases de la partición), **Mocar** (palabras y respuestas características para cada una de las clases), **Selec** (selección de las variables con miras a la descripción de las clases de la partición de individuos), **Decla** (descripción automática de las clases a partir de las variables ilustrativas nominales y continuas), finalmente **Posit** (posición de las variables nominales ilustrativas en los planos factoriales construidos, recordémoslo, con las palabras de las respuestas a las preguntas abiertas como variables activas).

```

Execution completed

=== Computation steps ===
-----

Step ArDat done (building archive dictionary and data)
Step ArTex done (building archive textual data)
Step Selox done (selecting an open question)
Step Numer done (numerical coding of texts)
Step Aspar done (direct CA of texts)
Step Clair done (description of axes in textual analysis)
Step Recip done (hierarchical clustering: reciprocal neighbours)
Step Parti done (partitioning by cutting a dendrogram)
Step Motex done (table categories x texts)
Step Mocar done (characteristic words)
Step Selec done (selecting active and illustrative elements)
Step Decla done (description of clusters)
Step Posit done (positioning categories in textual analysis)

- End of computation step -
-----

```

Presentación de las etapas de cálculo después de la ejecución

Nota: Una vez creado, es posible después de haber dejado Dtm-Vic, abrir de nuevo el archivo parámetro param_VISURECA.txt en el menú principal **Command File** con el procedimiento **Open an existing command file** luego de ejecutar el archivo **Execute**.

Los usuarios experimentados pueden modificar los parámetros directamente con el editor propuesto **Open an existing command file** o con otro editor de texto fuera de Dtm-Vic (ver el botón "help about parameters", menú principal).

III.3.3 Archivo de resultados.

Los resultados pueden consultarse en la sección **Result Files** del menú principal (MP).

- Haga Clic en **Basic numerical results** para navegar en el archivo en formato html y luego sobre **Return** para salir y volver al menú principal (MP).

```

DtmVic: Main basic numerical results
-----
Table of content
ArDat (building archive dictionary and data)
ArTex (building archive textual data)
Selox (selecting an open question)
Numer (numerical coding of texts)
Aspar (direct CA of texts)
Clair (description of axes in textual analysis)
Recip (hierarchical clustering: reciprocal neighbours)
Parti (partitioning by cutting a dendrogram)
Motex (table categories x texts)
Mocar (characteristic words)
Selec (selecting active and illustrative elements)
Decla (description of clusters)
Posit (positioning categories in textual analysis)

List of commands
==== DtmVic ==== Assignments: --> listf = no, listp = yes
listing of parameters
-----

```


Recordatorio: el archivo resultado "imp.txt" (como su homólogo "imp.html") también se guarda bajo el nombre "imp" ordenado por la fecha y la hora del análisis. Este archivo de copia guarda como archivos los resultados numéricos principales mientras que la carpeta "imp.txt" (resp. "imp.html") es sobrescrito por cada nuevo análisis ejecutado en el mismo directorio.

III.3.4 Visualización de los resultados e interpretación.

Esta segunda fase fundamental de Dtm-Vic aporta las herramientas de visualización necesarias para la validación y la interpretación de los resultados.




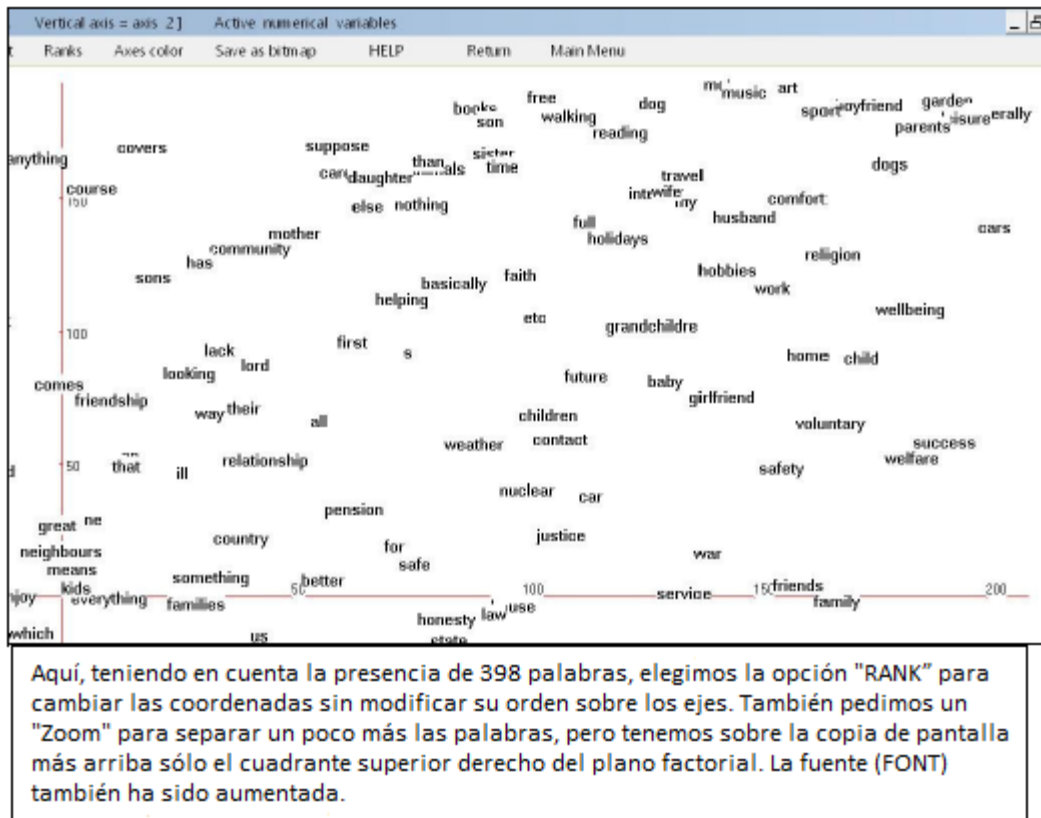
1 – Ejes Factoriales

- Haga Clic sobre  **AxesView**

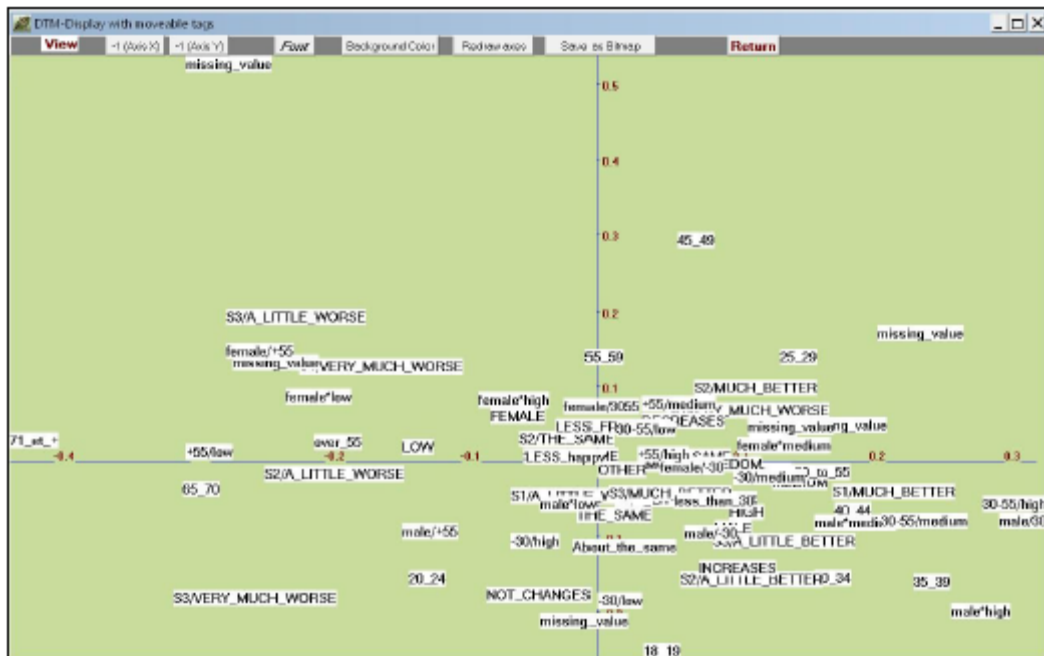
La utilización de **AxesView** es perfectamente similar a la de los análisis precedentes. Consultarlos para navegar en esta herramienta.

2 – Planos Factoriales

- Haga Clic en  **PlaneView**
 - Una ventana se exhibe proponiendo diferentes visualizaciones de los planos factoriales.
 - Elegir la sección "Actives columns (variables)", adaptada a este análisis. En efecto, ella se refiere a las palabras utilizadas. Las contigüidades entre palabras significan que estas palabras son utilizadas en las mismas respuestas, y a menudo en las mismas frases. Hay un componente sintáctico más pronunciado en las asociaciones que en el caso del análisis precedente que agregaba a las palabras utilizadas por las mismas categorías de encuestado y, por lo tanto al interior de textos mucho más importantes.
 - Aparece entonces una ventana para seleccionar el plano factorial siguiendo el par de ejes deseados. Elegir los ejes 1 y 2 y hacer Clic en **display**. Aparece el plano factorial.



También podemos elegir la opción: "PLANEVIEW with moveable tags" que retoma algunas de las secciones precedentes y permite desplazar los puntos superpuestos para hacer más legible el gráfico.



Categorías ilustrativas con la opción "Etiquetas desplazables"

En el submenú propuesto por "PLANEVIEW with moveable tags", seleccionamos las Categorías ilustrativas, que constituyen el principal interés de este tipo de análisis directo de las respuestas. El gráfico más arriba nos muestra que la edad es una de las variables más

importantes dentro de la dispersión de las respuestas abiertas, así como el nivel de instrucción y el género (sexo).

Es una consecuencia de este tipo de análisis realizado sin "a priori" que se pueden elegir los criterios más pertinentes de reagrupación de las respuestas.

Las otras herramientas (ClusterView, Kohonen) pueden ser utilizadas según las recomendaciones de las secciones precedentes.

IV. Importación (creación y exportación) de los archivos al formato Dtm-Vic

Los archivos con formato interno de Dtm-Vic son los archivos diccionario, los archivos de datos numéricos y los archivos de textos, presentados en el párrafo I.3. Son necesarios para proceder a un análisis de datos numéricos o a un análisis de datos textuales. El caso más completo que pone en ejecución estos tres tipos de archivos es el de una encuesta que contiene respuestas, a la vez, a preguntas cerradas (archivos diccionario y datos) y a preguntas abiertas (archivo de texto).

Los archivos internos son unos archivos en formato ".txt" y se obtienen o de modo manual a partir de un modo de captura de importación integrado en Dtm-Vic o, la mayoría de las veces, a partir de archivos preexistentes en formato ".doc" para ciertos datos textuales o en formato ".csv" nacido de Excel para los datos numéricos y textuales, o todavía simplemente en formato texto (códigos ASCII).

El procedimiento de importación se produce sólo una vez, antes de los diversos procesos del análisis.

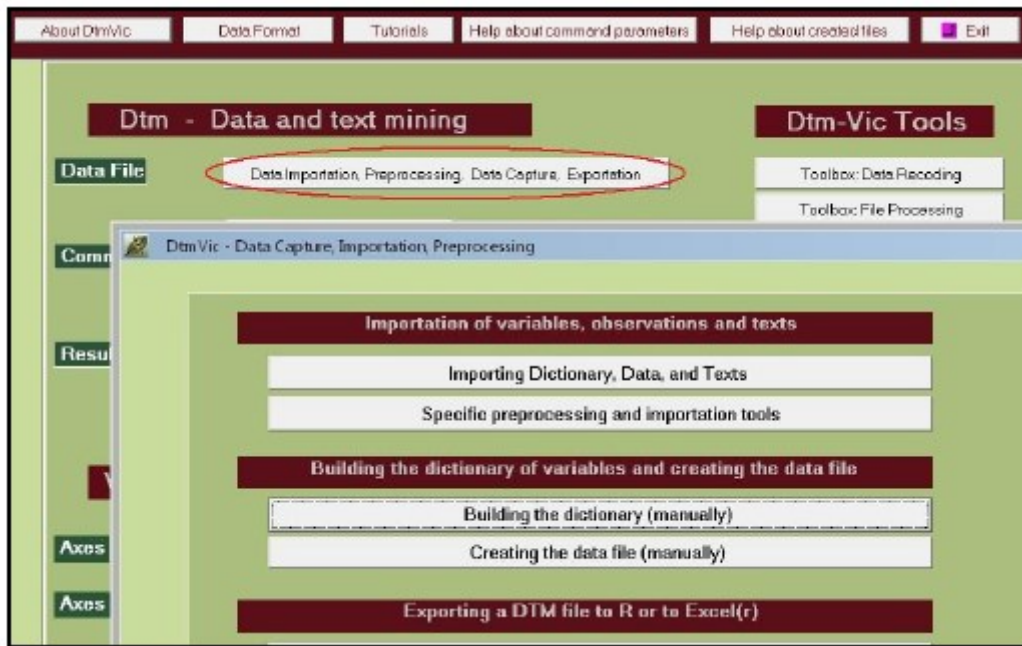
Aquí profundizaremos en la importación estándar, en formato "Excel", de datos numéricos y textuales, tales como los datos de encuestas compuestos por preguntas cerradas y abiertas, luego, en una segunda parte, presentaremos el procedimiento de captura directa de los datos.

Otros procedimientos son presentados en el Tutorial (en Inglés) integrado en Dtm-Vic. Los textos simples (formato interno tipo 1 descrito en la sección I.5, e ilustrado por el ejemplo III.1 del capítulo III) no dan lugar a un procedimiento de importación particular: es suficiente el insertar los separadores entre textos en los formatos usuales.

- Haga Clic sobre el botón **Data Importation, Preprocessing, Data Capture, Exportation**.

Se muestra una ventana que ofrece diferentes posibilidades para constituir un conjunto de datos numéricos o textuales en formato Dtm:

- **Importation of variables, observations and texts**: Importar datos numéricos o textuales en formato Excel, libre o fijo; datos textuales en formato libre; o también archivos XML que contienen datos numéricos o textuales.
- **Building the dictionary of variables and creating the data file**: Crear los archivos diccionarios y los archivos de datos numéricos o textuales manualmente a partir de un modo de captura de importación integrado en Dtm-Vic. Los otros dos procedimientos, **Exporting a DTM File to R or to Excel(r)** y **Dtm_tools**, se refieren a la exportación, la recodificación y el almacenamiento de los datos.



IV.1. Importación de archivos Excel®

IV.1.1 Presentación del archivo Excel.

Consideramos la tabla de datos de la encuesta "Life" presentada en los dos últimos ejemplos del capítulo III precedente. El archivo correspondiente dispone en fila 1.043 individuos y en columnas 17 variables: 9 variables nominales (el género, la edad recodificada, el nivel educativo y 6 variables de opinión), 1 variable continua (la edad), 3 variables textuales correspondientes a las 3 preguntas abiertas, y finalmente otras 4 variables nominales que corresponden a variables descriptivas recodificadas (la edad en 3 clases, los cruzamientos del género con la edad en 3 clases, el nivel educativo, el cruzamiento de la edad en 3 clases con nivel educativo).

ident	gender	age_code	age	education	important_life	important_probe	change_last_years	change_your_last_yrs	change_your_next_yrs	people_be_happier?	people_peace_of_mind.	more_or_less_freedom	culture	...
1	1	80	12	1	good health	happiness,	2	3	3	3	2	1		...
2	1	54	8	1	happiness in peop	contented with life as	1	1	3	1	1	1	education	
3	1	40	6	1	contentment	family	1	2	1	2	2	2	arts	
4	2	27	3	2	health	happiness, money, fa	1	2	1	1	1	1	the way british people	
5	2	39	5	2	to be happy	healthy, have enough	2	1	3	1	1	1		
6	1	80	12	1	my wife	music, holidays, I like	2	3	4	2	2	3	not much it's very imp	
7	2	48	7	2	health	happiness	4	3	0	0	2	1		
8	2	33	4	1	to be healthy	just to live long enou	3	4	1	2	3	1		
9	2	64	10	1	health,	keeping going, family	4	3	3	2	1	2	culture is good,	
10	2	65	11	1	husband	new baby grand dau	2	1	0	2	2	1	goodwill,	
11	1	58	9	3	companionship	job, good life, money	1	2	5	2	2	3	It's important, has exi	
12	2	74	12	1	good health	happiness, togethern	2	3	0	2	3	3	heritage, concerts, dr	
13	2	29	3	2	family	friends, pets,	2	2	2	3	2	1	theatre, national trust	
14	1	82	12	3	togetherness	peace of mind, good	3	3	0	2	2	2	music, poetry, ballet,	
15	2	68	11	1	my family really	health, walking	2	2	4	3	3	3	the beauty of our cou	
16	2	37	5	2	my children	my husband, my fam	1	2	1	3	0	1	can't think of anything	
17	1	34	4	2	my own time, not	my friends, plants, fo	2	4	3	0	2	2	the music of henry pu	
18	1	30	4	2	freedom of choice	sport, work, parents	2	1	2	1	2	1	literature, the theatre	
19	1	27	3	3	I suppose work	family, friends, gener	2	1	2	3	1	0	sausages, beefeaters	
20	1	85	12	1	health	family	0	3	3	2	1	2		
....														

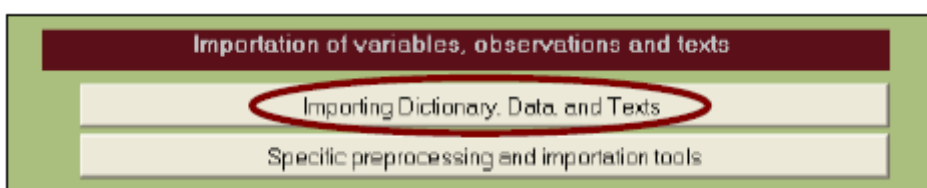
La primera fila y la primera columna contienen los identificadores de individuos y variables respectivamente. Todos los valores alfanuméricos, aquellos por ejemplo de los identificadores o también de las categorías de las variables nominales, deben estar compuestas de menos de 20 caracteres y preferentemente de menos de 10 y **no deben contener espacios vacíos**. Las respuestas a las preguntas abiertas son textos con menos de 8.000 caracteres. Por el contrario, los datos faltantes se muestran con espacios vacíos. Para una tabla de datos con n individuos y p variables, cualquiera sea su naturaleza, la tabla "Excel" dispone entonces de n+1 filas y de p+1 columnas.

El archivo se "guarda como" en formato ".csv", donde los separadores son los puntos-comas (*semicolon*)

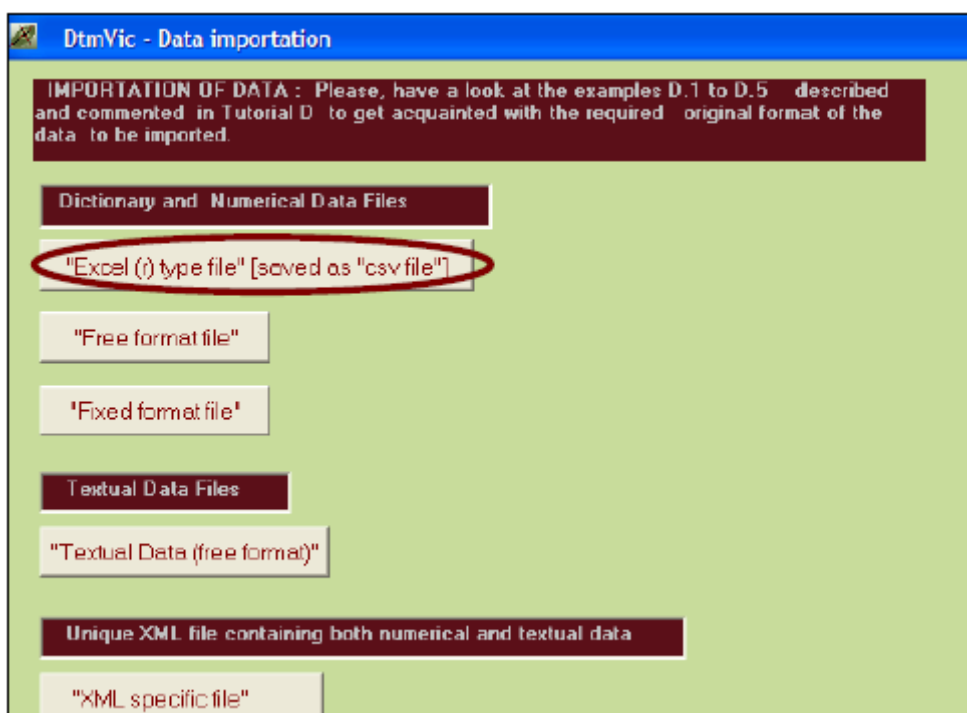
Este archivo que nos servirá de ejemplo tiene por nombre: `datbase_global.csv` y se encuentra en el directorio (carpeta): `DtmVic_Examples_D_Import\EX_D01.Importation.Num_Text` el mismo también en la carpeta `DtmVic-Examples` que se puede descargar con Dtm-Vic.

En ciertas versiones de Excel, especialmente en las versiones angloparlantes, el separador, para el formato ".csv", no es el punto y coma, sino la coma. El procedimiento de importación de DtmVic prevé una posibilidad de cambio de separadores. De hecho, todo como los espacios vacíos, los puntos y coma y los apostrofes en la expresión de los valores alfanuméricos no son autorizados y deben ser reemplazados por otro símbolo. Lo mismo para los valores numéricos, especialmente los números con más de 3 cifras, no deben contener blancos (escritura de los francófonos que dejan un semi-espacio para separar los millares). Finalmente, en la versión francesa y en algunas versiones europeas de Excel, "las comas decimales" deben ser reemplazadas por los puntos decimales habituales en las anotaciones anglosajonas y en los lenguajes de programación.

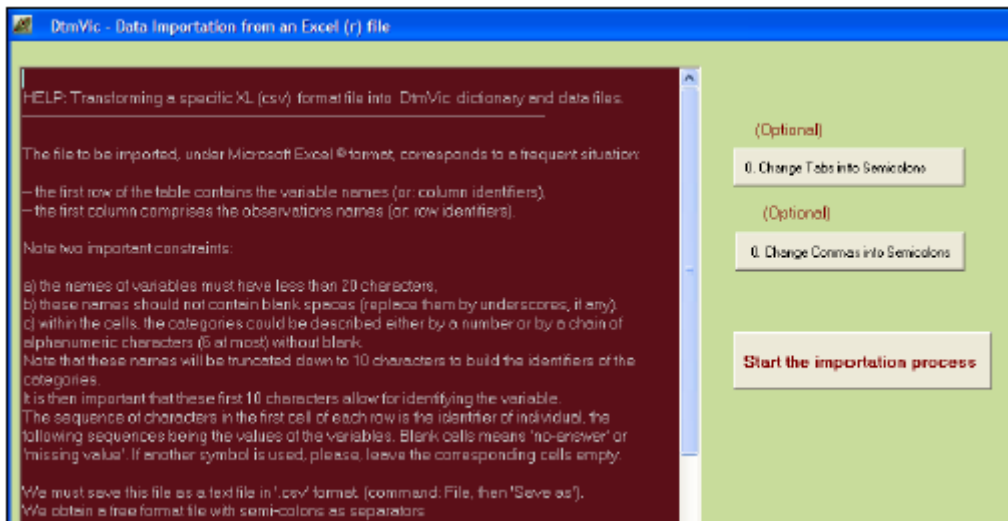
IV.1.2 Procedimiento de Importación



- Seleccione en el menu principal, **Data Importation, Preprocessing, Data Capture, Exportation**, luego **Importing Dictionary, Data and Texts** en **Importation of variables, observations and texts**
 - Aparece una ventana.



- Haga Clic después sobre **Excel (r) type file [saved as "csv file"]**
 - Aparece una ventana "Data Importation from an Excel ® file" en la que se proponen varias opciones.



Si el archivo Excel ha sido guardado utilizando "tabulaciones" o "comas" como separadores, haga Clic en uno de los botones opcionales:

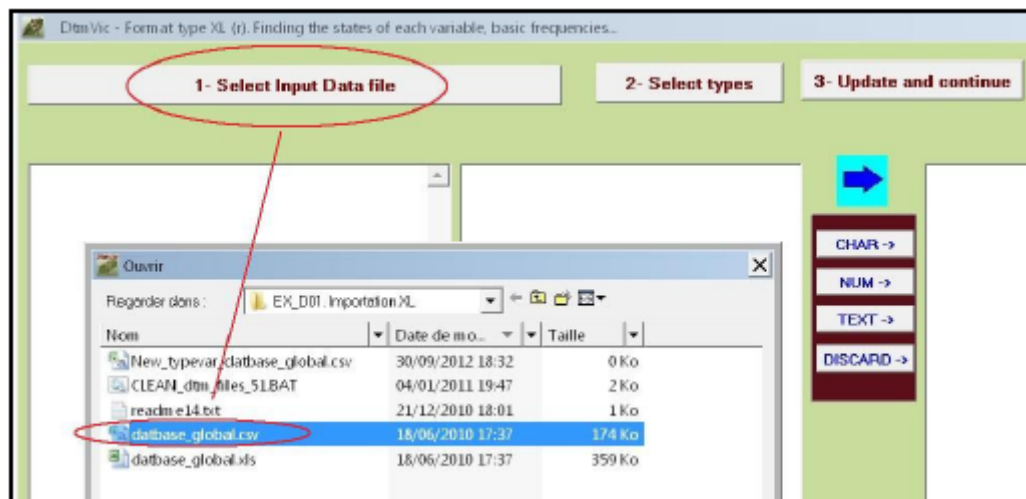
- **Change Tabs into Semicolons** convierte las tabulaciones en puntos-comas [después de haber verificado que el archivo original no contenía puntos-comas, y reemplazando éstos cuando sea necesario].

- **Change Commas into Semicolons** convierte las comas en puntos-comas. [Después de haber verificado que el archivo original no contenía puntos-comas, y reemplazado estas, cuando sea necesario].

En este caso, seleccione el archivo Excel guardado con tabulaciones o comas, y conviértalo. Un nuevo nombre es dado al archivo creado. El procedimiento de importación continuará empleando este nuevo archivo.

En todos los casos:

- Haga Clic en el botón **Start the importation process**.
 - Aparece una nueva ventana "Format type XL ®, Finding the states of each categorical variable, basic frequencies..."
- Haga Clic en **1.Select Input Data file** y abra el archivo XL en formato ".csv". Para el ejemplo, elegimos el archivo `datbase_global.csv` en el directorio: **DtmVic_Examples_D_Import\EX_D01.Importation.Num_Text**.
- Responder **OK** a la caja de mensaje.



El descriptivo de las variables se exhibe en la ventana de la izquierda. En la ventana central, podemos leer entre corchetes el número de valores distintos observados en el archivo y entre paréntesis una letra A o N.

La letra (A) significa que se observó valores no numéricos; la letra (N) indica que únicamente son valores numéricos. Es Entonces más fácil elegir el estado de las variables correspondiente a la segunda etapa de este procedimiento. Para esto:

- **2. Select types** Seleccione una o varias variables en la lista de la ventana central luego especifique su estado haciendo Clic sobre:

CHAR -> para una variable nominal (o categórica, aquí las variables de clasificación (1,2,4) y de opinión (7 - 12)

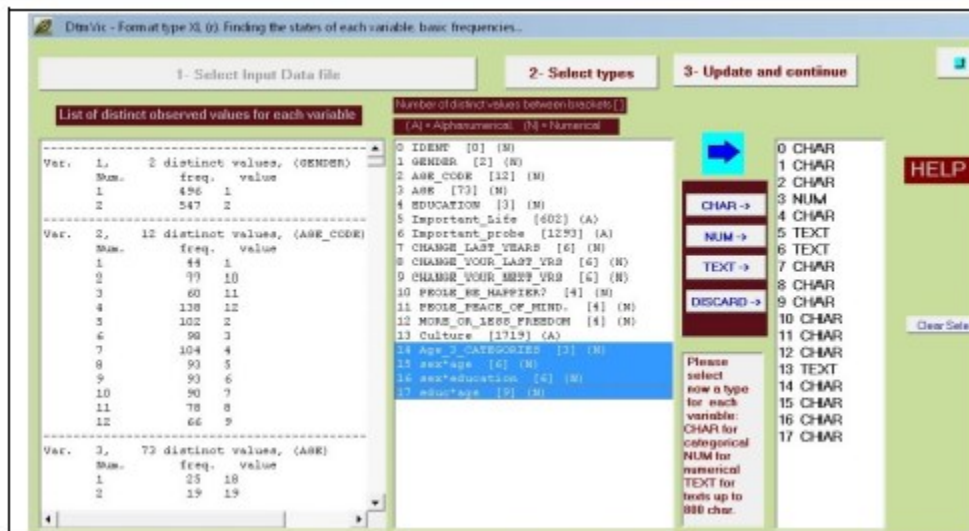
NUM -> para variable numérica (o continua, aquí la variable 3-edad)

TEXT -> para las variables textuales, las respuestas a las preguntas abiertas (variables 5, 6, 13).

DISCARD -> para abandonar variables (variables 14 - 17, por ejemplo)

- Una vez atribuido el estado, haga Clic en el botón **3.Updating and continue** luego responda **OK** sobre él "number of observations".

