

METODOLOGIA APROPIADA PARA EL ESTUDIO DE DATOS ECONOMICOS ESTRUCTURADOS EN DIVERSOS GRUPOS: EL ANALISIS FACTORIAL MULTIPLE

M. Isabel Landaluze Calvo
Amaya Zárraga Castro
Departamento de Economía Aplicada III
Facultad de CC.EE. y Empresariales. Universidad del País Vasco
{e-mail:il@alcib.bs.ehu.es}

Elena Abascal Fernández
Departamento de Estadística e Investigación Operativa
Universidad Pública de Navarra
{e-mail:eabascal@upna.es}

X Reunión Anual ASEPELT - ESPAÑA

Albacete , 20 - 21 de Junio de 1996

INTRODUCCION

Desde nuestra primera toma de contacto con el Análisis Factorial Múltiple (AFM), que se produjo casi de forma fortuita, perteneciendo en 1991 al equipo traductor del libro original en lengua francesa titulado: Análisis Factoriales Simples y Múltiples: objetivos, métodos e interpretación de B. Escofier y J. Pagès (1990), percibimos que se trataba de un método factorial con importantes innovaciones y aportaciones, tanto de aplicación como de interpretación, lo que nos impulsó a profundizar en él, Landaluze (1995).

El objetivo de este trabajo no es la presentación del método, sino dejar constancia de las buenas propiedades teóricas, así como de las ventajas que, desde el punto de vista de la interpretación, principalmente, presenta El Análisis Factorial Múltiple cuando los datos a analizar están estructurados en diferentes matrices. Estas pueden estar medidas en el mismo o en diferentes períodos de tiempo contener variables heterogéneas (tablas mixtas), con la restricción de que en cada matriz las variables sean de la misma naturaleza (cualitativa o cuantitativa) y hagan referencia al mismo colectivo de individuos. Se trata de un método con múltiples aplicaciones debido, no solo a la posibilidad que ofrece de analizar tablas mixtas, sino también a la diferente estructuración en grupos a que puede someterse un mismo conjunto de datos. Permite, por una parte, equilibrar la influencia de los diferentes grupos, sin modificar su estructura inicial, y, por otra parte, interrelacionar los diversos aspectos que los datos reflejan, pudiendo así extraer conclusiones desde diversas ópticas, lo que deriva en una riqueza interpretativa sin precedentes en el campo del análisis factorial.

El AFM es un método que combina elementos ya desarrollados en otros métodos de Análisis de Datos diseñados con anterioridad: yuxtaposición de tablas, ponderación de grupos de variables, medidas de

relación entre grupos de variables, tablas suplementarias, No obstante, se trata de un método diferente, con criterios propios.

PROPIEDADES Y VENTAJAS DEL AFM

En esta sección se presentan las propiedades tanto desde el punto de vista teórico, como desde el punto de vista de la interpretación del Análisis Factorial Múltiple, propiedades que hacen que sea un método apropiado para el estudio de datos que aparecen estructurados en diferentes grupos:

1) En el Análisis Factorial de Varias Tablas de Datos, en el que cada una de ellas define un grupo de variables, se pueden estudiar diferentes relaciones: las asociadas a los elementos que integran cada uno de los conjuntos definidos, individuos y variables, y las que se producen entre los diferentes grupos. Las preguntas que un investigador se puede plantear ante el análisis de un conjunto de datos de las características consideradas son las siguientes:

- -¿Qué relación existe entre los individuos cuando éstos están caracterizados por las variables de un solo grupo?
- -¿Qué relación existe entre los individuos caracterizados por el conjunto total de variables?
- -¿Qué relación existe entre las variables de un solo grupo?
- -¿Qué relación existe entre todas las variables que intervienen en el análisis?
- -¿Qué relación existe entre los individuos y las variables?
- -¿Qué relación existe entre los diferentes grupos de variables estudiados, considerados éstos como entes propios?

El Análisis Factorial Múltiple es un método que proporciona un análisis completo del conjunto de datos considerado, ya que permite dar una respuesta satisfactoria a cada una de las preguntas planteadas.

2) La posible estructuración en grupos de variables de los datos originales enriquece su estudio. En este caso los objetivos perseguidos no se limitan a la obtención de una tipología de los individuos definida a través del conjunto de variables, sino que se amplían a la búsqueda de posibles relaciones entre las estructuras obtenidas para cada uno de los grupos. La comparación de estas estructuras se puede abordar desde diferentes puntos de vista:

- - La obtención de una medida global de la relación existente entre todos los grupos estudiados.
- - La búsqueda de factores comunes a todos o a algunos de los grupos y de factores específicos de algún grupo.
- - La representación de una nube media, compromiso entre todos los grupos analizados.
- - La representación simultánea de las nubes de individuos y variables asociadas a los diferentes grupos en un marco de referencia común.

