

REFLEXION SOBRE UNA ESTRATEGIA METODOLOGICA COMPROMISO PARA EL TRATAMIENTO ESTADISTICO DE LAS ENCUESTAS DE OPINION

FRANCISCO JAVIER DÍAZ-LLANOS y SAINZ-CALLEJA

RESUMEN

Las técnicas de tratamientos de datos de encuestas han sido profundamente modificadas por el **Análisis de Datos**, principalmente: **análisis factorial en componentes principales**, **análisis factorial de correspondencias simples**, **análisis factorial de correspondencias múltiples**, **algoritmos de clasificación jerárquica y no jerárquica**, el cuál interviene en una fase preliminar para apreciar la calidad de la información y, orientar la sucesión de tratamientos.

De las dos fases principales de las que se compone un tratamiento estadístico, en profundidad: estructuración y modelización, tan sólo hemos incluido una reflexión sobre la primera.

La aplicación de una estrategia metodológica —**COMPROMISO**—, basada en la Estadística Exploratoria Multidimensional, a una **tabla de datos rectangular**, proveniente de una encuesta de opinión asociada al **cierre de ventas**, nos conducirá —sin duda alguna— a un conjunto de resultados satisfactorios, en cuanto a la **ayuda a la interpretación de los resultados**.

INTRODUCCION

Esta ponencia está dirigida preferentemente a los especialistas de investigación de mercados que tengan como principal objetivo analizar e interpretar de la forma más correctamente posible mediante técnicas actualizadas de **estadística exploratoria multidimensional** las encuestas de opinión asociadas al **cierre de ventas**.

El objetivo principal de ésta ponencia es el de presentar una estrategia metodológica —**COMPROMISO**—, basada en la **estadística exploratoria multidimensional**, constituida por cinco estrategias metodológicas que nos proporcionaran, sin duda alguna, una acción sinérgica en cuanto a la **ayuda a la interpretación de los resultados**, para aplicarla a una **tabla de datos rectangular** proveniente de una encuesta de opinión asociada al **cierre de ventas**.

El hecho de que hayamos incluido en ésta ponencia una reflexión filosófica sobre el método científico es para justificar que nuestra postura denominada «**realista crítica**» está en concordancia con la mantenida por el profesor Jean Pierre BENZECRI, en cuanto a la manera de establecer una **tipología concreta**. Más adelante, se comentará —más detenidamente— algunas opiniones de este insigne maestro dentro de la disciplina del **Análisis de Datos**.

REFLEXIONES EPISTEMOLOGICAS SOBRE EL METODO CIENTIFICO

Todo conocimiento, científico o de cualquier otro tipo, supone una relación entre dos elementos constantes: el objeto a conocer y el sujeto que conoce.

La relación intrínseca entre ellos —sea cual sea su contenido y naturaleza— dependerá en gran medida del tipo de disciplina así como, de su grado de desarrollo, límites estructurales, etc.

Aquí, fijaremos la atención en la estructura sujeto-objeto como modelo subyacente a cualquier conocimiento, sea cual sea su naturaleza.

Ahora bien, sería excesivamente ingenuo considerar que el objeto se limite a captar al sujeto, como una simple reacción inmediata.

La filosofía de la ciencia nos enseña justamente, a entender lo difícil que resulta creer en una relación, directa, del hombre y el mundo. Hasta el lenguaje considerado como instrumento, constituye una mediación, una transformación del objeto que es —así— adecuada a las condiciones subjetivas bajo las cuales puede conocer el sujeto.

Haciendo una abstracción, el Método sería precisamente el mediador entre el objeto y el sujeto.

Genéricamente, el Método empleado por la Ciencia se denomina Método Científico (MC) y se compone de una serie de elementos de diferentes clases: instrumentos de observación o de mediación, axiomas, supuestos científicos, conceptos instrumentales o hipótesis, etc

Según esto, se establecería la siguiente relación o Modelo 1:

$$S \rightarrow MC \rightarrow O$$

jugando MC su papel decisivo puesto que, en función de él actúan S y O.

— El sujeto (científico) debe atenerse a la legalidad interna de éste hasta el punto de asumirlo de un modo lo más completo posible.

Pensemos que la ciencia constituye un tipo de saber donde la libertad del investigador se encuentra bastante limitada por los supuestos, hipótesis, instrumentos, etc., con que debe contar a la hora de enfrentarse al objeto. El método científico considerado como instrumento facilita la labor de investigación, pero al mismo tiempo limita su naturaleza y su alcance. Es cierto que un investigador puede elegir —hasta

cierto punto— entre dos o más métodos científicos; pero, una vez elegido uno de ellos, el camino a seguir se halla, en buena medida limitado por el método elegido de antemano.

— Por parte del objeto, la importancia de MC es aún más decisiva: el objeto queda configurado y delimitado del resto de los objetos del universo por el propio MC.

Esto nos lleva a entender el planteamiento del siguiente problema, cuya envergadura justifica esta breve introducción filosófica:

Supongamos que «no» es indiferente la aplicación de un MC u otro sobre el objeto, es decir, que la naturaleza —concreta— del objeto se encuentra en función de la elección entre dos métodos inarticulables entre sí de modo que, sólo uno de los dos ha de ser asumido por el estudio del objeto. Nos quedaría la siguiente relación 2:

$$\begin{array}{l} \rightarrow MC1 \rightarrow O1 \\ S \rightarrow \\ \rightarrow MC2 \rightarrow O2 \end{array}$$

El problema, obviamente, es cual de los dos métodos ha de elegirse sabiendo que, nos encontramos ante las restricciones señaladas más arriba: tanto MC1 como MC2 son coherentes, son incompatibles, transforman el objeto en O1 y O2, etc. La respuesta no puede ser otra que la siguiente: depende de cuál sea la serie de preferencias y valores del sujeto o investigador, esto es, depende de **qué se esté buscando en el proceso de investigación**. En general, todas las preferencias y valores de investigación pueden agruparse en dos grandes categorías: **universalidad y exactitud**, valores en función de los cuales se encuentra MC1 y MC2. Es decir, que el investigador recurre a uno u otro método en función de que si desea un conocimiento más universal o un conocimiento más exacto. ¿Qué diferencia existe entre uno y otro valor y por qué son incompatibles?

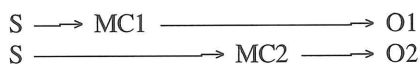
En el fondo de esta alternativa universalidad/exactitud subyace, como sabemos, la vieja problemática epistemológica entre racionalismo y empirismo. Por universal podemos entender aquél tipo de conocimiento que prefiere contener errores de mediación o de interpretación siempre que, ello le permita una visión global y fácilmente generalizable, esto es, una visión que contenga explicaciones de fenómenos inconexos entre sí, de fenómenos de diferentes clases.

Frente a ello, el conocimiento exacto pretende plegarse lo más fielmente posible a los fenómenos que estudia, aunque sus conclusiones sean inaplicables a otro tipo de fenómenos. Ni que decir tiene que, **la ciencia debe procurar articular lo más posible la exigencia de exactitud con la de universalidad**. La situación óptima sería englobar las explicaciones (e1,e1,e3,e4) de los fenómenos concretos (F1,F2,F3,F4), en un segundo nivel de explicaciones (ea,eb) más universales —por tanto menos exactos—, que a su vez, serían englobados en una teoría (T) completamente universal dado que, su nivel de abstracción (por tanto de inexactitud), es máximo.

Esta sería la situación óptima, pero no es, desde luego, aquélla en la que estamos. La razón puede deberse a que no existe aún en nuestra disciplina una teoría global T que contenga todos los niveles de explicación que se necesitan. Una prueba de ello es

—a nuestro juicio— que no están claras las funciones que unos y otros deben desempeñar.

El siguiente esquema sintetiza claramente las dos situaciones:



La diferencia esencial entre MC1 y MC2 radica en que: MC1, que se encuentra situado —más— cerca de S significará **mayor exactitud**: con MC conoceremos más fenómenos pero los conoceremos peor que con MC2, con el que conocemos un menor número de fenómenos, pero de un modo más exacto. La cuestión reside en saber qué grado de exactitud deseamos alcanzar de la investigación científica.

El caso que ahora nos ocupa, el el **Análisis de Datos** para adoptar una **tipología concreta**.

En un número apreciable de ocasiones, se han venido aplicando **hipótesis distribucionales** pero, creemos que debemos insistir en que este tipo de hipótesis —además de por llevar adosada una ley determinada de probabilidad, al prejuzgar de antemano la estructura de los datos a analizar— representan un rígido «a priori» cuya función mediadora va mucho más allá, —a nuestro juicio— que la producida por las mediaciones conceptuales (hipótesis).

Esto implica que el objeto —resultado— generado por la aplicación de la **hipótesis distribucional** no puede dejar de suministrar una información en exceso distorsionadora, en cuyo caso la universalidad de las conclusiones obtenidas ha supuesto la anulación de la tendencia a la exactitud.

Nuestra posición creemos que se ajusta a la denominada «realista crítica», esto quiere decir que intenta dirigirse a los datos de acuerdo con un modelo de conocimiento cuyo valor fundamental será la exactitud procurando, (en la medida que esto sea posible) articularlo a una «ley de probabilidad determinada» que es lo que dará la dimensión de universalidad y que: debe hallarse en función de la pretensión de exactitud, de tal manera que debe darse el trayecto MC2 → MC1.

De no ser posible, nos parece preferible esperar a disponer nuevos datos y objetos de estudio.

Así pues, dado que —para la interpretación de los resultados— la aplicación de un MC es «indispensable» y, puesto que, la deformación de la realidad mediante el mismo resulta «inevitable» (pues no hay mediaciones de grado «0» es decir, completamente transparentes), nos parece que, una mediación debe permitir la mayor fuerza explicativa. Esto es, que represente el compromiso más equilibrado posible entre lo abstracto (universal) y lo concreto (exacto).

Por esta razón, pensamos que lo que —en gran medida— se adecua es un modelo filosófico «realista-crítico» como punto de partida.

En suma, la posible/s solución/s (objetivo/s) a las tensiones —«campo de tensión»— establecido entre estas dos exigencias, aparentemente contrapuestas, depende-

rá de la propia naturaleza del objetivo/s por un lado y, del investigador (sujeto), por otro.

ELECCION DEL METODO CIENTIFICO PARA LA TOMA DE DECISION DE UNA ESTRATEGIA METODOLOGICA COMPROMISO PARA EL CIERRE DE VENTAS

Hemos tomado la decisión de elegir para la elaboración de las cinco estrategias metodologicas elementos instrumentales (**métodos factoriales** y **algoritmos de clasificación automática**), que funcionan sin la necesidad de la introducción de **hipótesis distribucionales a priori** ya que, los resultados obtenidos, aun con perdida de grado de universalidad serán mas fidedignos.

CREDIBILIDAD DE LOS ELEMENTOS INSTRUMENTALES

Los resultados que se obtengan mediante la aplicación de los elementos instrumentales elegidos, tendrán tanto mayor credibilidad cuanto se verifiquen dos aspectos fundamentales:

- 1) Que el cuestionario esté diseñado por especialistas en el **cierre de ventas**
- 2) Que el jefe de ventas tenga experiencia en el tema del **cierre de ventas**

En cuanto concierne al primer punto hemos de indicar que, el cuestionario ha sido diseñado por el profesor Neil RACKHAM especialista en el **cierre de ventas**.

En lo que respecta al segundo punto, hemos de recordar que, en los casos en los que el encuestador que se ha dirigido al jefe de ventas de una empresa para que rellene el cuestionario, ha advertido que éste no esta lo suficientemente enterado sobre dicho tema (lo cual ocurre en ocasiones), dicha empresa ha sido descartada automáticamente.

Cumplidos estos dos requisitos, la responsabilidad del establecimiento de una **tipología concreta** para llevar a cabo una **estrategia metodológica optima** para el **cierre de ventas**, va a depender en la «casi práctica totalidad», de los elementos instrumentales elegidos.

RECOGIDA Y ANALISIS ESTADISTICO DE LA INFORMACION

Esta parte está estructurada según la definición que propone el profesor W. WENTZ de Investigación de Mercados:

«La Investigación de Mercados es la recogida y análisis de la información para asistir a la dirección en la toma de las decisiones comerciales».

RECOGIDA DE LA INFORMACION

La **recogida de la información** es el **pilar fundamental** para que —a posteriori— las cinco estrategias metodológicas que vamos a proponer en la etapa: **análisis estadístico de la información**, tengan cierto grado de credibilidad.

Con carácter general hemos de indicar que, la **encuesta** que se proponga para responder a un objetivo de un tema concreto (por el que ha sido elaborada), ha de ser diseñada por verdaderos especialistas en el dicho tema.