[Este procedimiento crea un nuevo archivo de importación, denominado automáticamente **New_typevar_database_global.csv**, cuya segunda fila contiene los tipos de variables. Pero el usuario no tiene que preocuparse de este archivo.]



Precisiones sobre la particularidad de la importación:

El procedimiento de importación consiste en la construcción de un diccionario y de un archivo de datos de DtmVic a partir del archivo original de datos. Los nombres de las variables serán extraídos a partir de los identificadores de las variables en el archivo de inicio. El número de categorías para cada variable nominal y los nombres de estas categorías se establecerán a partir de este archivo.

Para cada variable, todas las diferentes secuencias de los caracteres observadas en el archivo de datos son detectadas y contadas. Las categorías de las variables nominales están organizadas según el orden alfabético de sus identificadores.

Las filas del archivo de datos DtmVic comenzarán con el identificador que figura en la primera columna "identificador" del archivo Excel.

Las modalidades de las variables nominales serán unos números enteros consecutivos que comenzarán con valor "1", en lugar de un símbolo alfanumérico (el orden de las modalidades será el orden alfabético de sus símbolos en el archivo de origen). Los valores faltantes (compartimientos vacíos en el archivo de inicio) dan lugar a una modalidad particular, identificada en el diccionario Dtm-Vic por la letra "b" (como "blanco").

Los valores de las variables numéricas serán idénticos a las del archivo original de datos, los valores faltantes (compartimientos vacíos en el archivo de inicio) son reemplazados, en esta versión de DtmVic, por el valor convencional "999".

Las variables textuales (respuestas a las preguntas abiertas) dan lugar a un archivo textual separado (formato textual de tipo 2, ver capítulo I, sección I.5).

- Aparece una segunda ventana "format type XL. Finding the states of each categorical variable, basic frequencies..."
- Haga clic sobre **Values and counts**.

El nombre de las variables se exhibe en la ventana de izquierda. La ventana de derecha presenta las estadísticas elementales de estas variables. Se trata solamente de permitirle al usuario verificar que los estados que eligió para las variables son correctos.

DtmVic - Format type XL (r). Finding the states of each categorical variable, basic frequencies..

1) Values and Counts (as a global check of the whole file content) 2) Create dictionary and data

total number of variables 17

0, IDENT, Char, 30, 1
 1, GENDER, Char, 6, 1
 2, AGE_CODE, Char, 6, 1
 3, AGE, Num, 6, 1
 4, EDUCATION, Char, 6, 1
 5, Important_Life, Text, 8000, 1
 6, Important_probe, Text, 8000, 1
 7, CHANGE_LAST_YMARE, Char, 6, 1
 8, CHANGE_YOUR_LAST_YRS, Char, 6, 1
 9, CHANGE_YOUR_NEXT_YRS, Char, 6, 1
 10, FEELS_BE_HAPPIER?, Char, 6, 1
 11, FEELS_PEACE_OF_MIND., Char, 6, 1
 12, MORE_OR_LESS_FREEDOM, Char, 6, 1
 13, Culture, Text, 8000, 1
 14, Age_3_CATEGORIES, Char, 6, 1
 15, sex*age, Char, 6, 1
 16, sex*education, Char, 6, 1
 17, educ*age, Char, 6, 1

Var. 1, 2 distinct values, (GENDER)
 Num. freq. value
 1 496 1
 2 547 2

Var. 2, 12 distinct values, (AGE_CODE)
 Num. freq. value
 1 44 1
 2 77 10
 3 60 11
 4 138 12
 5 102 2
 6 98 3
 7 104 4
 8 93 5
 9 93 6
 10 90 7
 11 78 8
 12 66 9

Var. 3, numerical, (AGE)
 mean sd min max
 45.968 18.393 18.000 90.0

- Haga Clic en **Create dictionary and data**
 - Una ventana "creating a dictionary and a data file" aparece sobre la pantalla.

DtmVic - Creating a dictionary and a data file

Name for the new dictionary

Name for the new data file

Name for the new text file

Creates new dictionary

Create data and text files

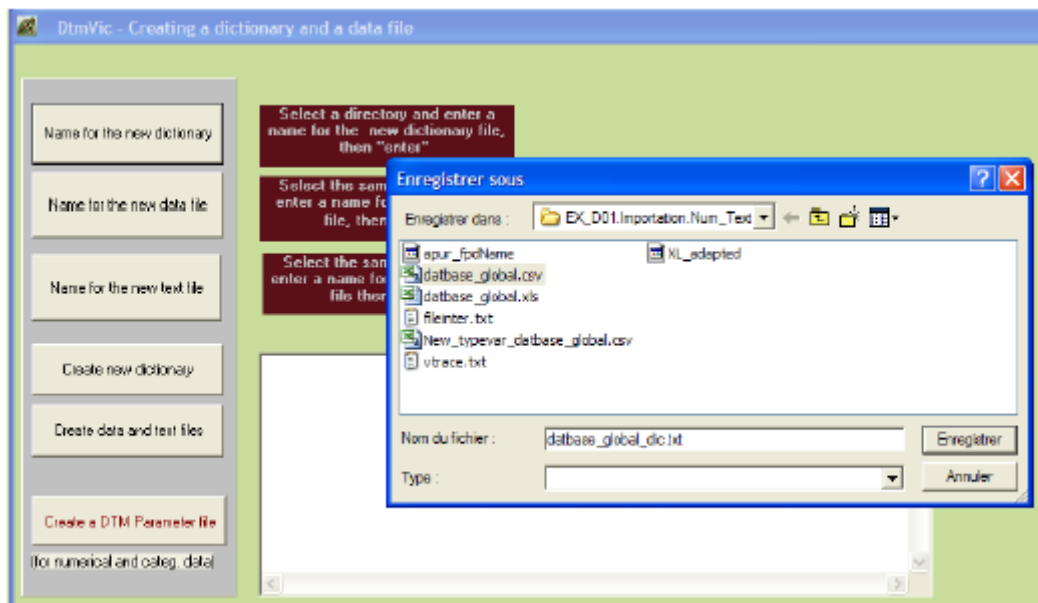
Create a DTM Parameter file
 (for numerical and categ. data)

Select a directory and enter a name for the new dictionary file, then "enter"

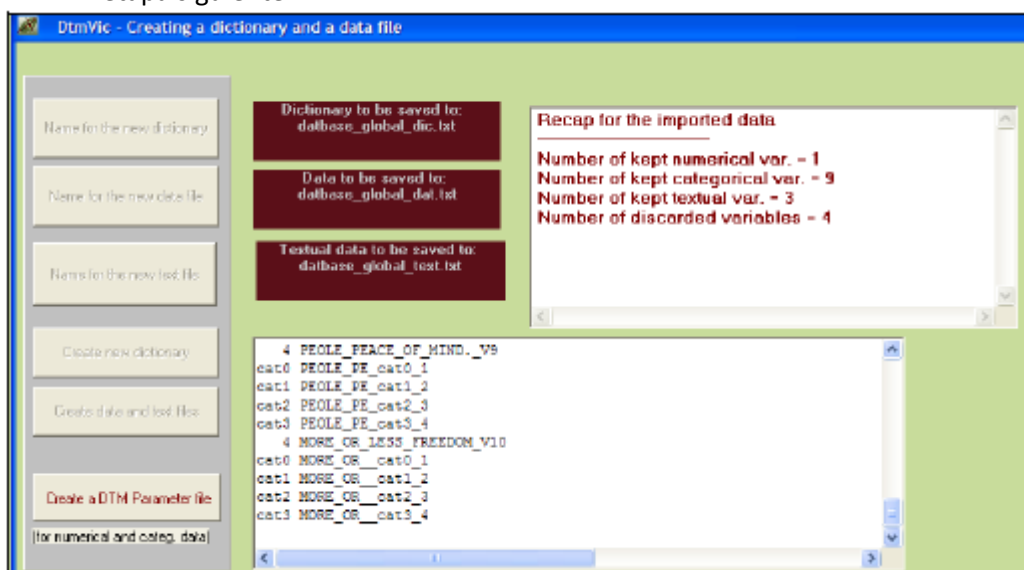
Select the same directory and enter a name for the new data file, then "enter"

Select the same directory and enter a name for the new textual file then "enter"

- Haga Clic en **Name for the new dictionary**. Ingrese el nombre del archivo diccionario `Database_global_dic.txt` (por ejemplo) y registre.



- Después, haga Clic en **Name for the new data file**. Ingrese el nombre del archivo de dato **Dabase_global_dat.txt** (por ejemplo) y registre.
- Haga Clic en **Name for the new text file**. Ingrese el nombre del archivo de texto **Dabase_global_text.txt** (por ejemplo) y registre. Si no hay datos textuales, pase a la etapa siguiente.

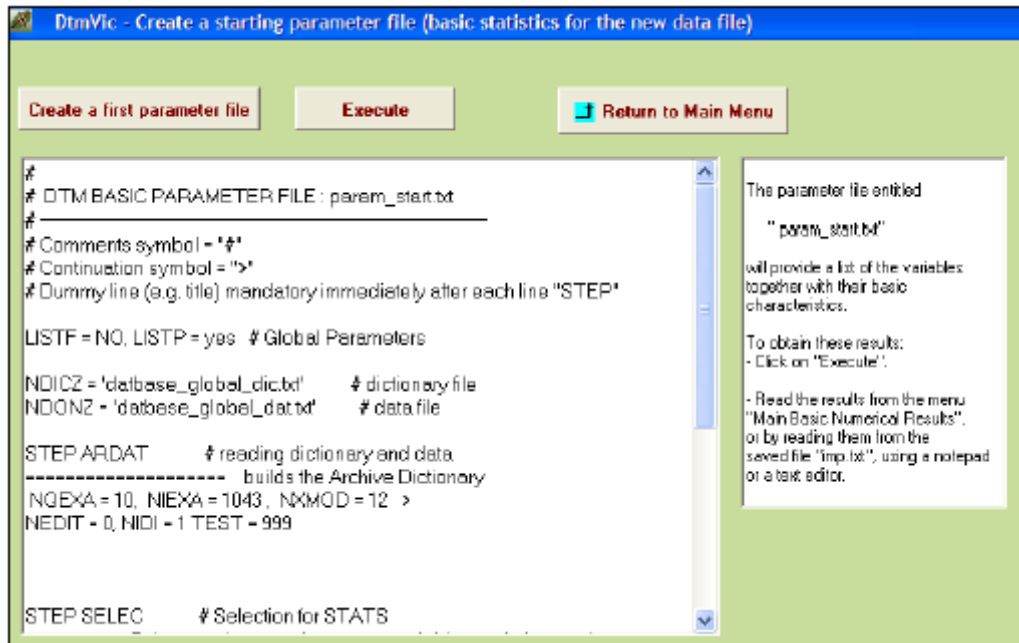


- Haga Clic en **Create new dictionary**. El archivo diccionario de DtmVic es creado automáticamente y se exhibe en la ventana. Responder **OK** a "New Dictionary completed". Del mismo modo haciendo Clic en **Create new data file**, se crea el archivo de datos de DtmVic. Una caja de mensaje muestra el número de individuos. Responder **OK**. En caso de presencia de preguntas abiertas, haga Clic en **Create new text file**

Un resumen de los datos importados aparece en una nueva ventana.

- Haga Clic finalmente en el botón **Create a DTM Parameter file**.
 - Una ventana "create a first parameter file" aparece sobre la pantalla.

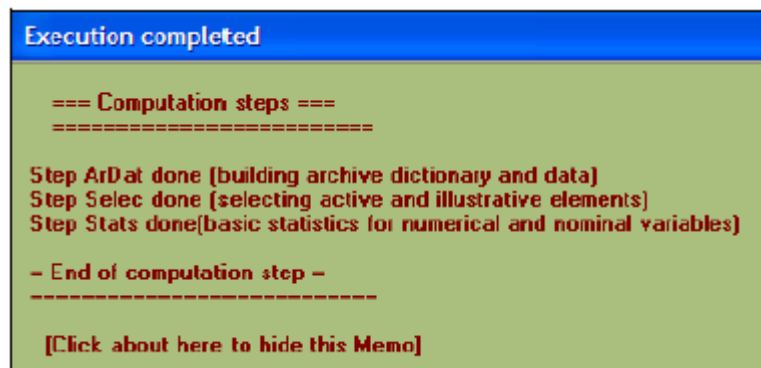
- Haga Clic entonces en **Create a first parameter file**. Un archivo de comando de DtmVic se muestra en la ventana inferior (en DtmVic, las expresiones "archivo de parámetro" y "archivo de comando" son equivalentes). Las operaciones y los comentarios quedan idénticos a los de la introducción.



El archivo parámetro automáticamente es guardado bajo el nombre de **param_start.txt** en la carpeta **DtmVic-Examples_D_Import/EX_D01.Num_Text**.

El archivo parámetro no incluye ningún comando de análisis estadístico elaborado. Se limita al cálculo de las estadísticas básicas de las variables. Simplemente sirve de control a la importación de los *datos numéricos*.

- Haga Clic finalmente en **Execute**
 - La ventana de ejecución, idéntica a todos procedimientos de análisis, aparece en la ventana del menú principal.



Los procedimientos se exhiben en bloque al final de la ejecución: la etapa **Ardat** archiva los datos y el diccionario. La etapa **Selec** elige las variables para el tratamiento siguiente; en este

caso, todas las variables disponibles son elegidas. La etapa **Stats** calcula las estadísticas generales.

Los resultados pueden ser consultados en la etapa **Result Files**

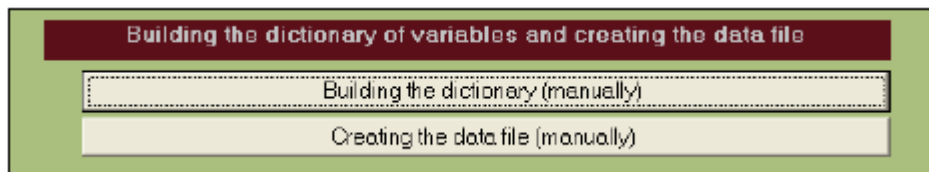
- Haga Clic en **Basic numerical results** para abrir el archivo en formato html luego sobre **Return** para salir y volver al menú principal



- haga Clic en **Basic numerical results (text format)** para abrir el archivo resultado en formato texto. La importación está terminada.

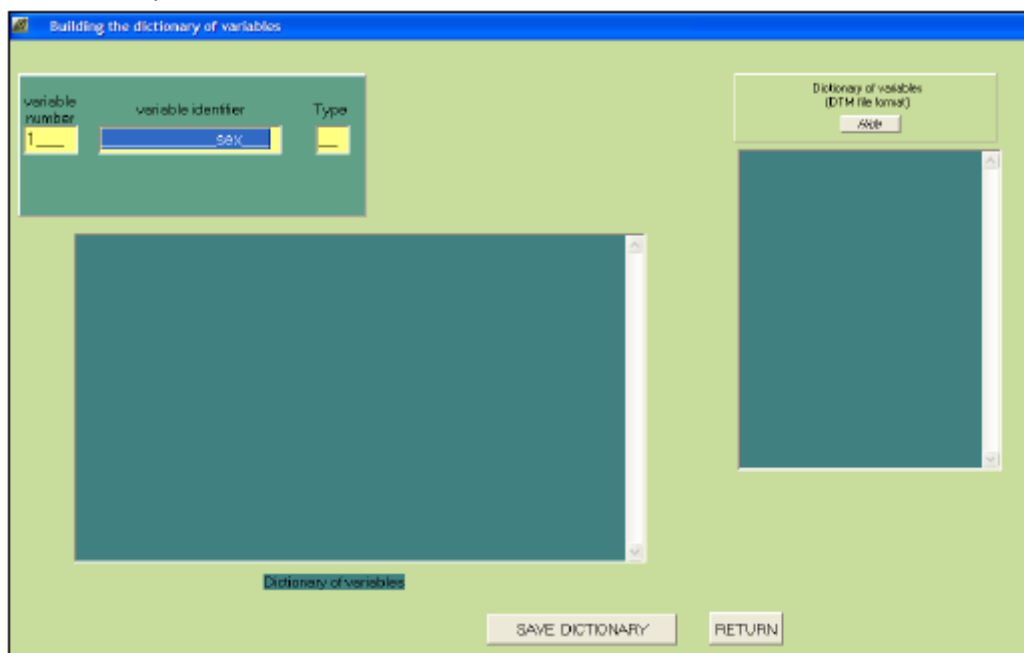
IV.2. Captura manual

DtmVic propone un módulo de captura de **datos numéricos**. Es sobre todo utilizable en un contexto pedagógico, para ingresar pequeños juegos de datos numéricos. Este módulo no permite ingresar sin embargo preguntas abiertas. Es deseable el pasaje por un archivo "Excel".



IV.2.1. El archivo diccionario

- Seleccione, en menú principal, Data Importation, Preprocessing, Data Capture, Exportation luego Building the dictionary en **Building the dictionary of variables and creating the data file**
 - Aparece una ventana dedicada a la construcción del diccionario.

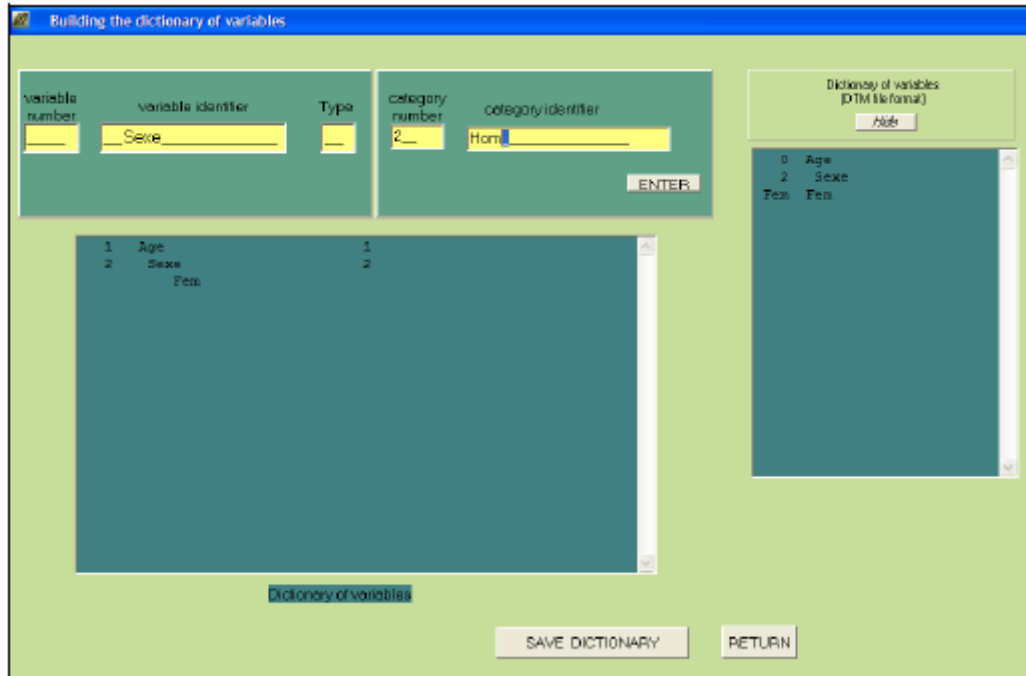


La primera subventana, arriba a la izquierda, permite ingresar el número, el nombre y el tipo de cada una de las variables.

- La 1a ventana amarilla muestra: "1", el número de la 1er variable que hay que ingresar.
En la segunda ventana, escribir el nombre de la variable luego en la 3era ventana, proporcionar el "Tipo" de la variable es decir el número de modalidades si la variable es nominal o escribir "0" si la variable es continua. Un botón **ENTER** se exhibe a partir del ingreso del tipo de la variable. Si ésta es continua, continúe el ingreso. Si es nominal, una ventana aparece para ingresar los números y las modalidades de la

variable nominal. Una vez las modalidades registradas, haga Clic en **ENTER** (o presione la tecla "Enter" en el teclado). Continúe ingresando el conjunto de las variables.

El resultado de la captura del diccionario de las variables aparece en la ventana inferior así como en la de la derecha, en la cual aparece en el formato interno de DtmVic.



Por ejemplo, una primera variable "edad" ha sido ingresada. Siendo una variable continúa el tipo es "0". Una segunda variable "Sexo" es ingresada. Teniendo dos modalidades, el tipo "2" es ingresado. A continuación aparecerá una ventana contigua junto a la cual se introducen las etiquetas de las dos modalidades.

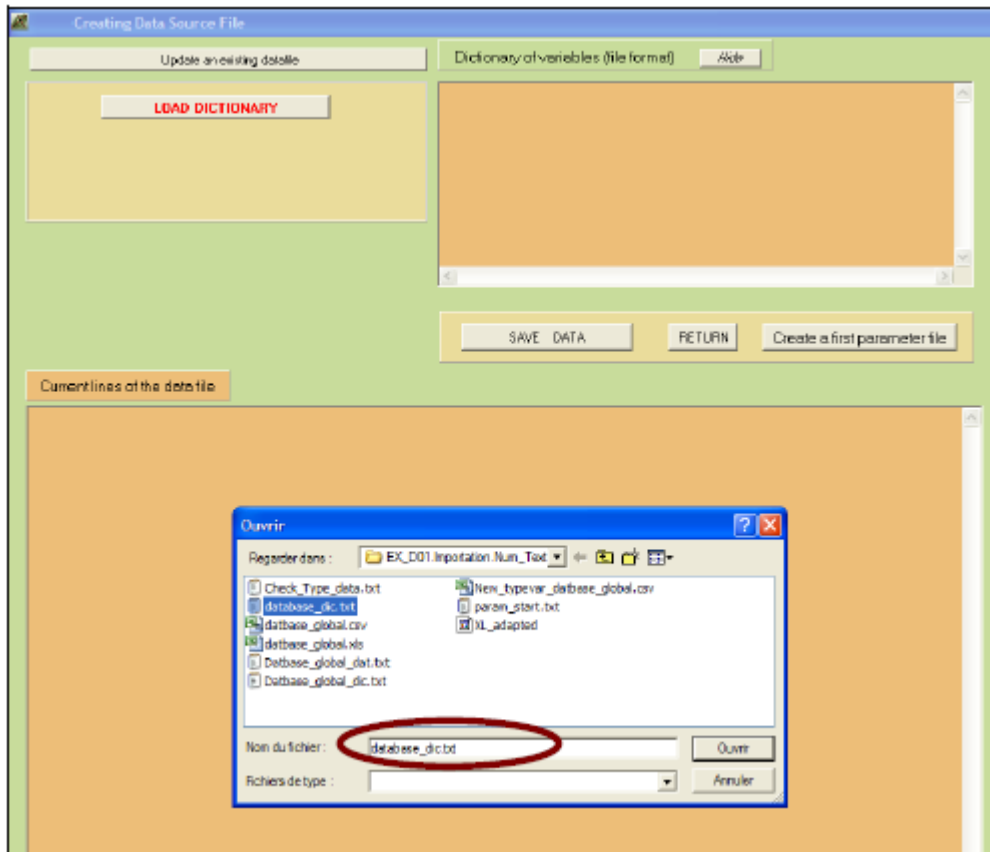
- Haga Clic en **ENTER** (o presione la tecla "Enter" en el teclado) después de cada ingreso.
- Una vez que el conjunto de las variables sean ingresadas, haga Clic en **SAVE DICTIONARY** y registre un nombre para el archivo del diccionario.

Podemos denominarlo: `Database_dic.txt`. Haga Clic luego en **RETURN**.

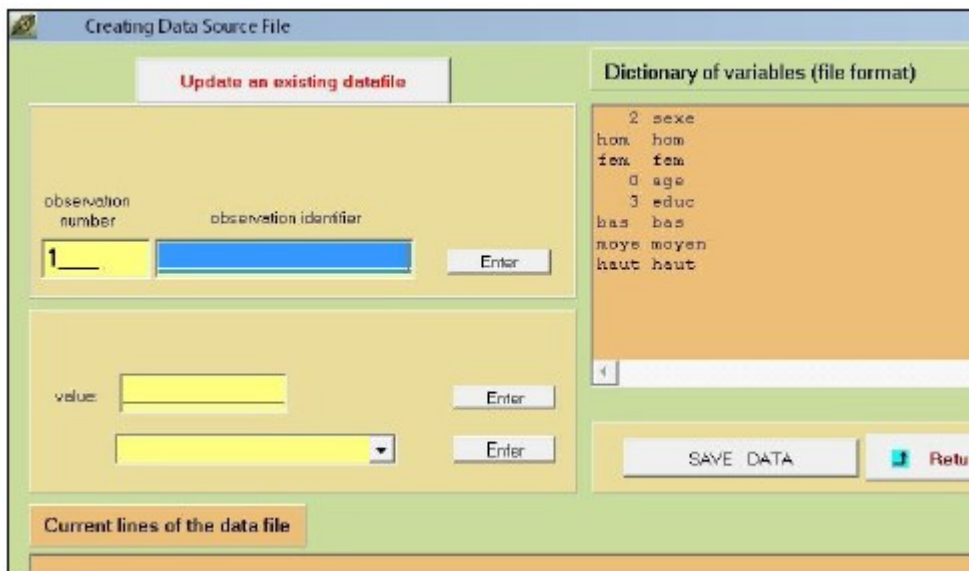
IV.2.2. El archivo de los datos

Una vez el archivo diccionario ha sido creado:

- Seleccione, **Creating the data file** en **Building the dictionary of variables and creating the data file**
 - Aparece una ventana para la construcción del archivo de datos.



- Haga Clic en **LOAD DICTIONARY** y abra el archivo diccionario creado anteriormente **Database_dic.txt**.
 - Aparece una ventana para la captura de datos. El diccionario de las variables se exhibe en la ventana derecha.



- Introducir el identificador del individuo y haga Clic en **Enter** (o presionar la tecla "Enter" en el teclado). La 1era variable se exhibe en la ventana.

- Seleccione la modalidad que corresponde al individuo con el menú que se abre luego haga Clic en **Enter** (o presionar la tecla "Enter" en el teclado).

La 2a variable se exhibe. Se trata de capturarla del mismo modo. Una vez ingresadas las variables para el individuo, aparece el individuo siguiente.

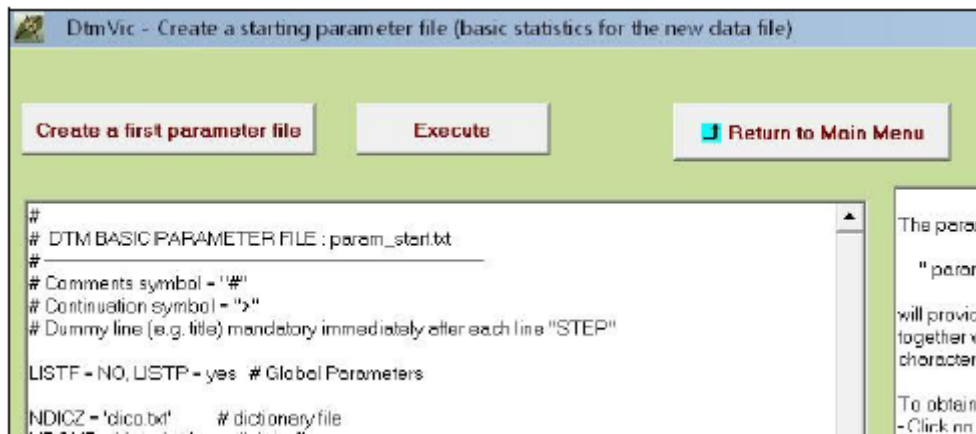
El diccionario se exhibe en la ventana superior y derecha y el archivo de los datos en la ventana de abajo.

IV.2.3. Creación de los archivos DtmVic

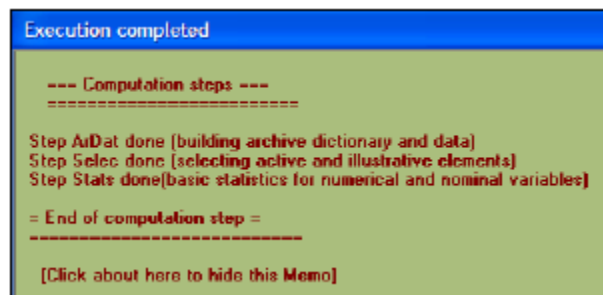
Una vez el ingreso ha sido terminado:

- Guarde el archivo haciendo Clic en **SAVE DATA** y registre el nombre del archivo de datos: **Database_dat.txt** (por ejemplo) relativo al archivo diccionario creado anteriormente luego:

- Hacer Clic sobre, **Creating a first parameter file**.
 - Aparece una ventana para la creación del archivo parámetro.
- Hacer Clic en el nuevo botón: **Create a first parameter file**. El archivo parámetro aparece en la ventana inferior.



- Hacer Clic en **Execute**.
 - La ventana de ejecución aparece, idéntica a la del procedimiento de importación (comprobación simple y estadísticas de base para los datos registrados).