El Análisis Factorial Múltiple propone soluciones, óptimas en cierto sentido, para todos y cada uno de estos puntos de vista. La técnica básica subyacente en este método es una técnica factorial bien conocida

y aplicada en muy diversos campos de investigación: el Análisis en Componentes Principales, heredando de ella la simplificación de cálculo y la riqueza interpretativa de los resultados alcanzados.

3) El tratamiento simultáneo de varios grupos de variables conlleva un problema: existe el riesgo de que un análisis factorial global esté determinado de manera preponderante por uno de estos grupos.

El método AFM busca una ponderación que equilibre la influencia de todos y cada uno de los grupos. Se desea que la ponderación tenga propiedades interesantes teniendo en cuenta, principalmente, dos puntos de vista: no modificar la estructura inicial de los grupos, realizando cada uno de ellos una función similar en el análisis global. En el papel que puede desempeñar un grupo de variables en el análisis simultáneo de varios grupos intervienen varios elementos:

- - El número de variables, ya que cuanto mayor sea éste mayor será la influencia del grupo.
- - La estructura del grupo, ya que cuanto más fuerte sea ésta, es decir, cuanto más relacionadas están las variables que lo constituyen más determinante será su influencia en la construcción de los principales factores de variabilidad de la nube global.

Teniendo en cuenta el primer punto de vista de los mencionados anteriormente, la ponderación debe ser la misma para todas las variables de un grupo, de esta forma la estructura interna no se verá modificada. Atendiendo, por otra parte, el segundo punto de vista, se puede considerar que la determinación de un factor se apoya solamente sobre la inercia en una dirección dada. Es natural, por tanto, normalizar esta última para igualar el peso de los diferentes grupos, esto es, para que los grupos participen de modo equilibrado en la determinación, fundamentalmente, del primer factor. El primer valor propio de los análisis separados de cada grupo recoge la máxima inercia de las proyecciones de los puntos de cada grupo, es decir, es la inercia de la primera componente principal de cada grupo. Por tanto, el inverso de este primer valor propio parece la ponderación ideal: no sólo tiene en cuenta el número de variables que constituyen cada grupo, sino también las relaciones entre éstas.

Una ponderación que no tuviera en cuenta las relaciones entre las variables, igualando las inercias totales de las nubes de variables de cada grupo, por ejemplo, haría débil relativamente la inercia de cada dirección de un grupo formado por muchas variables independientes. En cambio, tal ponderación haría fuerte, relativamente, la inercia en una dirección de un grupo compuesto por pocas variables muy correladas, aumentando, por tanto, su importancia en la determinación de los ejes factoriales.

Con la ponderación elegida un grupo compuesto de dos variables muy correladas será equivalente a un grupo de una sola variable, mientras que un grupo compuesto por dos variables no correladas conservará una inercia igual a 1 en las dos direcciones ortogonales y, por tanto, una inercia total igual a 2.

La ponderación de todas las variables de un grupo es idéntica, por lo que no se ve modificada la forma de estas nubes, pero sí normalizada en el siguiente sentido: la inercia de cada nube queda multiplicada por esa ponderación en cada dirección del espacio. Con la ponderación elegida, la inercia de la primera componente principal de cada grupo se iguala a 1, por tanto ningún grupo puede ser preponderante en la primera dirección de inercia de la nube media, manteniendo la estructura interna de cada uno de ellos. No obstante, el número de direcciones sobre las que influye un grupo crece, generalmente, con el número de variables que lo integran. Además, dos nubes homotéticas se harían iguales, lo que facilita la tarea de comparar las mismas.

4) El Análisis Factorial Múltiple puede quedar estructurado, al igual que la mayor parte de los métodos de Análisis Simultáneo de Varias Tablas de Datos existentes, en las tres etapas consolidadas desde la aparición del método STATIS: intra-estructura, compromiso e inter-estructura.

El compromiso y la intra-estructura de referencia corresponden al análisis factorial de la nube media: se obtienen, una vez ponderados los diferentes grupos de variables, los factores generales representativos del conjunto a analizar. Sobre estos factores generales se proyectan las nubes parciales asociadas a cada

uno de los grupos como elementos suplementarios. El estudio de la inercia de los puntos de estas nubes, con respecto a su centro de gravedad, da una idea clara de la intra-estructura, además de obtener una visión compromiso.

