

ACM y Statis dual ponderado. Dos técnicas complementarias para analizar una visión de la cultura de la Universidad(1)

por

K. FERNÁNDEZ AGUIRRE,
J. I. MODROÑO HERRÁN

Departamento de Economía Aplicada III
(Econometría y Estadística)
Universidad del País Vasco

M. I. LANDALUCE CALVO
Departamento Economía Aplicada
Universidad de Burgos

RESUMEN

Para realizar un análisis exploratorio de un conjunto de datos de naturaleza cualitativa estructurados en distintas tablas se pueden plantear varias alternativas. La más tradicional es la aplicación de un Análisis de Correspondencias Múltiples (ACM), bien actuando una de las tablas como activa y el resto como ilustrativas o mediante el uso de variables ilustrativas. En este trabajo se propone como alternativa una extensión de la metodología STATIS que va a permitir el análisis

(1) Este trabajo ha sido financiado por el Proyecto de Investigación PB98-0149 de la Dirección General de Enseñanza Superior del Ministerio de Educación y Cultura, por el grupo de Investigación Consolidado UPV 038.321-13631/2001 y por el Proyecto de Investigación FIT-080000-2001-0063

de varias tablas activas de datos nominales. Para ello, se parte de las tablas disyuntivas completas asociadas a cada tabla inicial y se ponderan, de forma adecuada, las modalidades asociadas a las variables cualitativas consideradas.

Palabras clave: variables cualitativas, tablas múltiples, ACM y STATIS

Clasificación AMS: 62H25

1. INTRODUCCIÓN

Para analizar, desde un punto de vista exploratorio, un conjunto de datos de naturaleza cualitativa que, en función de determinados objetivos, pueda interesar estructurar en distintas tablas (por ejemplo, hombres y mujeres, parados y no parados, etc.) se pueden plantear varias alternativas. Quizás la más tradicional sea el uso del Análisis de Correspondencias Múltiples (ACM), (Escofier *et al.*, 1992, Greenacre, 1984). En este caso, existen dos posibilidades, por un lado, una de las tablas actúa como activa y el resto, necesariamente, tiene que intervenir en el análisis con carácter ilustrativo y, por otro lado, introducir la variable discriminadora como ilustrativa y ver su relevancia. No obstante, en los últimos años, han proliferado diversas técnicas que permiten el análisis de tablas múltiples, esto es, métodos que permiten que todas las tablas tengan carácter activo en el estudio, enriqueciendo los objetivos del mismo, pudiendo ser uno de estos métodos otra alternativa. Centrándonos en la Escuela Francesa de Análisis de Datos se pueden mencionar métodos como STATIS, Doble Análisis de Componentes Principales (DACP), Análisis Factorial Múltiple (AFM), (Dazy *et al.*, 1996), Análisis Simultáneo de tablas de contingencia, (Goitisoló, 2001 y Zárraga y Goitisoló, 2000). Sin embargo, no todas estas técnicas han sido desarrolladas para el análisis de tablas de naturaleza cualitativa. Así, la metodología STATIS, (L'Hermier des Plantes, 1976), permite el análisis de distintos conjuntos de tablas (referidas a los mismos o a distintos individuos y a las mismas o a distintas variables) pero únicamente de naturaleza cuantitativa.

En este trabajo se plantea una extensión de la metodología STATIS que va a permitir el análisis de varias tablas de datos nominales. Para ello, se parte de las tablas disyuntivas completas asociadas a cada tabla inicial y se ponderan, de forma adecuada, las modalidades asociadas a las variables cualitativas consideradas. La ponderación va a posibilitar el tratamiento como variables cuantitativas de las variables iniciales de naturaleza cualitativa. Además, se analizan los resultados que ambos métodos proporcionan en un estudio, presentado en este trabajo con carácter meramente ilustrativo, sobre las características dominantes de la cultura

en la universidad del País Vasco en la actualidad. Los datos proceden de una encuesta realizada a profesores de la citada universidad por un grupo multidisciplinar de trabajo en el ámbito de un proyecto de investigación, en fase de realización, sobre gestión del conocimiento en la universidad. El análisis se ha realizado íntegramente con el programa SPAD, versión 4.5.