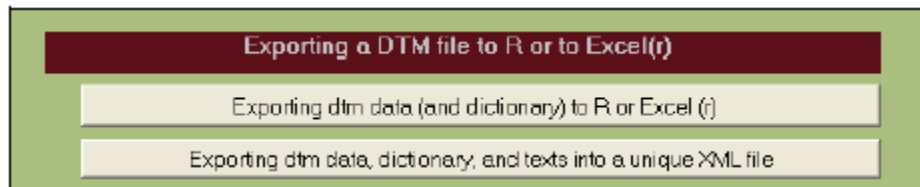
Los archivos elaborados (diccionario y datos) ahora están listos para los análisis.

IV.3. Exportación de archivos de datos en formato "Excel[®]" (o: XL)

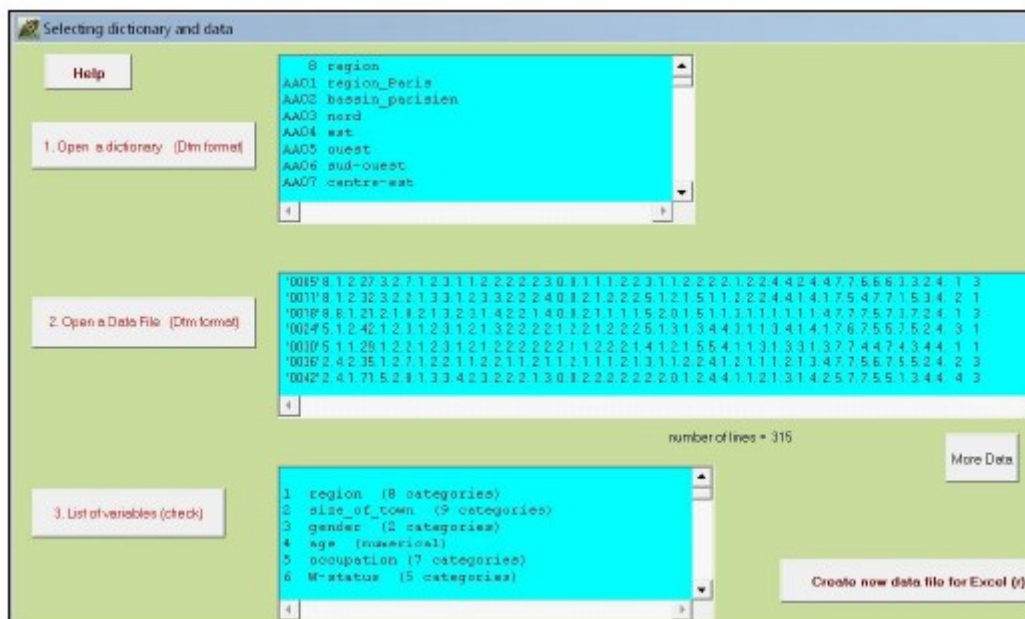
El procedimiento de exportación presenta el interés en variables recodificadas y sobre todo coordenadas factoriales archivadas o una partición calculada y archivada (los procedimientos del almacenamiento son tratados al capítulo V).

Proponemos aquí exportar el archivo de datos nacido del ejemplo del análisis de correspondencias múltiples del capítulo II. La exportación puede hacerse hacia un formato Excel o hacia un formato vecino aceptable por el procedimiento "read.table" del lenguaje R (archivo cuyo formato es idéntico al formato Excel, a excepción de la primera fila).

IV.3.1. Procedimiento de exportación



- Hacer Clic en **Exportation dtm data** en **Exporting a DTM file to R or to Excel**.
 - Una ventana aparece.
- Haga Clic en **Open a dictionary**. Abrir, por ejemplo, el archivo **MCA_dic.txt** "en" **EX_A03.MultCorAnalysis**
 - Una primera ventana anuncia la etiqueta de las variables y de las modalidades.
- Haga Clic luego sobre **Open a Data file** y abre el archivo "MCA_dat.txt" en "EX_A03.MultCorAnalysis". Luego haga Clic en **List of variables**.



Es posible exportar ya sea en formato Excel[®] o de formato R. Aquí, elegimos un archivo Excel.

- Seleccionar a y responder a la caja de mensaje: "New data file created".

Un nuevo archivo `MCA_d_dtm_XL.csv` es creado en el directorio `EX_A03.MultCorAnalysis`.

Un extracto de este archivo Excel (14 individuos, 4 variables) figura más abajo.

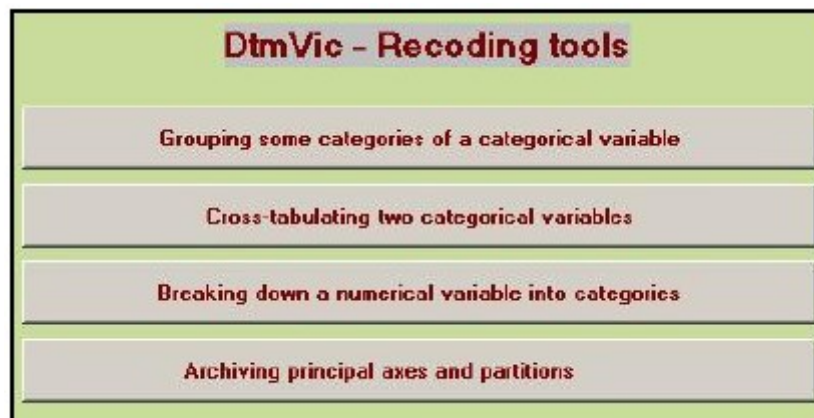
Identifiers	region	size_of_town	gender	age
5	mediterranee	<2000	female	27.000000
11	mediterranee	<2000	female	32.000000
18	mediterranee	>200000	male	21.000000
24	ouest	<2000	female	42.000000
30	ouest	<2000	male	29.000000
36	bassin_parisien	10001-20000	female	35.000000
42	bassin_parisien	10001-20000	male	71.000000
48	ouest	<2000	male	62.000000
54	ouest	20001-50000	male	24.000000
60	est	<2000	male	52.000000
66	est	10001-20000	female	42.000000

V. Recodificación, almacenamiento, herramientas diversas

La explotación de los datos estadísticos es un proceso interactivo que a menudo necesita varias iteraciones. Entre las operaciones más corrientes, la reagrupación de las modalidades de una variable nominal, el cruzamiento de dos variables nominales, la división en clases de una variable continua son frecuentemente producidos por los resultados de un análisis anterior. El almacenamiento de las particiones o de los ejes factoriales es también útil para avanzar en la comprensión de los datos permitiendo realizar análisis que los toman en consideración. Estas etapas de recodificación son bastante fundamentales. Aunque Dtm-Vic no sea un software de gestión de datos, pareció necesario hacer estas operaciones accesibles a partir de la caja de herramientas (*Toolbox*).

V.1. Recodificación

- Haga Clic sobre Toolbox Data Recoding
 - El menú que aparece concierne la recodificación de los datos y el almacenamiento de ciertos resultados.



Creación o recodificación de variables nominales:

- i) Reagrupación de modalidades
- ii) Creación de una variable nominal por cruzamiento de dos variables nominal
- iii) Transformación de una variable continua en variable nominal
- iv) Almacenamiento de los ejes factoriales y de las particiones.

Ya sea para la reagrupación de modalidades de una variable nominal, para la creación de una variable por cruzamiento de dos variables nominales o para la transformación de una variable continua en una variable nominal, la primera etapa consiste en:

- Abrir el archivo diccionario: 1. Open a dictionary

- Luego el de los datos: 2. Open a data file

- Listar las variables: **3. List of variables**

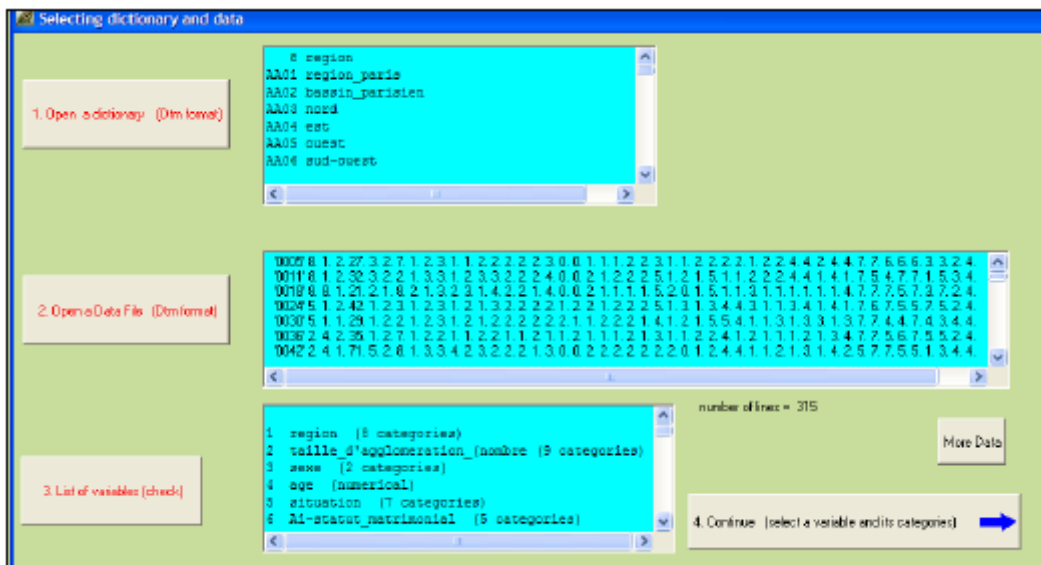
- Luego haga Clic sobre: **4. Continue**

Las operaciones siguientes son efectuadas a partir del juego de datos del ejemplo **EX_A03.MultCorAnalysis** en la carpeta **DtmVic_A_Start**.

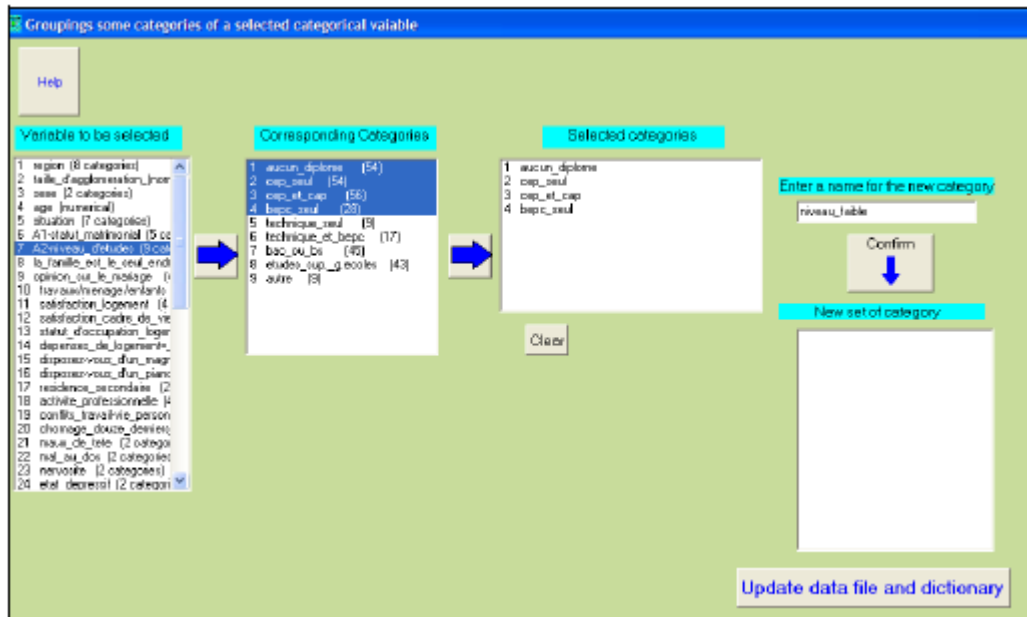
V.1.1 Reagrupación de modalidades de una variable nominal

En el momento del análisis de datos de encuesta y a la ocasión de frecuencias simples efectuadas sobre las variables nominales, debemos reagrupar a veces ciertas modalidades de una variable nominal para satisfacer, dentro de lo posible, ciertas reglas de recodificación: evitar modalidades de números bajos, equilibrar el número de modalidades de las variables nominales, reagrupar categorías similares o demasiado delgadas.

- Haga Clic en **Grouping some categories of a categorical variable**.
 - Aparece la ventana de selección de los archivos diccionario y de los datos.
- Abrir el archivo **MCA_dic.txt** y **MCA_dat.txt** en la carpeta **EX_A03.MultCorAnalysis**, poner en una lista las variables y hacer Clic en **4. Continue**.



- Aparece una nueva ventana.



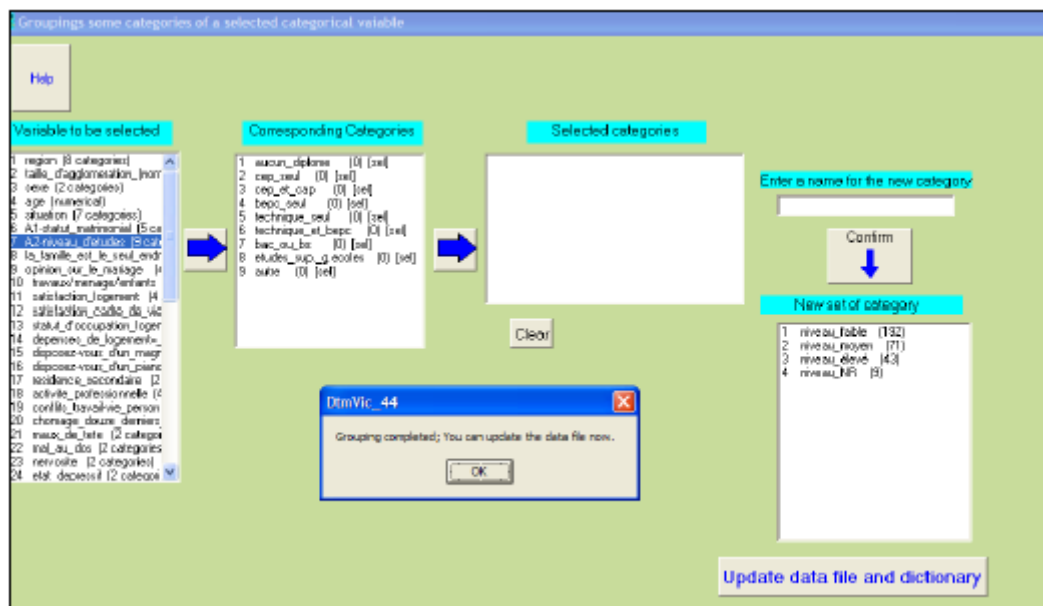
- Seleccione la variable que hay que recodificar. Aquí elegimos, en la 1era ventana, la variable "7-nivel de estudios" con 9 categorías. Las categorías (modalidades) de esta variable se exhiben en una 2a ventana. Seleccione el conjunto de las modalidades que hay que reagrupar quiénes aparecen en una 3a ventana. Ingrese el nombre de la nueva modalidad en la 4a ventana luego confirme. La nueva modalidad aparece en la 5a ventana.
- Empiece de nuevo el procedimiento para todas las modalidades de la variable. Si una modalidad no se puede reagrupar, seleccionarla y atribuirle una etiqueta.

En el ejemplo, reagrupamos las 4 primeras modalidades en "nivel inicial", las 3 otras en "nivel medio", la 8a modalidad en "nivel superior" y la 9a en "nivel NR" (no respuesta).

Las modalidades de la nueva variable aparecen en la 5a ventana.

Se coloca esta variable al final del archivo y se denomina "var7-4cat".

- Una vez que se acaban las reagrupaciones, responder: **OK** luego haga Clic sobre: **Update data file and dictionary**.



Dos nuevos archivo diccionario y de datos son creados `dtm_dic_newG7.txt` y `dtm_dat_newG7.txt`, siempre en la misma carpeta **EX_A03.MultCorAnalysis**.

- Una ventana se exhibe para presentar estos nuevos archivos (para los cuales el usuario podrá elegir nuevos nombres, si lo considera de utilidad).
- Haga Clic sobre: **Return**. La operación de reagrupación de las modalidades está terminada.

V.1.2. Cruzamiento de dos variables nominales

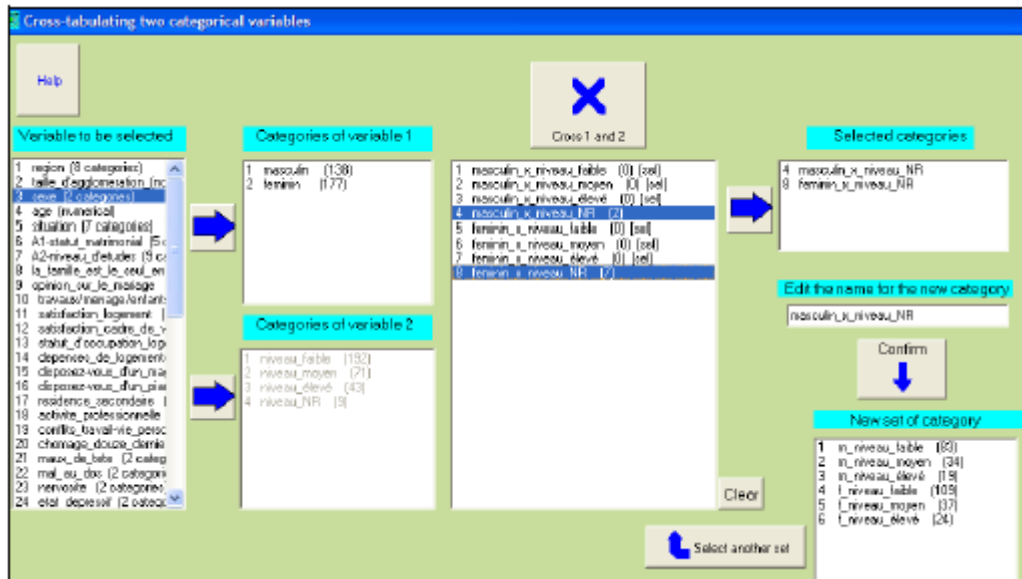
En este caso deseamos aumentar las posibilidades de análisis y de interpretación creando una nueva variable nominal a partir del cruzamiento de dos variables nominales (Ejemplo: sexo X edad).

- Haga Clic sobre: **Cross-tabulating two categorical variables**.
 - Aparece la ventana de selección de los archivos diccionario y de datos.

Seleccionar las dos variables que queremos cruzar y hacer Clic en el botón "Cross 1 and 2"

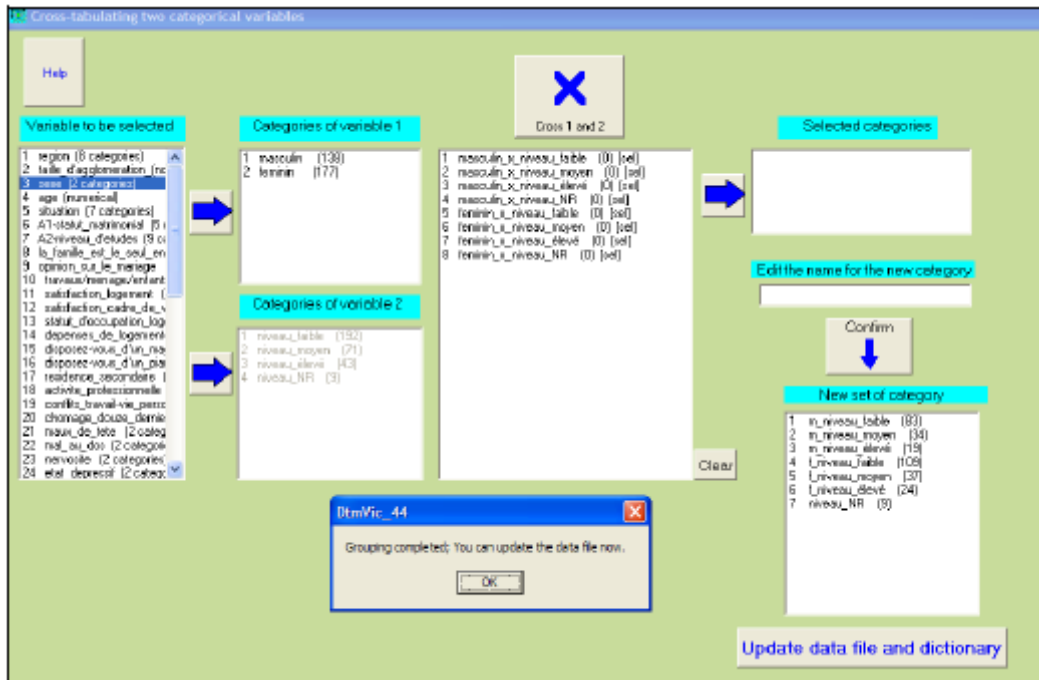
- Abrir los archivos diccionario y de datos en cuestión (para el ejercicio, podremos abrir los archivos anteriormente creados en la carpeta **EX_A03.MultCorAnalysis: dtm_dic_newG7.txt** y `dtm_dat_newG7.txt`), poner en una lista las variables, luego: **Continue**

- Aparece una ventana (ver más abajo)



- Seleccione las modalidades que hay que reagrupar o que hay que validar quiénes aparecen en una 3a ventana.
- Ingrese la etiqueta de la nueva modalidad en la 4a ventana luego confirme. La nueva modalidad aparece en la 5a ventana.
- Empiece de nuevo el procedimiento de etiquetado para todas las nuevas modalidades. Si una modalidad no está reagrupada, seleccionarla y atribuirle una etiqueta.
- Una vez se termine con las reagrupaciones, responder: **OK** a la caja de mensaje, luego haga Clic en **Update data file and dictionary**.

Dos nuevos archivos diccionario y de datos son creados: **dtm_dic_newCr3x52.txt** y **dtm_dat_newCr3x52.txt** en la carpeta **EX_A03.MultCorAnalysis**. Una ventana se exhibe para presentar estos nuevos archivos.

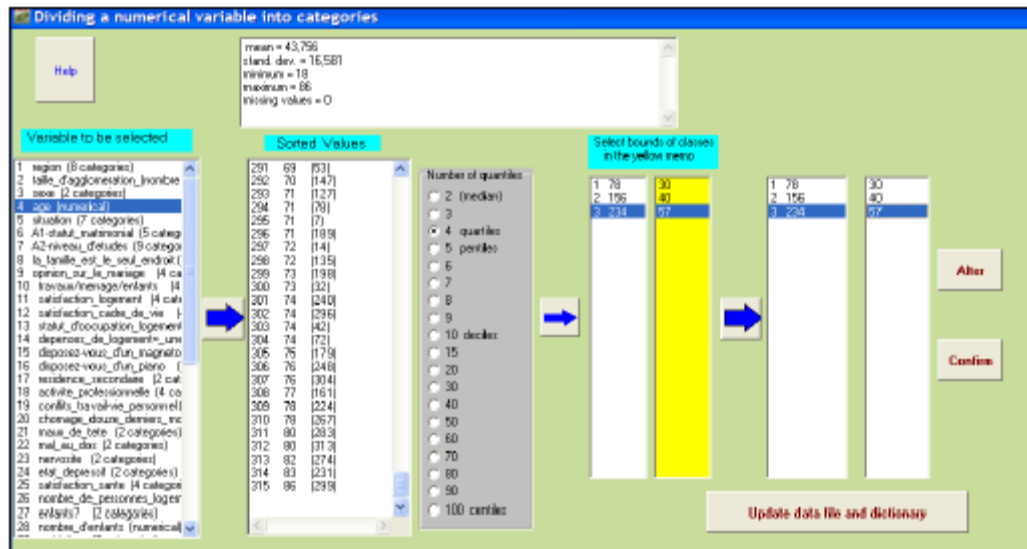


- Haga Clic en **Return** una vez finalizada la operación, puede modificar los nombres (por defecto) de los archivos si éstos no son adecuados.

V.1.3. Transformación de una variable continúa en variable nominal

Este procedimiento permite transformar una variable continua en una variable nominal, reagrupando los valores numéricos en clases. Esta reagrupación en k clases se hace a partir de un recorte previo en n cantiles (n clases de igual tamaño), no siendo n mucho más grande que k. Este recorte es útil porque “elimina el carácter lineal” del papel de la variable en los cálculos (las relaciones no lineales pueden entonces ser tomadas en consideración).

- Haga Clic en **Breaking down a numerical variable into categories**.
 - Aparece la ventana de selección de los diccionarios y de los datos.
- Abrir, en la carpeta **EX_A03.MultCorAnalysis**, los archivos diccionario y de datos **MCA_Fr_dic.txt** y **MCA_dat.txt**.
 - Aparece una ventana.



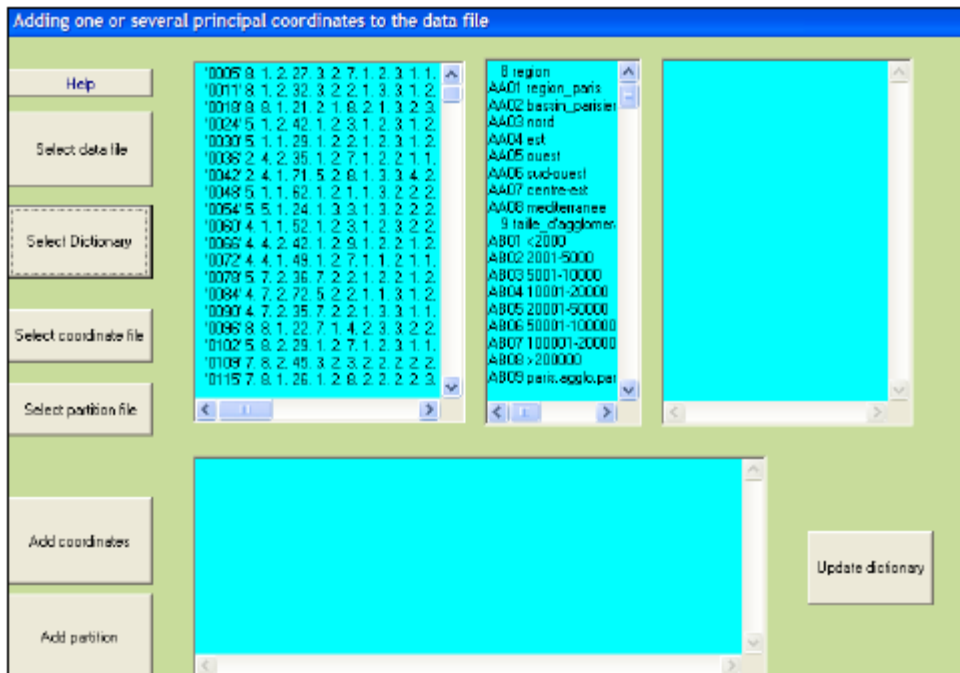
- Seleccione la variable continua (V4_age) y la traslada a la 2a ventana **Sorted Values**. Elegir el número de cantiles (5 por ejemplo, podemos también elegir 20 (o 100) cantiles para controlar mejor los límites de clases).
- Traslade haciendo Clic sobre **➡**. Confirmar y responder **OK** cuando el número de modalidades se exhibe.
- Una vez se finalicen las reagrupaciones, responder **OK** luego haga Clic en **Update data file and dictionary**. Son creados dos nuevos archivos diccionario y de datos: **dtm_dic_newD4.txt** y **dtm_dat_newD4.txt** así como un archivo "Dissecting_Check" que presenta los detalles de la operación. Haga Clic en **Return** para volver al menú principal.

V.1.4. Almacenar los factores o las particiones

Podemos querer mejorar el archivo inicial de datos con los resultados de un análisis factorial o de una clasificación. Los factores o las particiones entonces se consideran como nuevas variables.

Atención: no podemos almacenar factores o particiones si el análisis que los produjo utilizó un filtro interno sobre los individuos (en el momento de la creación del archivo de comando). Por el contrario, podemos utilizar un filtro externo (antes de cualquier análisis) tal como el que se define a continuación en el párrafo V.2.1.

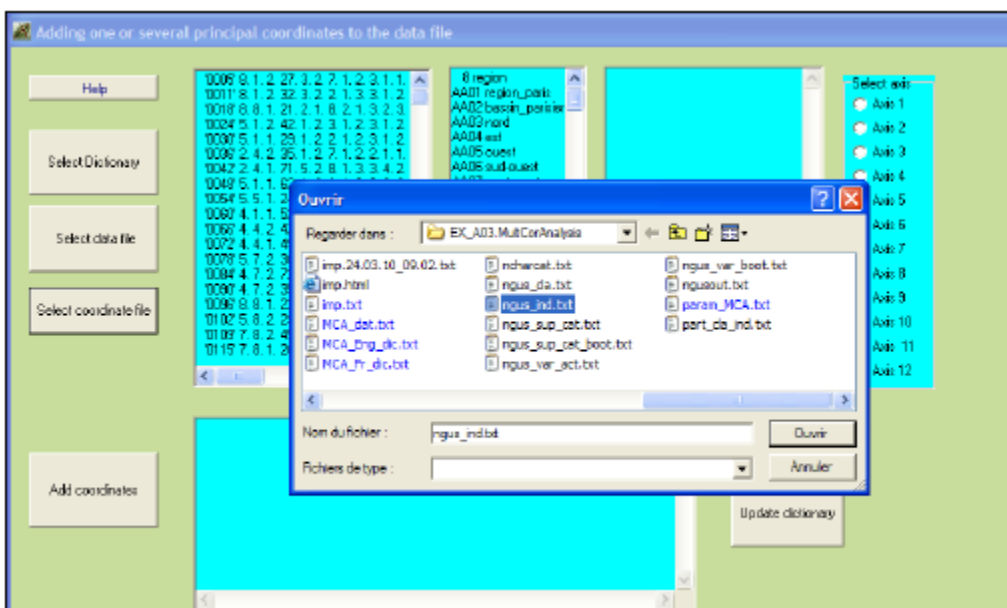
- Haga Clic en **Archiving principal axes and partitions**
 - Aparece una ventana.