En nuestro caso, hemos elegido el excelente cuestionario sobre el «**cierre de ventas**», diseñado por el gran especialista: profesor Neil RACKHAM, que se encuentra recogido en su libro titulado: «Making Major Sales» (comercializado bajo el título: «Como vender productos de Alta Inversión», en los países de habla hispana).

Sobre la traducción actual española, hemos introducido las modificaciones morfológicas y semánticas —que hemos estimado imprescindibles— para la correcta comprensión del mensaje completo, contenido en cada una de las sentencias (preguntas), de dicho cuestionario.

El cuestionario puede dirigirse a las 509 empresas que figuran en el ranking por ventas.

Dicho cuestionario, será debidamente complementado por los jefes de ventas de estas empresas, contiene 15 preguntas codificadas según una escala de Likert, es decir, la respuesta a cada pregunta esta caracterizada por cinco opciones que se puntuaran de la siguiente manera:

<i>Totalmente de acuerdo</i>	<i>De acuerdo</i>	<i>Indeciso</i>	<i>En desacuerdo</i>	<i>En total desacuerdo</i>
1	2	3	4	5

Por lo tanto, la **información recogida** —para ser susceptible de ser analizada en la siguiente etapa: **análisis estadístico de la información**— esta caracterizada por:

Una **tabla de datos rectangular** cuyas filas representan 509 empresas y sus columnas 15 preguntas, todas ellas a 5 modalidades.

A continuación, vamos a mostrar las 15 preguntas que contempla el cuestionario diseñado por N.RACKHAM:

C01) La fase de «cierre» es la mas valiosa y delicada de todas las técnicas para conseguir e incrementar las ventas.

C02) Tratar de cerrar una venta con mucha frecuencia disminuye las posibilidades de éxito.

C03) A menos que se posea un profundo conocimiento sobre múltiples técnicas de cierre, no será posible vender con eficacia.

C04) Incluso al principio de una venta, **jamás perjudica utilizar un cierre de prueba.**

C05) La deficiencia en el cierre es la causa más común de ventas fallidas, sin ningún efecto.

C06) La probabilidad de que el cliente realice la compra disminuye si éste detecta que ustedes, «los vendedores» realizan técnicas de venta.

C07) Al vender, no deben cerrar con demasiada frecuencia.

C08) Las técnicas de cierre no surten efecto con compradores profesionales.

C09) El ABC de las ventas es «Always be closing» (Estar siempre cerrando).

C10) Su otro comportamiento, «el principio de la visita» y **no** su técnica de cierre, es determinante en la decisión de comprar, por parte del cliente

C11) Deben tratar de cerrar cada vez que vean una señal de compra.

C12) A partir del momento en que entren en la oficina del cliente, deben actuar como si la venta ya hubiera sido hecha.

C13) Si el cliente se resiste al cierre de prueba, es señal de que con una alta probabilidad, hubiera debido cerrar con más firmeza.

C14) Independientemente de lo buenas que sean sus otras habilidades, jamás tendrán éxito a menos que tengan buenas técnicas de cierre.

C15) Utilizar técnicas de cierre al principio de la visita es una forma de provocar el stress y la hostilidad de los clientes.

DISCURSION ENTRE LAS ESCUELAS DE ANALISIS DE DATOS A LA HORA DE TOMAR LA DECISION DE LA PROPUESTA DE UNA ESTRATEGIA METODOLOGICA

Existen dos métodos científicos basados en diferentes filosofías para llevar a cabo un proceso metodológico.

Un filosofía se fundamenta en la consideración del cumplimiento de **hipótesis distribucionales a priori** y, otra, el de la no consideración del cumplimiento de **hipótesis distribucionales a priori**.

Estas filosofías estan, en cierta medida, asociadas a dos grandes escuelas que en la actualidad compiten entre ellas en cuanto al análisis y exploración de los datos. Nos referimos a la **escuela francesa** y a la **escuela anglosajona**.

Ciertamente que un compromiso entre las dos sería lo más acertado en el caso hipotético que se verificaran las hipótesis previas bajo un marco de estadística inferencial paramétrica.

A continuación vamos a comentar los rasgos más significativos de cada una de estas dos escuelas.

ESCUELA FRANCESA

Dentro de la **escuela francesa** existen varias subescuelas para abordar el mismo objetivo. Unas no consireran el cumplimiento de las hipótesis distribucionales a priori, otras si, y, por último, otras adoptan un comportamiento entre las dos primeras.

Hemos de indicar que entre los investigadores integrantes a la primera opción sostienen que el cumplimiento de las hipótesis distribucionales —por ejemplo, hipótesis de Laplace-Gauss multidimensional— no es más que un juego que está fuera de la realidad sobre todo cuando se desee establecer una «tipología concreta». Como representante principal de esta primera subescuela hemos de destacar a la personalidad indiscutible del profesor Jean Pierre BENZECRI de la Universidad de Pierre et Marie-Curie de Paris.

Jean Pierre BENZECRI es autor de múltiples artículos y libros de **Análisis de datos** y en uno de ellos titulado. «L'Analyse des Données. Tome 2. L'Analyse des Correspondances» contempla un principio que lo enuncia de la siguiente manera: «**Statistique n'est pas probabilité**. Sur le nom de **Statistique Mathématique**, des auteurs (qui, je vous le dis en française, n'écrivent guère dans notre langue..) ont édifié une ponpeuse discipline, **riche en hypothèses qui ne son jamais satisfaites dans la pratique**. Ce n'est pas de ces auteurs qu'il faut attendre la solution de nous problèmes typologiques».

Para el profesor Jean Pierre BENZECRI las bases del **Análisis de Datos** son mas bien algebraicas o geométricas que probabilísticas. Asi pues, el profesor Jean Pierre BENZECRI concentra su interes en la aproximación descriptiva, (postura «realista crítica») que nosotros sostenemos muy vivamente.

Sin embargo, dentro de la **escuela francesa** hay otras subescuelas en la cual plantean modelos en los cuales consideran el establecimiento de **hipótesis distribucionales a priori**. Entre ellos podemos destacar a los profesores Henry CAUSSINUS y Philippe BESSE de la Universidad de Toulouse III, Jeanne FINE de la Universidad de Toulouse II, Gilles CELEUX del INRIA, Alain POUSSE de la Universidad de Pau et de Pays de l'Adour.

En última instancia, los profesores Jean-Jacques DROESBEKE de la Universidad Libre de Bruxelles, Bernard FISHET de la Universidad d'Aix-Marseille y Philippe TASSI de Médiametrie ponen de manifiesto en la introducción del libro: «Modèles pour l'Analyse des Données Multidimensionnelles» la conveniencia de la utilización conjunta del Análisis de Datos tal como lo presenta el profesor Jean Pierre BENZECRI y desde el punto de vista inferencial clásico. Es decir, plantean un compromiso entre las dos subescuelas ya aludidas.

ESCUELA ANGLOSAJONA

En general, los investigadores pertenecientes a la **escuela anglosajona**, aunque provengan de distintas subescuelas son más afines para atacar un objetivo.

En este sentido, el profesor J.W.TUKEY en su artículo publicado en el año 1962 titulado: «The future of data analysis», pone de manifiesto no sólo el tratamiento de datos sin modelo probabilístico a priori, sino también a posteriori. Es decir, **adopta el compromiso, para el tratamiento de los datos, entre la Estadística Descriptiva y la Estadística Inferencial Paramétrica.**

En la página 2 de este artículo J.W.TUKEY escribe: «All in all, I have come to feel that my central interest is in data analysis, which I take to include, among other things: procedures for analyzing data, technique for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis essier, more precise and more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data «y añade más adelante: «Large parts of data analysis are inferential in the sample-to- population sense....»».

Así pues, el tratamiento estadístico de la información que propone el profesor Jean Pierre BENZECRI está contenido en el que propone el profesor J.W.TUKEY, aunque según el primero, la segunda componente de carácter probabilístico invalida cualquier «**tratamiento tipológico concreto**» que deseemos establecer a partir de una información de carácter multidimensional.

Por otra parte, las discrepancias en los métodos de estudio, existentes entre los distintos miembros, son apreciablemente menores que aquellas existentes entre las subescuelas francesas y sus correspondientes homólogos anglosajones.

En lo que respecta a este punto, podemos ver que existen notables diferencias entre la forma de analizar las **tablas a tres índices en la estadística** por las subescuelas anglosajonas y por las subescuelas francesas que no consideran el cumplimiento de **hipótesis distribucionales a priori.**

Esta diferencia se ve claramente reflejada en la forma de analizar este tipo de datos por el profesor P.M.KROONENBERG de la Universidad de Leiden y por el profesor Yves ESCOUFIER de la Universidad de Montpellier.

Así, el profesor P.M.KROONENBERG muestra, en su libro titulado: «Three-Mode Principal Component Analysis», diversos métodos de **Análisis de Datos**, analizados de forma muy diferente que la **escuela francesa**. En concreto, el profesor Yves ESCOUFIER analiza las **tablas a tres índices en la estadística** mediante su operador **WD**. La utilización de este operador está reflejada no sólo en numerosas tesis doctorales y artículos, sino también en el libro de la profesora Christine LAVIT de la Universidad de Montpellier prologado por Yves ESCOUFIER cuyo título es: «Analyse Conjointe de Tableaux Quantitatives».

JUSTIFICACION DE LA NECESIDAD DE LA APLICACION CONJUNTA DE, AL MENOS, UN METODO FACTORIAL Y UN ALGORITMO DE CLASIFICACION AUTOMATICA

La necesidad de la complementariedad entre los **métodos factoriales** y los **algoritmos de clasificación automática** permite pronunciarse, no sólo sobre la realidad de las clases, sino también, acerca de sus posiciones relativas, forma, densidad, así como, de su dispersión.

En concreto, el hecho de que tanto el **análisis factorial de correspondencias múltiples: MULTC** [método factorial] como el **algoritmo de clasificación cruzada: CROMUL** [algoritmo de clasificación no jerárquica] trabajen sobre las mismas nubes de puntos con las mismas métricas justifican su complementariedad en la búsqueda sistemática de **clases estables**.

El compromiso del número de clases que hemos de fijar a priori en el programa **CROMUL** nos proporcionará de manera simultánea tanto las empresas pertenecientes a cada una de las clases como las modalidades asociadas a las variables nominales.

Las delimitaciones de cada una de las clases vendrán reflejadas en los **planos factoriales** extraídos del **análisis factorial de correspondencias múltiples**.

En nuestro caso concreto, esta sería una de las etapas a tener en cuenta que nos conducirá, sin duda alguna, al establecimiento de una **tipología concreta** que proporcionará al departamento de la Dirección Comercial elementos suficientes para indicar a sus jefes de ventas la estrategia óptima que tiene que aconsejar a sus vendedores para el cierre de ventas.

La utilidad de la **complementariedad** entre los **métodos factoriales** y los **algoritmos de clasificación** está reflejada en numerosos artículos y libros de texto escritos por profesionales del **Análisis de Datos** de gran prestigio tales como: los profesores Jean Pierre BENZECRI de la Universidad de Pierre-et-Marie-Curie, Bernard BURTSCHY del ENST, Yves ESCOUFIER de la Universidad de Montpellier, Gérard GOVAERT de la Universidad de Metz, Michel JAMBU del CNRS-France-Telecom, Ludovic LEBART del CNRS, Alain MORINEAU del CISIA, Marie PIRON de ORSTOM, Henry RALAMBONDRAINNY del INRIA, Henry ROUANET y Maurice ROUX del CNRS, entre otros.

A su vez, esta conclusión de carácter práctico está fundamentada teóricamente. En este sentido, el profesor H.RALAMBONDRAINNY del INRIA pone de manifiesto desde un punto de vista teórico, en una monografía titulada «Problèmes d'optimisation relatifs aux tableaux multiples. Point de vue de la **clasificación automática**», la relación existente entre los **Métodos Factoriales** y los **algoritmos de clasificación automática** mediante el uso del **álgebra tensorial**.

El **álgebra tensorial** ha sido ya utilizada en el **Análisis de Datos**, según nuestro conocimiento, por el profesor J.P.BENZECRI en su primera intervención sobre el **análisis factorial de correspondencias** en el invierno del año 1963. Diez años más tarde publica un libro titulado «L'analyse des correspondences» en el cual incluye el **álgebra tensorial** en el desarrollo teórico del mismo. De lo que se desprende que

ademas de que «TEORIA y PRACTICA» no estan reñidas sino todo lo contrario, para poder entender de forma aceptable las técnicas de **Análisis de Datos** es fundamental tener un buen conocimiento del **álgebra tensorial**.

CRITERIOS DE ELECCION DE LOS METODOS FACTORIALES Y ALGORITMOS DE CLASIFICACION AUTOMATICA

Es obvio que el mercado de especialistas en el tema de **Análisis de Datos** está **segmentado** en pequeños mercados constituídos por especialistas que abordan el tema de la misma forma. Estos pequeños mercados estan mas separados entre si en la **escuela francesa** que en la **escuela anglosajona**.