2. METODOLOGÍA

En esta sección se van a presentar las dos técnicas objeto de estudio en este trabajo. La presentación del Análisis de Correspondencias Múltiples (ACM) se hará de forma breve ya que se trata de una metodología de análisis de datos más conocida y sobre la que existe numerosa literatura para consultar. Sin embargo, la presentación de la metodología STATIS será más detallada, ya que, por una parte, se trata de un método más novedoso y, por otra parte, este trabajo se centra en el desarrollo de una extensión de dicha metodología para el análisis de variables nominales.

2.1 Análisis de Correspondencias Múltiples (ACM)

El Análisis de Correspondencias Múltiples (ACM) permite estudiar un conjunto de individuos descritos por un grupo de variables cualitativas. Para transcribir numéricamente el conjunto de estos datos son posibles varias codificaciones. Por una parte, los datos pueden estar recogidos en una tabla del tipo “Individuos x Variables”, los primeros en filas y las segundas en columnas, en la intersección de una fila y una columna aparece un valor (codificación condensada) que no posee propiedad numérica alguna. Por otra parte, con estos mismos datos se puede construir una Tabla Disyuntiva Completa (TDC), en la que las filas representan a los individuos y las columnas a las modalidades de las variables (denominadas indicadoras de la modalidad) y en la intersección de una fila y una columna sólo puede aparecer un 1 o un 0 según que el individuo posea la modalidad o no, respectivamente. Dados los objetivos de este estudio vamos a adoptar esta última forma de codificación, es decir, vamos a trabajar sobre la TDC.

Los objetivos del ACM pueden ser presentados a partir de las tres familias de objetos que intervienen en el mismo: los individuos, las variables y las modalidades de las variables.

⇒ *Estudio de los individuos.* Uno de los objetivos del ACM es realizar una tipología de los individuos. Esta tipología debe basarse en una noción de semejanza tal que dos individuos están tanto más próximos cuanto mayor es el número de modalidades que poseen en común.

⇒ *Estudio de las variables*. Consiste en resumir el conjunto de variables nominales por un pequeño número de variables numéricas. El interés de estas variables sintéticas proviene de que están relacionadas con el conjunto de variables estudiadas.

⇒ *Estudio de las modalidades*. Consiste en establecer un balance de sus semejanzas. La semejanza entre dos modalidades debe basarse en su asociación mutua: dos modalidades se parecen tanto más cuanto mayor es su presencia o ausencia simultánea en un gran número de individuos.

2.1.1 *Análisis de la Tabla Disyuntiva Completa Ponderada*

Una variable cualitativa se puede considerar como una partición del conjunto de individuos. Esta variable está representada por el conjunto de las variables indicadoras de las clases de esta partición o por el subespacio que engendran. Este subespacio tiene como dimensión el número de modalidades, ya que las variables indicadoras correspondientes a una misma variable son ortogonales entre sí. Es, por tanto, el subespacio de funciones numéricas que toman el mismo valor para los individuos que han elegido la misma modalidad. Así consideradas, un conjunto de variables cualitativas está codificado bajo la forma de una TDC a la que tradicionalmente se aplica un ACM.

Se puede comprobar que los resultados obtenidos al aplicar un ACM son equivalentes a los obtenidos a partir de un Análisis de Componentes Principales (ACP) normado de la TDC cuando se han ponderado las variables indicadoras por la proporción de individuos que no las han elegido. Esto es, siendo n el número total de individuos estudiados y n_i el número de individuos que han elegido la modalidad i -ésima de una determinada variable cualitativa considerada, la ponderación de la

variable indicadora asociada a dicha modalidad vendrá expresada por $\frac{n - n_i}{n}$. La equivalencia entre ACP Normado de la TDC ponderada y ACM aparece demostrada en Escofier y Pagès (1990) y Landaluce (1997), destacando, entre otros, los siguientes resultados:

1. La nube de modalidades en ACM y la nube de indicadoras ponderadas en ACP tienen las mismas propiedades inerciales.

2. Existe equivalencia entre las operaciones de centrado del ACM y del ACP cuando se efectúan sobre las mencionadas nubes.

3. Las matrices analizadas en ACM y ACP normado de la TDC (matriz de correlaciones) son equivalentes, excepto por un coeficiente, el inverso del número de variables cualitativas analizadas.

