

Estudios de Economía Aplicada
Nº 15, 2000. Págs. 103-124

La importancia de la estructuración en tablas de los datos: Una aplicación

M. I., LANDALUCE CALVO
S., APARICIO CASTILLO
*Dpto. de Economía Aplicada
Universidad de Burgos*

RESUMEN

El objetivo principal de este trabajo es poner de manifiesto la importancia que tiene, para todo análisis de carácter exploratorio, la consideración o no de la estructura en diferentes matrices que, en muchas ocasiones, está asociada a la naturaleza de los datos a describir. Para ello se presentan los resultados de un estudio empírico que abarca los dos puntos de vista: tabla única *versus* tabla múltiple. Para el análisis de la tabla única se ha utilizado el Análisis en Componentes Principales (ACP) y para el análisis de la tabla múltiple el Análisis Factorial Múltiple (AFM).

Palabras clave: Tabla única, tabla múltiple, análisis factoriales simples y múltiples.

ABSTRACT

The main objective of this research is to state the importance, for any exploratory analysis, of considering or not the structure in a set of data matrices. Such structure is associated, in many cases, to the nature of the data. In this paper, we present the empirical results obtained from the two points of view: single table vs. multiple table analysis. For the single table case we have used Principal Components Analysis and Multiple Factor Analysis for the multiple table case.

Código UNESCO: 120903, 120909

Artículo recibido en octubre de 1999. Revisado en enero de 2000.

Introducción: objetivos y datos

El núcleo central de este trabajo lo constituye la presentación, en términos comparativos, y a partir de un estudio empírico, de dos metodologías exploratorias que permiten el análisis de un mismo conjunto de datos desde dos puntos de vista diferentes, proporcionando resultados, en ocasiones, distintos aunque complementarios. Estos puntos de vista hacen referencia a la consideración o no de la estructura en diferentes matrices que, en muchas ocasiones, está asociada a la naturaleza de la información a analizar y que supone estudiarla desde dos vertientes distintas: tabla única versus tabla múltiple.

El estudio empírico se ha centrado en el análisis de las cajas de ahorro del actual sistema financiero español, obteniendo una caracterización de las mismas en función de distintas magnitudes, relacionadas con los siguientes cinco aspectos: volumen de negocio, eficacia y productividad, gestión y rentabilidad, cuenta de resultados e inversión crediticia por sectores. Los datos seleccionados para el estudio aquí presentado corresponden a los publicados por la revista de carácter económico RANKING, en el número 116 del mes de junio de 1998.

Las herramientas estadísticas utilizadas para los fines perseguidos y dada la naturaleza cuantitativa de la información considerada, han sido las siguientes: para el estudio de la tabla única el Análisis en Componentes Principales (ACP) y para el análisis de la tabla múltiple el Análisis Factorial Múltiple (AFM). En este trabajo se van a poner de manifiesto las diferencias y semejanzas, así como las ventajas y los inconvenientes que presenta cada una de las dos metodologías.

Metodología

Como ya se ha señalado, los métodos de análisis exploratorio que se han utilizado en la realización de este trabajo son el Análisis de Componentes Principales (ACP), para el análisis de los datos estructurados en una sola tabla, y el Análisis Factorial Múltiple (AFM), para el análisis de los datos estructurados en diferentes tablas. Ambas metodologías se han utilizado con la filosofía característica de la escuela francesa de análisis de datos, y los resultados que se presentan se han obtenido a través del paquete estadístico SPAD- versión 3.21.

No se ha considerado necesario, dado el general conocimiento que existe sobre el ACP, realizar una presentación del mismo. Sin embargo, la técnica de análisis de

tablas múltiples, AFM, es una metodología relativamente nueva y, por ello, más desconocida, por lo que se va a proceder en esta sección a la exposición, de forma breve, de sus principios teóricos.

El AFM, desarrollado por B. Escofier y J. Pagès (1992), es un método factorial adaptado al tratamiento de tablas de datos en las que un mismo conjunto de individuos se describe a través de varios grupos de variables. Los grupos de variables pueden surgir de la utilización conjunta de variables de diferente naturaleza, cuantitativas y cualitativas, del empleo de tablas que provienen de otras de tres dimensiones o del manejo de un mismo conjunto de variables medidas en distintos periodos de tiempo. 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 en el seno de cada uno de los grupos.

Respecto a los métodos factoriales clásicos, esta técnica presenta algunas ventajas como la de equilibrar la influencia de los grupos en el análisis global, o la de trabajar con tablas de naturaleza mixta (grupos de variables cuantitativas o cualitativas), ofreciendo además resultados típicos de la estructura en grupos del conjunto de variables (la relación existente entre los grupos, los factores comunes a todos ellos, la representación simultánea de individuos y variables desde todos los puntos de vista, grupos, en un marco de referencia común).

El objetivo principal de este método es analizar la estructura común de las distintas tablas de datos objeto de estudio, poniendo de manifiesto cuáles son los elementos heterogéneos, es decir, aquéllos que tienen un comportamiento diferente al resto.

Hay que señalar que para las variables continuas esta técnica se comporta como un Análisis en Componentes Principales (ponderando las variables); y para las variables nominales se comporta como un Análisis de Correspondencias Múltiples (ponderando las modalidades) (Landaluce, 1997).

En una primera etapa del AFM, en la que se realizan los Análisis de Componentes Principales de cada grupo por separado, análisis parciales, se estudian los valores propios que ponen de manifiesto la dimensionalidad de la estructura interna de cada uno de ellos, comprobando, a la vez, el número de dimensiones de cada grupo que intervendrán con cierto peso en el análisis general.

En una segunda etapa, se realiza el análisis simultáneo de los grupos previamente ponderados, que se denomina análisis global, esto es, se analiza la tabla múltiple que surge al yuxtaponer de forma horizontal los diferentes grupos considerados. La ponderación seleccionada es el inverso del primer valor propio obtenido en los análisis separados de cada grupo. La ponderación de todas las variables que integran un grupo es idéntica, por lo que no se ve modificada la forma de las nubes asociadas a los distintos grupos. Esta ponderación no equilibra la inercia total de los grupos, pero

normaliza las nubes asociadas en el sentido de que la inercia de cada una de ellas queda multiplicada por esa ponderación en cada dirección del espacio (eje factorial). De esta forma, la inercia de la primera componente principal de cada grupo se iguala a 1.