Ni que decir tiene que, aún dentro de la misma subescuela: **para abordar un objetivo, el resultado será distinto si no existe una uniformidad estricta en las subrutinas que contenga el programa o paquetes de programas para la realización de los cálculos**. A este respecto, una de las preguntas clave que hay que cuestionarse a la hora de la elección de un programa es, **¿cuál es el algoritmo que contiene el programa para el cálculo de los valores propios y vectores propios?** Por supuesto, «la elección de dicho algoritmo va a intervenir directamente en la representación de los **puntos-individuos** y **puntos-variables** en los **planos factoriales** extraídos de cualquiera de los **métodos factoriales**».

Dada la estructura de la **tabla de datos** que va a someterse a las tecnicas de **Análisis de Datos**, entre los diversos **métodos factoriales** y **algoritmos de clasificación automática** existentes en el mercado, hemos tomado la decisión de elegir a) entre los **métodos factoriales** el: **análisis factorial de correspondencias simples**, el **análisis factorial de correspondencias multiples** y el **análisis factorial discriminante lineal**, y, b) entre los **algoritmos de clasificación automática**: el **algoritmo de clasificación jerárquica ascendente** cuya distancia será la **distancia euclidea clásica** y la del χ^2 segun sea la estructura de la **tabla de datos**, y, cuyo criterio de agregación será el de maximizar el momento central de orden dos de la reunión de dos clases y el **algoritmo de clasificación no jerárquica** conocido con el nombre de **algoritmo de clasificación cruzada**.

CONSIDERACIONES DE INTERES SOBRE LOS PROGRAMAS Y PAQUETES DE PROGRAMAS DE ANALISIS DE DATOS

Dado que los programas o paquetes de programas utilizados por la **escuela francesa** y la **escuela anglosajona** estan concebidos con diferentes filosofías sobre todo en cuanto concierne a la **ayuda a la interpretación de los resultados** las conclusiones finales seran, por lo tanto, diferentes. Hemos de destacar que muchos de los programas diseñados recientemente por la **escuela francesa**, que presentan una gran utilidad, a nivel de la **ayuda a la interpretación de los resultados**, ni siquiera se abordan por la **escuela anglosajona**.

Este hecho es debido a que los paquetes de programas norteamericanos a diferencia de los franceses han sido concebidos, en su primera versión, en un periodo de tiempo muy anterior y, en consecuencia, no contemplan las **técnicas de Análisis de Datos** más

recientes contenidas no sólo en los programas sino también, en los paquetes de programas que se están desarrollando en la actualidad tanto en centros de investigación como en sociedades privadas en colaboración con investigadores pertenecientes al ámbito universitario.

A título orientativo mostraremos, al final de esta ponencia, un conjunto de programas y paquetes de programas provenientes de Francia y Norte-América, que consideramos más representativos.

ELECCION DEL PROGRAMA O PAQUETE DE PROGRAMAS PARA LA BUSQUEDA SISTEMATICA DE UNA TIPOLOGIA CONCRETA

Hemos considerado que, la elección del programa o paquete de programas que vamos a utilizar en la **búsqueda sistemática de una tipología concreta**, asociada —en nuestro caso— al **cierre de ventas**, va a estar en función de la calidad científica del centro de investigación en donde se han desarrollado estos programas o paquetes de programas, y por tanto, de sus responsables científicos.

Al final de esta ponencia, hemos reflejado una lista tanto de los centros de investigación como, de los profesores responsables de los programas o paquetes de programas.

A la vista de esta lista, sería difícil tomar la decisión de, cuál de los centros de investigación y, por consiguiente, el responsable o los responsables científicos del programa o paquete de programas, presentan mayor calidad científica. Así pues, vamos a exponer las razones por las cuales elegiremos: por una parte, los paquetes de programas provenientes de Norte-América o de Francia y, por otra, dentro de éstos, aconsejaremos cuáles —a nuestro juicio— serían los más apropiados —según el caso— para llevar a cabo una estrategia metodológica «**compromiso**».

En cuanto al primer punto, en primera instancia, nos inclinaremos por los concebidos en Francia para su aplicación al **cierre de ventas** por los siguientes motivos:

- a) Están más depurados en cuanto a la **ayuda a la interpretación de los resultados**. En éste sentido el paquete de programas **STATlab** presenta una excelente **ayuda a la interpretación de los resultados**. En concreto, en cuanto a los **algoritmos de clasificación automática que contiene su ayuda a la interpretación de los resultados**, es difícilmente mejorable por otro paquete de programas proveniente de Norte-América. A nuestro juicio, el único competidor podría ser el **S-Plus** por su similitud con el **STATlab**.
- b) El **algoritmo de clasificación no jerárquica: CROMUL** adaptado a una **tabla de datos** con variables nominales, (cuya utilización indicaremos más adelante), que nosotros sepamos, tan sólo es contemplado por los paquetes de programas **SICLA** y **MODULAD**.
- c) El **algoritmo de clasificación jerárquica ascendente: CAHVR** contenido en el paquete de programas **MODULAD**, presenta propiedades muy interesantes frente a los concebidos en los paquetes de programas de Norte-América. El

programa CAHVR construye una jerarquía por agregaciones sucesivas, según el criterio del momento de orden dos de una partición. Este programa trabaja directamente sobre una **tabla de datos rectangular** y, calcula las distancias a medida que tengamos necesidad de ellas utilizando, a su vez, el algoritmo rápido «**de los vecinos recíprocos**». Estas dos particularidades, le permiten tratar —en un tiempo razonable— grandes **tablas de datos cuantitativos**.

- d) El **análisis factorial de correspondencias múltiples**: MULTC adaptado a una **tabla de datos** con variables nominales, que nosotros sepamos, lo contemplan muy pocos paquetes de programas de origen Norte-Americano. No obstante, tenemos noticias que en el año 1994 es posible adaptar un módulo de **Análisis de Datos «a la francesa»** en la nueva versión del STATGRAPHIS.Plus.windows.
- e) El **análisis factorial discriminante lineal**: DISC, presenta propiedades muy interesantes frente a los concebidos en los paquetes de programas de Norte-America. En el año 1990, se ha implementado un módulo de **validación cruzada**. El criterio de **validación cruzada** no sólo es más rápido que el de **bootstrap** sino que, además, es más adecuado cuando el tamaño de la muestra no es excesivamente elevado. El paquete de programas MODULAD contiene el programa DISC.
- f) El **algoritmo de formas fuertes o clases estables**: FF, presenta propiedades muy interesantes frente a los concebidos en los paquetes de programas de Norte-América. El paquete de programas MODULAD contiene, el programa FF.

En lo que concierne al segundo punto, dado que, es muy difícil discernir entre: la calidad científica de los centros de investigación, y por consiguiente, por ende, la de sus responsables científicos, hemos considerado pertinente, introducir un indicador diferencial entre ellos que consiste en retener, en primera instancia —en cuanto a los paquetes de programas se refiere— aquellos en los que exista una **interconexión**.

Pues bien, entre los que figuran en la mencionada lista, los que presentan esta condición, son los que a continuación mostramos: MODULAD, SICLA y SPAD. Estos tres paquetes de programas contemplan ampliamente el espectro de la Estadística Exploratoria Multidimensional. Ni que decir tiene, tanto los **métodos factoriales** como, los **algoritmos de clasificación**, que aconsejaremos para llevar a cabo una estrategia metodológica compromiso, están contenidos «al menos», en uno de ellos.

Entre los **métodos factoriales** caben destacar:

- a) El **análisis factorial en componentes principales**: COMPL debido a los profesores Ludovic LEBART y Alain MORINEAU.
- b) El **análisis factorial de correspondencias simples**: CORBI debido a los profesores Ludovic LEBART y Alain MORINEAU.
- c) El **análisis factorial de correspondencias múltiples**: MULTC debido a los profesores Ludovic LEBART y Alain MORINEAU.

- d) El **análisis factorial discriminante lineal: DISC** debido a los profesores C.CALLANT y G.CELEUX.

Entre los **algoritmos de clasificación** caben destacar:

- a) El **algoritmo de clasificación cruzada: CROMUL** debido al profesor Gérard GOVAERT.
- b) El **algoritmo de clasificación no jerárquica: KMEAN** debido al profesor P.VILLOING.
- c) El **algoritmo de formas fuertes o clases estables: FF** debido al profesor S.BOCHI.
- d) El **algoritmo de clasificación jerárquica ascendente: CAHVR** debido a Maurice ROUX.

TABLAS DE DATOS TRATADAS EN LAS ESTRATEGIAS METODOLOGICAS

La estrategia metodológica será función no sólo de la dimension de la **tabla de datos**, sino también de su estructura.

A continuación presentamos cinco tipo de **tablas de datos** que seran susceptibles de someterlas a análisis en las cinco estrategias metodológicas que mas adelante mostraremos:

TIPOLOGIA DE TABLAS DE DATOS	DIMENSIONES DE LA TABLA DE DATOS	
	FILAS	COLUMNAS
Tabla de orden	509 [empresas]	15 [variables]
Tabla de frecuencias	15 [variables]	5 [modalidades]
Tabla de rangos	15 [variables]	5 [modalidades]
Tabla disyuntiva completa	509 [empresas]	75 [modalidades]
Tabla de contingencias	X [grupos]	75 [modalidades]

Por consiguiente, presentaremos cinco estrategias metodológicas que, utilizadas conjuntamente generarán una acción sinérgica entre ellas que, nos conducirá a una «**tipología concreta**» para el **cierre de ventas**. Esto va a permitir, al departamento de Dirección Comercial, proporcionar indicaciones valiosas a los jefes de venta para que transmitan a los vendedores **cual** debe ser el momento crítico para el **cierre de ventas**.

Ni que decir tiene, durante el proceso de cada una de las cinco estrategias metodológicas que a continuación presentamos, se hace imprescindible la presencia de un profesional que conozca la problemática inherente al **cierre de ventas** para que, de esta manera, las conclusiones finales sean lo más coherentes posibles.

No obstante, antes de desarrollar las cinco estrategias metodológicas haremos una pequeña discusión, no sólo sobre las categorías de **algoritmos de clasificación automática** y dentro de éstas, de la distancia y el criterio de agregación sino también, acerca del **análisis factorial discriminante**.

CATEGORIAS Y COMENTARIOS SOBRE LOS ALGORITMOS DE CLASIFICACION AUTOMATICA

Categorías de Algoritmos de Clasificación Automática

Existen dos categorías de **algoritmos de clasificación automática**:

1.^a **Algoritmos de clasificación jerárquica.**

2.^a **Algoritmos de clasificación no jerárquica.**

Entre los **algoritmos de clasificación jerárquica** existen dos subcategorías:

1.1. **Ascendentes**

1.2. **Descendentes**

En principio, la elección de una u otra categoría va a estar en función de las dimensiones de la **tabla de datos original**.

De tal manera que:

a) Cuando la dimensión de la **tabla de datos** no sea elevada, utilizaremos **algoritmos de clasificación jerárquica**, y, b) cuando lo sea, es decir, cuando se trate de clasificar un grupo considerable de objetos (más de 300), identificados por un conjunto de variables, es aconsejable utilizar la técnica conocida con el nombre de: «**hybrid clustering**» (**clasificación mixta**). Dicha técnica consiste —fundamentalmente— en la aplicación conjunta de un **algoritmo de clasificación**: en primer lugar —**no jerárquica**— a la **tabla de datos original** y, en segundo lugar, uno de la —**jerárquica**— considerando como objetos, los baricentros de las **formas fuertes (clases estables)** establecidas con anterioridad.

En cuanto concierne a la categoría de **algoritmo de clasificación** a utilizar, el profesor MOLLIERE indica en el artículo: «Stratégies de classification pour de grand ensembles de données» que si disponemos de un número importante de unidades a clasificar, la elección de un **algoritmo de clasificación jerárquica** presenta el inconveniente de: la imposibilidad de hacer aparecer —claramente— una estructura en clases —más significativa— en los últimos niveles de agregación del **dendrograma** que son justamente esos que interesan en la práctica.

Consejos prácticos sobre el Algoritmo de Clasificación Automática a aplicar

La **distancia** y el **criterio de agregación** van a ser los dos elementos fundamentales tanto en los **algoritmos de clasificación jerárquica** como **no jerárquica** a la hora de la construcción del **dendrograma** y de las **clases fijadas a priori**, respectivamente. A continuación vamos a exponer un conjunto de consejos prácticos tanto sobre la distancia como sobre el criterio de agregación.

