

SOFTWARE DE ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES: UNA REVISIÓN COMPARATIVA

Rubén Ledesma

Dr. en Psicología.

CONICET / Universidad Nacional de Mar del Plata

Río Negro 3922, Mar del Plata (7600), Argentina

+54 223 471 6015

rdledesma@gmail.com

RESUMEN. El Análisis de Correspondencias Múltiples (ACM) resulta una técnica de análisis multivariado de gran utilidad en la investigación por encuestas, tanto por su potencial en términos exploratorios como por su adecuación para el tratamiento de variables categóricas. En este trabajo se realiza un análisis comparativo de diferentes implementaciones informáticas del ACM, incluyendo las versiones de SPSS, SPAD, Statistica, ViSta, LispStat, *R* y XLStat. Se toman como referencia ciertos criterios comparativos que pueden resultar de utilidad en la evaluación y selección de un software de ACM para su aplicación en el ámbito de la investigación por encuestas.

PALABRAS CLAVE. ACM, HOMALS, software estadístico, análisis comparativo

ABSTRACT. Multiple Correspondence Analysis (MCA) is a valuable multivariate tool in survey research. That because it is an exploratory powerful tool, as well as it is appropriate for categorical data analysis. In this paper we review some statistical software that incorporate MCA, such as SPSS, SPAD, Statistica, ViSta, LispStat, *R* and XLStat. The comparison is made by taking into account some dimensions that would be useful for evaluating and selecting MCA software in the area of survey research.

KEYWORDS. Multiple Correspondence Analysis software: a comparative review.

Recibido: 15 de octubre 2007

Revisado: 21 de marzo 2008

Aceptado: 4 de abril 2008

Introducción

El Análisis de Correspondencias Múltiples (ACM)

El ACM es una técnica de análisis factorial exploratorio para datos categóricos multivariados. En esencia, esta técnica busca describir, en un espacio de pocas dimensiones o factores, la estructura de asociaciones entre un grupo de variables categóricas, así como las similitudes y diferencias entre los individuos a los cuales esas variables se aplican. Una revisión extensiva de los fundamentos de esta técnica y otras relacionadas puede encontrarse en Greenacre (1984), Escofier y Pagès (1990) y Gifi (1990), entre otros.

Una aclaración válida es que, como han puesto en manifiesto Tenenhaus y Young (1985), el ACM ha sido ‘reinventado’ en varias ocasiones, por diversos autores, y bajo nombres o enfoques diferentes. No obstante, la tradición más conocida fue establecida por la denominada *Escuela Francesa* (Benzécri, 1973), en la cual se adopta propiamente la expresión ‘ACM’. Dentro de esta tradición, la presentación del ACM se realiza apelando a un enfoque geométrico y enfatizando el carácter representacional de la técnica, esto es, como medio para la construcción de ‘mapas’ o ‘planos’ factoriales donde se representan las correspondencias – asociaciones – entre variables cualitativas. El ACM desde esta perspectiva ha sido implementado extensivamente en el programa SpadN (Lambert, Lebart, Morineau y Pleuvret, 1996) y en su versión más moderna denominada SPAD (Coheris Spad, 2008).

Alternativamente, se han propuesto otros enfoques y formas de cómputo del ACM. Entre ellos, el denominado sistema *Gifi* (Gifi, 1990) - implementado en SPSS por el grupo Teoría de Datos de la Universidad de Leiden - resulta seguramente el más conocido. Este sistema es un conjunto de métodos multivariados desarrollados en base al algoritmo denominado ALS (*Alternating Least Squares*). Entre estos métodos el llamado ‘Análisis de Homogeneidad’ proporcionaría un modelo equivalente al ACM. La solución del Análisis de Homogeneidad por ALS se conoce como HOMALS (*‘Homogeneity Analysis by ALS’*) y puede encontrarse, entre otros programas, en el módulo *Categories* de SPSS (Neulman, Heiser y SPSS Inc., 2005) disponible a través de la opción *Reducción de datos / Escalamiento Óptimo* del menú de análisis del programa. Cabe señalar que, si bien las versiones de SPSS posteriores a 13.0 incluyen la opción ‘Análisis de Correspondencias Múltiples’, se trata básicamente de un cambio en la denominación; en esencia el programa realiza un *Análisis de Homogeneidad* por ALS, como lo explica la misma ayuda del programa. Posiblemente, este cambio responde al hecho de que la denominación ‘ACM’ es más conocida por los usuarios, puesto que en realidad ambos métodos son equivalentes.

En efecto, desde el punto de vista aplicado el resultado por ambas vías es esencialmente el mismo. Es decir, existen programas y denominaciones diferentes para alternativas que, en términos prácticos, conducen al investigador al mismo resultado. No obstante, también es verdad que la vía por medio de ALS presenta, en relación al enfoque tradicional, la ventaja de poder trabajar directamente con

matrices de datos con valores faltantes. Como desventaja, es un proceso de cálculo iterativo que puede consumir más tiempo. Al contrario, el ACM desde el enfoque convencional se presenta como una extensión del Análisis de Correspondencias Simples (ACS) de tablas de contingencia, siendo esencialmente una aplicación del mismo algoritmo al caso de una matriz multivariada de datos categóricos codificados como una tabla de Burt (Greenacre, 1984). Este algoritmo proporciona una solución directa, no iterativa. Además, tiene la ventaja de que si se posee un programa de ACS, puede obtenerse un ACM de forma relativamente sencilla, convirtiendo la matriz de datos en una tabla de variables indicadoras.