En esta etapa, se estudian, en primer lugar, los valores propios obtenidos en el análisis global. Estos valores proporcionan una primera información sobre las relaciones existentes entre los grupos analizados. El valor máximo que puede tomar el primer valor propio coincide con el número de grupos activos analizados. Se alcanza cuando la dirección global de mayor inercia es común a todos los grupos. El valor alcanzado por el resto de los valores propios depende de la estructura interna de cada una de las tablas. La existencia de factores comunes puede ser detectada, igualmente, a través del cálculo de las correlaciones entre los factores globales y los factores del mismo orden correspondientes a cada uno de los grupos analizados, obtenidos en los respectivos análisis parciales. Cuando la correlación es fuerte el factor global traduce una tendencia que está presente en todos los grupos. En caso de que la correlación sea alta para uno solo de estos grupos será un factor específico de dicho grupo. Esta fase de la interpretación es importante ya que el interés del análisis simultáneo de varios grupos de variables radica en el estudio de las comunales que éstos mantengan entre sí. En caso de observar la presencia de un grupo que está únicamente correlado fuertemente con factores que le son específicos se procederá a la eliminación del mismo como elemento activo del análisis, debido a las diferencias respecto al resto de los grupos. En este sentido el AFM puede ser interpretado como un análisis multicanónico: sus factores globales son como las variables generales de un análisis multicanónico (variables relacionadas con el conjunto de los grupos de variables), mientras que los factores de las nubes parciales pueden ser interpretados como las variables canónicas (combinación lineal de las variables de un grupo más relacionada con la variable general), ya que constituyen las representaciones de los factores globales en los distintos grupos.

Posteriormente se procede al análisis de las coordenadas de los grupos sobre un determinado factor, pudiéndose considerar una medida de la importancia de la dirección de dispersión en cada uno de los grupos. Esta coordenada se interpreta como la contribución absoluta de las variables de una tabla al factor en cuestión, esto es, indica los grupos que han determinado en mayor medida los factores. Hay que señalar que, como consecuencia de la ponderación, la contribución de los grupos al primer factor está, generalmente, bastante equilibrada.

El AFM es un método que permite el estudio tanto de la intra-estructura, esto es, estudio de la inercia de los puntos de cada nube (grupo) con respecto a su centro de gravedad, como de la inter-estructura o estudio comparativo de la proximidad entre los diferentes grupos. La suma de estas dos inercias es lo que se denomina inercia total. La razón inercia inter/inercia total es un índice que refuerza las conclusiones

alcanzadas a través del análisis de las correlaciones antes mencionadas. En caso de estar próxima a 1 confirma el carácter de común para el factor en cuestión. Además, los puntos que representan al mismo individuo, desde los diferentes puntos de vista referentes a los grupos, estarán próximos entre sí.

Existen otros coeficientes que ayudan a la interpretación del análisis y que pueden presentarse en forma de matrices. Por un lado, la matriz L recoge las relaciones, en el sentido del AFM, esto es, después de la ponderación, entre los grupos estudiados. Esta relación se define como la suma ponderada de los cuadrados de las correlaciones entre las variables de cada grupo. La proximidad a 1, de los coeficientes que forman la diagonal principal de esta matriz pone de manifiesto la existencia de un primer eje preponderante en el análisis parcial de cada grupo por separado. Mientras que el máximo valor que pueden tomar, en el caso de incorrelación entre las variables de cada grupo, corresponde al número de éstas. Hay que tener en cuenta que estos coeficientes de relación están afectados por la dimensionalidad de cada uno de los grupos, esto es, número de factores. Por otro lado, y como consecuencia de esto, se define una segunda matriz, denominada RV, obtenida como una matriz de correlación a partir de la matriz L, tomando esta última como una matriz de varianzas y covarianzas entre grupos. De esta forma se tiene en cuenta la variabilidad de los grupos, obteniendo coeficientes comparables ya que su valor está comprendido entre 0 y 1.

Por último, señalar que el AFM proporciona resultados de gran interés desde el punto de vista de la interpretación. En un mismo gráfico pueden aparecer conjuntamente individuos, variables y, en su caso, modalidades. La riqueza interpretativa alcanza su máximo en los planos factoriales de las representaciones superpuestas de los individuos vistos a través de cada una de las tablas analizadas. 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 raros, es decir, aquéllos 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 (Landaluce, 1995).

Resultados del primer análisis: ACP de la tabla única

La información que se ha analizado, para realizar el estudio empírico que se presenta en este trabajo, aparece recogida en una matriz de naturaleza cuantitativa que posee 50 filas, correspondientes a cada una de las 50 cajas de ahorro del actual sistema financiero español, y 58 columnas, cada una de ellas asociada a las diferentes variables seleccionadas por su relevancia para obtener una caracterización de las entidades de interés en este trabajo.

En el anexo se recoge la descripción de algunas de estas características métricas, así como la composición de cada uno de los cinco grupos considerados. El hecho de trabajar con unidades de medida muy distintas, nos ha llevado a la realización de análisis normados, tanto en el estudio de la tabla única, como en el estudio de la tabla múltiple, equilibrando, en cierto sentido su influencia en la caracterización de las cajas de ahorro.