Consejos prácticos sobre la distancia a utilizar

Fruto de la amplia experiencia de los grandes expertos del **Análisis de Datos**, se derivan los siguientes consejos:

El tipo de distancia a utilizar, será función de la estructura de la **tabla de datos** sobre la cuál, se va a realizar el **algoritmo de clasificación automática**, tanto de los individuos como de las variables.

a) Si la **tabla de datos** de partida es una **tabla de contingencias**, es aconsejable utilizar la **distancia distribucional** —posteriormente denominada distancia del χ^2 , debido a una observación del analista K.KRICKEBERG— debido, a que ésta —a diferencia de las demás distancias— presenta la propiedad de la «**equivalencia distribucional**».

b) Si la **tabla de datos** de partida esta constituida por las **coordenadas de los individuos (empresas) en los ejes factoriales**, es aconsejable utilizar la **distancia euclídea clásica**.

Es aconsejable que la **tabla de datos** que vayamos a someter a un **Análisis Exploratorio Multidimensional**, sea lo más homogénea posible.

De tal manera que a tal efecto hemos de indicar que, así como, la **distancia distribucional** garantiza la homogeneidad de la **tabla de datos** —ya que actúa con **perfiles de frecuencias relativas**— la **distancia euclídea clásica**, tan sólo la garantiza cuando la **tabla de datos** esté constituida por las **coordenadas de los individuos en los ejes factoriales**.

Por consiguiente, actuar de una forma o de otra —en cuanto a la distancia se refiere— es casi igual en un caso que en otro.

La conveniencia y operatividad de la utilización de estas distancias según sea la **tabla de datos** serán analizadas más adelante.

Consejos prácticos sobre el criterio de agregación elegido

Si adoptamos la decisión de aplicar un **algoritmo de clasificación jerárquica ascendente**, el **criterio de agregación de maximizar el momento central de orden dos de la reunión de dos grupos** sería el más conveniente ya que, este criterio proporciona —simultáneamente— **clases homogéneas y bien separadas las unas de las**

otras. Es decir: el hecho de que, por una parte, las clases sean homogéneas no comporta más que contengan individuos semejantes entre ellos; y, por otra parte, que las clases estén bien separadas las unas de las otras, nos asegura que los individuos que pertenezcan a dos clases diferentes no se parezcan más, que los individuos que pertenezcan a la misma clase.

CATEGORIAS Y COMENTARIOS SOBRE EL ANALISIS FACTORIAL DISCRIMINANTE

Categorías de Análisis Factorial Discriminante

Existen tres categorías de **análisis factorial discriminante**:

- a) **Análisis factorial discriminante lineal.**
- b) **Análisis factorial discriminante cuadrático.**
- c) **Análisis factorial discriminante logístico.**

CONSEJOS PRACTICOS SOBRE LA ELECCION DEL ANALISIS FACTORIAL DISCRIMINANTE

Entre los tres métodos de **análisis factorial discriminante**, ya aludidos con anterioridad, es aconsejable **retener el análisis factorial discriminante lineal** dado que, se ha demostrado que —desde el punto de vista práctico— es el más adecuado en cuanto a la **ayuda a la interpretación de los resultados** si se realizan las debidas transformaciones a las variables, para garantizar, no sólo la estabilidad de las varianzas sino también, el que las poblaciones se comporten según una **ley de Laplace-Gauss**.

Este tipo de análisis presenta la particularidad de que si, las variables son ortogonales; es decir, los **ejes factoriales** extraídos de un **método factorial —análisis factorial de correspondencias simples, análisis factorial de correspondencias múltiples—** la **distancia de Mahalanobis** —que es la que se emplea en dicho análisis— entre los **baricentros asociados a cada una de las clases estables**, se transforma en la **distancia euclídea clásica** por el simple hecho de que, la **varianza entre los ejes factoriales** es igual a 1 y la **covarianza** es igual a 0. Esto nos permite interpretar los **planos factoriales** extraídos de un **análisis factorial discriminante lineal** más fácilmente, dado que la distancia entre los **puntos-individuos** en los **planos factoriales** será la medida por una regla.

COMETARIOS SOBRE LA ELECCION DEL NUMERO DE EJES

La necesidad de reducir el espacio de representación es directamente proporcional a la dimensión de la tabla de datos.

En cuanto a la elección del número de **ejes factoriales** —que nosotros sepamos— existen ocho procedimientos que se encuentran recogidos y explicados de forma didác-

tica en el libro de los profesores Ludovic LEBART, Alain MORINEAU y Marie PIRON titulado «Statistique Exploratoire Multidimensionnelle».

ESTRATEGIAS METODOLOGICAS

Antes de mostrar el proceso metodológico contenido en cada una de las cinco estrategias metodológicas, es conveniente realizar un **proceso de depuración previo**. De tal manera que, a tal efecto proponemos el siguiente esquema a seguir:

Proceso de depuración previo al análisis de datos

Los pasos a seguir son los siguientes:

1.º Eliminación de aquellas modalidades que presenten efectivo nulo. El programa **ELIMOD** contenido en el paquete de programas **SICLA** contempla esta opción.

2.º Eliminación de aquella /s modalidad /s que presente efectivo débil. El paquete de programas **SPAD**, permite fijar un umbral NMIN a partir del cual las modalidades de efectivos a NMIN son consideradas como raras.

3.º Las modalidades que no presenten ningún efectivo (no respuestas), las analizaremos a parte.

Vamos a considerar que, durante el **proceso de depuración**, la **tabla de datos originales** no ha sufrido ninguna modificación en sus dimensiones originales.

Nota: si como resultado del **proceso de depuración** tan sólo tuviésemos una sola modalidad para cualquiera de las variables nominales, prodeceríamos a eliminarla/s.

PRIMERA ESTRATEGIA METODOLOGICA

Las fases metodológicas que contempla la primera estrategia son las que a continuación mostramos:

PRIMERA FASE

La primera fase consiste en la transformación de la **tabla de orden** [509 (empresas), 15 (variables)] a la **tabla de frecuencias absolutas** [15 (variables), 5 (modalidades)].

Ni que decir tiene que, dado que, este tipo de transformación no es inmediata, en ocasiones no se contempla en los paquetes de programas de **análisis estadístico multidimensional** y, por consiguiente, nos veremos obligados a escribir un programa en fortran y, a posteriori, implementarlo en nuestro ordenador.

Una vez realizada esta transformación pasamos a la siguiente fase.

SEGUNDA FASE

La segunda fase consiste en la realización de aplicaciones sucesivas del **análisis factorial de correspondencias simples** a la **tabla de frecuencias absolutas** [15,5].

Esta fase la estructuramos de la siguiente manera:

— **En primer lugar**, aplicaremos el **análisis factorial de correspondencias simples** a la **tabla de frecuencias absolutas** [15 × 5].

— **En segundo lugar**, observaremos los **planos factoriales más representativos** para el estudio de nuestros datos, que no son necesariamente aquellos que absorban la mayor inercia de la nube de puntos, y, aquellos **puntos-individuos (preguntas:15)** y **puntos-variables** (modalidades:5), que representen un porcentaje considerable en la formación de los **ejes factoriales**, los consideraremos como: **puntos-individuos y puntos-variables suplementarios**, es decir, que **no interviene en los cálculos de ajuste y no participan en la formación de los ejes factoriales**.

— **En tercer lugar**, aplicaremos el **análisis factorial de correspondencias simples** a la **tabla de frecuencias absolutas** [15 × 5] pero, considerando como **puntos-individuos y puntos-variables activos** aquellos que **no se han considerado como suplementarios**.

Los **puntos-individuos suplementarios y puntos-variables suplementarios**, aunque no participen en los cálculos de los **ejes factoriales**, aparecerán representados en los **planos factoriales** por su gran utilización en la **ayuda a la interpretación de los resultados**.

Dado que, la dimensión de la **tabla de datos** es pequeña [15 × 5], con este tipo de análisis es suficiente para extraer un conjunto de resultados aceptables asociados a las 15 variables nominales y a sus respectivas modalidades.

SEGUNDA ESTRATEGIA METODOLOGICA

Las fases metodológicas que contempla la segunda estrategia son las que a continuación mostramos:

PRIMERA FASE

La primera fase consiste en la transformación de la **tabla de frecuencias absolutas** [15 × 5] a la **tabla de rangos** [15 × 5]

Ni que decir tiene que, dado que, este tipo de transformación es directo, cualquier paquete de programas de **análisis estadístico multidimensional** contempla esta opción.

SEGUNDA FASE

La segunda fase consiste en la realización de aplicaciones sucesivas del **análisis factorial de correspondencias simples** a la **tabla de rangos** [15 × 5].

El procedimiento de análisis será análogo al descrito en la **primera estrategia metodológica**.

TERCERA ESTRATEGIA METODOLOGICA

Esta estrategia metodológica, no es aconsejable aplicarla cuando la **tabla de datos** es de dimensiones elevadas ya que, la nube de puntos proyectada en los **planos factoriales** tendría la forma de una gran mancha y, por consiguiente, nos impediría dar cualquier tipo de conclusiones.

No obstante, vamos a desarrollarla para que el lector tenga una alternativa de análisis en el caso hipotético de que disponga de una **tabla de datos** cuyas dimensiones no sean elevadas; es decir, tal como indicamos en el apartartado **categorías de algoritmos de clasificación automática**, que no supere de 300 individuos.

Las fases metodológicas que contempla la tercera estrategia son las que a continuación mostramos:

PRIMERA FASE

La primera fase consiste en la realización de aplicaciones sucesivas del **análisis factorial de correspondencias múltiples** a la **tabla de datos: T** [509 (empresas),15 (variables nominales)].

Esta fase 1.^a estructuramos de la siguiente manera:

— **En primer lugar**, aplicaremos el **análisis factorial de correspondencias múltiples** a la **tabla de datos: T** [509,15].

Para ello proponemos tres alternativas equivalentes que a continuación mostramos:

Primera alternativa

La alternativa que vamos a proponer a continuación, será de utilidad en el caso hipotético de que, el paquete de programas de **análisis estadístico multidimensional** que utilicemos, no contemple el **análisis factorial de correspondencias múltiples**. En este caso concreto, este método factorial, lo podemos realizar mediante la aplicación del **análisis factorial de correspondencias simples**, que «si» lo contienen gran parte de los paquetes de programas, a la **tabla de datos: BD** [584,75], construída mediante la **yuxtaposición vertical** de la **tabla de Burt** [75,75] y la **tabla disyuntiva completa** [509,75].

En esta nueva **tabla de datos:BD** [584,75], hay que tener presente dos puntos básicos:

1.º las **filas** correspondientes a la **tabla de Burt** [75,75] actuarán como **individuos activos**.

2.º las **filas** correspondientes a la **tabla disyuntiva completa** [509,75] actuarán como **individuos suplementarios**.

Segunda alternativa

Dada la **equivalencia** existente entre el **análisis factorial de correspondencias múltiples** y el **análisis factorial en componentes principales**, esta alternativa se desarrollará mediante la aplicación de este último **análisis factorial**.

Tercera alternativa

Esta alternativa consiste en la aplicación del **análisis factorial de correspondencias múltiples** a la **tabla de datos T** [509,15].

Ni que decir tiene que, esta alternativa será la más indicada.

— **En segundo lugar**, observaremos los **planos factoriales** más representativos, que no son necesariamente aquellos que absorban la mayor inercia de la nube total y, aquellos **puntos-individuos** (empresas) y **puntos-variables** (modalidades asociadas a las preguntas), que representen un porcentaje considerable en la formación de los **ejes factoriales**. Los consideraremos como: **puntos-individuos suplementarios** y **puntos-variables suplementarios**.

— **En tercer lugar**, aplicaremos el **análisis factorial de correspondencias múltiples** a una **tabla de datos: T** [509,15] considerando como **puntos-individuos activos** y **puntos-variables activos** aquellos que **no** hemos considerado como **suplementarios**.

SEGUNDA FASE

La segunda fase consiste en la realización de aplicaciones sucesivas del **algoritmo de clasificación cruzada: CROMUL** a una **tabla de orden**.

La segunda fase, la estructuraremos de la siguiente manera:

— **En primer lugar**, aplicaremos el programa **CROMUL** a la **tabla de orden TFC - TFC** es la **tabla de datos** resultante después de la eliminación de los **puntos-individuos suplementarios** (empresas) y los **puntos-variables suplementarios** (modalidades asociadas a las preguntas).

— **En segundo lugar**, como resultado de la aplicación del programa **CROMUL** a la **tabla de orden: TFC** obtendremos simultáneamente, por una parte, las clases formadas por las empresas y, por otra, las clases formadas por las modalidades de las variables.

El **número de clases** constituido tanto para las empresas como, para las modalidades asociadas a las preguntas, han de ser **fijadas «a priori»**.

Suponiendo que estas clases están bien fijadas, obtendremos tanto las empresas como, las modalidades de las variables, dentro de las clases ya establecidas.