4. Las distancias entre individuos y entre modalidades definidas en ACP y ACM mantienen una estrecha relación.

5. En conclusión, se puede afirmar que los dos análisis proporcionan los mismos factores.

2.2 Metodología STATIS

El método STATIS para el análisis de tablas múltiples de naturaleza cuantitativa, introducido por Y. Escoufier y L' Hermier Des Plantes en 1976 y desarrollado por C. Lavit, Dazy et al. (1996), se presenta en tres versiones distintas. Los principios de funcionamiento de todas ellas, como se va a exponer a continuación, se basan en conceptos relativamente simples de geometría euclídea.

Versión 1: Analyse triadique de Jaffrenou: permite el análisis simultáneo de matrices de datos referidas a los mismos individuos (filas) y a las mismas variables (columnas). Para ello, se calculan las distancias euclídeas entre las tablas de datos, lo que supone considerar una tabla como un vector (yuxtaposición de columnas) y utilizar el producto escalar habitual entre vectores. A partir de la matriz de productos escalares entre tablas, se construye una nube en la que cada punto corresponde a una tabla.

Versión 2: STATIS: permite el análisis simultáneo de diferentes tablas numéricas referidas a las mismas o a distintas variables (columnas) y a un determinado conjunto de individuos (filas). En esta versión, se comparan las nubes de filas por medio de las matrices de productos escalares centrados. Supone, además, que la ponderación de los individuos es uniforme y que la distancia entre dos cualesquiera corresponde a la distancia euclídea clásica entre dos filas de una tabla. A partir de las matrices de productos escalares, asociadas a cada tabla, se obtiene una matriz compromiso, como media ponderada de las anteriores. Esta matriz compromiso es una matriz de productos escalares entre las filas, a partir de la cual se construye la nube compromiso que mejor resume las posiciones de los individuos. Este procedimiento proporciona una medida de relación entre tablas que coincide con el coeficiente RV de Y. Escoufier.

Versión 3: STATIS DUAL: permite el análisis simultáneo de distintas matrices numéricas que recogen información sobre las mismas variables (columnas), medidas en diferentes conjuntos de individuos (filas). En esta versión, se comparan las nubes de variables normadas por medio de las matrices de correlación. Los productos escalares entre matrices de correlación definen una configuración de diversos puntos, en la que cada uno de ellos representa a una de las matrices (nubes de puntos). Como en la versión anterior, se busca una matriz compromiso, próxima a todas las matrices de correlación. Esta se define como una media ponderada de

estas matrices, siendo, por tanto, una matriz de correlación. A partir de esta matriz compromiso se construye la nube compromiso de variables.

Todas las distancias vistas en cada una de las versiones se descomponen para poder, por un lado, ser interpretadas. En este sentido, si se considera el conjunto de las tablas, la suma de los cuadrados de las distancias entre cada par de tablas es una medida de la dispersión de los puntos en la nube global. Esta cantidad se descompone según los individuos y según las variables, dependiendo de cada opción, dando estas contribuciones una explicación global de la dispersión y de los principales responsables de la misma. Por otro lado, la descomposición de estas distancias permite obtener la nube compromiso, entre todas las nubes estudiadas, siempre que ello sea posible.

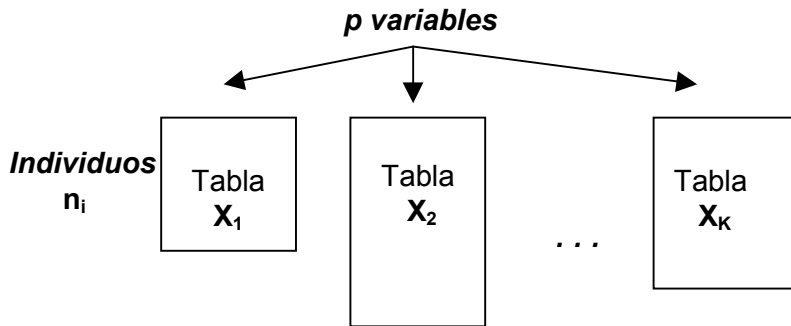
La metodología STATIS proporciona, además, en todas las versiones, y a partir de las matrices compromisos, una serie de resultados bajo la forma de nubes de puntos que serán explotadas de manera gráfica, a través de planos factoriales que (a diferencia de lo que ocurre en análisis en componentes principales y en análisis de correspondencias) no pasan forzosamente por el centro de gravedad de la nube.