- Ouvrir le fichier dictionnaire (**MCA_dic.txt**) puis celui des données (**MCA_dat.txt**) et sélectionner le stockage d'un facteur : **Select coordinate file** ou d'une partition : **Select partition file**

a. Almacener un factor

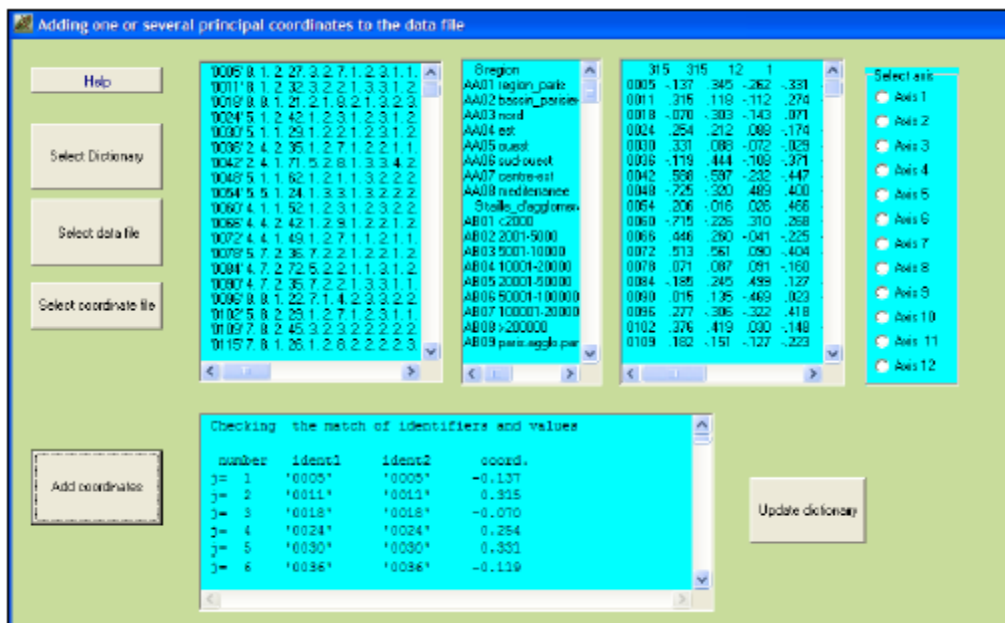
- Cliquer sur **Select coordinate file**
 - Apparaît une fenêtre qui montre le dossier **EX_A03. MultCorAnalysis** où se trouve le fichier **ngus_ind.txt** de coordonnées factorielles créées au moment de la procédure : **MCA – Multiple Correspondence Analysis**



- Abra el archivo **ngus_ind.txt**, luego seleccione el eje que hay que almacenar.
 - Las coordenadas factoriales aparecen en la 3a ventana.
- Haga Clic en **Add coordinates**.
 - Aparece una caja de mensaje: "Coordinate added. Please, update the dictionary". Responder **OK**. El almacenamiento de las coordenadas se exhibe en la ventana de abajo.
- Haga Clic en **Update dictionary** y responder **OK** en la caja de mensaje "Dictionary updated" que se exhibe.

Los archivos diccionario y de datos son creados en la carpeta **EX_A03. MultCorAnalysis** y son denominados: **dtm_dico_newA1.txt** y **dtm_data_newA1.txt**.

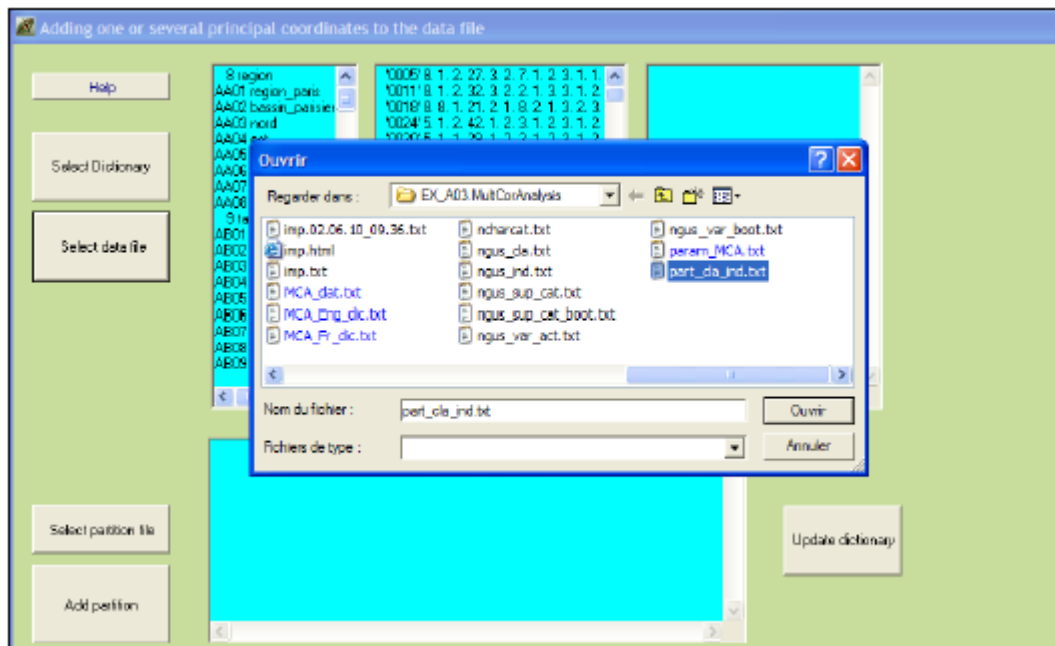
Para almacenar un segundo factor empezar de nuevo el procedimiento seleccionando los **nuevos** archivos diccionario y de datos: **dtm_dico_newA1.txt** y **dtm_data_newA1.txt**. A continuación siga el mismo procedimiento para almacenar una partición.



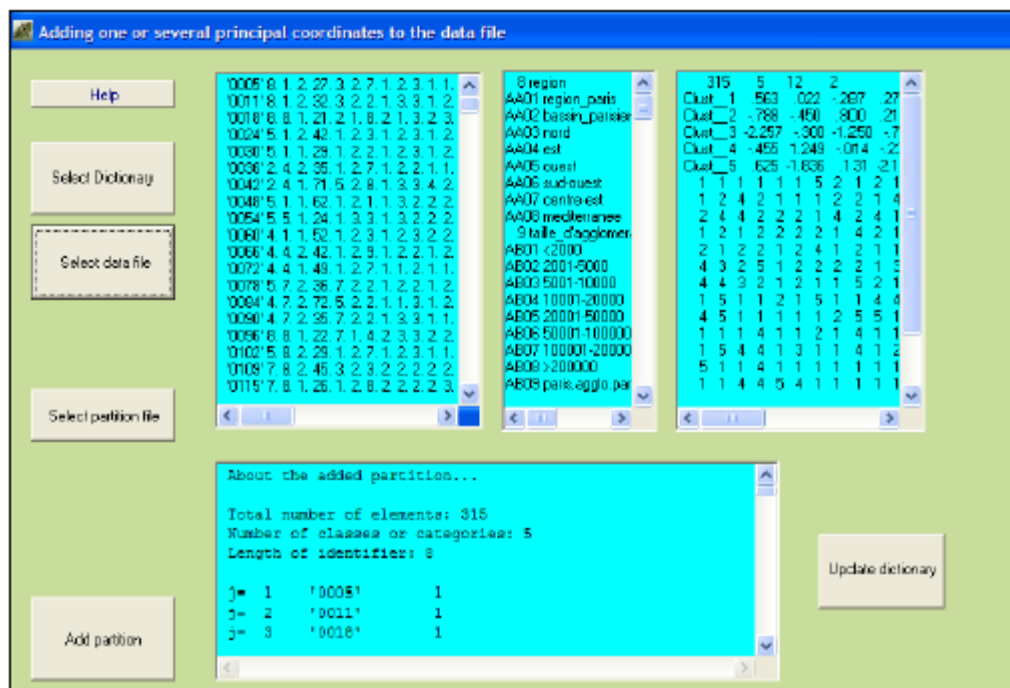
b. Almacenar una partición

- Haga Clic en **Select partition file**
 - Una ventana de la carpeta: **EX_A03. MultCorAnalysis** se exhibe donde figura el archivo: **part_cla_ind.txt** del almacenamiento de la partición creada en el momento del procedimiento: **MCA - Múltiple Correspondencia Analysis** y cuyo número de clases ha sido especificado en el momento de configurar el análisis.
- Abra, en la carpeta: **EX_A03. MultCorAnalysis**, el archivo: **part_cla_ind.txt** (archivo de la partición, ver los nombres de los diversos archivos de texto creados por Dtm-Vic en "Help about files" del menú principal.

- Haga Clic en **Add partition**.



- Aparece una ventana: “Partition added. Please, update the dictionary”. Responder: **OK**

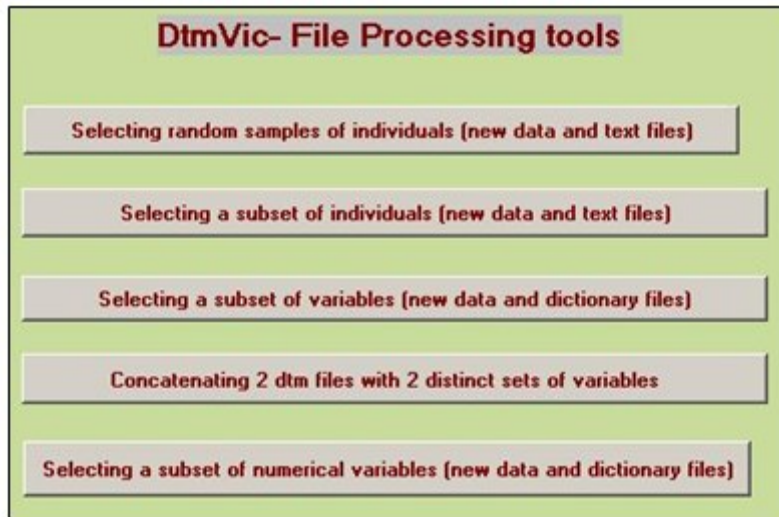


- El almacenamiento de la partición se exhibe en la ventana inferior
- Haga Clic en: **Update dictionary** y responder: **OK** en la ventana: “Dictionary update” que se exhibe.

Los archivos diccionario y de datos son creados en la carpeta **EX_A03. MultCorAnalysis** y son denominados: **dtm_dico_newP1.txt** y **dtm_data_newP1.txt**.

V.2. Intervenciones elementales sobre la base de datos

El segundo grupo de acciones se obtiene haciendo Clic sobre: **Toolbox: File Processing**



- i) Selección de un subconjunto aleatorio de individuos (filas);
- ii) Selección de un subconjunto de individuos (filas) a partir de un filtro;
- iii) Selección de un subconjunto de variables (columnas);
- iv) Concatenación de dos bases de datos (variables diferentes).
- v) Selección de un subconjunto de variables que tiene un peso máximo.

Las secciones i) y v) no serán tratadas aquí de un modo detallado. Contienen secciones "HELP" que deberían facilitar la tarea de los usuarios.

La sección i) permite dividir por 2 o 4 el tamaño original de la muestra (formado por la unión de los 2 o 4 grupos). Esto permite probar los análisis de un modo más económico, pero también validar las estructuras observadas.

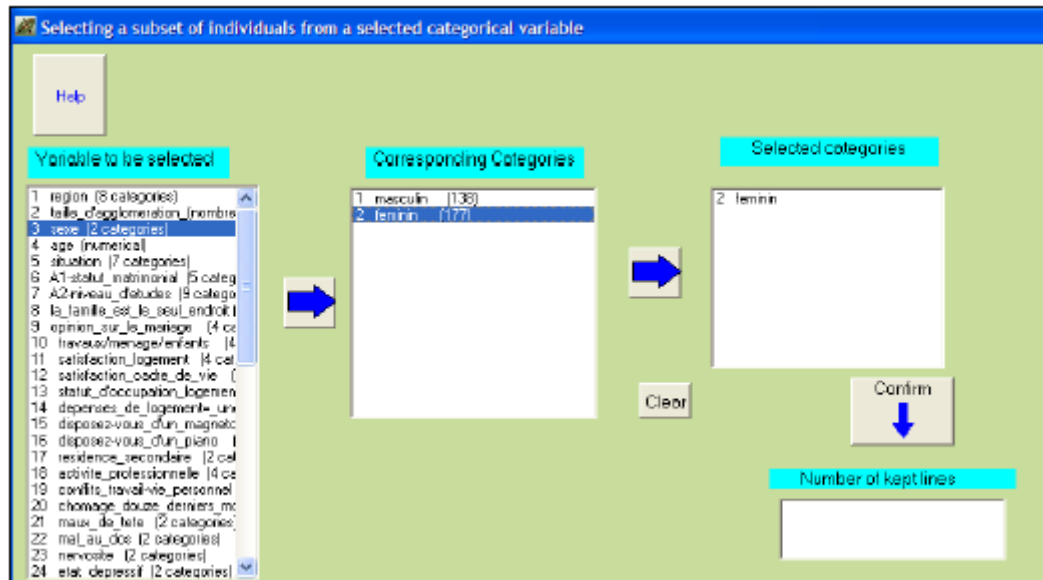
La sección v) es muy particular y responde a la situación práctica siguiente: si los datos contienen un gran conjunto homogéneo de n variables numéricas cuya suma sobre los individuos tiene un sentido, entonces podemos seleccionar p variables ($p < n$) de sumas más fuertes. Ejemplo: tenemos para 10.000 individuos 1.200 variables (número de visitas a 1.200 sitios Web). Podemos seleccionar los 400 sitios más visitados, para trabajar en este único subconjunto.

V.2.1 Selección de un subconjunto de individuos por filtrado

Es frecuente tener que trabajar de un modo más profundo con una subpoblación, por ejemplo las mujeres, las personas que tienen acceso a Internet en su domicilio, etc. Entonces es conveniente seleccionar una extensión de archivo Dtm-Vic, sin tener que reimportar los datos a partir de la base inicial.

- Haga Clic sobre **Selecting a subset of individuals**.
 - Aparece una ventana.

- Abrir los archivos diccionario (por ejemplo **MCA_dic.txt**), de datos (por ejemplo **MCA_dat.txt**), poner en una lista las variables, abrir, si existe, el archivo de texto de las preguntas abiertas, luego continuar.
 - Aparece una nueva ventana.



- Seleccione la variable nominal en la 1era ventana (por ejemplo 3 -Sexo), trasladarla a la 2a ventana.
- Seleccione la modalidad de filtrado (por ejemplo “femenino”).
- Haga Clic en **Confirm**. El número de filas (individuos) guardadas se exhibe en la ventana “Number of kept lines” y corresponde al número de individuos de la categoría establecido en la ventana "Corresponding Categories", categoría que no se exhibe más después del procedimiento de confirmación.
- Haga Clic en **Update data file and text file**.

Se crea un archivo cuyo nombre por defecto es: **dtm_data_Subset.txt** en la carpeta **EX_A03.MultCorAnalysis**. El archivo diccionario **MCA_dic.txt** queda sin cambiar. Se finaliza la operación.

V.2.2 Selección de un subconjunto de variables

- Haga Clic en **Selecting a subset of variables**. Aparece una ventana.
- Abrir los archivos diccionario y de datos de la base involucrada, poner en una lista las variables luego continuar. Una nueva ventana aparece.
- Seleccionar en la 1era ventana el conjunto de las variables que hay que conservar en la nueva base, trasladarlas a la 2a ventana.

- Hacer Clic en **Update data file and dictionary**.

Se crean dos archivos **dtm_dic_SELVAR.txt** y **dtm_dat_SELVAR.txt** en la carpeta **EX_A03.MultCorAnalysis**.

V.2.3 Concatenación de conjuntos de variables

Esta opción permite concatenar dos bases de datos de Dtm-Vic para crear una nueva base de datos que reúne dos conjuntos de variables (operación útil cuando los archivos recibidos son segmentados, como en el caso de las versiones de Excel para las cuales el número de columnas es limitado) **¡Atención! Ambas bases deben contener a los mismos individuos en las filas, clasificados en el mismo orden.**

- Haga Clic en **Concatenating 2 dtm files with 2 distinct sets of variables**.
 - Una ventana aparece.
- Abrir ambos archivos de datos luego de los diccionarios a concatenar. Se exhiben en cada una de las cuatro ventanas.
- Haga Clic en **Merge Sorted Files**.
 - Una serie de ventanas se muestran sucesivamente. Las dos primeras especifican la integración de ambos archivos de datos **In file, 0 individuals have no counterparts**: responder **OK** Una tercera ventana da al número de individuos del nuevo archivo: Responder **OK**

Finalmente, una cuarta ventana indica que se ha efectuado el procedimiento "merge" de ambos archivos de datos: responder **OK** Los identificadores de ambos archivos aparecen en la ventana de abajo.

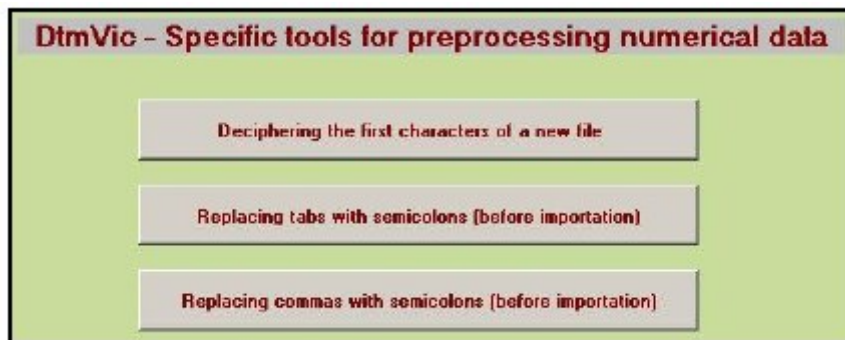
- Haga Clic en **Merge dictionaries**.
 - Una ventana indica que el procedimiento "merge" de los diccionarios se ha efectuado: responder **OK** y haga Clic en **Exit**.

Son creados dos archivos **dtm_dico_new** y **dtm_data_new**.

V.3. Herramientas específicas de pre tratamiento

V.3.1 Datos numéricos y textuales

El botón **Toolbox: Preprocessing (numerical)** propone herramientas elementales de pretratamiento para la importación o la utilización de datos numéricos y textuales



Cuando se recibe un archivo de datos (Internet, Pen Drive, DVD), es útil verificar la naturaleza de los caracteres presentes (numéricos, alfanuméricos, separadores, puntuación, tabulaciones eventuales, etc.).

El primer botón **"Deciphering the first characters of a new file"** nos da el código ASCII correspondiente a los 6.000 primeros caracteres de un archivo; una operación (a veces) tan útil como elemental.

El segundo botón, **Replacing Tabs with semicolons**, es útil en el momento de la importación de un archivo Excel[®]. En ciertas versiones de Excel, el separador del formato ".csv" es una coma (caso frecuente en los países para los cuales la notación decimal utiliza puntos en lugar de las comas, la coma puede entonces jugar un papel de separador de registro). El paso por la copia de seguridad, con tabulaciones como separadores, es entonces más práctico. Por lo tanto, es necesario utilizar este botón.

¡Atención! Si tal archivo ya contiene puntos y comas, la transformación no podrá efectuarse.

El tercer botón, **Replacing commas with semicolons**, es útil cuando el archivo proporcionado ya ha sido guardado con comas como separadores. Como anteriormente, si el archivo ya contiene puntos y comas, la transformación no podrá efectuarse. Conviene entonces reemplazarlos por otro símbolo antes de accionar el botón.

V.3.2 Datos textuales únicamente

El último botón **Toolbox: Preprocessing texts** propone algunos procedimientos en vista de la importación o la utilización directa de textos.

i) Conversión de los textos en minúsculas

El botón **"Conversion of all the characters of a text into lowercase"** transforma todos los caracteres en minúsculas. Esto permite ganar información en términos de frecuencias para el vocabulario trivial, pero serán necesarios tratamientos preliminares para lograr, por ejemplo,

la *homonimia* entre ciertos nombres propios (de lugares) y nombres comunes (Rosario, Granada, La Paz, Dolores, Salvador, Providencia, León). La etapa CORTEX (después del botón "Create" del menú principal) debe, en general, intervenir antes de este tipo de transformación.

ii) supresión de marcas XML abiertas y cerradas "<" y ">" y del texto que ellas pueden contener

El segundo botón "Suppress in a text the chains such as <...>" es útil si el texto transmitido contiene marcas que no se quieren tener en cuenta (textos formateados para el software Lexico3 por ejemplo). No obstante, este tipo de transformación debe realizarse después de que el texto haya sido segmentado a partir de ciertas marcas.

iii) Reimportación en DtmVic de un archivo de tipo Dtmic (tipo 1 o 2) que ha estado sujeto al software (gratuito) TreeTagger

El botón: Re-importing a Dtm text file after WinTreeTagger permite lematizar un texto (reemplazar las formas gráficas por el lema correspondiente). Igualmente permite suprimir ciertas categorías gramaticales (preposiciones, artículos, etc). Existen cuatro opciones disponibles para los textos ingleses, franceses, españoles e italianos, respectivamente. Esto supone la instalación del software (gratuito) WinTreeTagger.

TreeTagger: autor: Helmut Schmid, IMS, University of Stuttgart, TreeTagger es, en principio, un analizador morfosintáctico independiente de los idiomas. Las informaciones y la descarga se hacen a partir del sitio Web:

Http: // www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/

Se nota que TreeTagger no tiene interfaz gráfica. (Funciona con fila de comando). Así como sugeriera Helmut Schmid, se puede utilizar la interfaz Windows más amigable WinTreeTagger realizada por Ciarn O' Duibhin.:

Http: // [www.smo.uhi.ac.uk/~oduibhin/oideasra/interfaces/winttinter face.htm](http://www.smo.uhi.ac.uk/~oduibhin/oideasra/interfaces/winttinter%20face.htm)

Notese que el archivo que alimenta a WinTreeTagger debe, obligatoriamente, ser un archivo de texto en el formato Dtm-Vic: el nuevo archivo a importar, nacido de WinTreeTagger, contiene tres columnas separadas por tabulaciones. - Primera columna: palabra - Segunda columna: Etiqueta gramatical - Tercera columna: lema. Tal archivo contiene tantas filas como palabras que hay y signos de puntuación. (Ver "Help" de Dtm-Vic). Este es el archivo que el procedimiento devuelve en formato Dtm-Vic.

iv) Fragmentación de un texto en formato 1 (textos separados por **) en textos de formato 2**

El botón: Fragmentation of a Dtm text (format 1) into a Dtm Text (format 2) permite fragmentar los textos importantes en pequeñas unidades de longitudes variables. Estas unidades están formadas por una fila, dos filas... de los textos iniciales (se trata de algo similar a una fragmentación en unidades de contexto). Veremos más abajo que la longitud de las filas puede ser modificada dentro de ciertos límites. Se crea una variable nominal para conservar la información que relaciona a las unidades con los textos iniciales. (Ver "Help" in situ).

v) Cambio de longitud de las filas del texto.

El botón [Changing the size of the lines in a DtmVic text file](#) permite una importación o un reformateado de los archivos de texto. Al principio, disponemos de textos en formato DtmVic (1 o 2) sin limitación para la longitud de las filas. Al final: textos con filas de una longitud elegida por el usuario, (pero <200 caracteres). Este procedimiento permite importar textos con filas muy largas, pero también formatear las unidades de contexto (ver punto iv más arriba).

Vj) Procedimiento de las respuestas vacías

Finalmente el último botón [Add empty lines for ampty responses...](#) pone en marcha un procedimiento restringido y especializado que permite hacer respetar la limitación "una fila vacía por cada respuesta abierta vacía" para archivos que utilizarían dos separadores consecutivos. Esta es a veces útil después de la reimportación, según TreeTagger, de un archivo de tipo 2.

VI. Otros análisis con Dtm-Vic: Visualizaciones elaboradas, Contigüidad, Grafos, Imágenes

La orientación principal de Dtm-Vic es el análisis exploratorio multi-dimensional de datos numéricos y textuales, con validación sistemática de los resultados (por la complementariedad de enfoques diferentes y por los métodos de *Bootstrap*). En este capítulo se presentan otras aplicaciones y herramientas que permiten considerar análisis más elaborados.

En la carpeta: **DtmVic-Examples/DtmVic-Examples_C_NumData**, una serie de ejemplos retoma las técnicas de análisis de base sobre datos numéricos. Esta serie va a darnos la ocasión de profundizar las herramientas **Visualization** y **Contiguity** de la parte VIC de Dtm-Vic: **VIC steps**. Estudiaremos luego la aplicación de los análisis, en ejes principales, en las visualizaciones de grafos y en la compresión de imágenes

1. El ejemplo 1, en la carpeta **EX_C01.PCA_Semio**, describe un conjunto de variables numéricas (un extracto de datos semiométricos) mediante el análisis de componentes principales. Los ejes principales son completados por una clasificación y una descripción automática de clases (un archivo de comando totalmente preparado nos permite acceder, directamente, a la fase "VIC"). Presentaremos aquí sólo el submenú "Visualización" de la fase "VIC": Visualización de clases (o categorías) utilizando símbolos o colores, envolventes convexas o elipses de densidad para las clases, el trazado del árbol de longitud mínima (Minimum Spanning Tree), las visualizaciones de los grafos de los vecinos más próximos, clasificaciones de tipo k-means "sobre la marcha", etcetera

2. El ejemplo 2, en la carpeta **EX_C02.PCA_Contiguity**, analiza un conjunto clásico de variables numéricas (los datos IRIS de Anderson y Fisher, bien conocidas por los estadísticos) por el análisis en componentes principales, la clasificación, el análisis de contigüidad y el análisis discriminante. Este ejemplo retoma los procedimientos de base del precedente ejemplo 1: Análisis en componentes principales y clasificación (clustering) de un conjunto de datos numéricos, con diferentes herramientas de visualización, que implican también una variable nominal específica (la variable que identifica las 3 especies de iris). El ejemplo presenta a continuación las mejoras aportadas por el análisis de contigüidad, cuyo el análisis lineal discriminante es un caso particular.

3. El ejemplo 3, en la carpeta **EX_C03-Graphs** describe tres tipos simples de grafos de planos simétricos, principalmente por medio del análisis de correspondencias. Contrariamente a los ejemplos anteriores el directorio contiene varios juegos de datos: un grafo en forma de damero, un ciclo, y grafos empíricos que representan regiones de Japón y de Francia. Estos ejemplos quieren tender un puente entre las diferentes posibilidades del software Dtm-Vic: un mismo grafo puede provenir de datos de entradas diferentes: datos numéricos, datos textuales y también en este caso un "formato externo" específico para los grafos.

El ejemplo 4, en la carpeta **EX_C04.Images** tiene más bien una vocación pedagógica: mostrar las propiedades de compresión numérica de los métodos en ejes principales (y de las series discretas de Fourier, en calidad de comparación). A las imágenes que necesitan un

formato específico, esta aplicación no se ajusta a las cadenas de tratamiento más comunes de Dtm-Vic. Una interfaz especializada se obtiene con el botón **SVD and CA of Images** de la sección "DtmVic Images " del menú principal.

Los análisis de base a los cuales los ejemplos 1 - 3 recurren son presentados en el capítulo II. No volvemos entonces sobre la preparación interactiva del archivo de comando (o: archivo parámetro) y de los análisis. Directamente presentamos aquí estos análisis a partir del archivo de comando ya preparado y proporcionado con cada ejemplo.

VI.1. Datos numéricos: "Semiotría"

El ejemplo 1, en la carpeta **EX_C01. PCA_Semio**, analiza un conjunto de variables numéricas (extraído de "*datos semiométricos*") mediante el análisis en componentes principales. Los principales ejes de visualización se completan por una clasificación, con una descripción automática de las clases. El procedimiento "Vizualisation" propone diferentes herramientas de visualización de las envolventes convexas o de las elipses de densidad para las clases, el trazado del árbol de longitud mínima (*Minimum Spanning Tree*) y la visualización de los grafos de los vecinos más próximos. Una nueva clasificación de las variables (o de las observaciones o individuos) a través de un método de tipo k-means puede ser obtenida y visualizada, iteración después de iteración, a partir del submenú "Visualización".

VI.1.1. Los datos semiométricos

En la inmensa mayoría de las encuestas de marketing, es frecuente incluir informaciones sobre los modos de vida y los valores de las personas interrogadas. Estas informaciones son generalmente obtenidas a través de una serie de preguntas que describen las actitudes y las opiniones.