A continuación se exponen e interpretan los resultados obtenidos en el ACP de la tabla única:

- **El histograma de los 10 primeros valores propios** proporciona una primera visión de la relación global que mantienen las características estudiadas en este trabajo. Se observa un primer eje claramente preponderante, que recoge un 38.74% de la variabilidad total de los datos, proyectándose en el plano principal más de la mitad de la inercia, en concreto un 53.27% de la misma. Nuestro interés se centra en poner de manifiesto una realidad global, que corresponda a la mayor parte de los individuos aquí estudiados, por ello la interpretación de resultados la vamos a limitar a estos dos primeros ejes, ya que el resto recogen realidades correspondientes a entidades financieras concretas y a unas variables determinadas, no siendo el objetivo principal de este trabajo.

| | | | | |
|--|------------------|--------------------|-----------|-------|
| VALEURS PROPRES | | | | |
| APERCU DE LA PRECISION DES CALCULS: TRACE AVANT DIAGONALISATION58.0000 | | | | |
| SOMME DES VALEURS PROPRES58.0000 | | | | |
| HISTOGRAMME DES 10 PREMIERES VALEURS PROPRES | | | | |
| NUMERO | VALEUR PROPRE | POURCENT CUMULE | POURCENT. | |
| 1 | 22.4666 | 38.74 | 38.74 | ***** |
| 2 | 8.4307 | 14.54 | 53.27 | ***** |
| 3 | 6.5495 | 11.29 | 64.56 | ***** |
| 4 | 4.1808 | 7.21 | 71.77 | ***** |
| 5 | 3.4974 | 6.03 | 77.80 | ***** |
| 6 | 2.2090 | 3.81 | 81.61 | ***** |
| 7 | 1.7508 | 3.02 | 84.63 | ***** |
| 8 | 1.4006 | 2.41 | 87.04 | **** |
| 9 | 1.0281 | 1.77 | 88.82 | *** |
| 10 | 0.8991 | 1.55 | 90.37 | *** |

- **Análisis de la nube de puntos-variable (Gráfico 1).** El análisis de la nube en la que cada punto representa a una de las características consideradas, va a poner de manifiesto cuáles son las que más contribuyen a la formación de los primeros ejes de variabilidad, es decir, qué variables tiene mayor peso en la definición de las posibles tipologías existentes entre las cajas de ahorro. Merecen ser destacados los siguientes resultados:

1. Las variables Recursos de clientes, Fondo provisión para insolvencias, Caja y Depósitos en Bancos Centrales, Entidades de crédito de pasivo, Cartera de valores de renta fija, Número de empleados, Número de oficinas, Productos Financieros, Beneficio Neto y Dotación para obras de beneficio social, son las que más contribuyen a la formación del primer factor. Como puede comprobarse, todas ellas son características relacionadas principalmente con el volumen de negocio y la cuenta de resultados. Además, del gráfico correspondiente, se deduce que la práctica totalidad de las variables tienen coordenada de igual signo, lo que traduce un efecto talla, con la excepción de las tres únicas variables que tienen correlación positiva: Otros productos y cargas de explotación, Absorción gastos de explotación y Rentabilidad de la inversión certificada, que además, contribuyen muy débilmente a la formación de este factor. Por tanto, puede decirse que este primer eje pone de relieve, de izquierda (parte negativa) a derecha (parte positiva) una clasificación general de las cajas de ahorro, aunque la caracterización subyacente de las entidades de ahorro es parcial, dado el gran peso que, como ya se ha comentado anteriormente, algunas de las variables relativas al volumen de negocio y a la cuenta de resultados, tienen sobre la misma. No hay que olvidar que la realidad que refleja este eje de variabilidad representa un 38.74% de la información total, y que el efecto talla es muy habitual en la metodología de componentes principales, por lo que para vislumbrar otras caracterizaciones, también importantes, aunque más específicas, hay que estudiar e interpretar ejes posteriores.
2. En lo que se refiere al segundo eje señalar que éste pone de manifiesto una segunda caracterización de las entidades financieras, distinta a la reflejada por el factor anterior. Las variables que más contribuyen a su formación son: Activo por empleado, Depósitos por empleado, Beneficio por empleado, Activo por oficina, Depósito por oficina, Beneficio por oficina, Inversión crediticia por oficina, esto es, ratios relacionados con la eficiencia y productividad. En menor medida y enfrentadas, en el plano, a las anteriores destacan los indicadores de Absorción gastos de explotación (%) y Rentabilidad de la inversión crediticia (%), ambos relacionados con la gestión y rentabilidad. Por lo que esta segunda dirección de inercia va a poner de manifiesto, principalmente, cuáles son las entidades que destacan, entre las 50 estudiadas, por su comportamiento en lo que a eficiencia y productividad se refiere.

3. Por todo lo anteriormente dicho, se comprueba que las características que más discriminan entre los comportamientos de las entidades de ahorro del actual sistema financiero español son las relacionadas con el tamaño de las mismas, volumen de negocio, y con los resultados alcanzados a través del conjunto de sus estrategias de actuación, cuenta de resultados. Una vez obtenida esta tipología de las cajas de ahorro, que coincide básicamente con el ranking obtenido atendiendo a su activo total, se alcanza una nueva caracterización, más parcial y, por tanto, más específica que la anterior, a través de la que se puede vislumbrar cuales son aquellas entidades que destacan, tanto por tener unos ratios elevados, como débiles, en unos aspectos concretos de su política, aquéllos que tienen relación con la eficiencia y productividad, sobre todo de las oficinas, pero también de los empleados.
4. Hay que destacar que las variables relacionadas con el grupo relativo a la inversión crediticia por sectores no tienen peso alguno en este plano factorial. Hay que acudir al tercer eje para ver reflejada su realidad, pero como ya hemos indicado este factor no va a ser analizado ni presentado aquí, ya que carece de relevancia.

• **Análisis de la nube de puntos-individuo (Gráfico 2).** El identificador de la proyección de cada caja de ahorros en este plano está compuesto, además del nombre, por un número que corresponde al orden que ocupa dicha entidad en el ranking obtenido atendiendo al activo total del año 1997. Esta información será de gran ayuda en la interpretación del primer plano factorial, sobre todo del primer eje de variabilidad.