— **En tercer lugar**, procederemos a la delimitación de las clases en los **planos factoriales**, extraídos mediante la aplicación del **análisis factorial de correspondencias múltiples** a la **tabla de orden: TFC**.

TERCERA FASE

La tercera fase, consiste en la aplicación del **análisis factorial discriminante lineal** teniendo en cuenta que, se verifican las condiciones ya mencionadas en el apartado: **consejos prácticos sobre la elección del análisis factorial discriminante a una tabla de datos rectangular**, particionada en clases ya establecidas, cuyos elementos sean las **coordenadas de las empresas en los ejes factoriales más representativos**, es decir, que absorban un porcentaje de la inercia aceptable de la nube total.

CUARTA FASE

La cuarta fase, consiste en la validación de las clases.

Entre los tres criterios de validación de clases: **validación cruzada**, **jackknife** y **bootstrap**, hemos retenido el de la **validación cruzada**, por las razones —ya mencionadas— en el apartado: **elección del programa o paquetes de programas para la búsqueda sistemática de una tipología concreta**.

QUINTA FASE

Esta fase, consiste en la aplicación del **algoritmo de clasificación jerárquica ascendente: CAHVR** a la **tabla de datos**, constituida por las **coordenadas de los baricentros de las clases estables en los ejes factoriales** ya retenidos.

Uno de los criterios prácticos para la **elección del número de ejes factoriales** es el de la **regla empírica**. Haciendo uso de este criterio, fruto de la observación del **histograma de los valores propios**, extraídos del **análisis factorial de correspondencias múltiples**, retendremos los **ejes factoriales** que expliquen un porcentaje de la inercia de la nube razonable, respecto al total.

Acto seguido, aplicaremos el programa **CAHVR** a la **tabla de datos** —ya aludida— considerando como distancia: la **distancia euclídea clásica**.

El hecho de que, trabajemos con las **coordenadas de las clases estables en los ejes factoriales más representativos**, es una práctica interesante dado la importancia que supone, **no sólo la reducción de la dimensión del espacio de representación sino también, porque presenta la ventaja de eliminar las fluctuaciones aleatorias que**

constituyen —en general— lo esencial de la varianza recogida en las direcciones de los p-q (p y q representan el conjunto de las p variables y q factores) últimos ejes (variaciones no sistemáticas contenidas en los datos).

Esta observación la ponen de manifiesto los profesores Ludovic LEBART, Alain MORINEAU y Marie PIRON en su libro: «Statistique Exploratoire Multidimensionnelle».

SEXTA FASE

Esta fase, consiste en hacer uso de un conjunto de elementos de interés para la toma de decisión:

1.º Tipología de empresas y de las modalidades asociadas a las variables nominales.

2.º Variables más representativas en cada clase.

En cuanto concierne al primer punto, haremos uso de los siguientes elementos:

a) **planos factoriales** más significativos en el sentido de las representaciones simultáneas de los **puntos-individuos** y **puntos-variables**.

b) **dendrograma** obtenido por un **algoritmo de clasificación jerárquica ascendente**.

c) criterios que nos indicarán por donde realizar el **corte al dendrograma** tales como:

c1) la curva de los índices de agregación frente a los nudos.

c2) la curva de $M^2(v)/M^2(I)$ frente a los nudos.

d) criterio debido a la experiencia. Este criterio nos dirá como tenemos que hacer el corte al **dendrograma**: ¿**línea horizontal o línea sinuosa?**

Esta reflexión la ponen de manifiesto los profesores F.BENZECRI, J.P.BENZECRI, Y.L.CHEUNG y S.MAÏZA en el artículo titulado: «Aides à l'interprétation en Classification Ascendante Hiérarchique».

En lo que se refiere al segundo punto, una práctica interesante propuesta por el profesor Michel JAMBU, para conseguir resolver el problema de la segmentación de otra forma diferente a la tradicional es: construir una representación gráfica de la **clasificación jerárquica** explicada por las variables. Esta práctica consiste en, colocar el nivel de cada uno de los nudos en un recuadro que contenga el número de dichos nudos y, las variables o ejes explicativos de la subdivisión del nudo.

Dicha práctica se contempla en el libro titulado: «Exploration Informatique et Statistique des Données».

CUARTA ESTRATEGIA METODOLOGICA

El elemento diferenciador de esta estrategia metodológica con respecto a la tercera es que, en lugar de proceder a la aplicación de un **análisis factorial discriminante lineal** a las clases ya retenidas, realizaremos un **análisis factorial de correspondencias simples** a la **tabla de contingencias** cuyas filas representan los grupos y cuyas columnas representan las modalidades asociadas a las 15 variables nominales. Los valores de la **tabla de datos** serán la suma de los unos —contenidos en la **tabla disyuntiva completa restringida**— en cada clase para cada una de las 75 modalidades.

Este tipo de análisis nos conducirá a resultados muy interesantes, y, por simplicidad frente al **análisis factorial discriminante lineal**, es muy aconsejable practicarlo.

QUINTA ESTRATEGIA METODOLOGICA

Esta estrategia metodológica es aconsejable aplicarla, cuando partamos de una **tabla de datos** de elevadas dimensiones.

La quinta estrategia metodológica contiene las fases típicas de un **algoritmo de clasificación mixta**:

PRIMERA FASE

Esta fase consiste en, la aplicación de un **algoritmo de clasificación no jerárquica**: **NUBES DINAMICAS, KMEAN** etc.

Si tomamos la decisión de aplicar el **algoritmo de clasificación no jerárquica: KMEAN** a una **tabla de datos rectangular**, las columnas de ésta, deben estar constituidas bien por **variables cuantitativas**, o bien por las **coordenadas de los individuos en los ejes factoriales**.

En nuestro caso concreto, la **tabla de datos** contendrá las **coordenadas de los individuos (empresas)**, en los **ejes factoriales** extraídos de un **análisis factorial de correspondencias múltiples**.

SEGUNDA FASE

Esta fase consiste en, la aplicación de un **algoritmo de clasificación jerárquica ascendente**:: **CAHVR**

La distancia que utilizaremos será la **distancia euclídea clásica**.

TERCERA FASE

Esta fase consiste en, la realización de un **corte al dendrograma** para la obtención de clases, mediante la utilización de las técnicas, ya mencionadas con anterioridad.

CUARTA FASE

Esta fase, consiste en la consolidación por reafectación.

La homogeneidad de las clases puede ser optimizada por reafectación.

QUINTA FASE

Esta fase, consiste en la construcción de una función de decisión a partir de la **tipología** ya establecida.

Dicha función de decisión se establecerá mediante la aplicación del **análisis factorial discriminante lineal** a una **tabla de datos rectangular**, cuyas columnas serán los **ejes factoriales más representativos** y cuyas filas serán el **número de clases** ya establecidas con anterioridad.

OBSERVACIONES DE INTERES EN CUANTO A LA OPERATIVIDAD DE LAS ESTRATEGIAS METODOLOGICAS

1.^a Aquellos investigadores que deseen llevar a cabo no sólo la **primera alternativa de la tercera estrategia metodológica**, sino también la técnica de los **puntos-individuos suplementarios** y **puntos-variables suplementarios**, les aconsejamos — muy vivamente— que consulten la publicación de los profesores Luis CAMPOS SANCHEZ y Fco Javier DIAZ-LLANOS y SAINZ-CALLEJA titulada: «Procedimientos de gestión informática utilizando el STATlab y sus aplicaciones en la **Estadística Exploratoria Multidimensional**».

2.^a Es muy importante hacer notar que: **durante cada una de las cinco estrategias metodológicas, aplicaremos —tantas veces como sea preciso— los métodos factoriales como, los algoritmos de clasificación contenidos en ellas, con el fin de llegar a unos resultados los más coherentes posibles con la realidad.**

3.^a Tanto los **puntos-individuos como los puntos-variables suplementarios**, los analizaremos siguiendo —fiélmemente— las indicaciones que contempla el profesor Pierre CAZES en sus múltiples trabajos, asociados al **análisis factorial de correspondencias**, cuyas referencias figuran en la bibliografía.

UTILIZACION DEL PROGRAMA CROMUL CONTENIDO EN EL PAQUETE DE PROGRAMAS SICLA

Hemos creído conveniente la introducción de este epígrafe, no sólo porque el **algoritmo de clasificación cruzada: CROMUL [clasificación no jerárquica]** es de gran utilidad en el **tratamiento estadístico de las encuestas**, sino también, porque dicho algoritmo no se encuentra con frecuencia ni en los libros de texto de **Análisis de Datos** ni en los paquetes de programas.

El hecho de que mostremos la aplicación del programa **CROMUL** con el paquete de programas **SICLA** es porque, hemos trabajado más intensamente con éste que con el paquete de programas **MODULAD**.

En última instancia, antes de mostrar la utilización del programa **CROMUL**, hemos de recordar que es necesario introducir al ordenador —si es que no lo tiene ya implementado— un **editor**, dado que no esta incluido en el paquete de programa **SICLA**.

El **editor** que hemos introducido en el ordenador nos lo ha proporcionado el investigador Jean Baptiste DENIS del INRA. Dicho **editor** recibe el nombre de **editeur**.

En esta aplicación vamos a utilizar los programas **ELIMOD** y **CROMUL** contenidos en **SICLA**.

Designaremos por la letra F la orden escrita en francés, y por la letra E su traducción al castellano.

Las letras que estan en **negrita**, son las ordenes introducidas por nosotros desde el ordenador y el resto, son las ordenes que aparecen en pantalla como respuesta a las nuestras.

Desde la pantalla de nuestro ordenador nos disponemos a introducir las ordenes pertinentes para la utilización del paquete de programas **SICLA**.