La "Semiotría" es una técnica introducida por Jean-François Steiner⁹. La idea básica consiste en insertar en el cuestionario, una serie de preguntas compuestas únicamente de palabras (una lista de 210 palabras es actualmente utilizada, pero acá se usa una lista abreviada que contiene un subconjunto de 70 palabras). Las personas interrogadas deben anotar estas palabras según una escala que comporta siete niveles. El nivel más bajo (1), se relaciona a un sentimiento "muy desagradable" (o displacentero) frente a frente con la palabra presentada, "el nivel más alto (7) se refiere a una sensación "muy agradable (o placentera) para el sujeto.

El tratamiento de los cuestionarios por el Análisis en Componentes Principales pone en evidencia una estructura estable (la estabilidad referida al espacio de los primeros 8 ejes principales). Propiedades muy similares son observadas en diez países diferentes, a pesar de los problemas que surgen de la traducción de la lista de las palabras. En cuanto a los estudios de "estilos de vida" los espacios obtenidos se utilizan para posicionar los productos, las marcas o servicios a través de estudios de investigación de mercados

Los tres archivos que componen este ejemplo se encuentran en el directorio **DtmVic-examples/DtmVic-Examples_C_NumDa/ EX_C01.PCA_Semio**.

⁹ Para más información, referirse a la obra: "La sémiotrie" por L. Lebart, M. Piron, JF Steiner; Ed. Dunod, París, 2003. Este libro puede ser descargado desde el sitio: www.dtmvic.com (sección "Publicaciones").

1. El archivo de datos: `PCA_semio.dat.txt`

Este ejemplo es de tamaño reducido y comprende a 300 respondentes (en lugar de 1.000 o 2.000 que son los tamaños usuales de las muestras de encuesta semiométrica) y 76 variables: 70 palabras (las calificaciones asignadas a estas palabras están consideradas aquí como variables numéricas) y 6 variables nominales que describen las características de los respondentes.

2. El archivo de diccionario: `PCA_semio.dic.txt`

El archivo diccionario contiene los identificadores de las 76 variables. En el diccionario interno de DtmVic, los identificadores de categorías deben comenzar: "columna 6" [una fuente a intervalo fijo tal como "courrier" puede ser útil para facilitar este tipo de formato].

3. El archivo de comando: `EX_C01_Param.txt`

La fase de cálculo del análisis se descompone en "etapas". Cada etapa necesita algunos parámetros descriptos brevemente en el menú principal DtmVic (botón: `Help about command parameters`). Notemos que un archivo de comando similar al archivo de comando `EX_C01_Param.txt` puede también ser generado haciendo Clic en el botón: `Create` del menú principal (pasos básicos), como lo señalado en el capítulo 2 de este manual.

Se exhibe una ventana "Select a basic analysis".

Haga Clic luego en: `Principal Components Analysis` situado en la sección "Numerical Data", y siga las instrucciones

VI.1.2. Cálculos de base (PCA y clasificación)

(Ejecución del ejemplo C.01 "semiometria" y lectura de los resultados).

a. Apertura del archivo parámetro

- Haga Clic en el botón: `Open an existing command file` de la sección `Command File` (menú principal)

Luego, busque la carpeta `DtmVic-Examples_C_NumData` en `DtmVic-examples`. En este directorio (o carpeta), abra el directorio `EX_C01.PCA_Semio`

Abra el archivo de parámetros: `EX_C01_Param.txt`.

- El archivo parámetro se exhibe en la ventana del editor de texto:

```

#-----
LISTP = yes, LISTE = no, LERFA = yes # global parameters
#
NDICZ = 'PCA_semio.dic.txt' # Dictionary file
NDONZ = 'PCA_semio.dat.txt' # Data file

STEP ARDAT
===== Reading data and dictionary
NIDI = 1, NIEXA = 300 NQEXA = 76

STEP SELEC
===== Selecting active and supplementary variables
LSELI = TOT, IMASS = UNIF, LZERO = NOREC, LEDIT = short
CONT ACT 1--70
NOMI ILL 71--76
END

STEP STATS
===== Basic descriptions
LHIST=no

STEP PRICO
===== Principal component analysis
LCORR = 2, .....

```

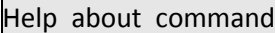
Verificar que los archivos de datos y diccionarios inscriptos en el archivo parámetro son coherentes con los del directorio.

Son efectuadas diez "etapas":

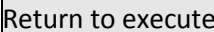
- ARDAT (Almacenamiento de los datos),
- SELEC (Selección de los elementos activos e ilustrativos),
- PRICO (análisis en componentes principales)
- DEFAC (descripción breve de los ejes factoriales),
- RECIP (Clasificación ascendente jerárquica – método de los vecinos recíprocos),
- PARTI (Corte del dendrograma produce por la etapa precedente, y optimización de la partición obtenida)
- DECLA (Descripción automática de las clases de la partición),
- SELEC (Selección de una variable específica),
- EXCAT (Extracción de la variable específica, seleccionada en la etapa precedente SELEC, para ser utilizada en la siguiente)

En este archivo de comandos, la etapa SELEC desempeña como siempre un papel fundamental para decidir cuáles variables serán activas o ilustrativas. La etapa RECIP efectúa una clasificación jerárquica de las observaciones utilizando el algoritmo "de la búsqueda en cadena de vecinos recíprocos" y la etapa PARTI corta el árbol obtenido según el número de clases fijado *a priori*, luego optimiza la partición por iteraciones de tipo "k-means" (RECIP y PARTI ejecutan un algoritmo "híbrido" de clasificación¹⁰).


¹⁰ "Statistiques Exploratoire Multidimensionnelle" (4a edición, L. Lebart, M. Piron, A. Morineau, Ed. Dunod, París, 2006).

El editor de texto interno de Dtm-Vic contiene también un botón  **Help about command parameters** que da brevemente (en Inglés) el significado de cada uno de los parámetros.


Nosotros no modificaremos el archivo de comando

- Haga Clic en  en la barra para volver al menú principal.

b. Ejecución del archivo de comando (archivo parámetro)

- Haga Clic en:  de **Command File**



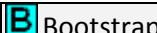

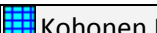
Las etapas básicas de cálculo presentes en el archivo de comando son ejecutadas: almacenamiento de datos y el diccionario, la elección de los elementos activos e ilustrativos, estadísticas elementales, el análisis en componentes principales de la tabla seleccionada, replicaciones "bootstrap" de la tabla, breve descripción de los ejes, la clasificación, descripción detallada de las clases. Las 9 etapas descritas más arriba se muestran al final de la ejecución. Para examinar los resultados numéricos, como anteriormente:


- Haga Clic en:  de **Result Files**

Los resultados numéricos son del mismo tipo que los presentados en la sección II.1.3 (Análisis en componentes principales, capítulo II).

VI.1.3. Visualización y lectura de los resultados

Procedemos en primer lugar como en el capítulo II a propósito de la visualización de los resultados utilizando las posibilidades ofrecidas por la segunda fase: **VIC: Visualization, Inference, Clasificatinn steps.**

El análisis realizado permite examinar los ejes y los planos factoriales: botón  **AxesView** y  **PlaneView**, la validación de las posiciones de los puntos sobre los gráficos por *Bootstrap*, con:  **Bootstrap**, la clasificación con el botón:  **ClusterView** y los mapas auto-organizados con:  **Kohonen Map**.

Las funcionalidades de estos cuatro primeros botones han sido descritas cuando se trataron los ejemplos de los capítulos II y III. En esta sección vamos a enfocarnos en las funcionalidades del botón  **Visualization**.

Esta opción propone herramientas de visualizaciones complementarias de los planos factoriales y de la clasificación: elipse de densidad o envolventes convexas de las clases;

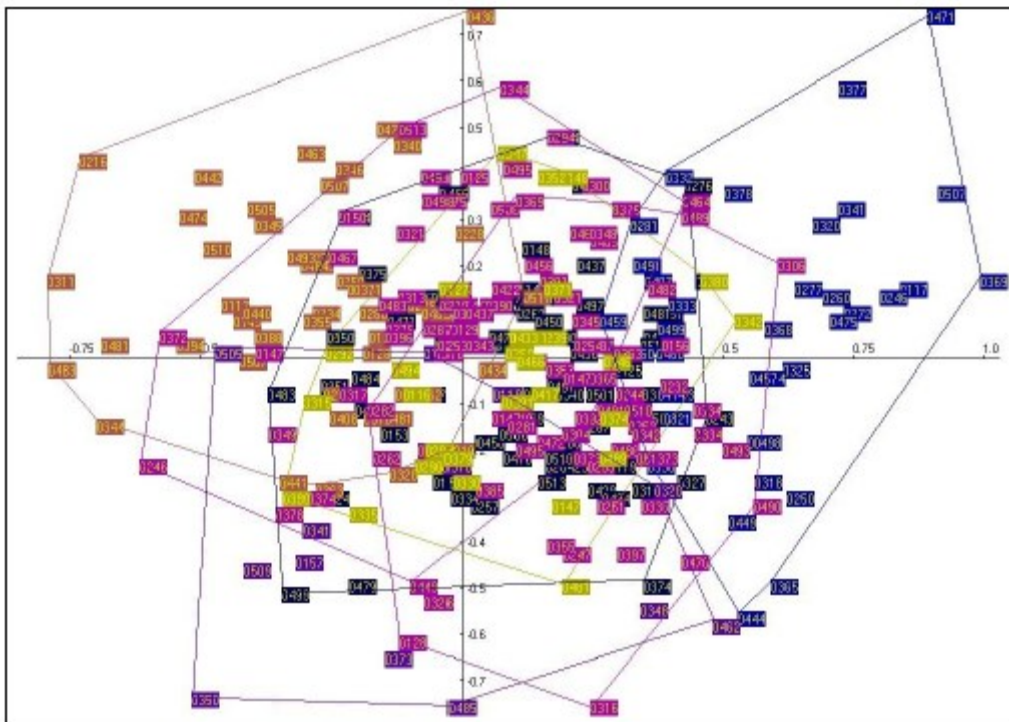
trazado del árbol de longitud mínima, trazado de los vecinos más próximos en los planos factoriales; visualización pedagógica de la construcción progresiva de las clases (casos del procedimiento k-means); visualización en los planos factoriales de los mapas de Kohonen y de ciertos grafos.

a. Visualización que utiliza la partición pedida en el archivo de comando (etapas RECIP y PARTI)

- Haga Clic en el botón **V** Visualization
 - Aparece una ventana titulada "DTM-visualization: loading files, selecting axes".
- Haga Clic en **Load coordinates**. En el submenú correspondiente, elegir, en primer lugar, el archivo: **ngus_ind.txt**. Son seleccionadas las principales coordenadas de los individuos (filas).
 - Una subventana da las características del archivo.
- Haga Clic luego en **Load or create a partition**. En el submenú correspondiente, seleccione la partición obtenida anteriormente en la etapa de cálculo. Elegir entonces **Load partition File** y abrir el archivo **part_cla_ind.txt** (clases de la partición para los individuos).
- Haga Clic en **Graphics**, luego en la ventana "Sélection des axes", elegir los ejes 2 y 3 (que constituyen el primer "plano semiométrico", porque el eje 1 es un "eje de notación").
- Haga Clic luego en **Continue** luego en **DISPLAY**.
 - Se exhibe el plano factorial (2, 3)

En la barra vertical, a la izquierda de la ventana "Graphics" figura una serie de botones: apretamos un botón para activarlo (color rojo), y pulsamos de nuevo para desactivarlo (color negro).

- El botón **C.Hull** (*Convex Hull* = Envoltente convexa) traza la envoltente convexa de cada clase. Presione este botón: la figura que se observa más abajo representa a los 300 individuos en el plano (2, 3), con un color por clase y una envoltente convexa por clase.



Envolturas convexas (Convex Hulls) de las 7 clases en el plano (2, 3) después de activación del botón: "C.Hull" y luego del botón: "Colours"

b. Visualización a partir de una variable nominal

La visualización precedente va a ser retomada, pero en lugar de utilizar una partición prevista por un algoritmo de clasificación, vamos a utilizar la partición inducida por las categorías de una variable nominal específica. Se trata de la variable número 76 (sexo), seleccionada y extraída a través las dos etapas SELEC y EXCAT (al final del archivo de comando).

- Haga Clic de nuevo en **Visualization**
- En la ventana titulada "DTM-visualization: Loading files, Selecting axes", haga Clic en **Load coordinates**

En el submenú correspondiente, elegir de nuevo el archivo: "ngus_ind.txt". Las coordenadas de los individuos (filas) son seleccionadas.

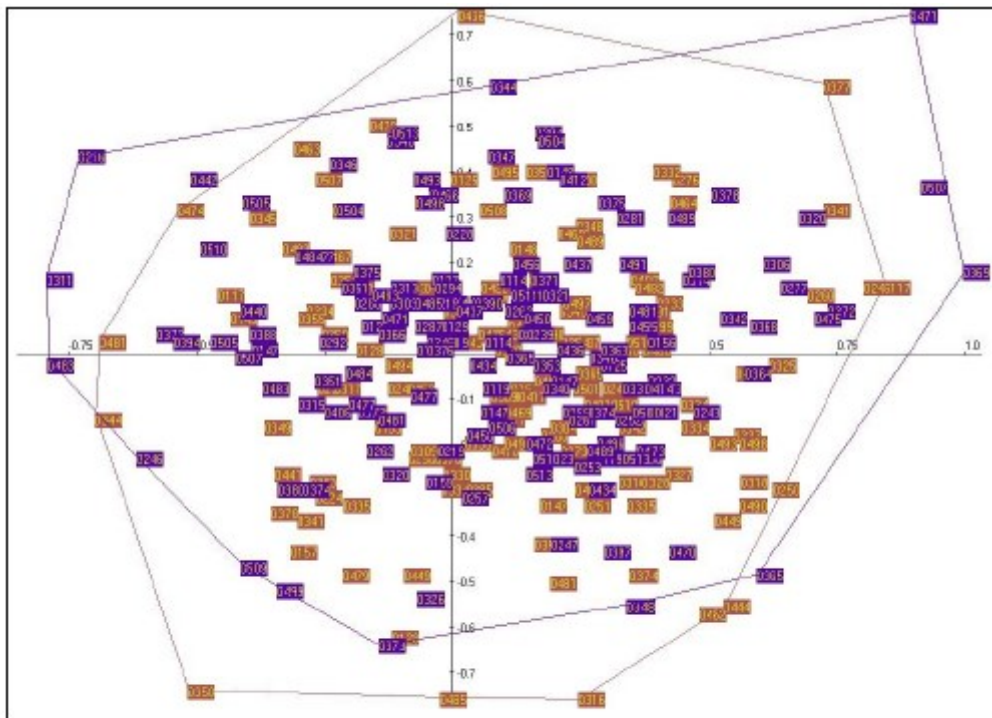
- Haga Clic luego en **Load or create a partition**

En el submenú correspondiente, elija el archivo "part_cat.txt".

La partición inducida por las categorías de la variable 76 (sexo) es cargada.

- Haga Clic en **Graphics** luego elija, todavía, los ejes 2 y 3 haga Clic en **Continue** luego en **DISPLAY**. El plano factorial (2,3) se exhibe.

- Haga Clic en el botón **C.Hull** (*Convex Hull* = Envoltente convexa). La figura más abajo representa entonces a los 300 individuos en el plano (2, 3), con un color por clase y un envoltente convexa por clase.



Envoltentes convexas de dos agrupaciones de hombres/mujeres en el plano semiométrico (2, 3) (después de uso del botón "Colours" para hacer contrastar a ambas subpoblaciones).

Comentario: Las dos categorías "Hombre" [violeta] y "Mujer" [marrón] de hecho están estrechamente vinculadas al eje vertical 3 (podemos verificarlo a partir de las zonas de confianza *bootstrap*). Pero este vínculo es a penas visible cuando directamente se mira las envoltentes convexas de las dos agrupaciones correspondientes a estas dos categorías de respondentes. Este resultado (casi) paradójico ilustra la diferencia entre "estadísticamente significativo" (que es el caso aquí tratado) y el "netamente distinto" (que no es el aquí tratado).

C. *Árbol de longitud mínima y vecinos más próximos en el espacio de las variables (palabras)*

- Haga Clic en **Visualization**
 - Aparece una ventana denominada "DTM-visualization: loading files, selecting axes"
- Haga Clic en **Load coordinates**. En el submenú correspondiente, elegir el archivo: **ngus_var_act.txt** para una clasificación de **variables**; se seleccionan las coordenadas principales de las variables activas.

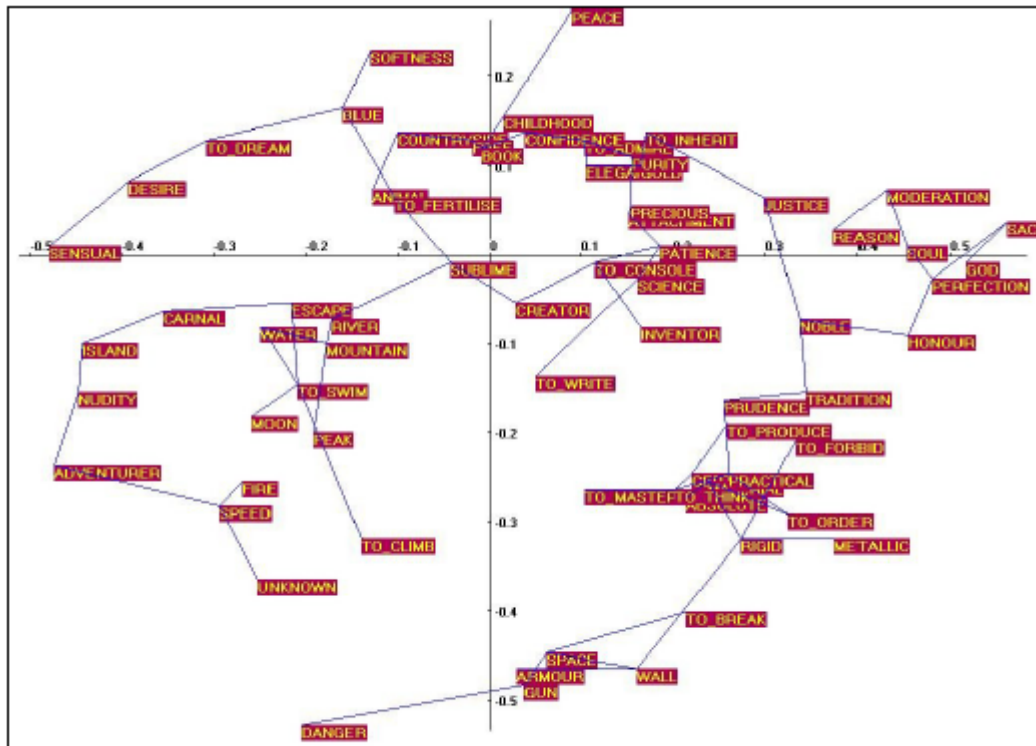
- Una subventana proporciona las características del archivo.
- Haga Clic luego en **Load or create a partition**. Elija entonces **No partition**
- 1. Haga Clic en **Min. Span. Tree** (Minimum Spanning Tree). Elija el número de ejes que servirán para calcular el árbol de longitud mínima; por ejemplo aquí los 3 primeros ejes. Confirmar haciendo Clic **OK** en el número de ejes conservados.
- 2. Haga Clic en **N.N** (búsqueda de vecinos más próximos (*Nearest Neighbours*) limitado en 20 NN). Responder **OK** a la búsqueda de los vecinos más próximos.
- 3. Haga Clic en **Graphics** a continuación elija los ejes 2 y 3 (que constituyen el primer "plano semiométrico", porque el eje 1 es "eje de notación ") en la ventana "Selection des axes", y haga Clic en **Continue** luego en **DISPLAY**.
- El Plano factorial (2, 3) se exhibe.

En la barra, a la izquierda de la ventana "Graphics", figuran cuatro familias de botones:

Sobre la barra de herramientas vertical a la izquierda, activamos un botón para acelerarlo (de color rojo), y tocamos de nuevo para desactivarlo (color negro).

- El botón **MST** (*Minimum Spanning Tree*) traza el árbol de longitud mínima.
- El botón **N.N** (*Nearest Neighbours* = vecinos más próximos) junta cada punto a sus vecinos más próximos. El botón **N.N.up** permite incrementar el número de vecinos más próximos (≤ 20).

La figura a continuación muestra el espacio de las palabras (plano 2,3) con el trazado del árbol de longitud mínima. Este árbol que está calculado en el espacio de los tres primeros ejes, aporta un complemento con relación al plano. Las figuras obtenidas a partir de los vecinos más próximos son análogas

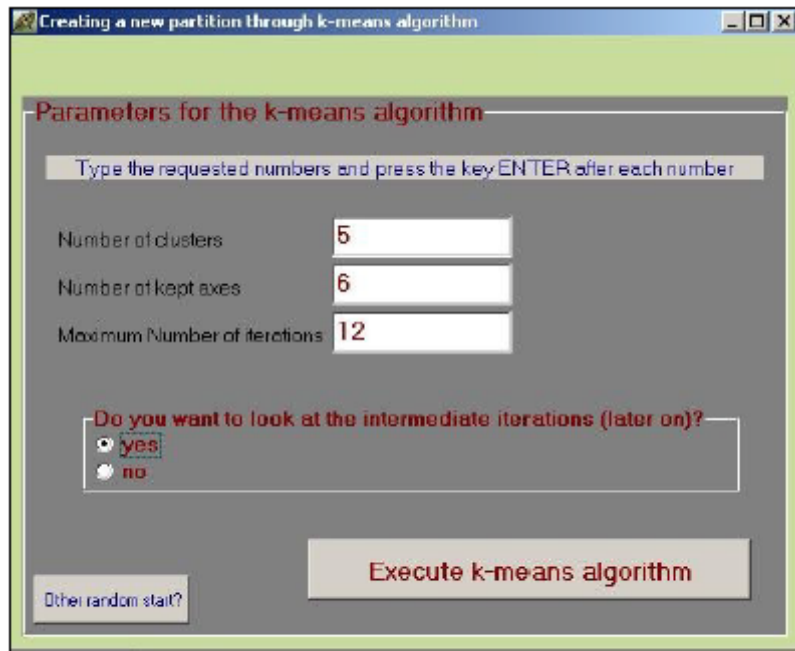


Trazado del Árbol de longitud mínima en el plano semiométrico (2, 3) (después de haber activado el botón " Colours ").

d. Cálculo directo de una partición en el menú "Visualización"

Dtm-Vic permite construir "al vuelo" (es decir por fuera del "archivo de comando") una "partición k-means" de variables (o individuos).

- Haga Clic en Visualization
 - Aparece una ventana denominada " DTM-visualization: Loading files, Selecting axes"
- Haga Clic en Load coordinates En el submenú correspondiente, elija el archivo: **ngus_var_act.txt** para una clasificación de las variables activas (para una reagrupación de individuos, seleccione el archivo: **ngus_ind.txt**).
- Haga Clic luego en Load or create a partition. En el submenú correspondiente, seleccione la opción "Create a new k-means partition". Debe luego seleccionar (figura más abajo) el número de clases deseadas, el número de coordenadas principales para los cálculos de distancias, el número máximo de iteraciones (generalmente <12) y debe marcar "yes" si usted desea visualizar las iteraciones.

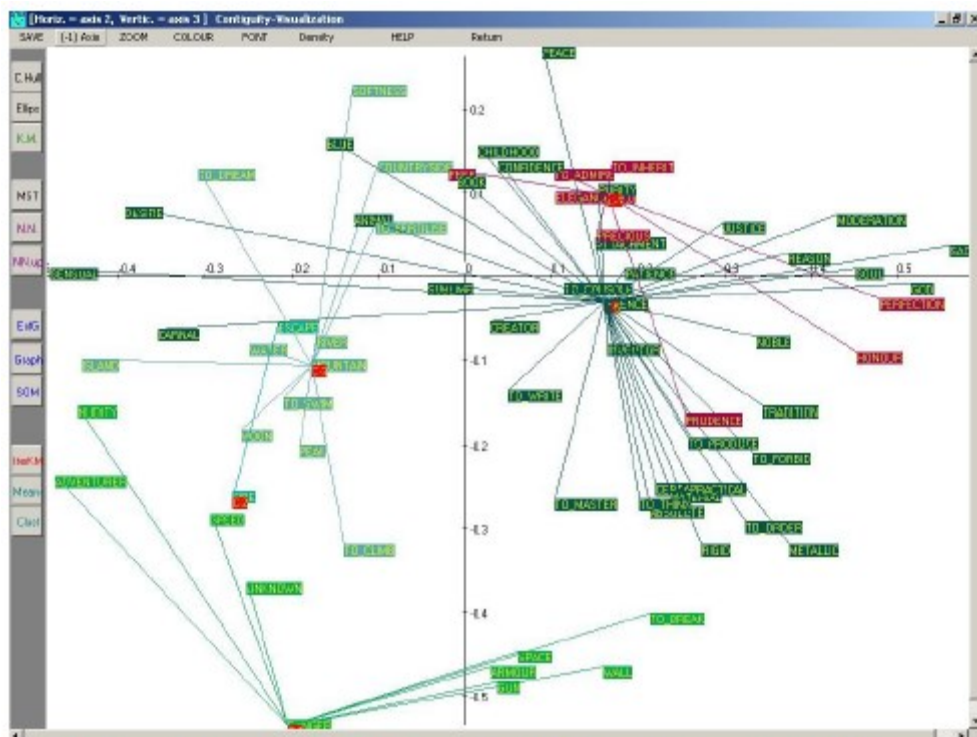


Ejemplo de la elección de 5 clases, calculadas con 6 ejes, en 12 iteraciones como máximo

Después de haber hecho Clic en **Graphics**, con carácter pedagógico, podemos visualizar en la ventana, las diferentes etapas de construcción de la partición. Luego hay que seleccionar los ejes 2 y 3, luego hacer Clic en **Continue** y finalmente hacer Clic sobre: **DISPLAY**.

En la barra vertical a la izquierda, hay, entonces que hacer Clic en **IterKM**, luego hacer Clic alternativamente en **Means** (cálculo de los centros de las clases) y sobre **Clust** (asignación de los elementos a los nuevos centros de clases) hasta que la convergencia sea lograda. Note que la partición obtenida por este algoritmo clásico de las k-medias no coincidirá, en general, con la partición inducida por los parámetros del archivo de comando

Ver el cuadro de la sección VI.1.2 precedente a propósito de los cálculos realizados por las instrucciones del archivo de comando (etapa RECIP y PARTI).



Ejemplo de visualización de la primera iteración de la construcción de la partición en 5 clases. Las variables (aquí: las palabras) están unidas por segmentos de líneas a los centros provisionales de clases a los cuales ellas son asignadas (las 5 palabras que sirven de centros provisionales de clases son identificables por un cuadrado rojo).

VI.2. Datos numéricos y contigüidad: Iris

Esta sección se refiere al análisis exploratorio de un conjunto de variables numéricas (Los datos "Iris" de Anderson y Fisher, juego clásico de datos para los estadísticos) mediante el análisis en componentes principales y la clasificación (con una descripción automática de las clases obtenidas). Añade a estos enfoques de base el análisis de contigüidad y el análisis discriminante.

La primera parte de este ejemplo es muy similar al ejemplo VI.1 de la sección anterior: análisis en componentes principales y clasificación (clustering) de un conjunto de datos numéricos, con herramientas diversas de visualización, también implicando la presencia de datos nominales.

Los párrafos que siguen presentan las mejoras aportadas por el análisis de contigüidad.

VI. 2.1 Recordatorio sobre el Análisis de Contigüidad

En el análisis de la contigüidad, consideramos el caso de un conjunto de observaciones multidimensionales (n objetos descritos por p variables, que conducen a una matriz X (n, p)). Las observaciones tienen *a priori* una estructura de grafo. Las n observaciones son así los n vértices de un grafo simétrico G , y la matriz asociada simétrica (n, n) es la matriz M ($m_{ij} = 1$ si los vértices i e i' están unidos por una arista, $m_{ij} = 0$ de otra manera).

Tal situación se produce cuando los vertices representan los puntos de una serie cronológica o de las zonas geográficas. El Análisis de contigüidad confronta las variancias locales y globales, y generaliza así el análisis discriminante, que confronta las variancias internas y globales (o, de modo equivalente las variancias internas y externas). Permite poner en evidencia los niveles de responsabilidad de los modelos observados (locales o globales). El grafo constituye pues una información externa.

En este ejemplo, vamos a tratar la situación en la cual la matriz M y la estructura del grafo no son externas, sino que provienen de la misma matriz de los datos X , siendo G por ejemplo el grafo simetrizado de los k vecinos más próximos (que provienen de una distancia entre las observaciones).

(El caso de un grafo externo forma parte de las funcionalidades del software Dtm-Vic, pero no es presentado en este manual de uso).

Se trata pues aquí de un análisis de contigüidad "intrínseca", donde se abren posibilidades interesantes de exploración de datos. La idea de deducir de los datos una métrica susceptible de poner en evidencia la existencia de clases ha sido sugerida por Art y al. (1982) y Gnanadesikan y al. (1982).

Algunas referencias para la sección VI.2.1

Art D., Gnanadesikan R., Kettenring J.R. (1982) Data Based Metrics for Cluster Analysis, *Utilitas Mathematica*, **21** A, 75-99.