1. Señalar que, después de analizar e interpretar la nube de puntos-variable, el primer factor de variabilidad de la nube de los individuos refleja, con algunas excepciones que merecen ser destacadas, el ranking general de las cajas de ahorro, que puede comprobarse con el número que acompaña al nombre de la entidad en su coordenada. Una de estas excepciones a las que se hace referencia es el hecho de que la CA y MP de Madrid, ocupa, a lo largo del eje horizontal, la primera posición (presenta la mayor coordenada negativa), por delante de la CA y Pens de Barcelona. Este hecho que destacamos ahora, va a ser comentado y explicado posteriormente con mayor detalle, en la sección dedicada al análisis de los resultados del AFM.
2. La distribución, sobre los cuatro cuadrantes del primer plano factorial, de los puntos asociados a cada uno de los individuos estudiados pone de manifiesto que, aunque el primer eje representa un destacable porcentaje de la realidad estudiada, algunas entidades tienen un comportamiento específico que merece ser interpretado con minuciosidad. Así, en el segundo eje destacar el protagonismo de dos entidades, en la parte positiva del mismo, que son CA y

MP de Guipúzcoa y San Sebastián y Bilbao Bizkaia Kutxa, que junto a otras entidades que tienen menos peso y, por lo tanto, menos protagonismo, como CA de Navarra, CA Municipal de Burgos, CA de Vitoria y Alava y CA de Asturias, encabezan un nuevo ranking caracterizado, como ya se ha comentado en el análisis de las variables, por los ratios relacionados con la eficiencia y productividad, sobre todo los que recogen los activos, depósitos, beneficios e inversión crediticia por oficina. Hay que destacar, además, que éstas son entidades que en el ranking general ocupan posiciones medias.

Resultados del segundo análisis: AFM de la tabla múltiple

En este análisis las 58 variables consideradas aparecen agrupadas en tablas relacionadas, cada una de ellas, con cada uno de los cinco puntos de vista seleccionados para obtener la caracterización multidimensional de las cajas de ahorro españolas: volumen de negocio, eficacia y productividad, gestión y rentabilidad, cuenta de resultados e inversión crediticia por sectores. Esta estructuración, que se ha realizado siguiendo un criterio totalmente objetivo dada la naturaleza de la información analizada, va a enriquecer, respecto al análisis anterior, los objetivos, ya que, además de obtener resultados globales, relacionados con todas las tablas, se van a obtener resultados propios de cada uno de los cinco grupos.

A continuación se exponen e interpretan los resultados más destacados obtenidos al analizar, a través de la metodología del Análisis Factorial Múltiple (AFM), las cinco tablas antes mencionadas:

- **El histograma de los cinco primeros valores propios**, del análisis de las cinco tablas yuxtapuestas, pone de manifiesto la existencia de un primer eje global preponderante, en el sentido de que recoge un importante porcentaje de inercia, prácticamente un 32%, frente al 13.72% de variabilidad que se proyecta en el segundo factor global. Aunque la tercera componente recoge una inercia nada despreciable, un 10.66%, nuestra atención se va a centrar, al igual que en el análisis anterior, en el plano principal generado por las dos primeras componentes, ya que en él se vislumbran las principales caracterizaciones de las entidades financieras, esto es, las diferencias y semejanzas más relevantes.

- Los coeficientes L, como se ha visto, están influenciados por la dimensionalidad interna de cada grupo, por lo que no son directamente comparables. Por este motivo, una medida más apropiada para estudiar el grado de relación entre los grupos son **los coeficientes RV**, que se obtienen como coeficientes de correlación, a partir de la matriz de coeficientes L, tomando estos últimos como varianzas y covarianzas. Los coeficientes RV, así obtenidos, ponen de manifiesto los grupos que presentan una estructura interna con mayores similitudes o diferencias, siendo, además, directamente comparables. En nuestro caso, se comprueba que los grupos relativos al volumen de negocio y a la cuenta de resultados (grupos 1 y 4) están estrechamente relacionados. En cierta medida este resultado es lógico ya que son, a priori, los dos puntos de vista con mayor relación causa-efecto, ya que recogen información relativa al tamaño de la entidad y a los resultados globales de su estrategia general de actuación. A su vez, estos dos grupos, junto con el grupo asociado a la inversión crediticia, son los que presentan mayores diferencias con el grupo 3, que hace referencia a la gestión y rentabilidad de las cajas de ahorro. El resto de relaciones se pueden situar en un nivel intermedio entre las anteriormente descritas, ya que toman valores próximos a 0.5.

| COEFFICIENTS RV DE LIAISON ENTRE GROUPES | | | | | |
|--|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | 1.00 | | | | |
| 2 | 0.56 | 1.00 | | | |
| 3 | 0.15 | 0.44 | 1.00 | | |
| 4 | 0.93 | 0.55 | 0.19 | 1.00 | |
| 5 | 0.45 | 0.34 | 0.15 | 0.48 | 1.00 |
| | 1 | 2 | 3 | 4 | 5 |

- No obstante, esta es una primera aproximación al estudio de las similitudes y diferencias existentes entre las tablas. A continuación se examinarán diversos resultados que van a poner de manifiesto cuáles son las variables, dentro de cada grupo, responsables de las mismas.
- La siguiente tabla recoge un índice, **cociente entre inercia inter e inercia intra**, que señala, de forma resumida, la similitud o diferencia que presentan los grupos considerados en los diversos ejes de variabilidad globales. Se comprueba que el único eje en el que las tablas manifiestan una cierta estructura similar es en el primero, alcanzando este indicador, para el resto de los ejes, valores parecidos y muy débiles. Este resultado pone de manifiesto que en el análisis comparativo de los cinco grupos considerados, el único eje que refleja una

realidad correspondiente a todas las tablas, esto es, a los cinco puntos de vista considerados, es la dirección de mayor variabilidad, el primer factor. No obstante, ya se ha comprobado que el porcentaje de inercia que se proyecta en los dos siguientes factores es destacable, por lo que van a reflejar similitudes y diferencias de carácter parcial, pero que merecen ser analizadas. Por tanto, y al igual que en el anterior análisis, se va a proceder al estudio del plano factorial definido por los dos primeros factores globales.