La inter-estructura o estudio comparativo de la proximidad entre las diferentes nubes corresponde tanto a la representación de los grupos de variables como al cálculo de la correlación entre los factores parciales, obtenidos en el análisis parcial de cada nube, con los factores comunes o generales.

5) El Análisis Factorial Múltiple permite el tratamiento simultáneo de variables activas cuantitativas y cualitativas de una manera simplificada y sin alterar su naturaleza. Las variables cuantitativas no sufren modificación alguna, mientras que las variables indicadoras, asociadas a las modalidades de variables cualitativas, son debidamente ponderadas. La ponderación elegida no modifica su naturaleza y permite construir una nube con las mismas propiedades inerciales que la nube de modalidades en Análisis de Correspondencias Múltiples. Analiza, sin modificar la naturaleza de las variables originales, tablas mixtas, esto es, tablas formadas por variables cuantitativas y cualitativas, interviniendo todas ellas como activas en el análisis global. La única restricción que impone el AFM para el tratamiento de las tablas mixtas es que las variables sean homogéneas en el interior de cada grupo.

En concreto, nosotros hemos comprobado, Landaluce 95, que las matrices analizadas en Análisis de Correspondencias Múltiples y Análisis en Componentes Principales normado de la Tabla Disyuntiva Completa así ponderada son equivalentes, excepto por un coeficiente, el inverso del número de variables cualitativas analizadas. Además, las distancias definidas en Análisis en Componentes Principales y en Análisis de Correspondencias Múltiples, aún siendo diferentes, guardan una estrecha relación.

En conclusión, un ACP normado de las variables indicadoras ponderadas, correspondientes a las modalidades de las variables cualitativas, conduce a los mismos factores sobre el conjunto de individuos que un ACM. Es posible, por tanto, extender a las variables cualitativas la metodología desarrollada para grupos de variables cuantitativas. Esta extensión del campo de aplicación del AFM está reforzada por la equivalencia entre un ACM y un AFM sobre un conjunto de variables cualitativas, considerando en este último análisis que cada grupo está formado por las indicadoras de una sola variable.

Este último resultado deriva del siguiente hecho: cuando se aplica el AFM a un conjunto de variables indicadoras, previamente ponderadas por la proporción de individuos que no las han elegido, respectivamente, los coeficientes de ponderación, propios del AFM, son todos iguales a 1. En efecto, la nube de las indicadoras asociadas a un variable cualitativa posee una inercia de 1 en todas las direcciones del subespacio que engendran. De donde se obtiene que los valores propios de los ACP de cada uno de los grupos son todos iguales a la unidad. El AFM conduce a los mismos factores que el ACP normado y es, por tanto, equivalente a un ACM.

6) El Análisis Factorial Múltiple permite el tratamiento de datos ausentes y modalidades de débil efectivo, dando soluciones a los problemas que estos tópicos conllevan.

Cuando se trabaja con variables cualitativas no es infrecuente encontrarse con modalidades que se refieren a muy pocos individuos (modalidades raras) y/o con falta de respuesta en el caso de encuestas (datos ausentes). Ambas situaciones provocan, cuando se analizan los datos, distorsiones en los resultados, lo que conduce a su tratamiento especial previo al análisis factorial.

Los procedimientos habituales, cuando se utiliza un ACM, consisten en la construcción de una modalidad suplementaria "dato ausente". Esta solución no es del todo satisfactoria ya que conduce, por una parte, con frecuencia a una modalidad de efectivo débil, y, por otra parte, puede perturbar los resultados, en caso de no tener esta nueva modalidad una interpretación particular. Desde el punto de vista multicanónico esto supone introducir una nueva dimensión al subespacio asociado a la variable cualitativa. Todas las direcciones de este subespacio tienen el mismo papel en la determinación de los factores, esta dirección no significativa ejerce una influencia similar al resto. No obstante, puede

considerarse válida cuando el dato ausente tenga algún significado, como el caso de la no respuesta voluntaria en las encuestas de opinión.

Existen otras soluciones a estos problemas: se pueden eliminar los individuos que han elegido la modalidad rara, pero existe el riesgo de que los factores sean diferentes ya que la dimensión del subespacio asociado a la variable decrece. Asimismo, se pueden suprimir las modalidades raras y no introducir la no respuesta. No obstante, la tabla resultante con estas omisiones, aún siendo disyuntiva, es incompleta, por lo que no tiene las mismas propiedades que una tabla disyuntiva completa. En concreto, la marginal sobre el conjunto de individuos no es constante, se pierde formalmente una parte de la coherencia del análisis y la doble interpretación de los factores.