Burtschy B., Lebart L. (1991) Contiguity analysis and projection pursuit. In: *Applied Stochastic Models and Data Analysis*, R. Gutierrez and M.J.M. Valderrama, Eds, World Scientific, Singapore, 117-128.

Gnanadesikan R., Kettenring J.R., Landwehr J.M. (1982) Projection Plots for Displaying Clusters, in *Statistics and Probability, Essays in Honor of C.R. Rao*, G. Kallianpur, P.R. Krishnaiah, J.K.Ghosh, eds, North-Holland.

Lebart L. (1969) *Analyse statistique de la contiguité*. Publicaciones del ISUP. XVIII, 81-112.

Lebart, L. (2000): Contiguity Analysis and Clasification, In: W. Gaul, O. Opitz and M.Schader (Eds): *Data Analysis*. Springer, Berlín, 233 - 244.

Lebart L. (2006): Assessing Self Organizing Maps vía Contiguity Analysis. *Neural Networks*, 19, 847-854.

VI.2.2 Los datos "Iris" de Fisher / Anderson:

Para los datos numéricos en formato texto de Dtm-Vic, busque el directorio **DtmVic_Examples**. En este directorio, abra la carpeta **DtmVic_Examples_C_NumData**. Luego abra la carpeta del ejemplo C.2, denominado **EX_C02.PCA_Contiguity**.

Como es habitual, se recomienda utilizar un directorio para cada aplicación, porque Dtm-Vic produce muchos archivos de texto intermedios vinculados a la aplicación.

Al principio, el directorio debe contener 3 archivos:

- a) el archivo de datos,
- b) el archivo diccionario,
- c) el archivo de comando

a) Archivo de datos: *iris_dat.txt*

El ejemplo contiene 150 observaciones y 5 variables: 4 medidas (estas variables numéricas son las longitudes de los diferentes componentes de las flores: Longitud de los sépalos, Anchura de los sépalos, Longitud de los pétalos, anchura de los pétalos) y una variable nominal que describe la pertenencia a las especies (tres especies de iris: *setosa*, *versicolor*, *virginica*). Referencia: Anderson, E. (1935). The irises of the Gaspé Peninsula, *Bulletin of the American Iris Society*, **59**, 2-5.

El archivo de datos *iris_dat.txt* comprende entonces 150 filas y 6 columnas (el identificador de filas [entre comillas] seguido por 5 valores [correspondientes a 4 variables numéricas y a una variable nominal, separadas por lo menos por un espacio]).

b) Diccionario: *iris_dic.txt*

El archivo-diccionario *iris_dic.txt* contiene los identificadores de estas 5 variables. En esta versión del diccionario interno Dtm-Vic, los identificadores de las categorías deben comenzar en la columna 6 [una fuente a intervalos fijo - *courrier*, por ejemplo - es necesaria para representar claramente este género de formato].

c) Archivo de comandos: **EX_C02_Param.txt**

La fase de cálculo del análisis se descompone en "etapas". Cada etapa necesita algunos parámetros descriptos brevemente en el menú principal de Dtm-Vic (botón: [Help about parameters](#)).

Notemos que otro archivo de comando similar (pero no forzosamente idéntico) al archivo de comando: **EX_C02_Param.txt** puede también ser generado haciendo Clic en el botón [Create](#), sección **Command File** del menú principal (" Basic Steps "). Proceder entonces como muestra el primer ejemplo de la sección II.1 destinado al análisis en componentes principales.

VI.2.3 Cálculos de base (ACP y clasificación)

(Ejecución del ejemplo C.2 "Iris" y lectura de los resultados)

a. Apertura del archivo parámetro

- Haga Clic en el botón: [Open an existing command file](#) de la sección **Command File** (menú principal). Busque en **DtmVic_Examples** el subdirectorio **DtmVic_Examples_C_NumData**. En este directorio, abrir el del ejemplo C.2 denominado **EX_C02.PCA_Contiguity**.
- Abrir entonces el archivo de comando: **EX_C02_Param.txt**
 - El archivo parámetro se exhibe en una ventana (que es también un editor de texto).

En este archivo de comandos, podemos leer, después de haber identificado ambos archivos (datos y diccionario), que se efectúan 9 "etapas":

- ARDAT (almacenamiento de los datos),
- SELEC (selección de los elementos activos e ilustrativos),
- PRICO (análisis en componentes principales),
- DEFAC (descripción breve de los ejes factoriales),
- RECIPI (clasificación jerárquica),
- PARTI (corte del dendrograma producido por la etapa precedente, y la optimización de la partición obtenida),
- DECLA (descripción automática de las clases de la partición),
- SELEC (selección de una variable nominal, en este caso),
- EXCAT (extracción de una variable nominal (3 especies de iris) seleccionada en la etapa SELEC)

Note que el botón: [Help about parameters](#) está accesible a partir de este editor de texto para aclarar (en Inglés) los parámetros de cada etapa.

b. Ejecución del archivo de comando (archivo parámetro)

Volver al menú principal y ejecutar las etapas de cálculo básico.

- Haga Clic en [Return to execute](#) en la barra para volver al menú principal.

- Hacer Clic en el botón: de: **Command File**

Esta operación ejecuta las etapas de cálculo del archivo de comando

c. Lectura de los resultados

- Hacer Clic en el botón: de: **Result Files**

El navegador abre el archivo HTML denominado "imp.html" que contiene los principales resultados de las etapas precedentes de cálculo básico. Después de la lectura de estos resultados numéricos, vuelva al menú principal.

VI.2.4. Visualización y lectura de los resultados

Así como en el ejemplo anterior C.1 que se refiere a la semiometría, vamos ahora a utilizar las funcionalidades del botón .

a. Visualización a partir de una partición inducida por una variable nominal (especie de iris)

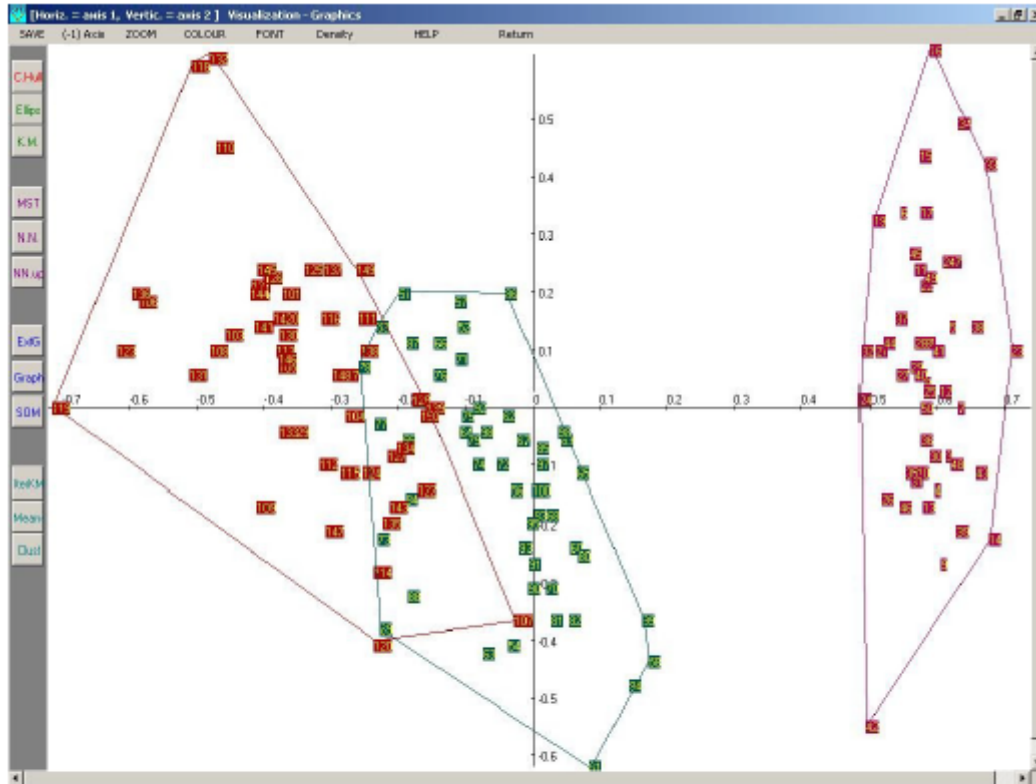
Vamos a visualizar las diferentes especies de flores (variable n ° 5) en el plano generado por los primeros ejes principales del ACP.

- Haga Clic en
 - Aparece una ventana denominada "DTM-visualization...".
- Haga Clic en . En el submenú correspondiente, elegir, al principio, el archivo: **ngus_ind.txt**. Las principales coordenadas de los individuos (filas) son seleccionadas.
- Haga Clic luego en . En el submenú que corresponde, elija entonces "Load partition files" y abrir el archivo **part_cat.txt**, la partición inducida por las 4 categorías de la variable 5 (las 4 especies de iris). Esta partición ha sido elegida y extraída a través de las 2 últimas etapas SELEC y EXCAT del archivo de comando de más arriba.
- Haga Clic en luego elija los ejes 1 y 2 (por defecto) en la pequeña ventana "Selections of axes" y haga Clic en "Continue" luego en .

En la nueva ventana denominada "Visualization - Graphics" se muestran los individuos en el plano generado por los ejes seleccionados. Un color aleatorio se atribuye a cada categoría. El botón permite probar un nuevo juego de colores.

Sobre la barra de herramientas vertical a la izquierda, apretamos un botón para activarlo (color rojo), y pulsamos de nuevo para desactivarlo (color negro)

- El botón **Density**, para mayor claridad, permite reemplazar los identificadores de los individuos por un solo carácter que recuerda su clase (el identificador y el número de la clase se obtienen haciendo Clic en el botón izquierdo del mouse en la cercanía de los puntos).
- Apriete el botón **C.Hull** (*Convex Hull* = envolvente convexa) que traza la envolvente convexa de cada clase. El trazado aparece más abajo.



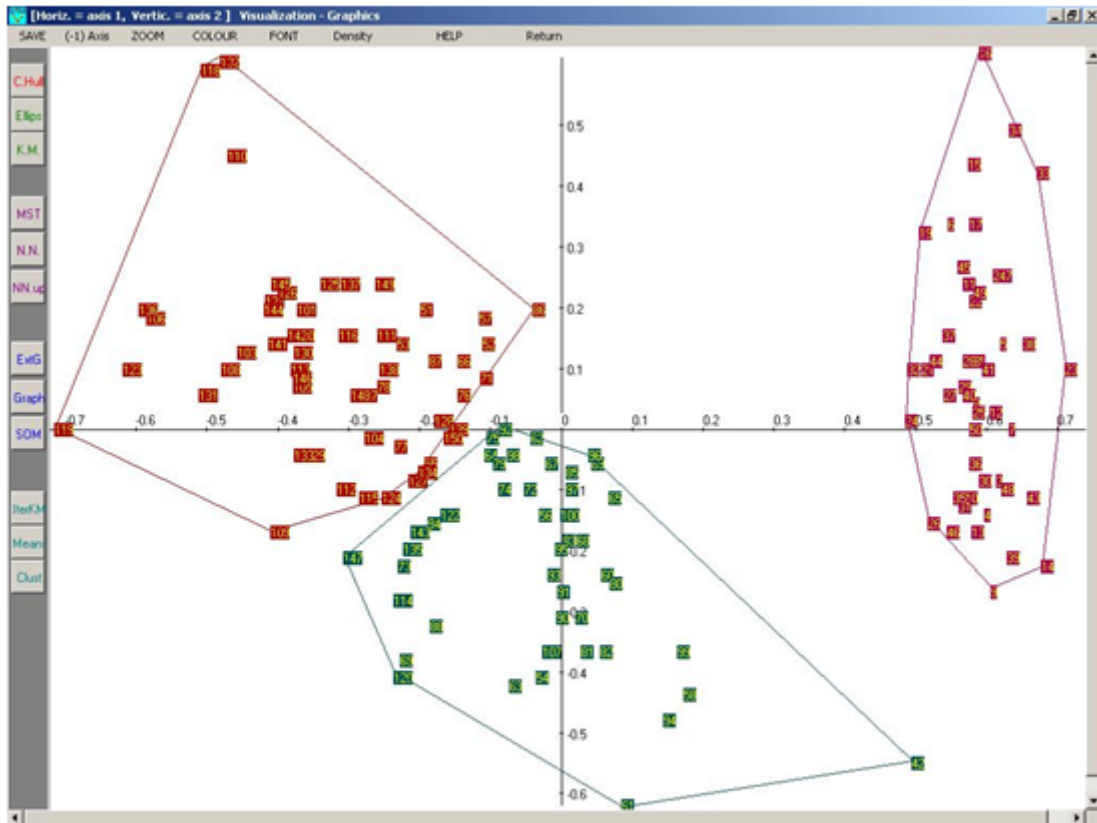
Plano principal del ACP de las 4 variables continuas (medidas) con trazado de las envolventes convexas correspondientes a las tres especies de iris. La identificación de las tres especies por colores diferentes es realizada *a posteriori*, después del análisis en componentes principales. Vemos que dos especies se superponen sobre este plano principal

En esta etapa, obtuvimos una visualización de los 150 individuos, con las envolventes convexas correspondientes a las tres especies. Esta es una muestra clásica, en el plano principal del ACP, que revela, sobre la derecha, que la primera especie *setosa* (cantidad = 50) está bien separada de las especies dos y tres, las que se superponen parcialmente.

b. Visualización de una partición en tres clases sin supervisión

Vamos ahora a volver al menú principal y a rehacer la visualización precedente, pero en lugar de cargar la partición inducida por las 4 categorías de la variable 5 (4 especies de iris), vamos a cargar una partición en tres clases producida por el algoritmo de clasificación contenido en las etapas de base: esta partición corresponde a las etapas RECIPI y PARTI (ver el archivo de comando). Ella no implica que es conocida la división en especies, de donde sale la denominación de partición sin supervisión.

- Haga Clic en **V Visualization**
 - Aparece la ventana "DTM-visualization...".
- Haga Clic en **Load coordinates**. En el submenú, elegir el archivo: **ngus_ind.txt**. Las principales coordenadas de los individuos (filas) son seleccionadas.
- Haga Clic luego en **Load or create a partition**. En el submenú correspondiente, elija entonces **Load partition File** y abrir el archivo **part_cla_ind.txt** (partición en 3 clases que resulta de las fases RECIP y PARTI).



Mismo plano principal que en la figura anterior. ¡Atención! Los colores diferencian las clases (que surgen del algoritmo de clasificación no supervisada) y tampoco las especies. La clasificación no supervisada en tres clases consigue aislar únicamente la clase de la derecha. Las otras dos especies son mezcladas en el seno de las dos clases restantes.

Así como se sospechaba, la partición obtenida directamente a partir de las medidas numéricas, ignorando la especie, no es capaz de separar las tres especies. Sólo la especie "setosa", bien separada de las dos otras, coincide con una de las clases (cluster) de la partición.


Vuelva a: **VIC: Visualization, Inference, Classification steps**

VI.2.5. Análisis de contigüidad

Dos análisis de contigüidad van a ser ejecutados. El primero, sin supervisión, utiliza el grafo de los vecinos más próximos. Es el análisis de contigüidad intrínseca. El segundo, supervisado, utiliza el grafo formado de tres conjuntos disjuntos correspondiente a las tres especies de iris (todos los pares de individuos que pertenecen a una misma especie son vecinos, dos pares que pertenecen a dos especies diferentes no son vecinos nunca). En este caso por el cual la pertenencia a una especie es conocida *a priori*, el análisis de contigüidad coincide con el análisis discriminante lineal.

a. Grafos de los vecinos más próximos

Vamos a efectuar un análisis de contigüidad que utiliza un "grafo de los vecinos más próximos" a partir de mediciones. La partición en tres especies no es tomada en consideración. Se trata pues de un enfoque no supervisado.

- Haga Clic en el botón:  Contiguity.
- Haga Clic en Parameter/Edit. Elija el elemento Create

La ventana siguiente aparece:

ncoord = input coordinate file

0 - No coordinate file (simple description of an external graph)

1 - File `ngus_ind` = coordinates of individuals / observations

2 - File `ngus_var_act` = coordinates of variables

3 - File `ngus_var_boot` = Bootstrap replication of variables

npart = partition file

0 - no partition file

1 - `part_cla_ind` = from clustering individuals (step Part)

2 - `part_cat` = categorical variable (steps Selec and Excat)

3 - `part_som` = from self-organizing map (Kohonen)

4 - `part_cla_var` = from clustering variables (step Permu Recip Part)

meth = method

1 - Contiguity graph defined by a distance threshold

2 - Contiguity graph defined by k nearest neighbours

3 - Classical linear discriminant analysis

4 - External contiguity graph

`npas` = increment from min to max

Min = first value for starting (min. number of edges if "nn")

Max = Maximum value (max. number of edges if "nn")

Luego, tendremos que ingresar los valores numéricos siguientes:

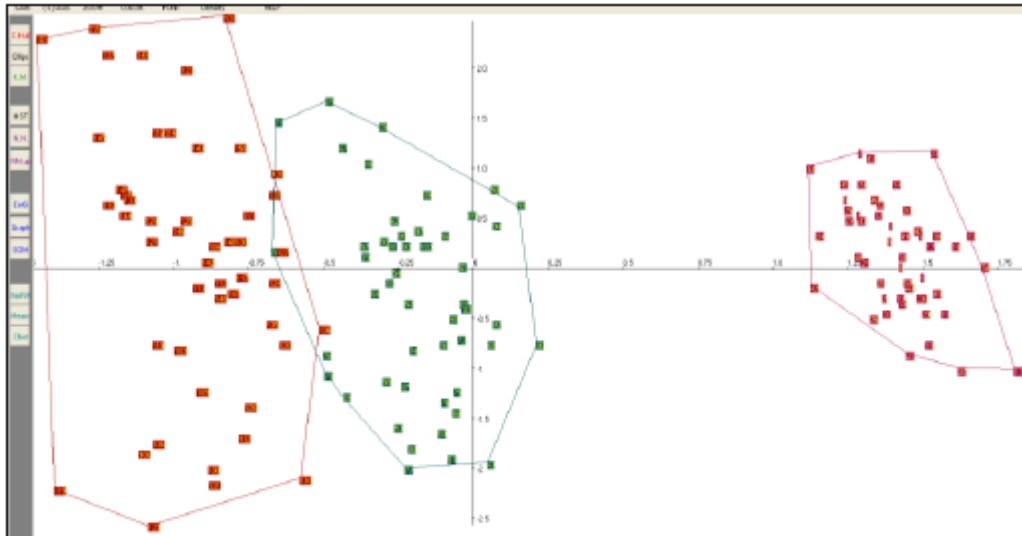
- npas = 2 (aumento del número de vecinos más próximos)
- Min = 4 (número mínimo de vecinos más próximos)
- Max = 8 (número máximo de vecinos más próximos)

Tres análisis de contigüidad serán entonces efectuados para los tres grafos que respectivamente corresponderán a 4, 6, 8 vecinos más próximos (de Min=4 a Max=8, con un incremento de npas =2).

- Haga Clic en **VALIDATE**.
- Aparece un resumen de los parámetros.
- En la barra superior de la ventana, haga Clic en **Execute**. Se efectúan los cálculos.
 - La sección **Results** permite consultar los detalles técnicos de los cálculos implicados en el análisis de contigüidad.
 - Haga Clic luego en **Contiguity View**.
 - Aparece la ventana "Visualization: loading files, selecting axes" que correspondía al botón **Visualization**.
 - En el menú **Load coordinates** de la nueva ventana, abra el archivo **ngus_contig.txt**. En lugar de utilizar las coordenadas principales del ACP (**ngus_ind.txt** como anteriormente), utilizamos ahora el resultado del análisis de contigüidad: **ngus_contig.txt**.
 - Haga Clic luego en **Load or create a partition**. En el submenú **Load partition File**, seleccione el archivo: **part_cat.txt**. (Con este archivo, vamos a identificar las especies). No podemos calcular el árbol de longitud mínima ("minimum Spanning Tree"), ni los vecinos más próximos a partir del archivo: **ngus_contig.txt**.
 - Haga Clic en **Graphics**. Elija luego los ejes 1 y 2 (que son por otra parte los valores por defecto).
 - Elija (marque) el número del nivel de contigüidad, por ejemplo 2, que corresponde a los 6 vecinos más próximos. (El nivel 1 corresponde a los 4 vecinos más próximos, y el nivel 3 a los 8 vecinos más próximos).
 - Haga clic en **DISPLAY**. Cambiar los colores, si necesario.
 - Haga clic sobre: **C.Hull**. Las tres especies están ahora mejor separadas.

Esto significa que el grafo (simetrizado) de los 6 vecinos más próximos permite calcular una matriz "local" de las covariancias que puede desempeñar el papel de una matriz "interna" de covariancias. En este ejemplo, el plano principal de un análisis de la contigüidad es similar al plano principal de un análisis lineal discriminante de Fisher.

Debemos tener en cuenta que el análisis de contigüidad no utiliza el conocimiento *a priori* de las especies. Es un método no supervisado, contrariamente al análisis discriminante, que intenta separar lo mejor posible a las especies.




El análisis de contigüidad consigue separar bastante correctamente las tres variedades de Iris. La matriz "local" de las covariancias calculada a partir de los vecinos más próximos proporciona aquí la estimación de una matriz "interna" de las covariancias. Los excelentes resultados se deben al hecho de que los vecinos más próximos son calculados en un espacio que tiene más de 2 dimensiones, y, para este ejemplo, al hecho de que las 3 clases son bastante bien separadas en este espacio.

b. Análisis discriminante

Vamos ahora a efectuar un "análisis de contigüidad" que coincide exactamente con un análisis discriminante lineal clásico.

El Análisis discriminante lineal en k clases es, en efecto, un caso particular del análisis de contigüidad. En tal caso, el grafo implicado en el análisis de contigüidad es hecho de k conjuntos (grafos completos) que corresponden a k clases del análisis discriminante. En nuestro caso particular, $k = 3$. Todos los pares de observaciones que pertenecen a la misma especie son unidos por una arista. Ninguna arista une dos observaciones que pertenecen a dos especies diferentes.

Vuelva al menú principal y haga Clic en  Contiguity

-
- Haga Clic en . Elija el elemento "Create".
- Marque:

"1" (*File ngus_ind: coordinates of individuals / observations*) en el primer bloque "ncoord = Input coordinate file"

"2" (*part_cat.txt, nominales*) en el segundo bloque "npart = partition file" (partición utilizada para construir el grafo).

"3" (*Classical Discriminant Analysis*) en el tercer bloque "meth = method".

En este caso particular de análisis discriminante, los parámetros siguientes no tienen sentido. Dtm-Vic le pide ignorarlos (Poner en 0 a los valores si necesario).

El análisis de contigüidad será efectuado utilizando el gráfico asociado a la partición en 3 especies de flores. (Todos los pares de individuos que pertenecen a la misma especie son unidos por una arista; No hay ninguna arista entre individuos que pertenecen a especies diferentes)

- Haga clic en **VALIDATE**
- En la barra superior de la ventana, haga Clic en **Execute**. Se efectúan los cálculos.

La sección "Results" de esta barra superior contiene los detalles técnicos de los cálculos implicados en el análisis de contigüidad. La matriz asociada al grafo con sus tres bloques diagonales de "1" y con valor "0" es por otra parte visible en esta presentación de los resultados.

- Haga Clic luego en **Contiguity View**.

Aparece la ventana "Visualization: loading files, selecting axes" correspondiente al botón **Visualization**.

- En el menú **Load coordinates** de la nueva ventana, abra el archivo **ngus_contig.txt**.
- En el menú **Load or create a partition** y en el submenú **Load partition File**, elija el archivo: **part_cat.txt** (vamos a identificar las tres especies de iris)

No podemos calcular el árbol de longitud mínima, ni los vecinos más próximos a partir del archivo de coordenadas nacido del análisis de contigüidad: **ngus_contig.txt**, pero podríamos cargar los resultados obtenidos anteriormente a partir del archivo **ngus_ind.txt**, nacido del análisis en componentes principales, resultados que son guardados.

- Haga Clic en **Graphics**. Elija luego los ejes 1 y 2 (valores por defecto)
- Haga Clic en **DISPLAY**. Cambiar los colores de la pantalla si es necesario para obtener un buen contraste entre las clases, luego bloquee los colores.
- Haga Clic sobre: **C.Hull**. Las tres especies están todavía bien separadas. Pero es no es tan una sorpresa, ya que el análisis discriminante lineal precisamente apunta a la separación de las clases. Estamos aquí en un caso "supervisado". El método utiliza el conocimiento *a priori* de la especie del iris para construir nuevas coordenadas (funciones discriminantes) que inducen a la mejor separación de las clases.

VI.3 Descripción de grafos

Contrariamente a los directorios de los ejemplos precedentes, el directorio **EX_C03.Graphs** contiene varios subdirectorios y varios ejemplos. Estos ejemplos describen algunos grafos de planos simétricos simples a partir de sus matrices asociadas, principalmente por análisis de las correspondencias.

VI.3.1 Visión de conjunto de las carpetas y los archivos

Los archivos relativos a los ejemplos de grafos están situados en la carpeta:

DtmVic-Examples/DtmVic-Examples_C_NumData / EX_C03.Graphs.

Esta carpeta consta de tres subdirectorios:

Chessboard (tablero o damero) se remite a la descripción de un grafo "en forma de damero" (49 vértices que corresponden a un damero cuadrado con 7 filas y 7 columnas, la matriz asociada es una matriz binaria 49x49).

Cycle corresponde a la descripción análoga de un *ciclo* (49 vértices).

Geography corresponde a la descripción de grafos asociados con los mapas geográficos (grafo de regiones contiguas de Japón registrado bajo forma textual y externa, grafo de los departamentos contiguos de Francia, registrado también bajo forma textual y externa).

a. El documento **Chessboard**

La descripción de un grafo en forma de damero puede ser obtenida a partir de varios archivos de datos y diccionarios diferentes:

a1. Un archivo de datos numéricos: **Chessboard_numerical**

En el subdirectorio **Chessboard**, abrir el subsubdirectorio **Chessboard_numerical**. Figuran allí los archivos de datos, el diccionario y los parámetros (formato numérico clásico de Dtm-Vic).

- El archivo de datos: **Chessboard_7x7_dat.txt** contiene la matriz de incidencia del grafo, con 49 filas y 49 columnas. Así como en todos los datos clásicos en el formato interno de DtmVic, cada fila comienza con su identificador. La celda $m(i, j)$ de tal matriz M vale 1 si i y j son vértices unidos por una arista, o 0 en caso contrario.

- Los identificadores de columnas se encuentran en el archivo diccionario asociado: **Chessboard_7x7_dic.txt**.

- Estos archivos serán analizados por el análisis de correspondencias (archivo de comando: **Chessboard_CA.Param.txt**) luego por el análisis en componentes principales (archivo de comando: **Chessboard_PCA.Param.txt**) con el fin de proceder a una comparación. La comparación no es favorable para el análisis en componentes principales en este caso particular¹¹.

¹¹ Ver por ejemplo: *Exploring Textual Data* (1998), por L. Lebart, A. Salem, L. Berry, ed. Kluwer-Spring. Esta comparación ya había sido hecha en el artículo: "Introduction à l'analyse des données", (L. Lebart) *Consommation*, n°4, 1969, p. 65-87, Dunod.

a.2. Un archivo de datos "externos": **Chessboard_Extern-7x7.txt**

Siempre en el directorio **Chessboard_numerical**, el archivo **Chessboard_Extern_7x7.txt** es otra codificación posible del grafo *Chessboard*, calificada de externa porque es diferente del formato interno general de Dtm-Vic. Proporciona, para cada vértice (fila), los números de los vértices contiguos. La primera fila contiene el número de vértice (49), luego la longitud de los identificadores (4) y el grado máximo del grafo (límite superior del número de aristas adyacentes a un solo vértice) (10). Note que cada fila de números se acaba con el valor convencional 0, que es el indicador de fin de fila para este formato.

Este formato específico, muy compacto, directamente puede conducir a una descripción del grafo en el submenú "contigüidad" de DtmVic.

a.3. Un archivo de datos textuales: **Chessboard_textual_7x7.txt**

El archivo **Chessboard_textual_7x7.txt**, en el subsubdirectorio **Chessboard_textual**, contiene las mismas informaciones de base bajo una forma completamente distinta: el formato es el de las respuestas a una pregunta abierta. Cada vértice del grafo está considerado como una persona interrogada que responde a la pregunta abierta ficticia: "¿cuáles son sus vecinos?". En lugar de una matriz binaria M , estamos tratando aquí un cuadro mucho más pequeño que contiene la dirección (número de columna) de los "1" en la matriz M . Los comandos de **Chessboard_Textual.Param.txt** conducen a los mismos resultados que el análisis de las correspondencias del aparte precedente, utilizando no obstante una secuencia de etapas muy distintas de Dtm-Vic. Es un "ejemplo pedagógico" de puente entre las medidas numéricas y textuales de DtmVic. ¡Cuidado! Con este tipo de datos, las cifras no son consideradas como números en el sentido matemático del término, sino como de secuencias simples de caracteres. [Ver más abajo el ejemplo de los mapas de Japón y de Francia, dónde los números de los vértices son reemplazados por los nombres de las regiones y de los departamentos en claro]. Esta carpeta también contiene el mismo archivo **Chessboard_Extern-7x7.txt** que la carpeta precedente.

b. El documento "Cycle"

Este subdirectorio **Cycle** es vecino del relativo al grafo *Chessboard*.