En todo caso, a nivel conceptual lo importante aquí es presentar el ACM como una técnica exploratoria multivariada para datos categóricos, que es lo que la convierte en una herramienta de utilidad en la investigación por encuestas. Primero, por el carácter muchas veces exploratorio de los estudios que utilizan metodología de encuesta en diversos contextos aplicados, por ejemplo, en la investigación de mercados. Segundo, por la posibilidad de abordar problemas multivariados que implican variables categóricas. Este tipo de datos es bastante habitual como resultado de la aplicación de técnicas de encuestas. Por otro lado, sabemos que el ACM en este contexto puede ser una técnica útil en combinación o como paso previo a la aplicación de otros análisis. Por ejemplo, puede utilizarse como medio para *cuantificar* variables categóricas que luego pueden ser utilizadas como *input* para nuevos análisis. Por caso, podemos mencionar la estrategia ACM seguido de *Análisis de Clusters*. Esta metodología es muy frecuente en la construcción de tipologías empíricas socio-demográficas o sustantivas basadas en variables categóricas - y es incluida como una opción básica en programas como SPAD -. En síntesis, podemos asegurar que el ACM se presenta como una técnica multivariada con múltiples aplicaciones en el ámbito de la metodología de encuestas.

Software estadístico y potenciales usuarios del ACM

La aplicación del ACM en la práctica, como acaso sucede con otras técnicas multivariadas, resulta casi imposible de pensar sin el apoyo de un software apropiado. De hecho, existe mucho y variado software para este propósito. No obstante, las implementaciones informáticas del ACM pueden variar considerablemente de un programa a otro, tanto en lo que respecta a la accesibilidad, como al tipo de interfaz de usuario, flexibilidad y opciones de análisis. Existen, por ejemplo, versiones muy conocidas disponibles en los programas comerciales más populares. También encontramos otras versiones eficientes y flexibles disponibles de forma gratuita, aunque en este caso se trata de opciones menos difundidas y utilizadas.

Algunas versiones informáticas del ACM pueden encontrarse en SPSS (Neulman, Heiser y SPSS Inc., 2005), Statistica (StatSoft Inc., 2004), SPAD (Coheris Spad, 2008), XLSTAT (Kovach Computing Services, 2006), ViSta (Young, 2005) y Ade4 (Thioulouse, Chessel, Dolédec y Olivier, 1997). Asimismo, se dispone de diferentes rutinas de ACM y HOMALS escritas en lenguaje LispStat (Tierney, 1990) y R (R Development Core Team, 2006). En LispStat encontramos

las versiones publicadas por Bond y Michailides (1996, 1997) y en *R* cabe una mención especial al reciente paquete FactoMineR (Husson, Josse, Le y Mazet, 2008), que integra varias opciones desarrolladas desde la óptica de la Escuela Francesa. En este lenguaje también tenemos la versión de HOMALS creada por deLeeuw (2005).

En este trabajo se analiza el software disponible desde diferentes criterios comparativos, cubriendo desde la etapa inicial de acceso al programa, hasta la última etapa de interpretación de resultados. Los programas no se describen exhaustivamente por separado, sino que se van mencionando en sus aspectos distintivos en función del desarrollo de los criterios comparativos. Por otro lado, en la revisión y comparación de los programas tendremos en mente dos tipos potenciales o ideales de usuario: uno con un *perfil aplicado* y otro con un *perfil metodológico*. El primero, manejaría datos y problemas reales, radicando su interés no tanto en el método de ACM en sí mismo como en su potencial para responder preguntas de investigación. Este usuario, no obstante, conoce los fundamentos conceptuales del método y es capaz de aplicar e interpretar de forma idónea el resultado de un ACM. Por otro lado, aunque menor medida, consideraremos la posibilidad de un usuario más avanzado, especialista en metodología y capaz de interactuar y experimentar con códigos y rutinas de programas. Este usuario, estaría más interesado por el aspecto informático, computacional o metodológico de la técnica, que por su aplicación concreta a datos reales. Esta distinción, creemos, es importante a la hora de valorar adecuadamente el software en sus diferentes dimensiones y de acuerdo a las posibles demandas del usuario.

En cuanto a las dimensiones de la comparación, analizaremos las siguientes cuestiones: a) *Accesibilidad*, b) *Interfaz de usuario*, c) *Capacidades de gestión, edición y manipulación de datos*, d) *Opciones de análisis*, y e) *Capacidades gráficas* de los programas. Esperamos que el trabajo contribuya a una valoración apropiada del software y proporcione a los usuarios elementos para una elección acorde a sus propias necesidades.

Análisis comparativo del software de ACM

Accesibilidad: programas comerciales y de libre distribución

En primer lugar, existe una cuestión básica en la valoración y elección del software que tiene que ver con el grado de accesibilidad del producto, en términos de que su política de distribución sea comercial o gratuita. En un trabajo previo hemos discutido como este aspecto puede afectar la accesibilidad del software estadístico en contextos de trabajo donde no se dispone de recursos económicos para costear la adquisición de software comercial (Ledesma, 2004). Por ello, una primera distinción dentro del software de ACM debe establecerse entre los programas comerciales y gratuitos.