En este método, ante la existencia de datos cuantitativos ausentes, se adopta la solución de sustituir dichos datos por la media aritmética, técnica utilizada habitualmente en Análisis en Componentes Principales. En el caso de variables cualitativas las modalidades de efectivo débil y los datos ausentes tienen, asimismo, un tratamiento específico que es igualmente válido para tablas disyuntivas completas e incompletas: se codifican con cero los valores desconocidos y se eliminan las modalidades raras. Se obtienen los mismos factores que en el Análisis de Correspondencias Múltiples en el primer caso y que el Análisis de Correspondencias con Marginal Modificada, propuesto por B. Escofier, en el segundo.

7) El Análisis Factorial Múltiple proporciona resultados de gran interés desde el punto de vista de la interpretación. En un mismo gráfico aparecen simultáneamente individuos, variables y modalidades. La riqueza interpretativa alcanza su máximo en los planos factoriales de las representaciones superpuestas de los individuos vistos a través de cada uno de los grupos analizados. Un estudio minucioso de estos planos permite detectar la tendencia general de los individuos, los grupos que definen en mayor medida esta tendencia así como los individuos cuyo comportamiento es diferente según el grupo que los describe. Esta tendencia queda definida a partir de todos los grupos de variables, sin que haya grupos que queden eclipsados por la presencia de otros con mayor peso inicial. No hay que olvidar que el AFM es un método diseñado para estudiar las diferencias entre varias tablas de datos que tienen suficientes caracteres comunes. Cuando estas tablas presentan grandes diferencias su estudio, a través del AFM, no tiene utilidad alguna.

En este método las reglas de interpretación referentes a individuos, variables e indicadoras de las variables cualitativas son esencialmente las mismas que en los métodos clásicos de Análisis de Datos, presentando, además, resultados específicos de la estructura en grupos, que tienen sus propias reglas de interpretación.

Otro de los objetivos del estudio de los grupos de variables es comparar la posición de un individuo en cada una de las nubes correspondientes a cada grupo, es decir, se desea analizar si dos individuos que son próximos desde un punto de vista general, son también próximos o no desde otros puntos de vista parciales. En AFM se efectúa una representación simultánea de estas nubes parciales en un espacio de reducida dimensión.

Las primeras componentes principales de los ACP parciales, análisis de cada uno de los grupos por separado, pueden situarse también en el espacio de los individuos y su representación simultánea resulta de gran utilidad para el estudio de las correlaciones inter-grupos. Con este fin se introducen estas componentes como elementos suplementarios en el análisis global. Para analizar las correlaciones intra-grupo es realmente fructífero observar los ángulos entre las variables de un mismo grupo.

El AFM proporciona un importante resultado: la representación del conjunto de los grupos de variables en la que cada uno de ellos aparece representado por un punto. La proximidad entre dos puntos se interpreta en términos de semejanza entre grupos. Sin embargo, este gráfico da una idea pobre de proximidad ya que no indica su descomposición. Para ello hay que recurrir a las medidas de relación entre factores parciales y generales, que señalan la existencia o no de factores comunes y específicos. La representación de los grupos de variables se utiliza en la práctica como ayuda a la interpretación.

El AFM permite el estudio de voluminosos ficheros de datos correspondientes a análisis en los que los individuos no interesan en sí mismos sino únicamente a través de los valores que toman en las variables

cuantitativas o las modalidades elegidas correspondientes a las variables cualitativas. Este método proporciona una representación en la que se proyectan, sobre los planos factoriales, los centros de gravedad de los individuos estudiados.

8) El Análisis Factorial Múltiple está situado a caballo entre la Escuela Francesa y la Escuela Anglosajona. Por una parte, es un método con una filosofía propia de la Escuela Francesa y, por otra parte, mantiene una estrecha relación con métodos de la Escuela Anglosajona.

El Análisis Factorial Múltiple y el Análisis Canónico Generalizado en el sentido de Carroll son dos métodos diferentes con un mismo planteamiento: la búsqueda de factores comunes a varios grupos de variables, pero que utilizan una medida de relación entre una variable y un grupo de variables diferente. Sin embargo, el Análisis Factorial Múltiple presenta ciertas ventajas, tanto en la obtención como en la interpretación de las variables generales y de las variables canónicas.