Encontramos allí de la misma manera que para el documento *Chessboard*, una codificación numérica y externa. Sólo la forma del gráfico es diferente. La codificación textual y el archivo de comando del Análisis en componentes principales han sido omitidos en este caso.

c. El documento *Geography*

Ambos subdirectorios del directorio **Geography** son los homólogos del ejemplo textual del documento **Chessboard**. El directorio **Japan_map** y **France_map** ilustran la "codificación textual" en el caso de los grafos que describen las diferentes regiones de Japón y de los departamentos de Francia. En el caso de Japón, por ejemplo, las dos primeras filas del archivo **Japan_map_textual.tex.txt** indican que las provincias de *Akita* y *de Iwate* son contiguas en la provincia de *Aomori*, etc. El archivo de comando correspondiente es el archivo **Japan_map_textual_Param.txt**. Es similar al archivo **Chessboard_Textual.Param.txt**

En el caso de Francia, por ejemplo, las dos primeras filas del archivo **France_Text.txt** indican que el departamento del *Ain* es contiguo a los departamentos *Isère*, *Jura*, *Rhône*, *Haute Saône*, *Savoie*, *Haute Savoie*. El archivo **France_Param.txt** es el archivo de comando que corresponde.

El archivo **France_extern.txt** representa el mapa de Francia en el formato externo definido en la sección a.2 aquí arriba. Permitirá trazar las aristas del grafo inicial en los planos factoriales.

VI.3.2 Ejecución del ejemplo "Chessboard_numerical"

(Directorio **Chessboard_numerical** en **EX_C03. Graphs / Chessboard**).

En esta carpeta, figuran los archivos de base:

- a) Archivo de datos: **Chessboard_7x7_dat.txt**
- b) Archivo Diccionario: **Chessboard_7x7_dic.txt**.
- c) Archivos de comando: **Chessboard_CA.Param.txt** [Analiza de las

Correspondencias] y **Chessboard_PCA.Param.txt** [análisis en componentes principales]

Es posible realizar o un análisis de correspondencias clásico o un análisis en componentes principales.

a. Apertura y Ejecución del archivo parámetro del AC

Empezaremos por ejecutar el análisis de las correspondencias.

- Haga Clic en el botón: **Open an existing command file** de **Command File** (menú principal). Luego busque la carpeta **Chessboard_numerical** en **DtmVic-examples / DtmVic-Examples_C_NumData**, luego el archivo de comando **Chessboard_CA.Param.txt**

Note de nuevo que estos "archivos de comando" pueden fácilmente ser generados haciendo Clic en el botón "Create" del menú principal (Basic Steps). Aparece una ventana "Select a basic analysis". Haga Clic luego en el botón: SCA - Simple Correspondence Analysis o sobre el botón: PCA - Principal Components Analysis - los dos situados en la sección "Numerical Data", y siga las instrucciones como está indicado en el capítulo II.

Después de haber identificado y haber verificado los archivos de datos y del diccionario, se van a efectuar tres etapas: ARDAT (almacenamiento de los datos), SELEC (selección de los elementos activos y adicionales), AFCOR (análisis de las correspondencias).

- Haga Clic en **Return to execute** en la barra para volver al menú principal.
- Hacer Clic en el botón: **Execute** de **Command File**
- Hacer Clic en el botón: **Basic numerical results** de **Result Files**

El botón abre el archivo HTML denominado "imp.html" que contiene los principales resultados de las etapas precedentes de cálculo básico. Después de la lectura de estos resultados numéricos, regrese al menú principal.

b. Visualización y lectura de los resultados

Ahora vamos directamente a visualizar el gráfico en la etapa **VIC: Visualization, Inference, Clasification steps**.

- Haga Clic en **V Visualization** (no utilizaremos aquí los botones "AxeView", "PlaneView", etc.)
 - Aparece una ventana denominada " DTM-visualization: loading files, selecting axes"

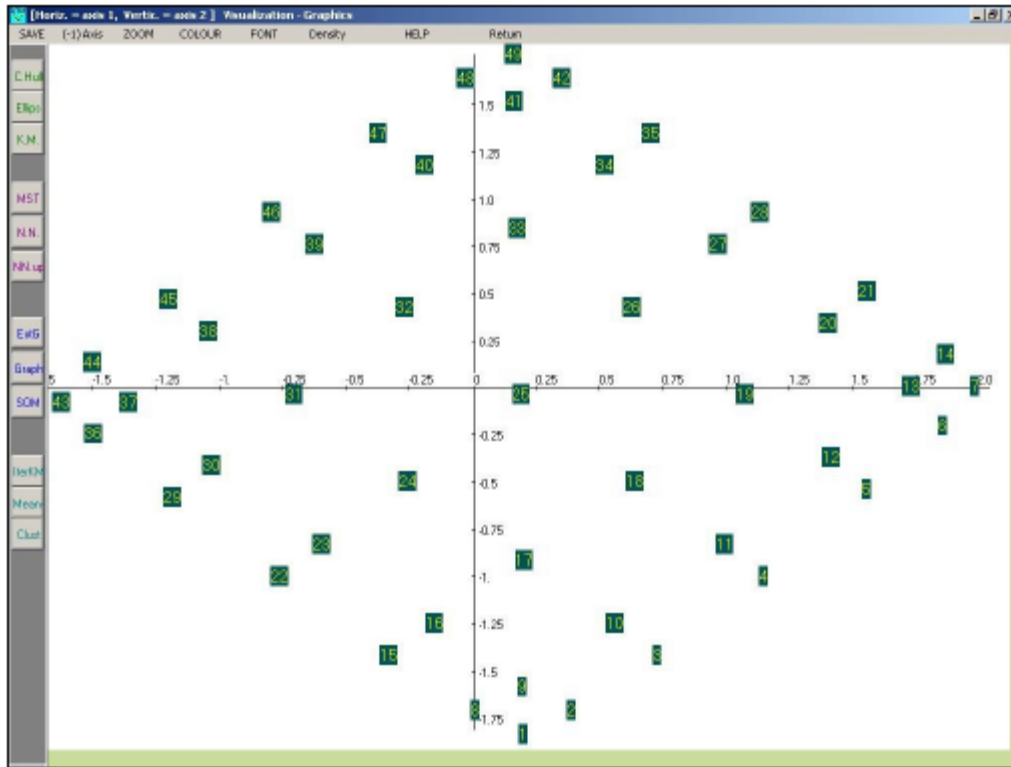


Figura VI.1 Plano factorial principal (Análisis de correspondencias) para el grafo "Damero" (después el cambio de fuente (botón "Font") y cambio de color (botón "Colour").

- Haga Clic en **Load coordinates**. En el submenú correspondiente, elegir el archivo: **ngus_ind.txt** (individuos u observaciones). Son seleccionadas las principales coordenadas de los individuos (filas). [En efecto, aquí, la matriz de datos es simétrica, es equivalente, en este caso muy particular, a elegir **ngus_var_act.txt**].
- Haga Clic luego en **Load or create a partition**. En el submenú que corresponde, seleccione **No partition**.
- Haga clic en **Graphics** luego elija los ejes 1 y 2 (por defecto) en la pequeña ventana "Sélection of axes" y haga Clic en "Continue" luego en **DISPLAY**.
- En una nueva ventana llamada "Vizualisation - Graphics", el plano factorial principal se exhibe (ver figura VI.1 precedente).

En la barra vertical de herramientas de la ventana "Graphics", el botón **ExtG** nos va a permitir trazar el grafo inicial a partir de la codificación externa.

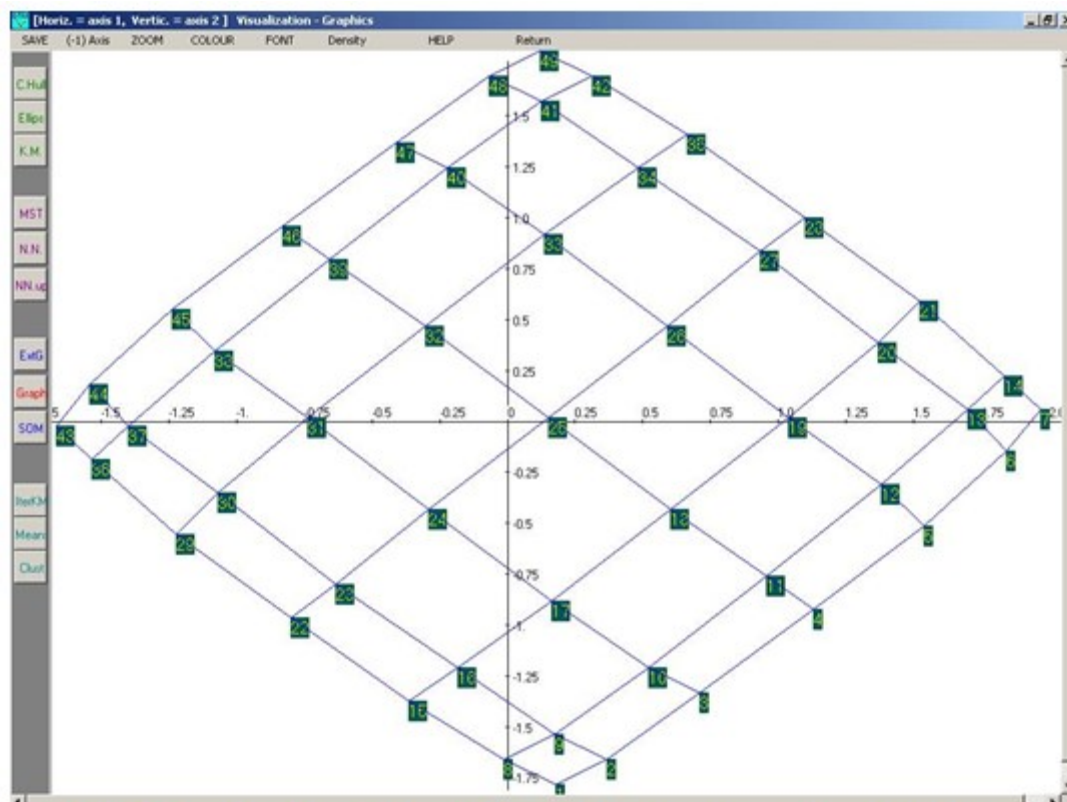


Figura VI.2. Mismo plano factorial principal para el grafo "Damero" con trazado del grafo inicial (después el cambio de fuente (botón "Font") y de color (botón "Colour").

Para representar las aristas del grafo de origen, haga Clic en el botón **ExtG** (grafo externo) de la barra vertical.

- Abra el archivo Chessboard_Extern_7x7.txt.
- Haga Clic en el botón **Graph**.

Obtenemos entonces una representación del grafo original con una representación de las aristas originales (Fig VI. 2).

Esta representación permite también observar las deformaciones del grafo plano en los espacios generados por los pares de ejes de rangos 3 - 12.

Observamos un efecto Guttman multidimensional¹².

- Regrese al menú principal dejando la ventana del plano factorial, luego haciendo Clic en **Return** luego abandone a Dtm-Vic.

c. Apertura y Ejecución del archivo parámetro del ACP

Retome las operaciones de las secciones a y b abriendo esta vez el archivo de comando: **Chessboard_PCA.Param.txt** (PCA: análisis en componentes principales). Repita todas las operaciones precedentes.

¹² Ver Benzécri, (1973) "L'analyse des données", Tome II B, capítulo 10, "Sur l'analyse de la correspondance définie par un graphe", pp 244 - 261]

Vemos a través del gráfico producido por este ejemplo que el Análisis en Componentes Principales describe de modo menos fiel la estructura del grafo que el Análisis de Correspondencias (Figura VI. 3).

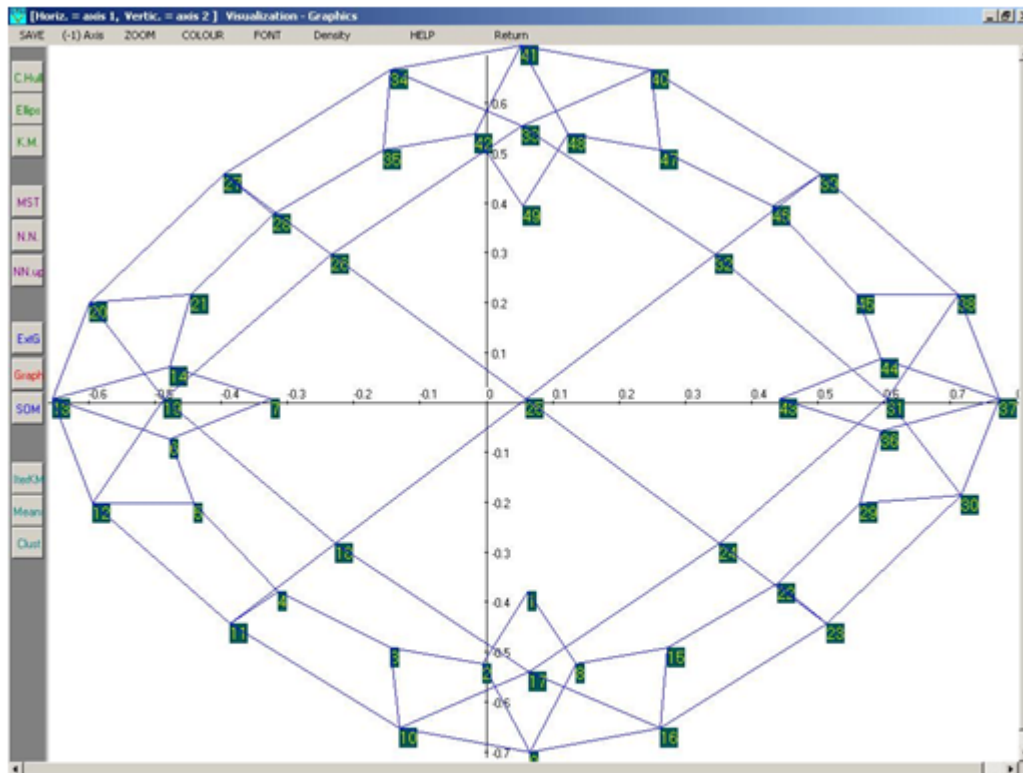


Figura VI.3. Caso del análisis en componentes principales. Plano factorial principal para el grafo "Damero" con trazado del grafo inicial (después de cambio de fuente (botón "Font") y cambio de color (botón "Colour"). El tratamiento asimétrico de las filas y las columnas y la normalización operada por el ACP no permiten obtener una descripción satisfactoria de este tipo de grafos.

VI.3.3 Ejecución del ejemplo "Chessboard_textual"

Esta sección se refiere a la ejecución del ejemplo **Chessboard_textual** del directorio **DtmVic-Examples_C_NumData / EX_C03. Graphs / Chessboard** y a la lectura de los resultados.

Estamos en el marco de un análisis textual similar al ejemplo el que describe las respuestas a una pregunta abierta en una encuesta por muestreo (Ejemplo III.2 del capítulo III).

Encontramos en este directorio el "archivo de texto" y el "archivo de comandos". (En este contexto particular, no hay archivos de datos ni archivo-diccionario: el cuestionario comprende una "pseudo pregunta abierta", hecha en cada vértice: "¿cuáles son sus vértices vecinas?").

1. Archivo de texto: **Chessboard_textual_7x7.txt**

El formato es el mismo que el descrito en el párrafo I.5 (Capítulo 1, §5, cuadro 4, en el caso de una sola pregunta abierta). Dado que las respuestas pueden tener longitudes muy diferentes, los separadores son utilizados para distinguir a los individuos (o: las personas interrogadas). Los individuos (aquí: los vértices) son separados por la cadena de caracteres "----" (a partir de la columna 1) eventualmente seguido por un identificador. Atención, los 49 números de

vértices están considerados aquí como palabras, podrían ser reemplazados por 40 nombres distintos con los mismos cálculos y el mismo resultado final para el trazado del grafo.

2. Archivo de comandos: **Chessboard_Textual.Param.txt**

La fase de cálculo del análisis se descompone en "etapas". Cada etapa necesita algunos parámetros descriptos brevemente en el menú principal de DtmVic (botón: "Help about parameters").

a. Apertura y Ejecución del archivo de comando

- Haga Clic en el botón: **Open an existing command file** de **Command File** (menú principal) y abra el archivo parámetro **Chessboard_Textual.Param.txt**

Son efectuadas cuatro etapas:

ARTEX (almacenamiento de textos), SELOX (selección de la pregunta abierta), NUMER (codificación numérica del texto), ASPAR (análisis de las correspondencias de la tabla de contingencia ["respondentes x palabras"]).

Note que este archivo de comando puede también ser generado haciendo Clic en el botón "Create a new command file" de la sección "Command file" del menú principal ("Basic Steps"). Aparece una ventana "Select a Basic Analysis". Haga Clic luego en el botón: VISURESP, situado en la sección "Textual Data", y siga las instrucciones como está indicado en los capítulos II y III. También note que en este caso de datos simples (una sola "pregunta abierta"), es posible considerar cada respuesta como un texto. En tal caso, el separador "----" debe ser reemplazado por el separador "*****", como en el ejemplo III.1 del capítulo III. En lugar del análisis "VISURESP" (Visualization of responses), es necesario entonces efectuar el análisis "VISUTEX" (Visualization of texts).

- Haga Clic en **Return to execute** en la barra para volver al menú principal.

- Hacer Clic en el botón: **Execute** de **Command File**

Esta fase ejecuta las etapas de cálculo en el archivo de comando: digitalización del "texto" y análisis de las correspondencias de la tabla léxica.

- Hacer Clic en el botón: **Basic numerical results** de **Result Files**

El botón abre el archivo HTML denominado "imp.html" que contiene los principales resultados de las etapas precedentes de cálculo de base.


La etapa NUMER nos hace conocer que, por ejemplo, tenemos 49 "respuestas", con un número total de palabras (casos = aquí: aristas del grafo) de 217, lo que implica 49 palabras distintas (aquí: los vértices vecinos sobre el damero). Note que cada vértice ha estado considerado también como su propio vecino.

Después de la lectura de estos resultados numéricos, vuelva al menú principal.

b. Visualización y lectura de los resultados

Vamos ahora a visualizar los resultados con las herramientas de la etapa



VIC: Visualization, Inference, Clasificación steps.

Para trazar el grafo: haga Clic en  Visualization

Todas las etapas de la sección precedente pueden ser realizadas del mismo modo. Los gráficos obtenidos son idénticos a los de la sección VI.3.2.b. No hay lugar para reproducirlos.

VI.3.4 Ejecución directa del ejemplo "Chessboard_Extern"

No hay ni archivo de comando, ni archivo de diccionario para este tipo de análisis que directamente utiliza el formato "Externo". Para este tipo de codificación del grafo ("codificación externa de grafo"), está prevista una entrada directa al menú "Contigüidad".




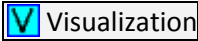
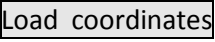

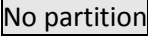
- Haga Clic en  Contiguity en la etapa **VIC: Visualization, Inference, Clasificación steps**
- Haga Clic en . Elija el elemento "Create"

Vamos a establecer los parámetros necesarios para una descripción gráfica:

- En el primer bloque denominado "*ncoord = Input coordinate file*", marque "0" (*File ngus_ind: coordinates of individuals / observations*). Ningún archivo de coordenadas (descripción simple de un grafo externo).

- En el segundo bloque denominado "*npart = partition file*" marque "0" (sin partición)

- En el tercer bloque denominado "*meth = method*", marque (grafo externo de contigüidad).

- Haga Clic en 
- En la barra superior de la ventana, haga Clic en 
 - Una nueva ventana aparece, y ustedes están invitados a elegir el archivo del grafo externo **Chessboard_Extern_7x7.txt** del directorio **EX_C04. Graphs Graphs/Chessboard Chessboard/Chessboard-Extern.**
 - Aparece otra ventana "Reading an external graph"
- Haga Clic en 
 - Aparecen una serie de ventanas en las que se indican los detalles técnicos de los cálculos contenidos en el análisis de correspondencias de la matriz M asociada al grafo (Estos resultados son registrados en el archivo `imp_contig.txt`, guardado en el directorio de trabajo).
- Haga Clic en 
 - Aparece la ventana denominada "DTM-visualization".
- Haga Clic en . En el submenú correspondiente, elegir el archivo: **anagraf.txt**, que contiene las coordenadas factoriales para los análisis directos de grafos.
- Haga Clic luego en . En el submenú correspondiente, seleccione . Luego proceda como en el ejemplo *Chessboard*.

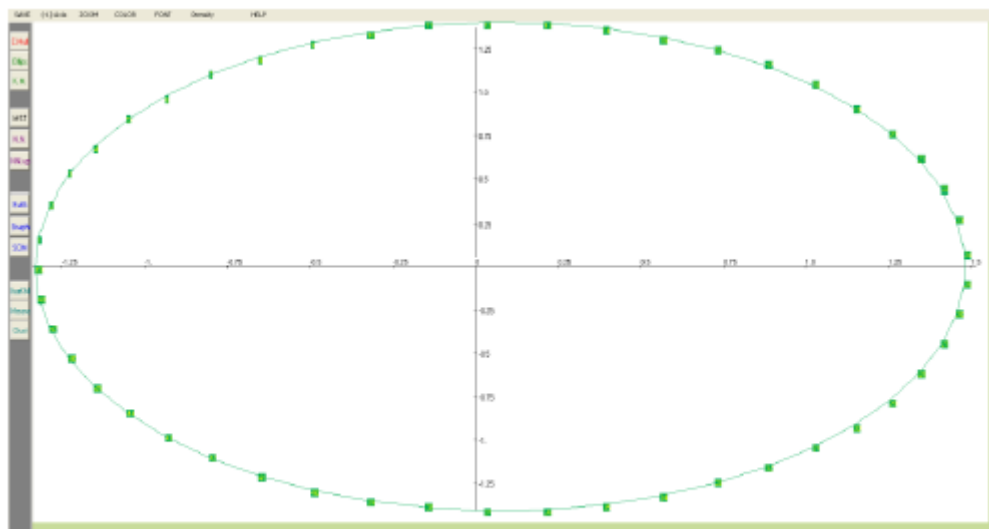
- Haga Clic en **Graphics** luego elija los ejes 1 y 2 (por defecto) en la ventana "Selección de los ejes" y haga Clic en **Continue** luego en **DISPLAY**.
 - En una nueva ventana denominada "Vizualisation - Graphics", se exhibe el plano factorial principal.

Una vez más, podrán realizarse todas las etapas de la sección precedente. Los gráficos obtenidos siguen siendo idénticos a los de la sección VI.3.2.b. Entonces no se los reproduce.

VI.3.5 Ejecución de los ejemplos "Ciclo"

Esta sección es desde todo punto de vista idéntica a la sección VI.3.2 (ejecución del ejemplo "Chessboard_Numerical") y a la VI.3.4. El gráfico tiene la forma de un ciclo, con el mismo número de vértices.

Los archivos homólogos de los archivos **Chessboard_7x7_dat.txt**, **Chessboard_7x7_dic.txt**, **Chessboard_Extern_7x7.txt** y **Chessboard_CA_Param.txt** son ahora respectivamente **Cycle_49_dat.txt**, **Cycle_49_dic.txt**, **Cycle_Extern_49.txt** y **Cycle_CA_Param.txt**. Ellos pueden encontrarse en el directorio **Ciclo**.



Plano factorial principal para el grafo "Ciclo" con trazado del grafo inicial (después del cambio de fuente (botón "Font") y el cambio de color (botón "Colour").

VI.3.6 Ejecución del ejemplo "France_map" (Carpeta: **Geography**)

Esta sección es idéntica a la sección VI.3.3 (Ejecución del ejemplo *Chessboard_Textual*). El gráfico es ahora una cartografía de un mapa de Francia, presentada como una continuación de respuestas a la pregunta abierta: "¿cuáles son sus departamentos vecinos?", las "personas encuestadas" son los departamentos franceses.

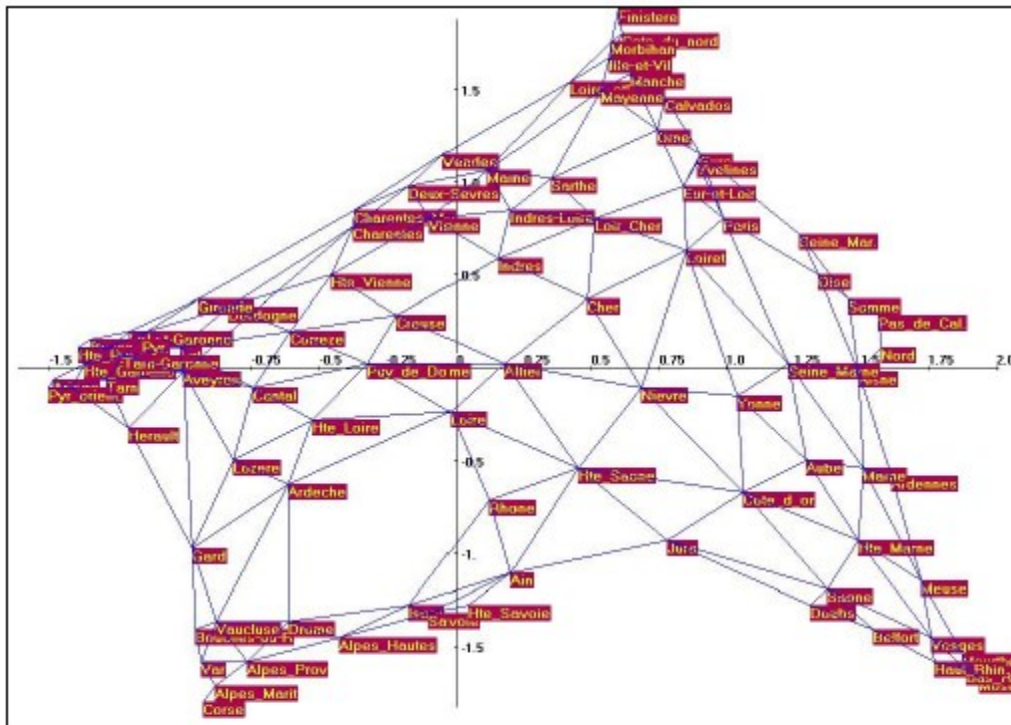

```

****   Ain
      Ain Isere Jura
      Rhone Hte Saone Savoie Hte Savoie
****   Aisne
      Aisne Ardennes Marne
      Nord Oise Seine_Marne Somme
****   Allier
      Allier Cher Creuse
      Loire Nievre Puy_de_Dome Hte_Saone

```

Extracto del archivo de datos textuales: France_Text.txt (tres primeros departamentos)

El homólogo de la carpeta **Chessboard_Textual** es: **France_map**, mientras que los equivalentes de los tres archivos **Chessboard_textual_7x7.txt**, **Chessboard_extern_7x7.txt** y **Chessboard_textual_Param.txt** son respectivamente los tres archivos: **France_Text.txt**, **France_extern.txt** y **France_Param.txt**.



Plano factorial principal para el grafo "Francia" con trazado del grafo inicial (después del cambio de fuente (botón "Font") y cambio de color (botón "Colour"). El signo de los ejes (arbitrario) puede ser cambiado, para recobrar la orientación inicial.

VI.3.7 Ejecución del ejemplo "Japan_map"

(Carpeta: **Geography**)

Esta sección es idéntica a la precedente, así como a la sección VI.3.3 (Ejecución del ejemplo "Chessboard_Textual"). El gráfico es ahora un bosquejo de un mapa de Japón, codificado como las respuestas a la pregunta abierta "Cuales son sus regiones próximas", los "respondentes (ficticios)" son las mismas regiones de Japón. La carpeta **Japan_map** contiene los tres archivos

homólogos de los precedentes (texto, externo y parámetro): `Japan_map_Textual.tex.txt`, `Japan_map_Extern.txt` y `Japan_map_Textual.Param.txt`.

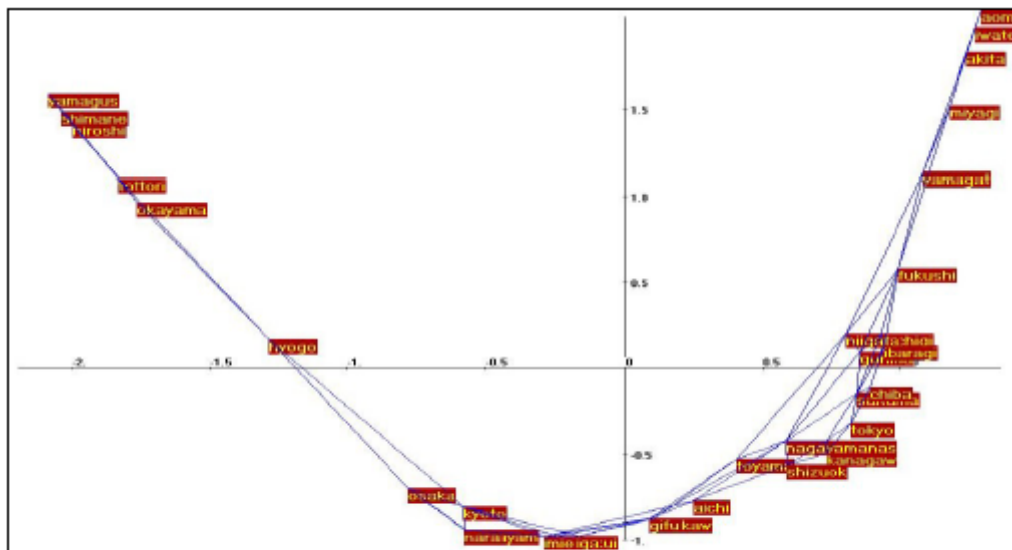
```

---- aomori
akita iwate
---- akita
aomori iwate yamagata miyagi
---- iwate
aomori akita miyagi
---- yamagata
akita miyagi niigata fukushima

```

Extracto del archivo de datos textuales: `Japan_map_Textual.tex.txt` (tres primeras regiones). Aquí, las regiones están consideradas como individuos (separador ----) mientras que los departamentos han estado considerados como textos (separador ****). Ambas codificaciones son posibles en esta configuración simple.