Entre los programas comerciales más conocidos que incorporan ACM podemos mencionar SPSS (con su módulo '*Categories*'), Statistica, SPAD y XLSTAT – el 'add-in' estadístico de Microsoft Excel -. Se trata, en todos los casos,

de productos que requieren la adquisición de licencias de uso, aunque su costo puede variar considerablemente de un programa a otro. Por ejemplo, XLSTAT es un programa muy accesible, al menos si lo comparamos con SPAD, que es más completo pero también mucho más costoso. Por otro lado, entre los programas de ACM a los que puede accederse de forma gratuita, encontramos las versiones de este método disponibles en *R* o LispStat, así también en programas como Ade4 y ViSta.

Puede que los programas comerciales sean, en ciertos aspectos, mejores para un usuario estándar con intereses aplicados, pero también vale destacar que algunas de las implementaciones más sofisticadas disponibles en la actualidad se encuentran en LispStat y *R*, aunque quizás de una manera no tan amigable para el usuario, como comentaremos más adelante. Por otro lado, también debemos mencionar que en la categoría del software gratuito, los programas escritos en *R*, LispStat y ViSta proporcionan versiones del ACM con acceso abierto al código del programa, lo que puede resultar atractivo para los usuarios más avanzados. En todo caso, en este apartado corresponde destacar que, cuando no se dispone de recursos para adquirir software comercial, la alternativa de las implementaciones informáticas de libre acceso constituye una posibilidad interesante para el usuario, no solo si se quiere realizar una aplicación, sino también si se desea experimentar con los códigos.

Interfaz de usuario

La naturaleza de la interfaz de usuario puede ser determinante en la elección de un software de ACM, dependiendo su importancia de la pericia y formación del usuario. En general, los usuarios que buscan realizar aplicaciones reales se sentirán más cómodos si el programa tiene una interfaz completamente gráfica, que les permita realizar la aplicación de forma automática mediante menús, cuadros de diálogo, etc. Ciertamente, los programas más populares y extendidos, como SPSS y SPAD proporcionan este tipo de ambientes de trabajo, ofreciendo al investigador una IGU simple e intuitiva para interactuar con el programa, así como sistemas de ayuda y documentación completa.

Por otro lado, los usuarios más avanzados o especializados pueden requerir implementaciones más flexibles y abiertas. En este caso, lo más apropiado sería acudir a las rutinas escritas en lenguajes como LipStat o *R*, que si bien no tienen una interfaz muy amigable, permiten que el usuario experto pueda adaptar los programas a sus necesidades o realizar nuevos aportes y desarrollos a los programas. En ambos lenguajes, LipStat y *R*, existen versiones del ACM clásico y de HOMALS.

También es importante comentar que se están realizando esfuerzos importantes para crear interfaces de usuarios más interactivas para *R* y facilitar así el acceso a ciertos grupos de usuarios. La posibilidad de correr paquetes de *R* a través de la web es otra posibilidad interesante. Por ejemplo, cabe mencionar que Ade4 ha sido completamente reescrito en *R* y puede utilizarse on-line a través del sistema Rweb (Chessel, Dufour & Thioulouse, 2004). Otro esfuerzo importante por simplificar y hacer más amigable la interacción con *R* ha sido el llamado *R Commander* (Fox, 2005). Usando este desarrollo, puede ejecutarse el paquete

FactoMineR mediante una IGU más simple. En este sentido, también podemos decir que *ViSta* representa un caso interesante, ya que ofrece una IGU simple e intuitiva para interactuar con programas que en realidad están escritos en LispStat.

Por último, otra cuestión básica vinculada a la simplicidad de la interfaz de usuario es el idioma. En tal sentido, cabe señalar que en castellano solo encontramos versiones de programas comerciales como SPSS y XLSTAT. Esto, claro está, puede ser un punto a favor de estos programas, pues se simplifica así la interacción con el programa y, eventualmente, la comprensión de la documentación de usuario.

Capacidades de gestión, edición y manipulación de de datos

En general, el análisis de datos en el ámbito de los estudios por encuesta requiere el uso de entornos informáticos que permitan *portar* (importar/exportar), *editar*, *manipular* y *transformar* y *analizar* los datos haciendo uso de diferentes modelos estadísticos. En la práctica, es realmente raro que los datos puedan, sin más, estar preparados para un ACM; pero aún si lo están, suelen ser necesarias varias aplicaciones del análisis para arribar a una solución clara o satisfactoria. Es decir, generalmente el ACM se desarrolla en el marco de un proceso cíclico en donde se ‘itera’ entre los datos de entrada y los resultados. Este proceso implica la depuración progresiva de datos, la creación, re-codificación o eliminación de categorías, etc. Por esta razón, es necesario evaluar las capacidades de gestión, edición y transformación de datos de que dispone un programa de ACM. En este sentido, pueden encontrarse diferencias importantes entre los programas, lo que los hace más o menos aptos para realizar aplicaciones reales en el ámbito de la investigación por encuestas.