RAPPORT: INERTIE INTER/INERTIE TOTALE

AXES 1 A 5

| | RAPPORTS | | | | |
|------|----------|------|------|------|------|
| FAC. | 1 | 2 | 3 | 4 | 5 |
| | 0.76 | 0.35 | 0.38 | 0.28 | 0.25 |

- Llegados a este punto, se va a examinar qué grupos tienen mayor protagonismo en los dos primeros factores globales, que, como ya se ha comentado, son los únicos en los que nos vamos a centrar. En la siguiente tabla se presentan las **correlaciones entre las variables canónicas** (proyección de las nubes parciales) y **los cinco primeros factores globales**. Se comprueba que estas correlaciones son destacables para todos los grupos en el primer eje de dispersión, sobre todo para los grupos que han manifestado tener una estructura con mayores similitudes (grupos 1, 2 y 3). Por tanto, este factor pone de manifiesto una realidad que corresponde, en mayor o menor medida, a todas las tablas analizadas. Esto no ocurre en el resto de los ejes globales, que, como puede observarse, reflejan realidades parciales en el sentido de que no corresponden a la totalidad de los grupos de variables, sino a algunos de ellos, esto es, a los que presentan mayores correlaciones. Por su parte, el segundo factor revela una tipología de las cajas de ahorro españolas relacionada con dos grupos: eficacia, productividad y servicio al cliente (grupo 2) y gestión y rentabilidad (grupo 3). Señalar que este resultado se repite en el cuarto eje, mientras que en el tercero son destacables las correlaciones que presentan los cuatro primeros grupos, siendo bastante débil la correlación asociada al grupo que hace referencia a la inversión crediticia.

CORRELATIONS

| FAC. | 1 | 2 | 3 | 4 | 5 |
|------|------|------|------|------|------|
| GR 1 | 0.95 | 0.31 | 0.74 | 0.29 | 0.34 |
| GR 2 | 0.94 | 0.87 | 0.76 | 0.71 | 0.90 |
| GR 3 | 0.75 | 0.91 | 0.71 | 0.94 | 0.64 |
| GR 4 | 0.95 | 0.38 | 0.87 | 0.34 | 0.50 |
| GR 5 | 0.79 | 0.24 | 0.45 | 0.16 | 0.33 |

- La siguiente tabla aporta información adicional a la previamente analizada, que confirma las conclusiones expuestas. En ella se comprueba, en primer lugar, y a través de las coordenadas de los grupos en los diversos ejes, **la distribución de la inercia** de cada uno de ellos en los distintos grupos. Por otro lado, podemos deducir cuáles son los grupos con mayor contribución a la formación de los diferentes factores. Así, destaca la contribución de los grupos con mayores similitudes, volumen de negocio (grupo 1) y cuenta de resultados (grupo 4), seguidos de los grupos asociados a eficacia y productividad (grupo 2) y a la inversión crediticia (grupo 5), mientras que el grupo asociado a la gestión y rentabilidad (grupo 3), contribuye débilmente a la formación de este primer eje de dispersión. En lo que se refiere al segundo factor global, señalar que son los grupos segundo y tercero los que contribuyen a su formación, de una forma equilibrada.

COORDONNEES

[illegible]

- **Análisis de la nube de puntos-variable (Gráfico 1).** Dada la gran similitud de los planos factoriales (1-2) correspondientes a la nube de puntos-variable de los dos análisis realizados en este trabajo, ACP y AFM, no se ha considerado necesario la inclusión de ambos, por lo que el gráfico 1 sirve de referencia en ambos casos. Las contribuciones que se han analizado en el punto anterior son, como se ha indicado en el desarrollo de la metodología AFM, la suma de las contribuciones de todas y cada una de las variables que integran cada uno de los grupos. Ahora vamos a analizar cuáles son concretamente las variables que, en el seno de cada una de las tablas, más contribuyen a la formación de los ejes. En AFM el peso de los grupos ha sido equilibrado, limitando a 1 la contribución al primer factor. Este hecho tiene gran relevancia ya que las variables del grupo 5, que es el menos numeroso (4 variables) en el ACP tenían una contribución no destacable en los primeros ejes, mientras que ahora en AFM presentan las mayores contribuciones, aún no siendo el grupo con mayor contribución a la formación de este primer factor. El resto de las variables de los demás grupos tienen un papel similar al jugado en el ACP, destacando, sin variaciones dignas de ser mencionadas, las mismas variables. Por tanto, la interpretación realizada en el primer análisis de la tabla única corresponde igualmente a los resultados ahora obtenidos, en el análisis de la tabla múltiple, con la salvedad de incluir a las variables del grupo 5. Esto es, en lo que se refiere al primer factor, éste pone de manifiesto una caracterización global de las entidades de ahorro.
 - **Análisis de la nube de puntos-individuo (Gráfico 3).** En el gráfico 3 aparecen proyectados, sobre los dos primeros ejes parciales, tanto los individuos medios (posición media que ocupan las cajas de ahorro desde los cinco puntos de vista considerados), como los individuos parciales (diferentes posiciones que ocupan las entidades financieras desde cada uno de los puntos de vista), esto es, cada individuo está representado mediante seis puntos. Dado el elevado número de puntos representados en este plano, se ha decidido proyectar la información correspondiente a aquéllos con cierto protagonismo en el mismo. En este apartado vamos a analizar únicamente la proyección de la nube de los individuos medios. En ella se observan ciertas diferencias con respecto al plano correspondiente al ACP que merecen una atención especial y que serán comentadas posteriormente.
1. En lo que se refiere a los dos primeros factores señalar que lo expuesto en el apartado dedicado al ACP se puede recuperar aquí, ya que el primer eje sigue siendo un factor de talla, estableciendo un ranking, de carácter general, entre las entidades de ahorro. Sin embargo, sí merece la atención señalar que este ranking es más completo, ya que en él intervienen variables de todos los grupos considerados, algo que no ocurría en ACP. En el segundo eje, las diferencias entre los resultados de ambos análisis son menos reseñables.