La misma secuencia de operación conduce al gráfico siguiente, cuya forma parabólica es en parte atribuible a la forma del conjunto, pero también al efecto Guttman, ya mencionado en la sección VI.3.2.b, a propósito de los ejes 3 y los siguientes, y resaltado aquí por una diferencia de escala entre los ejes. Este efecto Guttman aparece evidentemente desde el segundo eje para los grafos en forma de cadenas o en forma de trenzas (el primer eje dominante, los ejes siguientes son funciones polinomiales del primero).



Plano factorial principal para el grafo "Japón" con trazado del grafo inicial (después de cambio de fuente (botón "Font") y cambio de color (botón "Colour"). El signo de los ejes es arbitrario. Puede también ser cambiado, para recobrar la orientación geográfica inicial.

VI.4. Reconstitución de imágenes

(Paréntesis metodológico)

Los ejemplos en esta sección VI.4 son principalmente unos ejemplos pedagógicos que sirven para ilustrar las propiedades de compresión de los análisis en ejes principales (guardando un número limitado de ejes principales que provienen de una descomposición a los valores singulares [SVD] o de un análisis de las correspondencias) en el uso del análisis de imágenes (uso poco familiar para algunos usuarios actuales de Dtm-Vic). Se hace una comparación con las series de Fourier discretas (guardando un número limitado de términos de la expansión) quienes toman en consideración las posiciones relativas de los píxeles.

VI.4.1 Formato de los archivos de imágenes

Este tipo de tratamiento no hace uso de los datos en formato-texto interno a Dtm-Vic, porque se trata de imágenes digitalizadas. Un cuadro simple y rectangular de números enteros basta: no es necesario tener identificadores de filas o columnas (diccionario).

De hecho, tres formatos específicos serán utilizados: cuadros rectangulares de niveles de gris (formato texto simple: ".txt"), formato ".pgm" (sigla de "Portable Gray Map" o "Portable Grey Map" en Inglés británico) y para las imágenes en color, el formato ".ppm" (sigla de "Portable Píxel Map").

Encontraremos los archivos de ejemplo en la carpeta **EX_C05 Images** incluida en la carpeta **DtmVic_Examples_C_NumData**. En esta carpeta, abra la carpeta del ejemplo **C.5: EX_C05. Images**. Cuatro subdirectorios corresponden a los cuatro ejemplos:

- "1_Cheetah_txt",
- "2_Baalbeck_pgm",
- "3_Cardinal_ppm_color",
- "4_Extra_pgm_ppm".

Todos los archivos contenidos en estos subdirectorios pueden ser examinados con un editor de texto (tal como "Notepad", incluido en Windows, "UltraEdit", o un software libre tal como "Notepad++" o "TotalEdit", etc.). Para las imágenes en niveles de gris, dos formatos de entrada están disponibles:

1 - El formato de texto simple. [Ver el ejemplo 1, es decir la imagen **cheetah.txt**¹³ de la carpeta **1_cheetah.txt**]. El cuadro de datos contiene los enteros positivos inferiores o iguales a 255 que son los valores del nivel de gris para cada píxel (ningún identificador). Este formato que explícitamente no contiene el tamaño de la imagen es el más simple. Debido a su rusticidad, no es utilizada ni proporcionada por los softwares usuales de tratamiento de imágenes.

2 - El formato pgm. ("Portable grey map") (ver el ejemplo 2, con la imagen **Baalbeck.pgm** de la carpeta **2_Baalbeck_pgm**, utilizando un editor de texto). El formato pgm es un formato simple y transparente de niveles de gris. La primera fila contiene el identificador de formato: P2. Las segundas y terceras filas contienen tres

¹³ Imagen adaptada del libro "Data Compression", Mark Nelson, M & T Publishing Inc, 1992

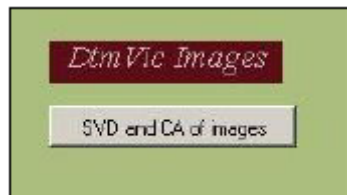
enteros: número de columnas, número de filas, y el valor (255) máximo. Luego, el cuadro es establecido por fila. Cada píxel de la tabla es representado como un número decimal que describe el nivel de gris (<255). Cada píxel de la tabla tiene por lo menos un espacio en blanco antes y después. Ninguna fila sobrepasa los 72 caracteres¹⁴

3 - El formato ppm. Para las (pequeñas) imágenes en color, el formato de entrada es el formato texto ppm ("portable píxel map"). Consulte el ejemplo 3 **Cardinal.ppm**, vía un editor de texto (carpeta **3_Cardinal_ppm**). Este formato es bastante semejante al de pgm, pero con tres enteros (3 niveles de RGB: Red, Green, Blue) sobre una misma fila por píxel. Este formato también es el del ejemplo 4.

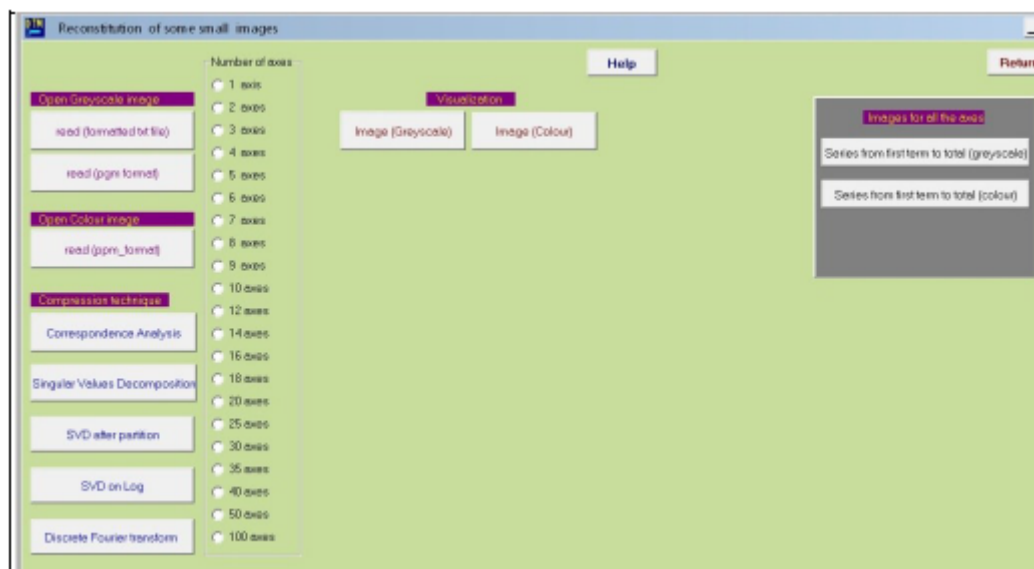
Los archivos pgm y ppm pueden ser obtenidos por una exportación a partir del software libre "Open Office" (precisar pgm, formato texto), utilizando un archivo JPEG en la entrada. [Atención, para este módulo, restricción a 1000 para el número de píxeles en fila o en columna].

VI.4.2 Análisis para la compresión de imágenes

- Haga Clic en el **SVD and CA of images** en la sección "DtmVic Images" del menú principal.



- Aparece una ventana, cuyo lado superior está representado aquí abajo



¹⁴ Para más información sobre tal formato, consulte (por ejemplo) <http://netpbm.sourceforge.net/doc/pgm.html>

Descripción de la ventana "Reconstitution of some small images"

A la izquierda figuran en columna tres botones (rojo oscuro) correspondiente a los tres formatos de archivos de imágenes descritos en el párrafo precedente (formato simple de niveles de gris, formato pgm de niveles de gris, formato ppm en color). Luego, más bajo, cinco botones (azules) que corresponden a los cinco métodos de compresiones elegidas (Análisis de correspondencias, SVD - Descomposición a los valores singulares, Análisis después de partición previa de la imagen¹⁵, Análisis logarítmico¹⁶, Series de Fourier discretas). Para los cuatro primeros métodos, el número de ejes guardado (de 1 a 100) tiene que marcarse en la segunda columna. Si el número retenido de ejes es 8, por ejemplo, son los 8 primeros términos de la fórmula de reconstitución de los datos que son utilizados para reconstruir la imagen. Los dos botones centrales activan una pantalla de imágenes (gris o en color). Los dos botones del panel gris a la derecha activan una exploración automática para todos los ejes propuestos. Todas las figuras intermedias son guardadas en formato *Windows bitmap* (.bmp).

Antes de examinar los ejemplos, sinteticemos la continuación de las operaciones que hay que hacer en el caso de los análisis en ejes principales (métodos factoriales):

- Haga Clic, según la extensión del archivo de imágenes, sobre uno de los botones **Read**. (formato txt, o: formato pgm, o: formato ppm). Responder **OK** a las cajas de mensaje **number of columns** y **number of rows** que se exhiben.
- Seleccionar uno de los métodos, el análisis de correspondencias **Correspondence Analysis** o la descomposición a los valores singulares **Singular Values Decomposition**. Responder **OK** cuando aparece la caja de mensaje **End of computation**.
- Seleccionar el número de ejes. Responder **OK** en la ventana **number of axes**.
- Hacer Clic en uno de los botones **Image** según sea la imagen elegida (negra y blanca o de color). De hecho, el botón "Help" permite obtener las informaciones necesarias (en Inglés). Los archivos de imágenes creadas (imagen original, e imágenes reconstituidas a partir de un número variable de ejes principales) automáticamente son guardados en formato "bmp".

El software "Paint", del componente "Accesorios" de los programas bajo Windows, (o el software gratuito "IrfanView" por ejemplo) permite visualizar estas imágenes y guardarlas en formato JPEG, con ahorro de espacio.

- Hacer Clic en **Exit**.

VI.4.3 Ejecución de un primer ejemplo

(formato de texto simple: Ejemplo: *Cabeza de cheetah*: **1_Cheetah_txt**)

¹⁵ Esta variante consiste en centrar previamente los niveles de gris dentro de p zonas rectangulares antes de la SVD, luego en añadir la p medias después de la SVD. (Podemos elegir p = 2 x 2, 3 x 3, 4 x 4, 5 x 5, etc.)

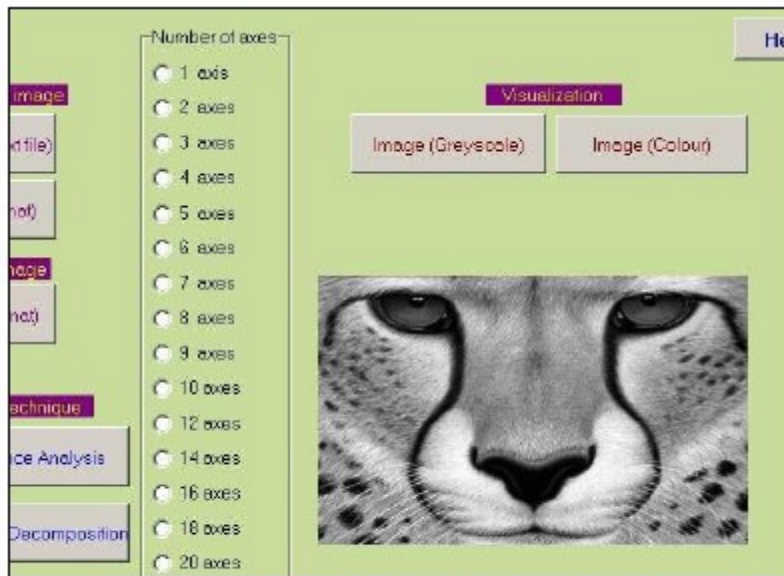
¹⁶ Esta variante consiste en hacer una transformación logarítmica previa, luego en proceder en un SVD del cuadro doblemente centrado en fila y en columna.

- Haga Clic en el botón: **SVD and CA of images**, en la sección **DtmVic- Images** del menú principal.
 - Aparece la ventana "Reconstitution of some small images" (ver aquí arriba).

a. Haga Clic en el primer botón **Read (formatted txt file)** en la sección **Open Greyscale image**

En el directorio **EX_CO4_Image**, abra el subdirectorio **1_Cheetah_txt**. En este directorio, abra el archivo **Cheetah.txt**. Una caja de mensaje recuerda las dimensiones del archivo de imágenes.

b. Si usted desea visualizar la imagen original, en la sección **Visualization**, hace Clic en **Image (Greyscale)**. La imagen aparece entonces en el centro de la ventana, como indicado más abajo



Porción de ventana que presenta la imagen original **Cheetah.txt** antes de la elección del número de ejes.

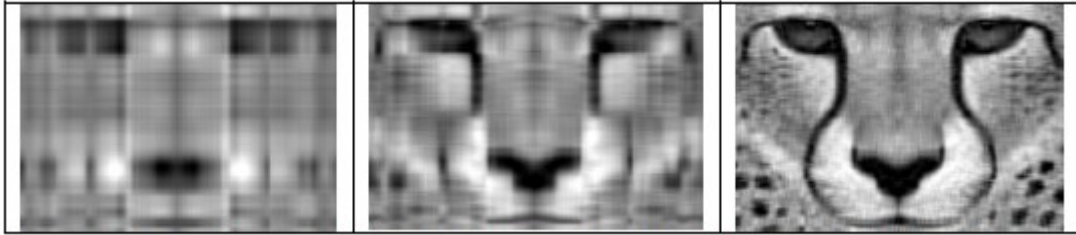
A continuación la sección "c" se dedica a los métodos factoriales de compresión (ejes principales), luego en la sección "d" que sigue se examinará en calidad de comparación la compresión obtenida reteniendo sólo los primeros términos de las series de Fourier discretas. No se trata aquí de optimizar la compresión, sino de comparar dos enfoques jerárquicos simples (bases de vectores *versus* bases de funciones trigonométricas).

c. El caso de los métodos factoriales

En la parte de abajo a la izquierda de la ventana, en la sección: **Compression technique**, haga Clic en el botón: **Correspondence Analysis** (para comenzar). Se efectúa el análisis.

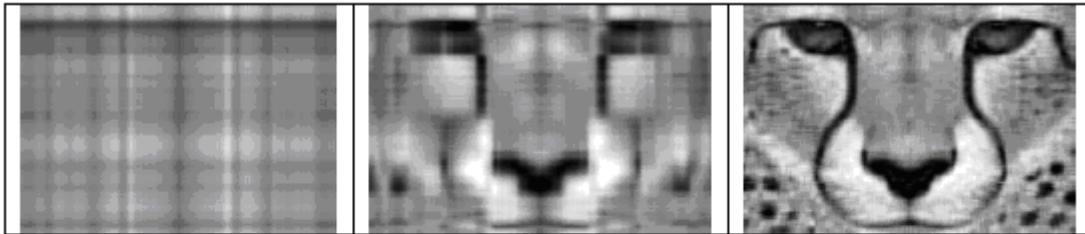
c1. Si desea obtener una visión de conjunto de la reconstitución de los datos, de 1 a 100 ejes, haga Clic directamente en el botón: **Series from first term to total (greyscale)** en el panel: **Images for all the axes**. Podemos entonces observar la reconstitución progresiva de la imagen.

c2. Si usted se interesa en un número específico de ejes, seleccione el número requerido en la lista vertical correspondiente, y visualice cada imagen con el botón utilizado en **b**.



Caso del análisis de correspondencias: imágenes reconstituidas sucesivamente con un eje principal, cuatro ejes y 16 ejes. En este caso, para un solo eje, la fórmula de reconstitución contiene dos términos: el término correspondiente a la hipótesis de independencia (eje 0) y el primer eje

c3. En lugar del análisis de correspondencias, puede elegir el método de "Singular Value Decomposition" (Descomposición a los Valores singulares), y rehacer las operaciones **c1.** y **c2.**



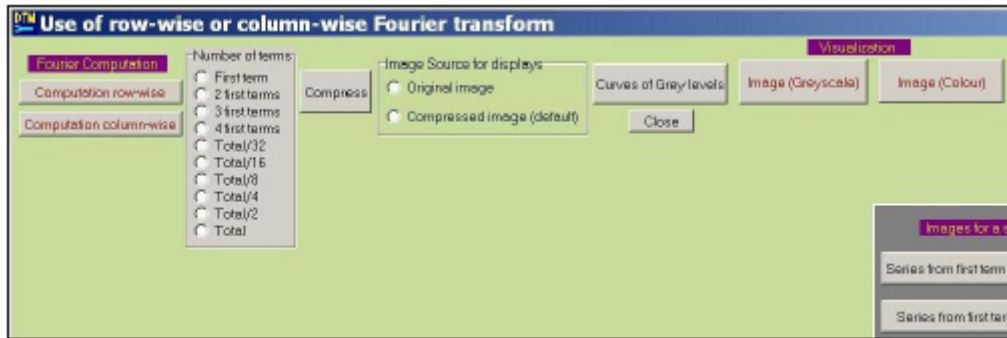
Caso de la descomposición a los valores singulares : Imágenes reconstituidas sucesivamente con un eje principal, cuatro ejes y 16 ejes. En este caso, para un eje, la fórmula de reconstitución contiene sólo un solo término, de donde hay un "retraso" con relación al análisis de correspondencias, el retraso que se desvanece a través de la acumulación de los ejes.

Nota: Todas las imágenes creadas son registradas sistemáticamente en el formato bitmap (extensión: ".bmp") en el directorio del archivo de la imagen analizada.

d. El caso de las series discretas de Fourier:

En la parte de abajo a la izquierda de la ventana, en la sección: **Compression technique**, haga Clic en el botón: **Discrete Fourier Transform**.

Una nueva ventana se exhibe



Porción de la ventana de comando de las compresiones por series discretas de Fourier.

d1. Usted debe luego seleccionar el modo de cálculo de la serie de Fourier, en fila o en columna ("Row-wise" o "columnwise"). Seleccione a "Row-wise", por ejemplo.

d2. Luego, como anteriormente, si desea obtener una visión de conjunto de la reconstitución de los datos cuando el número de términos aumenta, directamente haga Clic en el botón: **Series from first term to total (greyscale)** en el panel: **Images for a series of terms**. Podemos entonces observar la reconstitución progresiva de la imagen.

d3. Si se interesa por un número particular de términos (entre los términos de la selección sugerida), seleccione el número requerido en la lista vertical correspondiente, y visualice cada imagen con similar botón al utilizado en **b**



Caso de las series discretas de Fourier (opción: línea por línea): imágenes reconstituidas sucesivamente con dos términos, 9 términos y 19 términos. El análisis columna por columna da resultados diferentes, pero con un poder de compresión equivalente en el caso de esta imagen.

d4. Es interesante la comparación de la reconstitución obtenida (en función al número de términos conservados en la descomposición de Fourier) con la reconstitución precedente (con la ayuda de CA o con la ayuda de SVD)

Nota 1: una representación gráfica de los niveles de gris para cada fila puede ser obtenida a partir del botón "Curves of grey levels" (pulse varias veces para escanear toda la imagen).

Nota 2: todas las imágenes creadas son registradas en el formato bitmap (extensión: ".bmp") en el directorio del archivo de la imagen analizada.

Nota 3: la compresión por SVD o CA no depende del orden de las filas y de las columnas de la tabla (de manera contraria a la compresión de Fourier). Sin embargo, esta compresión por ejes principales que se puede calificar de "compresión estructural" (porque ignora las posiciones relativas de los elementos) da resultados satisfactorios.

VI.4.4 Ejecución de otros ejemplos

- Haga Clic en el botón: `SVD and CA of images`, en la sección **DtmVic-Images** del menú principal de Dtm-Vic.
 - Aparece (ver aquí arriba) la ventana "Reconstitution of some small images"

VI.4.4.1 Ejemplo "Baalbeck"

a. Haga Clic en el primer botón `Read (pgm format)` en la sección **Open Greyscale image**.

En el directorio **EX_CO4_Image**, abra el subdirectorio **2_Baalbeck_pgm**. En **2_Baalbeck_pgm**, abra el archivo **Baalbeck.pgm**. Una caja de mensaje recuerda las dimensiones del archivo de imágenes.

b. Si desea visualizar la imagen original, en la sección **Visualization** haga Clic en: `Image (Greyscale)`

c. Luego, en la parte de abajo a la izquierda de la ventana, en la sección: **Compression technique**, haga Clic (para comenzar) en el botón: `Correspondence Analysis` El análisis se efectúa.

Luego, rehacer todas las operaciones de c.1 a c.3, luego de d.1 a d.4.

Este ejemplo es interesante porque pone en evidencia el hecho que una fuerte estructura geométrica de la imagen (aquí: las columnas del templo de Baalbeck) puede contaminar la reconstitución en el caso de los ejes principales.

No es el caso de la reconstitución de Fourier fila por fila: en la que reconstituyendo una fila del lado superior de la imagen (el cielo), ignoramos que hay unas columnas más bajo en la imagen. En cambio es el caso para la reconstitución de Fourier columna por columna...



Templo de Baalbeck. Caso del análisis de correspondencias: imágenes reconstituidas sucesivamente con dos ejes principales, nueve ejes y 50 ejes. Las características estructurales captadas por los primeros ejes se reflejan sobre los ejes siguientes, y hay que alcanzar cerca de 50 ejes para obtener un cielo conforme con el de la imagen inicial.

VI.4.4.2 Ejemplo "Cardinal"

Para abrir el archivo en color del Cardenal de la isla Mauricio, haga Clic en el tercer botón `Read (ppm format)` en la sección **Open colour image**. En el directorio **EX_CO4_Image**, abra el

subdirectorio **3_Cardinal_ppm_color**, luego abra el archivo **Cardinal.ppm**. Una caja de mensaje recuerda las dimensiones del archivo de imágenes.

Nota: recordemos que en el formato ppm, los tres colores de base (Rojo, Verde, Azul) correspondientes a cada píxel tienen posiciones consecutivas sobre la misma fila (cuya longitud es pues tres veces el número de píxeles de la fila). La compresión por SVD o CA no depende del orden de las columnas, lo que significa que no utilizamos el hecho que los tres colores son relativos al mismo píxel!

Sin embargo, la "compresión estructural" funciona. En este caso, la serie de Fourier fila por fila evidentemente no es adaptada (el color aparece sólo con los últimos términos de las series).



Cardenal de la isla Mauricio. Caso del análisis de correspondencias: imágenes reconstituidas sucesivamente con dos ejes principales, 10 ejes y 100 ejes.

VI.4.4.3 Ejemplo "Extra_pgm_ppm"

Este último ejemplo contiene ambos formatos de imagen pgm y ppm.

En el directorio **EX_CO4_Image** abra el subdirectorio **4_Extra_pgm_ppm**, luego abra el archivo **broom.pgm**. Una caja de mensaje recuerda las dimensiones del archivo de imágenes.



Niño que barre un patio. Caso del análisis de correspondencias: imágenes en niveles de gris (pgm) reconstituidas sucesivamente con 2 ejes principales, 10 ejes y 100 ejes.

Ya sea en negro o en color, pulsando la tecla automática permitida por los botones **Series from first term to total**, constatamos que la imagen de la escoba no aparece antes del eje 20: las características estructurales diagonales están en desventaja con la fórmula de reconstitución de los datos...



Niño que barre un patio. Caso del análisis de correspondencias: imágenes en color (ppm) reconstituidas sucesivamente con dos ejes principales, 10 ejes y 100 ejes.

Breves referencias bibliográficas

- Becue M. (1991) *Analisis de Datos Textuales*. CISIA, Saint-Mandé.
- Benzécri J-P. (1973) *L'Analyse des Données*, Tome 1: *La Taxinomie*, Tome 2: *L'Analyse des Correspondances*, Ed. Dunod, Paris (2de. éd. 1976).
- Benzécri J-P. (1992) *Correspondence Analysis Handbook*. Marcel Dekker New York.
- Bouroche J.-M., Saporta G. (1980) *L'analyse des Données*. Coll. Que Sais-je ?, PUF, Paris.
- Bry X. (1995) *Analyses Factorielles Simples*. Economica, Paris.
- Efron B. (1979) Bootstraps methods: another look at the Jackknife, *Ann. Statist.*, 7, p 1-26.
- Escofier B., Pagès J. (1988) *Analyses factorielle simple et multiple*. Ed. Dunod, Paris.
- Gifi A. (1990) *Nonlinear Multivariate Analysis*. Wiley, Chichester.
- Gower J.C., Ross G. (1969) Minimum spanning trees and single linkage cluster analysis. *Appl. Statistics*, 18, 54-64.
- Gower J.C., Hand D.J. (1996) *Biplots*. Chapman and Hall, London.
- Greenacre M. (1984) *Theory and Application of Correspondence Analysis*. Academic Press, London.
- Greenacre M., Blasius J. (editors) (2006) *Multiple Correspondence Analysis and Related Methods*. Chapman and Hall/CRC, London.
- Habert B., Nazarenko A., Salem A. (1997) *Les linguistiques de Corpus*. Armand Colin, Paris.
- Hayashi C., Suzuki T., Sasaki M. (1992) *Data Analysis for Social Comparative research: International Perspective*, North-Holland, Amsterdam
- Jambu M., Lebeaux M-O. (1978) *Classification Automatique pour l'Analyse des Données*. Tome 1: *Méthodes et Algorithmes*, Tome 2: *Logiciels*. Ed. Dunod, Paris.
- Kohonen T. (1989) *Self-Organization and Associative Memory*, Springer-Verlag, Berlin.
- Lambert T. (1986) *Réalisation d'un Logiciel d'Analyse de Données*. (Thèse) Université de Paris-Sud, Dép. Statistique, Orsay.
- Le Roux B., Rouanet M. (2009) *Multiple Correspondence Analysis*. Vol. 163, Sage Publication Inc.
- Lebart L., Morineau A., Tabard N. (1977) *Techniques de la Description Statistique, Méthodes et Logiciels pour l'Analyse des Grands Tableaux*. Ed. Dunod, Paris
- Lebart L., Morineau A. Pleuvret P., Brian E., Aluja T. (1983) *SPAD Système Portable pour l'Analyse des Données*, Tome II. CESIA 82 rue de Sèvres, 75007 Paris
- Lebart L., Morineau A. Bécue M. (1989) *SPAD.T Système Portable pour l'Analyse des Données Textuelles*, Manuel de Référence. CISIA, Paris.
- Lebart L., Morineau A., Warwick K.W. (1984) *Multivariate Descriptive Statistical Analysis, Correspondence Analysis and Related Techniques for Large Matrices*. Wiley, New York.
- Lebart L., Salem A. (1994) *Statistique Textuelle*. Ed. Dunod, Paris.
- Lebart L., Salem A. Becue-Bertaut M (2000) *Analisis estadístico de textos*. Ed. Milenio Lleida
- Lebart L., Piron M., Morineau A., (2006) *Statistique Exploratoire Multidimensionnelle, Visualisation et Inférence en Fouille de Données*. Ed. Dunod, Paris. (4^{ème} édition, refondue). [consultar para una bibliografía más completa].
- Lebart L., Salem A., Berry L. (1998) *Exploring Textual Data*. Ed. Kluwer-Springer, Dordrecht
- Lebart L., Piron M., Steiner J.-F. (2003) *La Sémiométrie*, Ed. Dunod, Paris.
- Lerman I. C. (1981). *Classification et Analyse Ordinale des Données*. Ed. Dunod. Paris.
- Marano P. (1972) Applications de l'analyse factorielle des correspondances à la compression de signaux d'images. *Annals of Telecommunications*, vol. 27, n° 5-6, 163-172.
- Marchand P. (1998) *L'Analyse de Discours Assisté par Ordinateur*. Armand Colin, Paris.

- Murtagh F. (2005). *Correspondence Analysis and Data Coding with R*. Chapman and Hall, Boca Raton, USA.
- Roux M. (1985) *Algorithmes de Classification*. Masson, Paris.
- Salem A. (1987) *Pratique des segments répétés, Essai de statistique textuelle*, Klincksieck, Paris
- Saporta G. (1990) *Probabilités, Analyse des Données et Statistique*. Technip, Paris.
- Tenenhaus M. (2007) *Statistique*. Ed. Dunod, Paris.
- Tuffery S. (2006) *Data Mining et Statistique Décisionnelle*. Technip, Paris
- Volle M. (1980) *Analyse des Données*, Economica, Paris.

© **L2C** Mayo 2013

ISBN 978-2-953777-D-8

Descargado sobre el sitio : www.dtm-vic.com