En primer lugar, XLSTAT es posiblemente la opción más sofisticada en este sentido, si consideramos que como *add-in* de Excel el usuario tiene a su disposición todas las capacidades añadidas de este entorno para gestionar, editar y transformar sus datos durante la aplicación del ACM. Desde luego, muchas de estas capacidades también pueden encontrarse en los sistemas estadísticos completos como SPSS, Statistica y SPAD. A su vez SPAD, es interesante por sus capacidades generales, pero también por sus opciones integradas al ACM, por ejemplo, la capacidad de tratar con categorías con frecuencias bajas en la misma ejecución del análisis, filtrar o ponderar casos, etc. Asimismo, las últimas versiones de SPSS también han mejorado y ampliado las opciones específicas del ACM en este aspecto. Claro está, estos programas son de distribución comercial, como ya se ha mencionado.

Por otro lado, debe reconocerse que las versiones en LispStat y *R* pueden ser menos apropiadas o cómodas en cuanto a gestión y edición de datos; incluso pueden requerir mecanismos o formatos de entrada de datos poco habituales para el usuario. Una opción sería utilizar estos programas en combinación con una planilla de cálculo o base de datos, pero lo cierto es que esto podría restar bastante agilidad al proceso de análisis. Nuevamente, una situación intermedia entre los sistemas comerciales completos, como SPSS o SPAD, y los programas escritos en LispStat o *R*, podemos encontrar en *ViSta*. Si bien en este aspecto *ViSta* no es equiparable a los paquetes comerciales completos, ofrece varias herramientas de gestión, edición y

transformación de datos. Este programa permite importar datos en formato texto, posee una grilla o planilla de datos convencional y dispone de una buena gama de herramientas para manipular y realizar transformaciones de datos.

Opciones de análisis

Otro aspecto que distingue a los diferentes programas es su capacidad de análisis en el caso concreto del ACM. Por ejemplo, la posibilidad de incorporar variables o sujetos ilustrativos o suplementarios al análisis, algo que suele resultar bastante útil en el contexto de la investigación por encuestas. Algunos programas admiten esta posibilidad (por ejemplo, el ACM de *R* y *LispStat*, *SPAD*, *Ade4*, y las últimas versiones de *SPSS*), en tanto que otros carecen de dicha opción (por ejemplo, *ViSta*). No obstante, hay que mencionar que tal vez la diferencia más importante está en el tipo de enfoque, pues los programas que incluyen el enfoque francés ofrecen una salida de resultados más detallada y sofisticada. En este aspecto, *SPAD* es seguramente una de las opciones más completas, tanto por sus capacidades de análisis como por su salida de resultados (incluye medidas de ajuste de la solución y varios índices para interpretar el resultado - coordenadas, contribuciones, cosenos cuadrados, test de significación, etc. -). *XLSTAT* también ofrece una salida con índices de este tipo. Por su lado, las salidas de *HOMALS* (p.e. en *SPSS* y *ViSta*) suelen ser más limitadas y menos informativas si se las compara con las salidas del ACM en programas como *SPAD*. Obviamente, la valoración de la utilidad comparativa de las salidas e índices dependerá en buena medida del enfoque con el cual el usuario esté más familiarizado (*ACM* o *HOMALS*). Aquí, cabe nuevamente una aclaración para las últimas versiones de *SPSS*, ya que pueden inducir a cierta confusión. Si bien el nombre del procedimiento es “*ACM*”, el usuario encontrará que el método realmente es *HOMALS*, como ya lo hemos mencionado. En esencia, se trata del mismo módulo de Escalamiento Óptimo, con algunas mejoras y añadidos. No obstante, en algunos aspectos la salida utiliza denominaciones propias de *HOMALS* y en otros refiere a indicadores más asociados a un *ACM*. Así, el usuario posiblemente se encuentre en la necesidad de documentarse sobre la equivalencia de denominaciones al momento de comprender la salida de resultados.

En otros aspectos no hay grandes diferencias entre los programas, salvo por algunas implementaciones más sofisticadas existentes en *LispStat* o *R*. Por ejemplo, el módulo de *ACM* de *LispStat* incluye métodos *Bootstrap* para evaluar la estabilidad de la solución. Asimismo, *R* implementa algunas opciones avanzadas adicionales tanto en el paquete *HOMALS* como en el *ACM* tradicional. Aquí, lo importante es que el usuario evalúe, según sus necesidades, si estos beneficios compensan las dificultades que puede encontrarse al momento de interactuar con la interfaz del software, así como las mencionadas limitaciones de estos programas en lo referente a gestión de datos.

Por último, otro elemento a valorar del software es el tipo de algoritmo que utiliza para obtener la solución. Debe recordarse que *HOMALS* utiliza un procedimiento iterativo en todas las implementaciones informáticas (*SPSS*, *LispStat*, *ViSta* y *R*). Esto podría afectar la eficiencia del análisis en lo referente al tiempo

requerido, sobre todo si se trabaja con datos de gran dimensionalidad o volumen. Asimismo, HOMALS posee una característica distintiva en cuanto al tratamiento de datos faltantes. Estos pueden tratarse como en el enfoque clásico – borrado *listwise* de los individuos o codificación de los valores faltantes como categoría activa o ilustrativa en el análisis – pero también pueden ser directamente incluidos utilizando la información parcial que proporcionan. Así, los usuarios de SPSS pueden comprobar como HOMALS estima los puntajes factoriales o coordenadas de los individuos incluso para aquellos que poseen algún valor faltante.