2. Centrando el interés en el primer factor se observa que presenta una diferencia a destacar con el correspondiente al obtenido en el primer análisis: las cajas de ahorro que presentan las dos mayores coordenadas negativas, han intercambiado sus posiciones. Así, en el análisis de la tabla única la Caja de Ahorros y Monte de Piedad de Madrid tenía la mayor coordenada negativa en el primer factor, seguida de la Caixa, mientras que en el análisis ahora estudiado, AFM de la tabla múltiple, la mayor coordenada corresponde a esta última entidad. Esto es debido a que en el primer análisis tienen mayor influencia en la definición del ranking global, como ya se ha visto, las variables relativas a los grupos 1 y 4, es decir, volumen de negocio y cuenta de resultados, respectivamente. Mientras que, en el segundo análisis, la influencia de las cinco tablas estudiadas se ha equilibrado, por lo que intervienen todos los puntos de vista en la definición del ranking correspondiente. Además, no hay que olvidar que en este análisis se está considerando la posición media que las entidades de ahorro ocupan para los cinco grupos.
 3. En cuanto al segundo eje, la posición de las entidades con mayor peso, Caja de Guipúzcoa y San Sebastián y Bilbao Bizkaia Kutxa, en el mismo coincide en los dos análisis. Sin embargo, hay que destacar que la posición de estas dos entidades es más extrema en el enfoque de tabla única, que en el enfoque de tabla múltiple. Las causas y explicaciones de estas posiciones se verán más claramente en el siguiente apartado, cuando se analice la nube de puntos parciales. No obstante, sí que se puede deducir que este resultado es debido a que en el primer enfoque los grupos 2 y 3, eficacia y productividad y gestión y rentabilidad, que son los que más contribuyen a la formación del segundo factor en los dos análisis, tienen un mayor peso.
- **Análisis de la nube de puntos-parciales (gráfico 3).** Este plano en el que aparece cada individuo representado por seis puntos es de gran riqueza interpretativa. En él podemos observar, junto a la posición media relativa que alcanza cada entidad financiera, la posición parcial relativa, esto es, la correspondiente a cada uno de los cinco puntos de vista estudiados en este trabajo, y comprobar que hay determinadas entidades de ahorro con un comportamiento heterogéneo, en el sentido que la posición que ocupan en un determinado aspecto de los estudiados, no corresponde a la posición que presentan en otro u otros. Son muchos los resultados que merecen atención en este plano de puntos-parciales, sin embargo, para no alargar demasiado el trabajo, sólo se van a examinar e interpretar los resultados correspondientes a las entidades de ahorro con mayor peso en los dos primeros factores, es decir, con gran contribución en los mismos. Así:
 1. En el primer eje de dispersión podemos comprobar la posición extrema que ocupa Caja Madrid en los grupos 1 y 4, relativos al volumen de negocio y a la

cuenta de resultados, respectivamente, presentando coordenadas mayores que las correspondientes a la Caixa, que, como ya se ha visto, ocupa la primera posición en el ranking general, correspondiente a la posiciones medias. Este resultado asociado a las entidades mencionadas, coincide con el obtenido en el análisis de la tabla única, dado que los grupos 1 y 4 tenían gran peso en la determinación del primer factor. Además, hay que señalar que Caja Madrid en lo que se refiere a la inversión crediticia (grupo 5) ocupa una posición próxima al origen de coordenadas, reflejando su situación de entidad media en este punto de vista concreto. Sin embargo, su inmediata competidora, la Caixa, es la primera caja de ahorros en este aspecto, con una coordenada extrema, notablemente alejada del resto de cajas.

2. En el segundo eje señalar las posiciones destacadas que ocupan la Caja de Guipúzcoa y San Sebastián y la BBK en los grupos con mayor peso en esta segunda dirección de dispersión, los relativos a eficacia, productividad y atención al cliente y gestión y rentabilidad, grupos 2 y 3 respectivamente. Este resultado pone de relieve que son las entidades que más destacan en los ratios que miden estos dos puntos de vista, sobre todo en las características relacionadas con indicadores por oficina. Señalar que son entidades de tamaño medio, o, de otro modo, que ocupan posiciones medias en el ranking general. Podemos señalar que las dos mayores cajas de ahorro, protagonistas en el primer factor, presentan una situación media en los dos grupos aquí considerados, esto es, son entidades con un nivel de eficacia y productividad y de gestión y rentabilidad muy próximo a la media del conjunto de las cajas de ahorro españolas.
3. Para finalizar, y en relación con el comportamiento medio y parcial que las entidades financieras manifiestan en este análisis, se puede deducir cuáles son las que presentan un comportamiento similar en todos los aspectos considerados en este estudio y cuáles las que presentan un comportamiento muy distinto. Entre las primeras, esto es, entre las entidades homogéneas destacan, básicamente, todas aquellas que se pueden calificar de pequeñas (con un activo total que no supera los 500.00 millones de pts.). No obstante hay que destacar que en este grupo también se puede incluir Unicaja, que, por otro lado, ocupa la novena posición en el ranking obtenido a partir del activo total. Entre las entidades con un comportamiento heterogéneo se encuentran las entidades que podemos calificar de grandes, en el sentido anteriormente señalado, esto es, aquellas que se sitúan a la cabeza del ranking, con gran protagonismo de las cuatro entidades que más contribuyen a la formación de los dos ejes de mayor dispersión, y a las que se ha hecho referencia en los párrafos anteriores.