El Análisis Factorial Múltiple ofrece, asimismo, la posibilidad de estimar los parámetros que plantea el modelo INDSCAL. Estos son: factores y pesos asociados a estos factores en cada uno de los grupos de variables. Esta posibilidad se debe al planteamiento similar en ambos métodos: la búsqueda de una estructura común a todos los grupos de variables. Al igual que en el caso anterior, el Análisis Factorial Múltiple presenta ventajas respecto al modelo INDSCAL, que se refieren tanto a la sencillez de los cálculos y la no existencia de problemas de convergencia, como a la etapa interpretativa y estudio comparativo de los resultados alcanzados.

APLICACIONES DEL AFM

El AFM tiene múltiples aplicaciones debido, no solo a la posibilidad que ofrece de analizar tablas mixtas, sino también a la diferente estructuración en grupos a que puede someterse un mismo conjunto de datos. Cada una de las diversas agrupaciones puede corresponder a un punto de vista distinto. Los resultados que se desprendan de los diferentes análisis que resultan de cada estructuración se complementan entre sí, por lo que en la mayoría de los casos es interesante realizarlos todos, así como los análisis de cada grupo por separado.

Un ejemplo frecuente es el conjunto de variables observadas en diferentes momentos de tiempo, un grupo puede estar formado por la misma variable medida en diferentes períodos de tiempo o bien por todas las variables medidas en el mismo momento. Los resultados que se desprenden de estos dos análisis serán, seguramente, bastante diferentes, pero también se complementarán entre sí. Además, el AFM puede utilizarse para mostrar, de forma pedagógica, las diferencias que presentan los métodos factoriales ACP y Análisis Factorial de Correspondencias cuando se aplican al mismo conjunto de datos. En este caso habrá dos grupos de variables formados, cada uno de ellos, por los factores obtenidos a través de cada método.

Estos ejemplos ponen de manifiesto que la estructuración en grupos de variables no es sólo un problema técnico, sino que puede considerarse un problema metodológico que enriquece el campo del Análisis de Datos, enriquecimiento al que contribuye, en gran medida, el Análisis Factorial Múltiple.

Ante cualquier conjunto de datos susceptibles de ser estructurados en grupos es recomendable aplicar inicialmente un Análisis Factorial Múltiple a dicho conjunto. A partir de este análisis se obtiene una visión general de los datos a estudiar y de las posibles relaciones entre los grupos. Esta visión permitirá la toma de decisiones sobre posteriores análisis: incluir todos los grupos, seleccionar subconjuntos, elegir algún grupo para actuar como suplementario, etc. A su vez es de gran interés y, a menudo, de gran utilidad realizar los análisis separados de cada grupo. Estos análisis parciales pueden proporcionar resultados y conducir a conclusiones que de otro modo quedarían eclipsados en análisis más generales.

Por último, señalar que el AFM aplicado a un conjunto de variables homogéneas, cualitativas o cuantitativas, en el que se considera cada grupo formado por una sola de estas variables (en el caso de variables nominales cada grupo estaría integrado por las modalidades de una sola variable, ponderadas por la proporción de individuos que no la han elegido) proporciona, por una parte, los mismo resultados que un ACM o un ACP normado, respectivamente. Efectivamente, la ponderación de las variables numéricas coincidiría con su varianza, mientras que para las variables nominales la ponderación sería la unidad dado que las variables indicadoras correspondientes a cada una de las modalidades de una sola variable son ortogonales entre sí. No obstante, por otra parte, el AFM proporciona importantes resultados adicionales, sobre todo en lo que se refiere a las representaciones gráficas: proyecciones de los individuos y de los grupos de variables en el caso cualitativo y proyecciones de los puntos parciales, que representan a cada individuo (y/o centro de gravedad de las clases de individuos en el caso cualitativo) según cada una de las variables (grupos).

REFERENCIAS BIBLIOGRAFICAS

- Escofier B., Pagès J. (1982a), "Comparaison des groupes de variables définies sur le même ensemble d'individus", INRIA, Institut National de Recherche en Informatique et en Automatique, 149
- Escofier B., Pagès J. (1989), "Multiple Factor Analysis: Results of a Three Year Utilization", Multiway Data Analysis, 277-285
- Escofier B., Pagès J. (1990), "Análisis Factoriales Simples y Múltiples: Objetivos, Métodos e Interpretación", Lejona-Bilbao, Servicio Editorial de la Universidad del País Vasco, 1992
- Landaluce M.I. (1995), "Estudio de la Estructura de Gasto Medio de las Comunidades Autónomas Españolas. Una Aplicación del Análisis Factorial Múltiple", Facultad de Ciencias Económicas y Empresariales. País Vasco