C:\>**CD ESICLA**

F C:\ESICLA>**EDITEUR**

E C:\ESICLA>EDITOR

F SPECIFIER LE NOM DU FICHER DE TEXTE À EDITER

E ESPECIFICAR EL NOMBRE DEL ARCHIVO DE TEXTO PARA EDITAR

F SI LE FICHER N'EXISTE PAS,IL VA ÊTRE CREE

E SI EL ARCHIVO NO EXISTE, SERA CREADO

F APPUYER SUR ESC POUR QUITTER LE PROGRAMME

E PULSAR LA TECLA ESC PARA ABANDONAR EL PROGRAMA

F NOM DU FICHER? **MAR.DI**

E NOMBRE DEL ARCHIVO? MAR.DI

F C:\ESICLA\MAR.DI.....FICHER INTROUVABLE

E C:\ESICLA\MAR.DI.....ARCHIVO NO ENCONTRADO

F CREATION D'UN NOUVEAU FICHER (O/N)? **O**

E CREACION DE UN NUEVO ARCHIVO (O/N)? **O**

C:\ESICLA\MAR.DI

CAR=1 COU=1 ^J=AIDE

CONTROL J

F POUR ENTRER DU TEXTE

E PARA INTRODUCIR EL TEXTO

F DEPLACEMENT DU COURSEUR
E DESPLAZAMIENTO DEL CURSOR

F EFFACEMENT
E BORRAR

F FIN DE L'EDITION
E FIN DE LA EDICION

F APPUYER SUR ESC POUR CONTINUER
E PULSAR LA TECLA ESC PARA CONTINUAR

F VOULEZ-VOUS AFFICHER LE GUIDE DE REFERENCE RAPIDE (O/N)? O
E QUIERE UD VER LA GUIA DE REFERENCIA RAPIDA (O/N)? O

C:\ESICLA\MAR.DI

CAR=1 COU=1 ^J=AIDE

F DEPLACEMENT DU COURSEUR EFFACEMENT FIN DE L'EDITION

E DESPLAZAMIENTO DEL CURSOR BORRAR FIN DE LA EDICION

1234567891011

51525354

00 CV01 PREGUNTA 1

2 0 0

01 TO01 TOTALMENTE DE ACUERDO

02 DE01 DE ACUERDO

03 IN01 INDECISO

04 ED01 EN DESACUERDO

05 TD01 EN TOTAL DESACUERDO

00 CV02 PREGUNTA 2

2 0 0

01 TO02 TOTALMENTE DE ACUERDO

02 DE02 DE ACUERDO

03 IN02 INDECISO

04 ED02 EN DESACUERDO

05 TD02 EN TOTAL DESACUERDO

00 CV03 PREGUNTA 3

2 0 0

01 TO03 TOTALMENTE DE ACUERDO

02 DE03 DE ACUERDO

03 IN03 INDECISO

04 ED03 EN DESACUERDO

05 TD03 EN TOTAL DESACUERDO

00 CV04 PREGUNTA 4

2 0 0

01 TO04 TOTALMENTE DE ACUERDO

02 DE04 DE ACUERDO

03 IN04 INDECISO

04 ED04 EN DESACUERDO

05 TD04 EN TOTAL DESACUERDO

00 CV05 PREGUNTA 5

2 0 0

01 TO05 TOTALMENTE DE ACUERDO

02 DE05 DE ACUERDO

03 IN05 INDECISO

04 ED05 EN DESACUERDO			
05 TD05 EN TOTAL DESACUERDO			
00 CV06 PREGUNTA 6	2	0	0
01 TO06 TOTALMENTE DE ACUERDO			
02 DE06 DE ACUERDO			
03 IN06 INDECISO			
04 ED06 EN DESACUERDO			
05 TD06 EN TOTAL DESACUERDO			
00 CV07 PREGUNTA 7	2	0	0
01 TO07 TOTALMENTE DE ACUEDO			
02 DE07 DE ACUERDO			
03 IN07 INDECISO			
04 ED07 EN DESACUERDO			
05 TD07 EN TOTAL DESACUERDO			
00 CV08 PREGUNTA 8	2	0	0
01 TO08 TOTALMENTE DE ACUERDO			
02 DE08 DE ACUERDO			
03 IN08 INDECISO			
04 ED08 EN DESACUERDO			
05 TD08 EN TOTAL DESACUERDO			
00 CV09 PREGUNTA 9	2	0	0
01 TO09 TOTALMENTE DE ACUERDO			
02 DE09 DE ACUERDO			
03 IN09 INDECISO			
04 ED09 EN DESACUERDO			
05 TD09 EN TOTAL DESACUERDO			
00 CV10 PREGUNTA 10	2	0	0
01 TO10 TOTALMENTE DE ACUERDO			
02 DE10 DE ACUERDO			
03 IN10 INDECISO			
04 ED10 EN DESACUERDO			
05 TD10 EN TOTAL DESACUERDO			
00 CV11 PREGUNTA 11	2	0	0
01 TO11 TOTALMENTE DE ACUERDO			
02 DE11 DE ACUERDO			
03 IN11 INDECISO			
04 ED11 EN DESACUERDO			
05 TD11 EN TOTAL DESACUERDO			
00 CV12 PREGUNTA 12	2	0	0
01 TO12 TOTALMENTE DE ACUERDO			
02 DE12 DE ACUERDO			
03 IN12 INDECISO			
04 ED12 EN DESACUERDO			
05 TD12 EN TOTAL DESACUERDO			
00 CV13 PREGUNTA 13	2	0	0
01 TO13 TOTALMENTE DE ACUERDO			
02 DE13 DE ACUERDO			
03 IN13 INDECISO			
04 ED13 EN DESACUERDO			
05 TD13 EN TOTAL DESACUERDO			

```

00 CV14 PREGUNTA 14                2  0 0
01 TO14 TOTALMENTE DE ACUERDO
02 DE14 DE ACUERDO
03 IN14 INDECISO
04 ED14 EN DESACUERDO
05 TD14 EN TOTAL DESACUERDO
00 CV15 PREGUNTA 15                2  0 0
01 TO15 TOTALMENTE DE ACUERDO
02 DE15 DE ACUERDO
03 IN15 INDECISO
04 ED15 EN DESACUERDO
05 TD15 EN TOTAL DESACUERDO

```

CONTROL k → SALE ^K → TECLEAR X

```

F SAUVEGARDE DU FICHIER DE TEXTE C:\ESICLA\MAR.DI
E GUARDAR EL ARCHIVO DE TEXTO C:\ESICLA\MAR.DI

```

```

F C:\ESICLA>EDITER
E C:\ESICLA>EDITOR

```

```

F SPECIFIER LE NOM DU FICHIER DE TEXTE À EDITER
E ESPECIFICAR EL NOMBRE DEL ARCHIVO DE TEXTO PARA EDITAR
F SI LE FICHIER N'EXISTE PAS, IL VA ÊTRE CREE
E SI EL ARCHIVO NO EXISTE, SERA CREADO
F APPUYER SUR ESC POUR QUITTER LE PROGRAMME
E PULSAR LA TECLA ESC PARA ABANDONAR EL PROGRAMA

```

```

F NOM DU FICHIER? MAR.DO
E NOMBRE DEL ARCHIVO? MAR.DO

```

```

E C:\ESICLA\MAR.DO ..... FICHIER INTROUVABLE
F C:\ESICLA\MAR.DO ..... ARCHIVO NO ENCONTRADO

```

```

F CREATION D'UN NOUVEAU FICHIER (O/N)? O
E CREACION DE UN NUEVO ARCHIVO (O/N)? O

```

C:\ESICLA\MAR.DO

```

1234      10 12 14 16 18 20 22 24 26 28 30 32 34 36 38
E001      2  1  4  4  4  3  2  4  4  3  2  4  4  4  3
.         .  .  .  .  .  .  .  .  .  .  .  .  .  .
.         .  .  .  .  .  .  .  .  .  .  .  .  .  .
.         .  .  .  .  .  .  .  .  .  .  .  .  .  .
.         .  .  .  .  .  .  .  .  .  .  .  .  .  .
E509      1  3  1  2  1  4  4  3  2  2  3  5  4  3  3

```

CONTROL K → SALE ^K → TECLEAR X

F SAUVEGARDE DU FICHER DE TEXTE C:\ESICLA\MAR.DO
E GUARDAR EL ARCHIVO DE TEXTO C:\ESICLA\MAR.DO

C:\ESICLA>CONVER

F ENTREZ LE NOM DE LA COMMANDE [AIDE:? FIN:/
E INTRODUZCA EL NOMBRE DE LA ORDEN [AYUDA:? FIN:/]]

ENRFIC

F ENTREZ UN NOM POUR LA STRUCTURE DE DONNEES À CREER
E INTRODUZCA UN NOMBRE PARA LA ESTRUCTURA DE DATOS A
CREAR

MAR.SDO

F → FICHER:MAR.SDO
E → ARCHIVO:MAR.SDO

F ENTREZ UN TITRE POUR LES DONNEES
E INTRODUZCA UN TITULO PARA LOS DATOS

I.D.M. CIERRE DE VENTAS

F ENTREZ LE NOM DU FICHER DICTIONNAIRE
E INTRODUZCA EL NOMBRE DEL ARCHIVO DICCIONARIO

MAR.DI

F ENTREZ UN MAJORANT DU NOMBRE TOTAL DE VARIABLES TRAI-
TEES
E INTRODUZCA UNA COTA SUPERIOR DEL NUMERO TOTAL DE VA-
RIABLES TRATADAS

15

F ENTREZ UN MAJORANT DU NOMBRE TOTAL DE MODALITES TRAI-
TEES
E INTRODUZCA UNA COTA SUPERIOR DEL NUMERO TOTAL DE LAS
MODALIDADES TRATADAS.

75

F DESIREZ-VOUS UNE IMPRESSION DU DICTIONNAIRE POUR CONTROLE?
E DESEA UD IMPRIMIR EL DICCIONARIO PARA VERIFICARLO?

O

F FIN DE LECTURE DU DICTIONNAIRE
E FIN DE LA LECTURA DEL DICCIONARIO

F Y-A-T-IL DES DONNEES MANQUANTES?
E HAY DATOS AUSENTES?

F 0: PAS DE VALEURS MANQUANTES
E 0: NO HAY VALORES AUSENTES

F 1: VALEURS MANQUANTES POUR LES VARIABLES QUANTITATIVES
E 1: VALORES AUSENTES PARA LAS VARIABLES QUANTITATIVAS

F 2: VALEURS MANQUANTES POUR LES VARIABLES QUALITATIVES
E 2: VALORES AUSENTES PARA LAS VARIABLES CUALITATIVAS

F 3: VALEURS MANQUANTES POUR TOUTES LES VARIABLES
E 3: VALORES AUSENTES PARA TODAS LAS VARIABLES

0

F FIN DE ENRFIC
E FIN DE ENRFIC

F ENTREZ LE NOM DE LA COMMANDE [AIRE: ? FIN:]
E INTRODUZCA EL NOMBRE DE LA ORDEN [AYUDA: ? FIN:]

ENRI

F NOMBRE MAXIMUN D'INDIVIDUS?
E NUMERO MAXIMO DE INDIVIDUOS?

509

F NOMBRE DE CARACTERES DE L'IDENTIFICATEUR DES INDIVIDUS?
E NUMERO DE CARACTERES DEL INDENTIFICADOR DE LOS INDIVIDUOS?

4

F TYPE DE FORMAT DE LECTURE DES DONNEES
E TIPE DE FORMATO DE LECTURA DE DATOS

F 1: LECTURE DES DONNEES SUIVANT UN FORMAT FORTRAN
E 1: LECTURA DE DATOS SIGUIENDO UN FORMATO FORTRAN

F 2: LECTURE DES DONNEES SUIVANT UN FORMAT LIBRE
E 2: LECTURA DE DATOS SIGUIENDO UN FORMATO LIBRE

2

F ENTREZ LE NOM DU FICHER CONTENANT LES DONNEES
E INTRODUZCA EL NOMBRE DEL ARCHIVO QUE CONTIENE LOS DATOS

MAR.DO

F FIN DE ENRI
E FIN DE ENRI

F ENTREZ LE NOM DE LA COMMANDE [AIDE: ? FIN:\]
E INTRODUZCA EL NOMBRE DE LA ORDEN [AYUDA: ? FIN:\]

/

SICLA FIN DU MODULE CONVER LISTING LIS17_22

C:\ESICLA>DEL LIS17_22

C:\ESICLA>SICLA2

F ENTREZ LE NOM DE LA COMMANDE [AIDE: ? FIN:\]
E INTRODUZCA EL NOMBRE DE LA ORDEN [AYUDA: ? FIN:\]

ELIMOD

F NOMBRE DE MODALITES D'EFFECTIFS NULS:0
E NUMERO DE MODALIDADES DE EFECTIVOS NULOS:0

F TOUTES LES MODALITES ONT ETE CONNSERVEES
E TODAS LAS MODALIDADES HAN SIDO CONSERVADAS

F FIN DE ELIMOD
E FIN DE ELIMOD

F ENTREZ LE NOM DE LA COMMANDE [AIDE: ? FIN:\]
E INTRODUZCA EL NOMBRE DE LA ORDEN [AYUDA: ? FIN:\]

/

C:\ESICLA>MSICLA

F ENTREZ LE NOM DE LA COMANDE [AIDE: ? FIN:\]
E INTRODUZCA EL NOMBRE DE LA ORDEN [AYUDA: ? FIN:\]

CROMUL

F SELECTION DE VARIABLES
E SELECCION DE VARIABLES

1A\$

F SELECTION DES INDIVIDUS
E SELECCION DE INDIVIDUOS

1A\$

F VARIABLES SELECTIONNEES
E VARIABLES SELECCIONADAS

CVO1 CVO2 CV03 CV04 CV05 CV06 CV07 CV08
CV09 CV10 CV11 CV12 CV13 CV14 CV15

F NOMBRE DE CLASSES EN LIGNES?
E NUMERO DE CLASES EN LAS QUE SE AGRUPAN LAS FILAS?

X: REPRESENTA EL NUMERO DE CLASES

F NOMBRE DE CLASSES EN COLONNES?
E NUMERO DE CLASES EN QUE SE AGRUPAN LAS COLUMNAS?

Y: REPRESENTA EL NUMERO DE CLASES

F NOMBRE DE TIRAGE AU HASARD?
E NUMERO DE EXTRACCIONES AL AZAR?

1

ACCEPTEZ-VOUS LES CLASSES VIDES?
E ACEPTA UD LAS CLASES VACIAS?

0

F VOULEZ-VOUS EDITER UNIQUEMENT LE MEILLEUR RESULTAT?
E DESEA USTED EDITAR UNICAMENTE EL MEJOR RESULTADO?

0

F VOULEZ-VOUS L'EDITION REORDONNEE DU TABLEAU DISJONCTIF?
E DESEA UD LA EDICION REORDENADA DE LA TABLA DISYUNTIVA.
F COMPLET ASSOCIE AU TABLEAU INITIAL?
E COMPLETA ASOCIADA A LA TABLA INICIAL?

0

F VOULEZ-VOUS L'EDITION DU TABLEAU DISJONCTIF
E DESEA UD LA EDICION DE LA TABLA DISYUNTIVA
F COMPLET REGROUPESUIVANT LES PARTITIONS EN LIGNES?
E COMPLETA REAGRUPADA SIGUIENDO LAS PARTICIONES EN
FILAS?

0

F VOULEZ-VOUS L'EDITION DU TABLEAU DISJONCTIF
E DESEA UD LA EDICION DE LA TABLA DISYUNTIVA

F COMPLET REGROUPE SUIVANT LES PARTITIONS EN COLONNES?
E COMPELTA REAGRUPADA SIGUIENDO LAS PARTICIONES EN COLUN-
NAS?