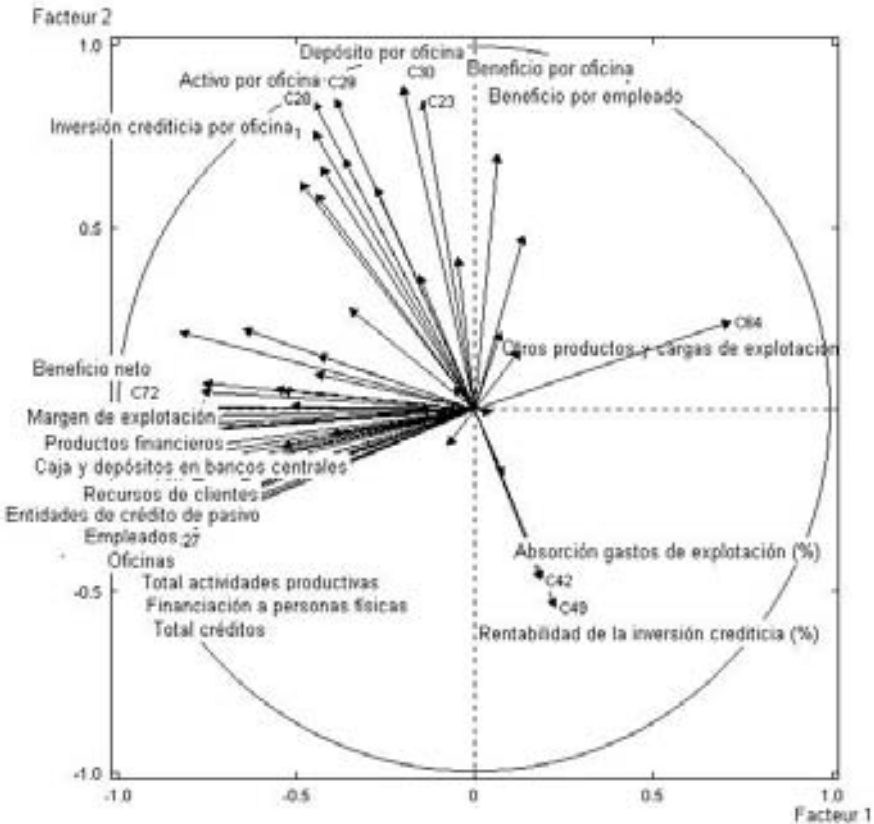
Conclusiones

En este trabajo se ha realizado el estudio de un mismo conjunto de datos desde dos perspectivas distintas, aunque complementarias: tabla única y tabla múltiple. Una vez analizados con detalle los resultados obtenidos en ambos estudios, se puede deducir que: ante el estudio de un conjunto de información, referida al mismo colectivo de individuos, susceptible de ser analizada como una sola matriz o como un grupo de subtablas, los objetivos planteados en dicho trabajo deben ser los que determinen, en primer lugar, cuál de las dos posibles estructuras seleccionar. Así, si se desea obtener una visión global, aunque simplificada, de la realidad estudiada, el análisis de la tabla única es suficiente. Si se desea obtener, además, una visión parcial de cada posible submatriz, esto es, de cada uno de los distintos puntos de vista recogidos en los datos, así como una medida de la relación entre estos grupos de variables, será necesario optar por el análisis de la tabla múltiple. En cualquier caso, siempre es aconsejable realizar los dos análisis, e incluso completar con el análisis de cada subtabla por separado.

Para finalizar, señalar que son importantes las ventajas que presenta la metodología del análisis de la tabla múltiple, entre las que destacan:

- Proporciona una visión global de los datos y una medida de la relación entre los distintos grupos que se pueden extraer de ellos. Estas medidas ayudan a tomar decisiones sobre los grupos a eliminar del análisis, y analizarlos por separado, dado su comportamiento específico y distinto al resto.
- Los resultados que se obtienen poseen una gran riqueza interpretativa, ya que un estudio minucioso de los mismos permite detectar la tendencia general de los individuos, los grupos que definen en mayor medida esta tendencia, así como los individuos raros, es decir, aquéllos cuyo comportamiento es diferente según el grupo que los describe. Además, esta tendencia queda definida por el conjunto de los puntos de vista, sin que haya grupos que queden eclipsados por la presencia de algunos con mayor peso inicial.
- Hay conjuntos de datos que pueden estructurarse en grupos de distintas maneras, correspondiendo cada una de ellas a un objetivo diferente. Un ejemplo típico es aquél en el que se observan las mismas variables a lo largo del tiempo: un grupo puede estar constituido por la misma variable en diferentes períodos o por el conjunto de las variables observadas en el mismo período. El análisis de las dos estructuras de datos dan resultados distintos, pero complementarios.

Gráfico 1 (ACP): Plano Factorial (1-2) de la nube de puntos-variable



Nota:
El punto “Fondo para Provisión de Insolvencias” está debajo de “Caja y Depósitos en Bancos Centrales”
El punto “Cartera de Valores de Renta Fija” está debajo de “Entidades de Crédito Pasivo”
El punto “Dotación para Obras de Beneficio Social” está debajo de “Margen de Explotación”
El punto “Activo por Empleado” está debajo de “Activo por Oficina”
El punto “Depósitos por Empleado” está debajo de “Inversión Crediticia por Oficina”

Anexo

Se adjunta a continuación una explicación¹ de las características utilizadas en este trabajo:

Absorción gastos de explotación: gastos de explotación respecto del margen operacional.

Activos totales: total activo al 31-XII-1997, según circular 4/91 del Banco de España, deducidas pérdidas de ejercicios anteriores.

Aportación productos de servicios: es el porcentaje que representan las comisiones netas sobre el total margen operacional.

Cartera de valores renta fija: suma de los epígrafes de activo, deudas del estado, y obligaciones y otros valores de renta fija.

Cobertura de la inversión: fondo de provisión de insolvencias respecto del total de créditos sobre clientes.

Comisiones netas: diferencia entre comisiones percibidas y pagadas.

Coste por empleado: gastos de personal por empleado.

Costes financieros: epígrafe del debe de la cuenta de pérdidas y ganancias: intereses y cargas asimiladas.

Costes de recursos de clientes: coste financiero de acreedores más costes financieros de empréstitos respecto a recursos de clientes medios.

Cuota de mercado: porcentaje de los recursos de clientes de cada caja respecto del total.

Depósitos por empleado: recursos de clientes por empleado.

Depósitos por oficina: recursos de clientes respecto al número de oficinas.

Gastos de explotación: gastos de personal más otros gastos administrativos.