La ponderación que, en ocasiones, utiliza este método no equilibra la influencia de las diferentes tablas, sino que asigna mayor peso a aquéllas que presentan una estructura similar a la estructura común, penalizando, en cierto sentido, al resto.

2.2.1 STATIS DUAL

En este trabajo planteamos con carácter complementario al ACM un método que permita el análisis de varias tablas de datos de naturaleza cualitativa referidas a las mismas variables y a distintos individuos. Además, proponemos trabajar con la TDC ponderada de manera adecuada (como se ha mencionado en la sección 2.1.1) para que sea susceptible de ser analizada como si de una tabla cuantitativa se tratara, a través de las correlaciones entre las variables indicadoras ponderadas asociadas a las diversas modalidades. Por todo ello, se puede comprobar que la versión de la metodología STATIS, mencionada anteriormente, más apropiada, dados los objetivos perseguidos, es la tercera, correspondiente a STATIS DUAL. En esta sección se va a hacer una presentación más detallada de dicha técnica, (Dazy *et al.*, 1996).

Figura 1
 TABLAS DE DATOS EN STATIS DUAL



Partiendo de K tablas de datos X_k ($k= 1, \dots, K$) referidas a las mismas p variables (columnas) y a distintos conjuntos de individuos n_i ($i= 1, \dots, k$) (filas), el método consiste en comparar las K nubes de variables y se desarrolla, básicamente, en las siguientes etapas:

1. *Estudio de la interestructura.* Es el estudio de las relaciones entre diferentes tablas. Consiste en una comparación global de la estructura de las matrices de datos. Para ello, es necesario definir una distancia entre dos tablas y encontrar una representación gráfica de las mismas en la que la proximidad entre dos puntos corresponda a una semejanza entre tablas, en el sentido considerado.

Siendo V_k ($k= 1, \dots, K$) las matrices de correlación asociadas a las tablas iniciales, la distancia al cuadrado entre dos de estas matrices se define:

$$d^2(\mathbf{V}_k, \mathbf{V}_{k'}) = \sum_{j=1}^p \sum_{j'=1}^p ((\mathbf{V}_k)_{jj'} - (\mathbf{V}_{k'})_{jj'})^2.$$

Distancia que puede descomponerse en contribuciones debidas a las variables.

Esto es, para dos matrices cualesquiera la cantidad $\sum_{j=1}^p ((\mathbf{V}_1)_{jj'} - (\mathbf{V}_2)_{jj'})^2$ representa el cuadrado de la distancia entre la columna j de V_1 y la columna j de V_2 . La columna j de V_1 contiene las correlaciones entre la variable j -ésima y las $p-1$ restantes (posición de dicha variable en la nube de variables). La cantidad $\sum_{j=1}^p ((\mathbf{V}_1)_{jj'} - (\mathbf{V}_2)_{jj'})^2$ mide la modificación de las relaciones lineales de la variable

j -ésima con el resto, entre las dos tablas. Se expresa en términos de contribución, como porcentaje del cuadrado de la distancia entre las dos matrices de correlación. Si se considera el conjunto de las K matrices de correlación, la suma de los cuadrados de las distancias entre cada pareja $(V_k, V_{k'})$ se descompone según las variables:

$$\sum_{k=1}^K \sum_{k'=1}^K \left[\sum_{j=1}^p (V_k)_{jj} - (V_{k'})_{jj} \right]^2 .$$

Esta cantidad (que suele expresarse en porcentaje) mide la parte de la dispersión entre los puntos de la nube de la interestructura debida a la variable j -ésima.

Para obtener una representación gráfica de la nube de la interestructura (nube de las matrices de correlación, en la que cada punto representa a una matriz) se diagonaliza la matriz W de productos escalares entre las matrices V_k :

$$W = \begin{bmatrix} \ddots & & \ddots \\ & \langle V_k | V_{k'} \rangle_{HS} & \\ \ddots & & \ddots \end{bmatrix} .$$

2. *Búsqueda de un compromiso.* Esta etapa tiene como objetivo resumir todas las tablas de datos en una sola V , denominada “compromiso”, que deberá ser representativa del conjunto de matrices. Esta matriz V se calcula como una media

ponderada de las K matrices iniciales de la siguiente manera:

$$V = \sum_k \beta_k V_k .$$

Con $\beta_k = u_1(k) / \lambda_1$, siendo u_1 el vector propio normado asociado al mayor valor propio λ_1 del análisis factorial de la matriz W y $u_1(k)$ su k -ésima componente.