Capacidades gráficas

Una vez ejecutado el análisis, es importante considerar las capacidades del software en lo que respecta a la salida gráfica de resultados. Aquí, un aspecto clave es la calidad y versatilidad de las salidas en formato gráfico. Recordemos que el ACM, como técnica exploratoria de reducción de la dimensionalidad, ofrece una salida numérica que puede ser bastante extensa y difícil de interpretar de forma directa. Por ello, la interpretación y comprensión del sentido de los resultados se basa, principalmente, en la inspección de salidas gráficas; generalmente, diagramas de dispersión con las coordenadas o puntajes factoriales de las categorías de las variables y/o los individuos. La exploración visual de este tipo de diagramas, llamados ‘planos’ o ‘mapas’, es una tarea fundamental en la aplicación del ACM.

Pues bien, los programas varían considerablemente en cuanto al tipo de salidas gráficas, desde los gráficos en formato texto, pasando por los gráficos convencionales de SPSS, SPAD o Statistica, hasta los gráficos dinámicos e interactivos de LispStat y ViSta. Este último tipo de gráficos constituyen el nivel más alto en materia de visualización de resultados estadísticos (Wainer y Velleman, 2001). Los gráficos dinámicos son incorporados en la mayoría de las rutinas de LispStat y, especialmente, en algunos programas escritos en este lenguaje como ViSta o Arc (Cook y Weisberg, 1999). De hecho, LispStat sigue siendo la plataforma más apropiada para el desarrollo de este tipo de gráficos, superando en este aspecto a *R* (Valero y Udina, 2005).

La *Figura 1* muestra un ejemplo de visualización de resultados de un ACM en ViSta, utilizando gráficos dinámicos e interactivos diseñados específicamente para este análisis. Este tipo de visualización posee algunas características de interés: a) permite visualizar varios aspectos del resultado en una misma pantalla (coordenadas, medidas de discriminación de las variables, etc.); b) permite interactuar de forma directa con los gráficos a diferente nivel (selección de las observaciones, ‘zoom’, cambio de planos factoriales, etc.), y c) los gráficos están ‘vinculados’ entre sí, de modo que al interactuar con un gráfico se producen cambios en el resto (por ejemplo, la matriz de diagramas de dispersión – arriba a la derecha – puede usarse como panel de selección de los ejes en el diagrama 3D ubicado en el centro de la representación). Estas capacidades gráficas pueden ser muy valiosas a la hora de explorar, interpretar y comprender el sentido de la solución del ACM. Lamentablemente, los programas de ACM más utilizados no incorporan gráficos dinámicos de esta naturaleza, siendo la prestación en este sentido bastante limitada.

Por supuesto, esta limitación no está reservada al caso del ACM. Si comparamos, por ejemplo, SPSS con programas como ViSta o DataDesk (DataDescription Inc., 2005) encontraremos una limitación importante en sus capacidades gráficas, al menos si se requieren técnicas multivariadas dinámicas.

No obstante, también hay que reconocer que si consideramos el caso particular del ACM, algunos programas tienen una excelente prestación gráfica, aunque no se implementen tecnologías de gráficos dinámicos en toda su extensión. Este es el caso de SPAD, que integra opciones específicas interesantes para las salidas gráficas del ACM y proporciona un editor muy completo para crear buenos gráficos de presentación (ver Figura 2). SPAD permite, por ejemplo, que el usuario defina muchas de las preferencias de la representación gráfica en los planos factoriales, como los símbolos y el color para el formato de visualización de las variables e individuos (activos e ilustrativos). XLSTAT también ofrece opciones de ese tipo para definir y editar los planos factoriales del ACM.

Figura 1.

Visualización de resultados en el programa (Ledesma, Valero-Mora y Young, 2002)

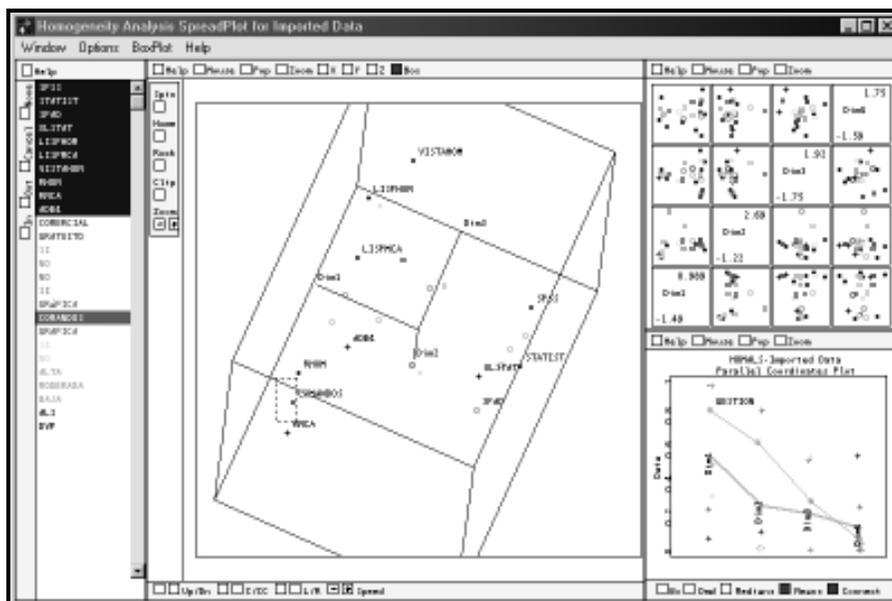


Figura 2.