Índice de cobertura de morosidad: fondo de provisión de insolvencias respecto del volumen de morosos.

Índice de economicidad: beneficio neto respecto de productos totales.

Índice de morosidad: volumen de morosos respecto del riesgo total.

Inversión certificada neta: epígrafe del activo: créditos sobre clientes.

Margen de explotación: margen ordinario, menos gastos de explotación, menos amortización y saneamientos de activos materiales e inmateriales, más otros productos y cargas de explotación.

Margen financiero: diferencia entre productos financieros y costes financieros.

Margen operacional: suma del margen financiero y las comisiones netas.

Margen ordinario: margen operacional más resultado neto de operaciones financieras.

Productos financieros: suma de los epígrafes del haber de la cuenta de pérdidas y ganancias: intereses y rendimientos asimilados y rendimientos asimilados y rendimiento de la cartera de renta variable.

Productos totales: total del haber de la cuenta de pérdidas y ganancias, deducidas las pérdidas del ejercicio.

Recursos de clientes: suma de los epígrafes de pasivo: débitos a clientes, débitos representados por valores negociables y pasivos subordinados.

Recursos propios: son los recursos computables con arreglo a la circular 5/1993 del Banco de España.

Rentabilidad activos totales: beneficio neto respecto activo total medio.

Rentabilidad de la inversión certifica: productos de la inversión certifica respecto a la inversión certifica media.

Rentabilidad de los recursos propios: beneficio neto respecto de los recursos propios medios.

Resultados extraordinarios: beneficios menos pérdidas extraordinarias.

Resultado neto de operaciones financieras: diferencia entre beneficios y pérdidas por operaciones financieras.

Riesgo total: inversión crediticia bruta más riesgos de firma.

Saneamiento del inmovilizado financiero: saneamiento de inmovilizaciones financieras menos fondos de saneamiento de inmovilizaciones financieras disponibles.

Volumen de morosos: deudores en mora, litigio o de dudoso cobro, se incluyen los créditos de firma de dudosa recuperación.

Composición de los grupos:

Grupo 1: Volumen de Negocio

- | | |
|--|---|
| 1. Recursos de Clientes | 6. Cartera de Valores de Renta Fija |
| 2. Fondo provisión de insolvencias | 7. Cartera de Valores de Renta variable |
| 3. Cajas y Depósitos en Bancos centrales | 8. Participaciones en Empresas Grupo |
| 4. Entidades de Crédito de Activo | 9. Inmovilizado Material |
| 5. Entidades de Crédito de Pasivo | 10. Recursos Propios |

Grupo 2: Eficacia, Productividad y Servicio al Cliente

- | | |
|---------------------------------------|---|
| 11. Empleados | 21. Beneficio por Oficina |
| 12. Activo por Empleado | 22. Inversión Crediticia por Oficina |
| 13. Depósitos por Empleado | 23. Cajeros Automáticos |
| 14. Beneficio por Empleado | 24. Tarjetas Emitidas |
| 15. Inversión Crediticia por Empleado | 25. Total Ingresos por Tarjeta |
| 16. Empleados por Oficina | 26. Terminales y Ordenadores Personales |
| 17. Coste por Empleado | 27. Datáfonos |
| 18. Oficinas | 28. Número de Cuentas (Impositores) |
| 19. Activo por Oficina | 29. Cifra media Depósito/Impositor |
| 20. Depósito por Oficina | 30. Número de Clientes |

Grupo 3: Gestión y Rentabilidad

- | | |
|---|---|
| 31. Productos Totales | 38. Índice Cobertura Morosidad (%) |
| 32. Productos Inversión Certifica | 39. Coste Recurso de Clientes |
| 33. Absorción Gastos de Explotación (%) | 40. Rentabilidad de la Inversión Crediticia |
| 34. Aportación Productos de Servicio | 41. Rentabilidad Recursos Propios (%) |
| 35. Cobertura de la Inversión Certifica (%) | 42. Rentabilidad Activo Total (%) |
| 36. Índice de Economicidad (%) | 43. Coeficiente Solvencia (%) |
| 37. Índice Morosidad (%) | |

Grupo 4: Cuenta de Resultados

| | |
|---|--|
| 44. Productos Financieros | 50. Dotac. al Fondo para Riesgos generales |
| 45. Rend. Neto de Operaciones Financieras | 51. Resultados Extraordinarios Netos |
| 46. Otros Productos y Cargas de Explotación | 52. Impuesto de Sociedades |
| 47. Margen de Explotación | 53. Beneficio Neto |
| 48. Amort. y Provisiones para Insolvencias | 54. Dotación para Obra Beneficio Social |
| 49. Saneamiento del Inmovilizado Financiero | |

Grupo 5: Inversión Crediticia por Sectores

| |
|-------------------------------------|
| 55. Total Actividades productivas |
| 56. Financiación a Personas Físicas |
| 57. Sin Clasificar |
| 58. Total Créditos |

Bibliografía

- ALUJA, T y MORINEAU, A. (1999): *"Aprender de los Datos: El Análisis de Componentes Principales. Una Aproximación desde del Data Mining"*. EUB. Barcelona
- DAZY, F. y LE BARZIC, J.F. (1996): *"L'analyse des Données Evolutives. Méthodes et Applications"*. Editions Technip. Paris
- ESCOFIER B., PAGES J. (1992): *"Análisis Factoriales Simples y Múltiples: Objetivos, Métodos e Interpretación"*. Universidad del País Vasco. Bilbao.
- LANDALUCE M.I. (1995): *"Estudios de la Estructura de Gasto Medio de las Comunidades Autónomas Españolas: Una Aplicación del Análisis Factorial Múltiple"*. Thesis. Universidad del País Vasco.
- LEBART, L., MORINEAU, A. y PIRON, M. (1997): *"Statistique Exploratoire Multidimensionnelle"*. Dunod. Paris
- SPAD-version 3.21 (1997): *Logiciel diffusé par CISIA*. 1 av. Herbillon 94160 Saint-Mandé.