3. *Estudio de la intraestructura.* La búsqueda de los vectores y valores propios de la matriz compromiso proporciona, a través de las proyecciones sobre los ejes principales, una representación gráfica compromiso aproximada de las variables. Además, considerando las variables de cada tabla como elementos suplementarios se pueden obtener las trayectorias de dichas variables en esta misma gráfica.

3. ESTUDIO EMPÍRICO

En esta sección se presentan e interpretan los resultados obtenidos en un análisis empírico realizado, entre otros, con el fin de servir de ilustración al principal objetivo de este trabajo: proponer la metodología STATIS Dual de la TDC ponderada como técnica factorial complementaria al ACM para el análisis de tablas cualitativas.

Los datos proceden de una encuesta realizada por un grupo multidisciplinar de trabajo en el ámbito de un proyecto de investigación, en fase de realización, sobre gestión del conocimiento en la universidad(2). Este grupo de trabajo aborda uno de los objetivos del proyecto: el estudio de las características dominantes de la cultura en la universidad del País Vasco en la actualidad.

De dicha encuesta se ha seleccionado la siguiente pregunta que da lugar a 9 variables con una escala de diferencial semántico.

(2) El proyecto dirigido por el Dr. Arturo Rodríguez Castellanos se titula: "Gestión del conocimiento en una universidad pública: El proceso de Investigación-Desarrollo-Transferencia del Conocimiento Científico Técnico". La encuesta ha sido diseñada por el grupo de trabajo dirigido por el Dr. Sabino Ayestarán Echeverría.

La cultura de la Universidad es:

Rígida	1	2	3	4	5	Flexible
Abierta hacia el exterior	1	2	3	4	5	Centrada en la propia Universidad
Individualista	1	2	3	4	5	Trabajo en equipo
Innovadora	1	2	3	4	5	Tradicional
Planificada	1	2	3	4	5	Imaginativa
Cooperación	1	2	3	4	5	Rivalidad
Resultados a largo plazo	1	2	3	4	5	Resultados a corto plazo
Valoración de la investig.	1	2	3	4	5	Indiferencia hacia la investigación
Investigación básica	1	2	3	4	5	Investigación Aplicada

Además, se tienen las siguientes variables sobre las características personales de los profesores encuestados, con sus correspondientes modalidades:

- Sexo: *varón – mujer*
- Edad: *menor o igual que 30 años – entre 31 y 40 años – entre 41 y 50 años – entre 51 y 60 años y mayor de 60 años*
- Rango: *Catedrático – Titular – Asociado*
- Participación en proyectos de investigación: *Sí– No*
- Areas de la Ciencia y la Tecnología (UNESCO): *Ciencias exactas y naturales – Ingeniería y Tecnología – Ciencias médicas – Ciencias Agrarias – Ciencias Sociales – Humanidades.*

La muestra inicial fue de 600 encuestados, sin embargo, la muestra finalmente analizada está integrada por 534 individuos después de haber eliminado aquéllos que presentaban un porcentaje de no respuesta superior al 86% en el conjunto de las 9 cuestiones activas. La información obtenida es de naturaleza cualitativa y se ha estructurado, dados los objetivos de este trabajo, en diferentes tablas de individuos x variables. Las características de cada una de estas matrices de datos, así como la metodología utilizada en cada caso han sido:

⇒ Una tabla de modalidades (con codificación condensada) de dimensiones 534 x 9, en la que cada columna representa a los diferentes aspectos sobre la cultura en la Universidad propuestos en la encuesta, con cinco modalidades posibles de respuesta. Esta tabla ha sido analizada a través de Análisis de Correspondencias Múltiples (ACM). Para estudiar la posible existencia de diferentes opinio-

nes atendiendo a las diversas características consideradas (sexo, edad, ..) se han introducido 5 columnas suplementarias con sus correspondientes modalidades.