Ejemplo de salida gráfica de SPAD (le Roux, Börjesson y Bonnet, 2006)

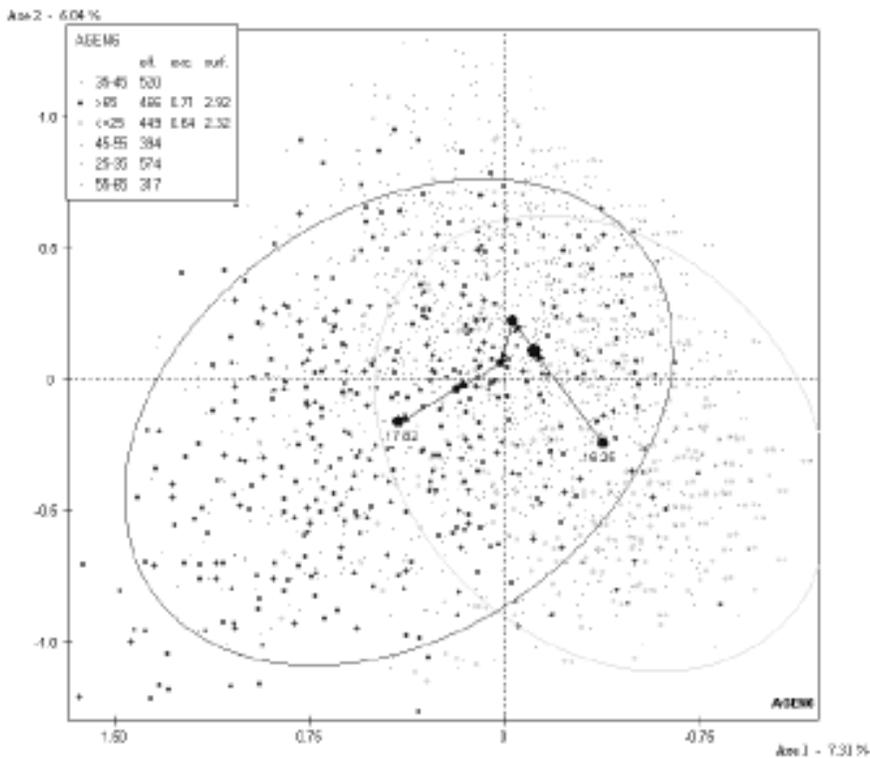


Tabla 1.a
Tabla comparativa de algunos programas de ACM

	<i>Accesibilidad</i>	<i>Capacidades de gestión de datos</i>	<i>Algoritmo / enfoque</i>
SPSS	Distribución comercial. El módulo <i>Categorías</i> - que incorpora HOMALS - no se incluye en el paquete base. Disponible en castellano.	Muy buenas capacidades para gestionar, editar y manipular datos.	HOMALS
SPAD	Distribución comercial. Su última versión es costosa. No está disponible en castellano.	Muy buenas capacidades para gestionar, editar y manipular datos	ACM
STATISTICA	Distribución comercial. El ACM está integrado a la versión Avanzada, no a la versión Base del producto. No disponible en castellano.	Muy buenas capacidades para gestionar, editar y manipular datos	ACM
ViSta	Gratuito y abierto. HOMALS se instala fácilmente como plug-in. Disponible solo en inglés.	Buenas capacidades, aunque no se equiparan a otros programas – como SPSS -	HOMALS
R	Existen varios paquetes que se obtienen de forma gratuita y que se corren en R. Disponibles en inglés.	Capacidades limitadas si quieren realizar operaciones de gestión y edición de datos mediante IGU	Existen versiones de HOMALS y del ACM
LispStat	Existen rutinas gratuitas y de código abierto desarrolladas en este lenguaje. Disponibles en inglés.	Capacidades limitadas si quieren realizar operaciones de gestión y edición de datos mediante IGU	Existen versiones de HOMALS y del ACM
XLSTAT	Distribución comercial, relativamente accesible. Disponible en castellano.	Muy Buenas capacidades, con las ventajas añadidas de Excel.	ACM

Tabla 1.b.
Tabla comparativa de algunos programas de ACM

	<i>Interfaz de usuario</i>	<i>Salida de resultados</i>
<i>SPSS</i>	IGU simple e intuitiva, con buena documentación de usuario.	Salida de resultados completa desde el enfoque HOMALS. Salida gráfica poco interactiva y sin opciones 'a medida' para el modelo en cuestión. Las versiones posteriores a 13.0 introducen importantes mejoras.
<i>SPAD</i>	IGU simple e intuitiva. Bien documentado.	Salida de resultados muy completa desde el enfoque francés (ACM). Buenos gráficos de presentación y opciones específicas para el ACM.
<i>STATISTICA</i>	IGU simple e intuitiva, con buena documentación de usuario.	Salida de resultados completa desde el enfoque francés (ACM). Buenos gráficos de presentación.
<i>ViSta</i>	IGU simple e innovadora. La documentación de usuario es algo limitada.	Salida de resultados completa desde el enfoque HOMALS. No admite variables o sujetos ilustrativos. Destaca por sus gráficos dinámicos a medida para el modelo.
<i>R</i>	Interacción por comandos en línea. Aunque existen desarrollos actuales de interfaz gráfica. Bien documentado.	Existen varias versiones sofisticadas y flexibles, que proporcionan salidas completas para ambos enfoques (HOMALS y ACM). Comparativamente, las implementaciones son limitadas en el aspecto gráfico.
<i>LispStat</i>	Son programas que pueden ejecutarse por comandos o mediante IGU. Están bien documentados.	Tanto HOMALS como el ACM son implementaciones sofisticadas y flexibles. Proporcionan salidas completas y algunas opciones no incluidas en otros programas (p.e. bootstrap). Ofrece salidas específicas con gráficos dinámicos.
<i>XLSTAT</i>	IGU simple e intuitiva. Documentación de usuario en castellano.	La salida de resultados es muy completa. Proporciona buenos gráficos de presentación y opciones específicas para el ACM.