O

F FIN DE CROMUL
E FIN DE CROMUL

F ENTREZ LE NOM DE LA COMMANDE [AIDE: ? FIN:]
E INTRODUZCA EL NOMBRE DE LA ORDEN [AYUDA: ? FIN:]

/

C:\ESICLA>EDITEUR LIS17_22 (visualizo el listado en la pantalla)

CONTROL K → X

Si nuestro ordenador no esta conectado a una impresora desde nuestro ordenador hemos de copiar el listado en un diskete de la siguiente manera:

C:\ESICLA>COPY LOS17_77 B:SAL

A continuación, en otro ordenador que este conectado a una impresora hemos de introducir la siguiente orden:

D:\>TYPE B:SAL>LPT1 (obtengo el listado por impresora).

Es conveniente dar las instrucciones precisas a la impresora para que el listado salga en sentido horizontal.

PROGRAMAS Y PAQUETES DE PROGRAMAS PROVENIENTES DE FRANCIA Y DE NORTE-AMERICA

Bajo este epígrafe vamos a poner en relieve no sólo un conjunto de programas y paquetes de programas provenientes de Francia y Norte-América, sino también indicaremos brevemente sus objetivos, y sus reponsables científicos.

- **En primer lugar**, mostraremos una lista por orden alfabético de los programas y paquetes de programas de origen francés.

- **ACPVI** [1983] (programme d'analyse en composantes principales de variables instrumentales). Responsable científico el investigador Robert SABATIER. Université des Sciences et Techniques de Languedoc.

- **ACT** [1994] (analyse conjointe de tableaux:méthode STATIS).

Responsable científico la profesora Christine LAVIT. Université des Sciences et Techniques du Languedoc.

— **ADM** [1988] (analyse discriminante sur variables continues et qualitatives). Responsables científicos los profesores Jean-Jacques DAUDIN y Mohamed SOUKAL. Département de Mathématique et Informatique. INAPG

— **AMADO** [1994] (logiciel interactif de traitement graphique de données). Responsable científico el profesor Jacques BERTIN. Laboratoire de Graphique de l'Ecole des Hautes Etudes en Sciences Sociales de Paris.

— **ANAVAR** (programme d'analyse de la variance). Responsable científico el investigador Jean Baptiste DENIS. INRA.

— **ANAVAR** (un nouveau programme d'ajustement de modèles linéaires par les moindres-carrés). Responsable científico el profesor Thierry CEMBRZYNSKI. INRIA

— **A.TEMP** [1995] (analyse de series temporelle). INRIA.

— **CHADOC.vs** [1984] (logiciel d'analyse des données bien adapté au traitement des enquêtes statistiques). Responsable científico el profesor Jacques LEMAIRE. Université de Nice.

— **COSI.windows** [1994] (logiciel de traitement statistiques de fichiers). STATIRO.

— **CSQlab** [1991] (logiciel de contrôle des processus). SLP.

— **DATAVISION** [1984] (logiciel d'analyse multivariable). Responsable científico el profesor Maurice ROUX. CNRS.

— **DESTIN.windows** [1994] (logiciel de traitement d'enquêtes). STATIRO.

— **DISQUAL** [1989] (logiciel d'analyse discriminante non paramétrique sur variables qualitatives). Responsable científico el profesor Gilbert SAPORTA. Mathématiques-Informatiques du Conservatoire National des Arts et Métiers.

— **EOLE3.windows** [1995] (logiciel le plus performant des étude. Ventilations et analyses de données).

— **LADDAD** [1981] (logiciel de l'association pour le développement et la diffusion de l'analyse des données, **addad**). Responsable científico el profesor Jean Pierre FENELON. L'ISUP.

— **LEAS** [1984] (logiciel d'enseignement et d'analyses statistiques). Responsables científicos los investigadores A.DELCAMP y T.MARTONO.

— **LTSM** [1982] (logiciel de traitement statistique multidimensionnels). Responsable científico el profesor Jean Pierre BENZECRI. Université de Pierre et Marie Curie. Paris.

— **MICRO-MARKETING** [1983] (étude de marche, analyse de données, micro-informatique, aide à la decision). Responsable científico el profesor Jean-Luis CHANDON. Institut d'Administration des Entreprises d'Aix.

- **MODALISA.versión 2.6** [1994] (enquêtes et sondages. Analyses de données)
- **MODULAD.versión 2.1** [1987] (bibliotheque fortan 77 pour l'analyse des données).INRIA.

Programas contenidos en **MODULAD** referenciados en esta ponencia:

- **COMPL.versión 2.0** [1985] (analyse en composantes principales). Responsables científicos los profesores Ludovic LEBART y Alain MORINEAU. CESIA.
- **CORBI.versión 2.0** [1985] (analyse de correspondances simples). Responsables científicos los profesores Ludovic LEBART y Alain MORINEAU.CE-SIA.
- **CROMUL.versión 1.0** [1986] (ce programme effectue une clasification simultanément sur l'ensemble I des individus et sur l'ensemble J de toutes les modalités d'un questionnaire qui peut être considéré comme un tableau de contingence multiple). Responsable científico el profesor Gérard GOVAERT.INRIA.
- **DISC.versión 1.0** [1986] (analyse discriminate linéaire pour k groupes a priori (k quelconque), les variables discriminantes étant quantitatives). Responsable científico el profesor Giles CELEUX.INRIA.
- **DISC** [1990] (programa que contempla la validación cruzada). Responsables científicos los profesores C.CALLANT y G.CELEUX.
- **FF.versión 1.0** [1985] (construit la partition de formes fortes ou classes stables). Responsable científico el profesor S.BOCHI. INRIA.
- **KMENS.versión 3.0** [1985] (procédure de classification non hiérarchique. Les variables sont quantitatives, et peuvent être les coordonnées factorielles). Responsable científico el profesor P.VILLOING. CELAR/CCSA.
- **MNDQL.versión 1.0** [1986] (nuées dynamiques qualitative). Responsable científico el profesor H. RALAMBONDRAINNY. INRIA.
- **MODLI.versión 2.0** [1986] (modèle linéaire). Responsable científico el profesor G.DECOUX y A.KOBILINSKY. INRA
- **MULTC.versión 2.0** [1985] (analyse des correspondances multiples). Responsables científicos los profesores Ludovic LEBART y Alain MORINEAU. CESIA.
- **MLP** [1991] (programme de reseau de neurones multicouches). Responsables científico el profesor Joseph PRORIOL. CNRS.
- **MULCO** [1990] (analyse des correspondances multiples conditionnelle). Responsable científico la profesora Brigitte ESCOFIER. IRISA.

— **NJACK** [1993] (programme de régression non linéaire). Responsable científico el profesor Richard TOMASSONE. Institut National Agronomique.Paris.

— **PAC.versión 1.5** [1995] (programme d'analyse des comparaisons). Responsables científicos los profesores Bruno LECOUTRE y Jacques POITEVINEAU. CNRS.

— **PANAVEA** [1984] (programme d'analyse de la variance). Responsable científico el profesor Giles CARAUX. Unité de Biometrie. Ecole Nationale Sup. d'Agriculture.

— **QUESTION.windows** [1995] (logiciel de traitement d'enquêtes et d'analyse de données).

— **RESID** [1991] (le programme permet une analyse factorielle comparative à deux tableaux). Responsable científico el profesor Roger LAFOSSE. Laboratoire de Statistique et Probabilités. Université Paul-SABATIER.

— **SICLA** [1988](système interactif de classification automatique). Responsable científico el investigador Yves LECHEVALLIER. INRIA.

Programas contenidos en **SICLA** referenciados en esta ponencia:

— **CROMUL** [1993] (classification croisée d'un tableau de variables qualitatives). Responsables científico el profesor Gérard GOVAERT. INRIA.

— **SICLA 1.4** [1994] (système interactif de classification automatique).

— **SIRENE** [1994] (logiciel de scoring par simulation de reseaux neuronaux). STATIRO.

— **SPAD.versión 1.0** [1985] (système portable d'analyse des données). Responsables científicos los profesores Ludovic LEBART, Alain MORINEAU y colaboradores. CISIA

— **SPAD.GF** [1995] (exploration graphiques de plans factoriels issus de SPAD.N et SPAD.T). CISIA

— **SPAD.MC** [1995] (logiciel d'analyse des mesures conjointe. Trade off). Responsable científico el profesor Gilbert SAPORTA. Mathématiques-Informatiques du Conservatoire National des Arts et Métiers.

— **SPAD.N.versión 2.0** [1991]

— **SPAD.N.versión 2.5** [1993]

— **SPAD.versión 3.0 windows** [1996]

— **SPAD.T.versión 1.5** [1994] (analyse textuelles, questions ouvertes, entretiens semi-directifs). Responsable científico el profesor Ludovic LEBART.

— **SPHINX.Plus windows** [1995] (logiciel d'enquête et analyse de données). Responsables científicos los profesores Yves BAULAC, Jean DE LA GARDE y Jean MOSCAROLA. Université de Savoie.

— **STATBOX** [1994] (la boîte à outils statistiques pour microsoft). Grimmer logiciels

— **STATIS.versión 3.0** [1986] (structuration des tables à trois indices en la statistique). Responsable científico la profesora Christine LAVIT. ENSA.INRA.USTL. Unité de Biométrie. Montpellier.

— **STAT-ITCF** [1990] (logiciel d'analyse de données). Institut Techniques de Cereales et des Fourrages. Paris.

— **STATlab.versión 2.1 windows** [1991] (logiciel d'analyse des données). Responsable científico el profesor Michel JAMBU. Université Paris IX- Dauphine.

— **STATlab.versión 3.0 windows** [1997] (logiciel d'analyse des données). Responsable científico el profesor Michel JAMBU. Université Paris IX- Dauphine.

— **STATVIEW 4.0** [1995] (modèles linéaires généralisés et aux plans d'expérience)

— **SUPERANOVA** [1995] (modèles linéaires généralisés et aux plans d'expérience)

— **SURVENT** [1994] (système informatique de gestion d'enquêtes téléphoniques). STATIRO.

— **TIMElab** [1991] (logiciel de prévisions). SLP Infoware.

— **TIMElab** [1997] (logiciel de prévision). SLP Infoware.

— **TRAMEN** [1988] (programa para el tratamiento estadístico de las encuestas de opinión). Responsables científicos los profesores Maria Enriqueta GARCIA MOUTON y Fco Javier DIAZ-LLANOS y SAINZ-CALLEJA. Escuela Técnica Superior de Ingenieros Agrónomos de Madrid.

• **En segundo lugar**, mostraremos una lista por orden alfabético de programas y paquetes de programas de origen norte-americano:

— **ALLOC80** [1982] (discrimination analysis program). Universidad de Leiden.

— **CART** [1984] (a software of clasification and regresion trees) California Statistical. Software, inc. Yorshire Ct. Lafayette, California.

— **EGRET** [1988] (Statistical package. Statistics and epidemiology research corp, Seattle, WA.

— **SPSS** [1978] (statistical packages for social sciences). Universidad de Chicago.

— **SPSS 6.1 windows** [1996]

— **SPSS 6.1.3** [1996]

— **SPSS 7.5 windows** [1997]: tercera versión traducida al castellano y disponible en CD-ROM.

— **OSIRIS III** [1973] (an integrate collection of computer programs for the management and analysis of social science data). Universidad de Michigan.

— **BMDP** [1977] (biomedical computer programs). Universidad de Carolina, Los Angeles.

— **STATGRAPHIS.Plus windows** [1994] (statistical graphics). Las novedades que contempla esta nueva version frente a las anteriores es que existe la posibilidad de implementar dos modulos de **Análisis de Datos**: uno «a la francesa», y otro, «a la anglosajona».

En cuanto se refiere al módulo **Analyse des Données UNISTAT pour Windows** aporta ocho programas de **Análisis de Datos «a la francesa»** a este paquete de programas:

- **Analyse en composantes principales.**
- **Analyse des correspondances.**
- **Analyse de correspondances multiples.**
- **Analyse discriminante pas à pas.**
- **Analyse discriminante stochastique.**
- **Analyse sur tableau de distances ou de dissimilarités.**
- **Classification hiérarchique.**
- **Ellipses de confiance et de tolérance.**

En lo que respecta al módulo **Méthodes Multivariées pour Windows** aporta cinco programas de **Análisis de Datos «a la anglosajona»** a este paquete de programas:

- **Analyse factorielle en facteurs communs et spécifiques.**
- **Analyse discriminante géométrique.**
- **Analyse canonique.**
- **Classification.**

— **SAS** [1979] (statistical analysis system). Universidad de Rayleigh, Carolina del Norte.

— S (an interactive environment for data analysis and graphics). Laboratoires Bell d'ATT.

— S.PLUS 3.2 [1995] (logiciel complet d'analyse de données)

S-PLUS permet une approche du type «exploration —analyse— nouvelle exploration- nouvelle analyse», que ne permettent pas les logiciels plus rigides.