Discusión

En primer lugar, hemos mencionado las implementaciones del ACM en algunos de los programas comerciales más conocidos, como SPSS, Statistica y SPAD. Se trata en estos casos de sistemas que poseen interfaces de usuario completamente gráficas, resultan simples de utilizar y proporcionan al investigador aplicado un entorno completo para gestionar, transformar y analizar sus datos desde diferentes modelos. Estas son, sin dudas, ventajas importantes para el usuario que busca realizar aplicaciones concretas en el ámbito de la investigación por encuestas. Entre estos programas comerciales completos o de propósitos múltiples destaca, sin lugar a dudas, SPAD, que ha sido concebido desde su origen como software de aplicación de métodos factoriales basados en la tradición francesa. En esta categoría de programas, cabe también una aclaración en relación a las últimas versiones (posteriores a 13). En ellas, se han ampliado y mejorado las opciones de análisis y la salida de resultados del ACM. No obstante, se advierte un cambio en la denominación que puede resultar algo confusa. Si bien el nombre del método es ahora 'ACM', el método que se utiliza es un Análisis de Homogeneidad por ALS. De hecho, la técnica se encuentra dentro del módulo general de Escalamiento Óptimo (Optimal Scaling), dentro del cual HOMALS es un caso particular que se aplica cuando todas las variables son definidas como categóricas. Salvo esta dificultad que puede inducir a cierta confusión parte del usuario, el módulo actual resulta algo más flexible y sofisticado si se usa la IGU del programa.

Por otro lado, hemos señalado que los programas anteriores pueden resultar difíciles de adquirir, ya que son de distribución comercial y sus costos asociados restringen el acceso a quienes no disponen de recursos económicos suficientes (esto se aplica especialmente a la última versión del programa SPAD, que resulta bastante costosa). Además, si tenemos en mente a los usuarios más sofisticados y con intereses metodológicos, hemos indicado que este tipo de programas puede resultar poco flexible y carente de interés desde el punto de vista experimental. Por último, las prestaciones gráficas en algunos casos resultan algo limitadas, incluso si se las compara con otros programas de ACM que se distribuyen de forma gratuita. Este es el caso, por ejemplo, si analizamos la implementación de HOMALS en SPSS en relación al mismo método en ViSta (Ledesma, Valero-Mora y Young, 2002).

Por su lado, las rutinas escritas en LispStat y *R* poseen ventajas en cuanto a su accesibilidad - libre distribución -, flexibilidad y capacidades de análisis. Estas características pueden ser atractivas para los usuarios avanzados o para quienes están interesados en experimentar o interactuar de modo más razonado con el ACM. No obstante, las interfaces de usuario, aún cuando en algunos casos son gráficas, resultan poco amigables o familiares para el usuario estándar. Además, este tipo de usuarios encontrará una limitación importante en materia de edición y manipulación de datos, al menos si se desean aplicar operaciones de esta naturaleza a través de una IGU. En el caso de LispStat, no obstante, se encontrará una importante ventaja en lo referente a las posibilidades en materia de gráficos dinámicos, lo cual puede resultar clave en la aplicación de un ACM. En esta categoría de software, también deben mencionarse algunos esfuerzos de desarrollo de interfaces más amigables, sobre

todo para *R*. Asimismo, el programa ViSta proporciona una interfaz gráfica ‘convencional’ para interactuar con el lenguaje LispStat. En todo caso, es de esperar que en el futuro se incrementen los esfuerzos destinados a desarrollar interfaces gráficas que faciliten el uso de rutinas escritas en estos poderosos lenguajes.

Por último, merecen una referencia especial los programas XLSTAT y ViSta. XLSTAT porque funciona como *add-in* estadístico de Excel, lo que brinda al programa la ventaja añadida de las herramientas de gestión propias de este entorno. Además, es el producto más económico entre los programas de distribución comercial. Por su lado, ViSta destaca por sus técnicas gráficas dinámicas, incluyendo representaciones y procedimientos específicos para el caso del ACM no disponibles en el resto de los programas. Como ventaja complementaria, se trata de un sistema gratuito que puede obtenerse on-line de forma libre.

En cualquier caso, las necesidades del usuario resultarán clave en la elección final de un programa de ACM. Por su lado, el investigador aplicado priorizará la simplicidad en el uso del software y valorará otras capacidades accesorias de los programas, como sus herramientas en materia de gestión de datos. Así, es razonable que opte por programas de ACM de alcance comercial, como SPAD o el módulo de SPSS, al menos si puede cubrir los costos asociados a la adquisición de licencias. Por otro lado, el especialista en metodología, más interesado quizás por la experimentación con los métodos en sí mismos que por sus posibles aplicaciones, debe valorar la flexibilidad de los lenguajes de programación y otros aspectos computacionales. En tal sentido, encontrará una mejor opción en las rutinas escritas en *R* o LispStat.