BIBLIOGRAFIA

- BARRENO RODRIGUEZ, E., CRESPO DE LAS CASAS, A., DIAZ-LLANOS y SAINZ-CALLEJA, Fco., TARAZONA LAFARGA, M. T. [1980]: «Estudio estadístico de la vegetación líquénica de los yesos del centro de España». Serie Recursos Naturales. INIA, n.º 4, pp. 139-163.
- BENALI, H. [1985]: Stabilité de l'Analyse en Composantes Principales et de l'Analyse des Correspondances Multiples en présence de certaines types de perturbations. Méthodes de dépouillement d'enquêtes. Thèse de Docteur en Troisième Cycle. Université de Rennes I.
- BENZECRI, J. P. [1973]: L'Analyse des Données. Tome 1. La Taxinomie. Dunod.
- BENZECRI, J. P. [1973]: L'Analyse des Données. Tome 2. L'Analyse des Correspondances. Dunod.
- BENZECRI, J. P. [1982]: «Sur la proportion des paires de voisins réciproques pour une distribution uniforme dans un espace euclidien. Cahiers de l'Analyse des Données, Vol. 7, pp. 185-188.
- BERNARD, M. L., DIAZ-LLANOS y SAINZ-CALLEJA, Fco., ESCOUFIER, Y. [1979]: La méthode STATIS. Une application à l'évolutions des campagnes Languedociennes. Rapport Technique n.º 7905 Montpellier. Ce rapport a fait l'objet d'une communication oral au cours du stage d'Analyse Spatio-Temporelle, organisé par la Formation Permanente de CNRS.
- BURT, C. [1950]: The factorial analysis of qualitative data. British J. of Statist. psychol. 3,3, pp 166-185.
- CALLANT, C., CELEUX, G. [1990]: Nouvelle version de DISC. La revue Modulad n.º 5, pp. 49-51.
- CAMPOS SANCHEZ, L., DIAZ-LLANOS y SAINZ-CALLEJA, Fco. [1997]: Procedimientos de gestión informática utilizando el STATlab y sus aplicaciones en la Estadística Exploratoria Multidimensional. Oficina provincial del registro de la propiedad intelectual. Madrid. Solicitud número: 63.787
- CAZES, P. [1982]: «Note sur les éléments supplémentaires en Analyse des Correspondances».
- a) Pratique et Utilisation, C.A.D., Vol. VII n.º 1, pp. 9-23
- b) Tableau Multiples, C.A.D., Vol. VII n.º 2, pp. 133-154
- CAZES, P. [1990]: «Quelques considerations sur l'utilisation des elements supplementaires en Analyse Factorielle. La revue de Modulad n.º 5, pp. 29-44.
- CELEUX, G. [1988]: «Classification et modèles. Revue de Statistique Appliquée, XXXVI (4), 43-58.
- CELEUX, G., DIDAY, E., GOVAERT, G., LECHEVALLIER, Y., RALAMBON-DRAIN, H. [1989]: Classification Automatique des Données. Environnement Statistique et Informatique. Dunod Informatique.
- CHANDON, J. L., PINSON, S. [1980]: Analyse Typologique. Théorie et applications. Préface de E.DIDAY. Masson.

- DAMBROISE, E., ESCOUFIER, Y., MAS-SOTE, P. [1987]: «Application de l'Analyse des Données à l'élaboration de minisondages d'opinion». *Revue de Statistique Appliquée*, pp. 9-23.
- DE RHAM, C. [1980]: «La Classification hiérarchique selon la méthode des voisin réciproques». *Cahiers de l'Analyse des Données*, Vol, 5,2, pp.135-144.
- DIAZ-LLANOS y SAINZ-CALLEJA, Fco. [1985]: Técnicas multidimensionales para el estudio de la evolución del sector agrario y afines a nivel provincial. Tesis doctoral presentada en la Universidad Politécnica de Madrid. Escuela Técnica Superior de Ingenieros Agrónomos. Departamento de Matemática 1. Director de la tesis. Dr.D. Jose Luís DE MIGUEL ARENAL. Catedrático de Matemática 1.
- DIAZ-LLANOS y SAINZ-CALLEJA, Fco. [1991]: Criterios de segmentación de la realidad empírica. Juicios de valor y mediciones técnicas. Discurso de entrada en la Real Academia de Doctores de Madrid. Depósito legal M-40621-1991.
- DIAZ-LLANOS y SAINZ-CALLEJA, Fco. [1995]: El tratamiento estadístico de las encuestas de opinión, pieza clave el la ingeniería de la demanda. Un enfoque didáctico y conceptual. Ediciones CEES.
- DIAZ-LLANOS y SAINZ-CALLEJA, Fco., ESTELLES ADAM, A., ORTIZ MACI-DE, J. M., TADEO LLUCH, J. L. [1983]: «Étude des huilles essentielles de feuilles du groupe de la mandarine clémentine. Utilisation Taxonomique». *Fruits*, vol 38, n.º 2, pp. 125-131.
- DIAZ-LLANOS y SAINZ-CALLEJA, Fco., GARCIA MOUTON, M. E. [1989]: «Efecto y utilidad de la métrica en el análisis de encuestas». *Revista de Estadística Española*, núm 121, pp. 253-280.
- DIDAY, E. [1972]: «Optimisation en Classification Automatique et Reconnaissances de Formes». *Revue Française d'Automatique Informatique Recherche Opérationnelle*. RAIRO. 6 ème année-Novembre. Vol. 3, pp. 61-96. DUNOD. Paris.
- DIDAY, E. [1974]: Optimisation in non hierarchical clustering. *Pattern Recognition*, Vol. 6, pp. 17-33- PERGAMON PRESS.
- ESCOFIER, B. [1987]: «Traitement des questionnaires avec non réponse, analyse de correspondances avec marge modifiée et analyse multicanonique avec contrainte. *Pub. Inst. Stat. Univ.XXXII*, fasc 3, pp. 33-69.
- ESCOFIER, B. [1990]: «Traitement des variables incompletes en Analyse des Correspondances Multiples». *La revue Modulad n.º 5*, pp. 1-11.
- ESCOFIER, B. [1990]: «Analyse des Correspondances Multiples Conditionelle». *La revue Modulad n.º 5*, pp. 13-27.
- ESCOFIER, B., PAGÈS, J. [1988]: *Analyses Factorielles Multiples*. Dunod.
- ESCOUFIER, Y. [1979]: *Cours d'Analyse des Données*. RT 7901. CRIG. Montpellier.
- ESCOUFIER, Y. [1981]: «L'Analyse des Tableaux de Contingence Simples et Multiples». *Comunicación al Congreso de Roma de Junio*.
- ESCOUFIER, Y. [1982]: «L'Analyse des Correspondances Simples et Multiples». *Metron*, 1-2, p. 53-78.
- ESCOUFIER, Y. [1985]: *L'Analyse des Correspondances, ses Propriétés, ses Extensions*. *Bull. of the Int. Statist. Inst*, 4, 28-2.
- ESCOUFIER, Y. [1988]: *Beyond Correspondence Analysis. In: Classification and Related Methodes et data Analysis*, H. H. Bock, Ed., North Holland, p. 505-514.
- FACY, F., LECHEVALLIER, Y. [1992]: «Place de l'Analyse Discriminante dans le traitement d'enquêtes psycho-sociales». *La revue de Modulad, n.º 9*, pp. 11-29.
- GARCIA MOUTON, M. E. [1988]: *Nuevos métodos evaluativos multidimensionales para el tratamiento de las encuestas en el sector agrario*. Tesis doctoral presentada en la Universidad Politécnica de Madrid.

- Escuela Técnica Superior de Ingenieros Agrónomos. Departamento de Matemática Aplicada. Director de la tesis. Dr.D. Fco Javier DIAZ-LLANOS y SAINZ-CALLEJA. Investigador del INIAA.
- GERARDIN, B., MOLLIERE, J. L. [1984]: «Macro for determining the optimal clusters in large data sets». SEUGI'84 - Proceedings - AMSTERDAM. April 4-6.
- GOVAERT, G. [1977]: «Algorithme de Classification d'un tableau de contingence. Premières journées internationales analyse de Données et Informatique. INRIA, Versailles.
- GOVAERT, G. [1983]: Classification Croisée. Thèse d'Etat. Paris VI.
- GOVAERT, G. [1989]: «La Classification Croisée». La revue Modulad n.° 4, pp. 9-36.
- GOVAERT, G. [1990]: «Classification binaire et modèles». Revue de Statistique Appliquée, 38,1, 67-82.
- GRANGE, D., LEBART, L. [1993]: Traitement Statistique des enquêtes. Préface de J.ANTOINE. Dunod. Paris.
- INRIA [1985]: Catalogue des prototypes de recherches au logiciel pour l'Analyse des Données.
- JAMBU, M. [1989]: Exploration Informatique et Statistique des Données. Dunod Informatique.
- JAMBU, M. [1997]: Initiation à l'Analyse des Données. à paraître. Dunod.
- JAMBU, M., LEBEAUX, M. O. [1983]: Cluster analysis and data analysis. Amsterdam. North-Holland.898 p.
- JUAN, J. [1982]: Le programme HIVOR de classification ascendente hiérarchique selon les voisins réciproques et le critère de la variance. Cahiers de l'Analyse des Données, Vol, 7,2, pp. 173-184.
- LEBART, L. [1975]: L'orientation du dépouillement de certaines enquêtes par l'Analyse de Correspondances Multiples. Con-sommation n.° 2, p. 73-96. Dunod.
- LEBART, L. [1987]: Analyse de Correspondances, Classification et traitement des données d'enquêtes. Formation et réflexion autour du logiciel SPAD. Blanes, 6-7-8 Mai.
- LEBART, L. [1990]: «Stratégies du traitement des données d'enquêtes». La revue Modulad n.° 3, pp.21-30.
- LEBART, L., MORINEAU, A. [1986]: «Specific clustering algorithms for large data sets and implementation in SPAD software». In: Classification as a tool of research, Gaul. W., Schader M., Eds., North-Holland, Amsterdam, p. 321-330.
- LEBART, L., MOURINEAU, A., PIRON, M. [1995]: Statistique Exploratoire Multidimensionnelle. Dunod.
- MODULAD [1987]: Bibliothèque Fortran 77 pour l'Analyse des Données. Version 2.1. Edité par l'Institut National de REcherche en Informatique et en Automatique. Domaine de Voluceau-Rocquencourt. BP. 105. 78153 Le Chesnay Cedex France.
- MOLLIERE, J. L. [1986]: «What's the real number of clusters?». In: Classification as a tool of research, Gaul, W., Schader, M. (eds), North-Holland, Amsterdam, p. 311-320.
- MOLLIERE, J. L. [1989]: «Stratégies de Classification pour de grand ensembles de Données». La revue de modulad n.° 3, pp. 31-69.
- MORINEAU, A. [1985]: «Note à propos de formules de l'Analyse Linéaire Discriminante». Bull. Tech. CESIA, 3, 35-42.
- MOSCAROLA, J. [1990]: Enquêtes et Analyse de Données. Vuibert Gestion.
- OK-SAKUN, Y. [1987]: Le système SICLA. Manuel de l'utilisateur. INRIA. Rapports Techniques, n.° 92.
- PERRUCHET, C. [1983]: «Une analyse bibliographique des épreuves de classifiabilité

- lité en Analyse des Données». Revue de Statistique et Analyse des Données. Vol. 8, n.º 2, pp. 18-41.
- RACKHAM, N. [1987]: Making Major Sales. Ed. Gower Publishing Company Limited, Gower House, Croft Road, Aldershort, Hampshire GU 11 3hr, England.
- ROUX, M. [1989]: «Construction ascendante hierarchique rapide: le programme CA-HVR. La revue de Modulad, n.º 3, pp. 1-6.
- SAS Institute [1982]: SAS. User's Guide: Statistics. Edition. SAS Institute, Cary, N.C.
- SICLA [1989]: Introduction à SICLA. INRIA.
- SICLA [1989]: Manuel de mise en oeuvre et exemples. INRIA.
- SICLA [1989]: Les commandes de SICLA. INRIA.
- SPAD (versión 3.0 windows) [1996]: Système Portable pour l'Analyse des Données. CISIA.
- STATlab (versión 3.0 windows) [1996]: CNET-France Telecom-SLP.
- SLP Infoware [1996]: STATlab by SLP. The software for exploratory Data Analysis. User' guide.
- SLP Infoware [1996]: STATlab by SLP. Getting Started. Traducido al castellano por Fco Javier DIAZ-LLANOS y SAINZ-CALLEJA.
- VILLOING, P. [1983]: «Méthodes de classification automatique non hiérarchique. CEAR/CCSA. Note d'Etude. L. A. n.º 56.
- WONG, M. A. [1982]: «A hybrid clustering method for identifying high density cluster». J. of. Amer. Statist. Assoc., 77, p. 841-847.