En síntesis, la elección de un software apropiado de ACM requerirá por parte del usuario una salida de compromiso entre varias dimensiones, entre las que se cuentan la accesibilidad, la naturaleza de la interfaz de usuario, las capacidades de gestión de datos y, por supuesto, la flexibilidad y capacidad del análisis propiamente. En este sentido, esperamos que a partir del presente trabajo los usuarios del ámbito de la investigación por encuestas dispongan de más información para poder valorar y elegir un software estadístico de acuerdo a sus propios intereses y necesidades.

Referencias

- Benzécri, J.P (1973). *Analyse des Données* (2 vols.). París: Dunod.
- Bond, J. y Michailides, G. (1996). Homogeneity Analysis in LISP-STAT. *Journal of Statistical Software*, 1, 2, <http://www.jstatsoft.org/>
- Bond, J. y Michailides, G. (1997). Interactive Correspondence analysis in a dynamic object-oriented environment. *Journal of Statistical Software*, 2, 8, <http://www.jstatsoft.org/>
- Chessel, D. Dufour, A. y Thioulouse, J. (2004). The ade4 package - I : One-table methods. *Rnews*, 4,1.
- Coheris Spad (2007) SPAD [programa informático]. URL: <http://spadsoft.com/>
- Cook, R. D. y Weisberg, S. (1999). *Applied regression including computing and graphics*. New York: John Wiley & Sons.

- DataDescription Inc. (2005). *DataDesk*, v. 5 [programa informático]. Disponible: <http://www.datadesk.com>
- de Leeuw, J. (2005) The homals Package . Version 0.2.4. [software] [On-line] Disponible: URL : cran.r-project.org/doc/packages/homals.pdf
- Escofier , B. y Pagès, J. (1990). *Analyses Factorielles Simples et Multiples: Objectifs, Methodes et Interpretation*. París: Dunod, 2da ed.
- Fox, J. (2005) The R Commander: A Basic-Statistics Graphical User Interface to R. *Journal of Statistical Software*, 14, 9. <http://www.jstatsoft.org/>
- Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Chichester: Wiley.
- Greenacre, C. M. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.
- Husson, F., Josse, J., Le, S. y Mazet, J. (2008). *The FactoMineR Package*. Disponible: <http://factominer.free.fr>
- Kovach Computing Services (2006) *XLSTAT* [software] [On-line] Disponible: URL: <http://www.kovcomp.co.uk/XLStat/downl.html>
- Lambert T., Lebart L., Morineau A. y Pleuvret P. (1996). *Manuel de référence de SPAD*. CISIA-CERESTA, Saint-Mandé.
- le Roux, B., Börjesson, M., y Bonnet, P. (2006). *Performing multiple correspondence analysis (MCA) using SPAD (version 6.5)*. Disponible: <http://www.skeptron.uu.se/broad/seg/p-gda-0609-spadguide-mca.pdf>
- Ledesma, R. (2004). Sistemas estadísticos de propósitos múltiples: Una revisión de programas gratuitos. *Metodología de Encuestas*, 6, 2, 105-117
- Ledesma, R. Valero-Mora, P. y Young, F. W. (2002). Análisis de Homogeneidad en ViSta " The Visual Statistics System". *Metodología de las Ciencias del Comportamiento*, 4, 1, 139-149
- Nenadic, O. y Greenacre, M (2005). *Computation of Multiple Correspondence Analysis, with code in R*. Univeristad Pompeu Fabra, Informe de trabajo 887, Disponible: <http://www.econ.upf.edu/docs/papers/downloads/887.pdf>
- R Development Core Team (2006). *The R Project for Statistical Computing*, v. 2.2 [software] [On-line]. Disponible: <http://www.r-project.org/>
- Meulman, J. J.; Heiser, W. J. y SPSS Inc. (2005) *SPSS Categories 14.0*. Chicago: SPSS Inc.
- StatSoft Inc. (2004) *Electronic Statistics Textbook*. Tulsa, OK: StatSoft. [on-line] Disponible: URL: <http://www.statsoft.com/textbook/stathome.html>
- Tenenhaus, M. y Young, F. (1985). An analysis and synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis and other methods for quantifying categorical data. *Psychometrika*, 50, 91-119.
- Thioulouse J., Chessel D., Dolédec S., y Olivier J.M. (1997). ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing*, 7, 1, 75-83.
- Tierney, L. (1990). *Lisp-Stat - An Object-Oriented Environment for Statistical Computing and DynamicGraphics*. NY: John Wiley & Sons.
- Valero Mora, P. y Udina [Eds], (2005). The health of LispStat. *Journal of Statistical Software*, special issue, 3, 8.

- Wainer, H. y Velleman, P. (2001). Statistical graphics: mapping the pathways of science, *Annual Review of Psychology*, 52, 305–35.
- Young, F. (2005). *ViSta "The Visual Statistics System"* (version 6.4) [software] [Online], Disponible: URL: <http://forrest.psych.unc.edu/research/index.html>