⇒ Una tabla disyuntiva completa (TDC) ponderada de dimensiones 534 x 45, en la que cada columna representa a las distintas modalidades de respuesta (de 1 a 5) de las 9 alternativas propuestas en la encuesta. A la vista de los resultados del anterior análisis (ACM) y con el fin de estudiar la existencia de una diferente visión de la cultura en la universidad atendiendo al área de la ciencia y la tecnología a la que pertenecen los encuestados se ha considerado esta tabla inicial como una yuxtaposición (en filas) de las siguientes cinco subtablas: Tabla 1, asociada a Ciencias exactas y naturales de dimensiones 118 x 45, Tabla 2 asociada a Ingeniería y Tecnología de dimensiones 103 x 45, Tabla 3 asociada a Ciencias médicas de dimensiones 79 x 45, Tabla 4 asociada a Ciencias sociales de dimensiones 162 x 45 y Tabla 5 asociada a Humanidades de dimensiones 72 x 45. Como puede comprobarse no hay ninguna tabla asociada al área de Ciencias Agrarias, debido a que no había ningún efectivo. Además, para poder analizar esta tabla múltiple como tabla cuantitativa, a través de las matrices de correlación de las variables indicadoras, cada una de las columnas ha sido ponderada por la proporción de individuos que no han elegido la modalidad a la que está asociada (como ya se indicó en la sección 2.1.1). Dado que finalmente se obtiene una tabla múltiple cuantitativa formada por cinco subtablas referidas a las mismas variables y a diferentes individuos, se ha elegido la metodología STATIS DUAL como la más apropiada para su análisis.

A continuación se exponen y analizan los principales resultados obtenidos en los análisis de estas dos tablas. Teniendo en cuenta que se trata de un estudio ilustrativo no se va a realizar una interpretación en profundidad.

3.1 Resultados del Análisis de Correspondencias Múltiples (3)

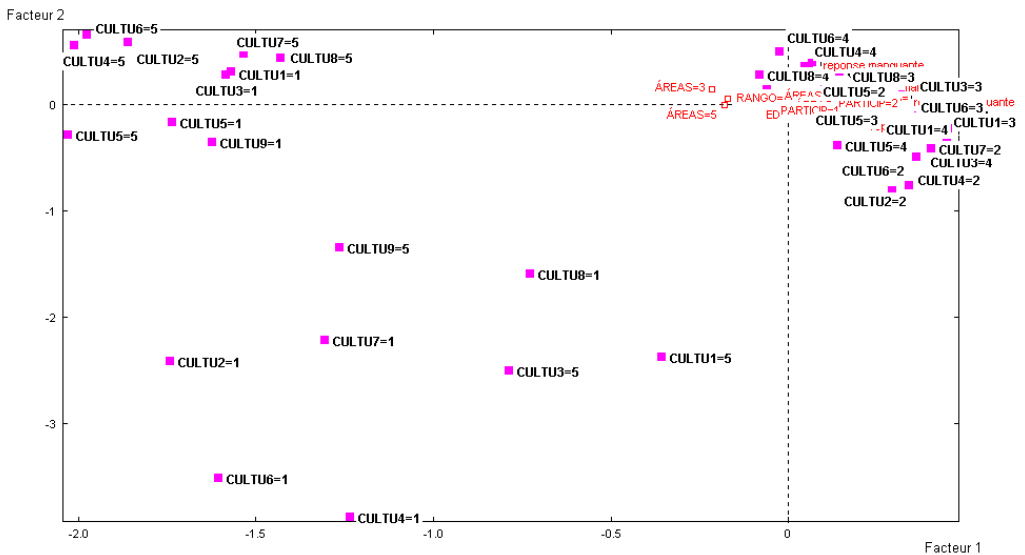
Con el fin de no hacer demasiado extensa la presentación de los resultados de este análisis, a continuación se exponen únicamente los planos correspondientes a los tres primeros ejes factoriales, en los que aparecen proyectadas las 45 modalidades de respuesta consideradas en este estudio. Estos ejes representan en conjunto un 23.11% de la inercia (dispersión) total(4), reflejando el primer eje un

(3) Los nueve aspectos considerados en la encuesta, sobre la cultura de la universidad, aparecen en los planos presentados con las etiquetas CULTU1 hasta CULTU9. El número que aparece a continuación corresponde a la modalidad, de 1 a 5.

(4) Es conocido que las tasas de inercia proyectadas en los diferentes ejes del ACM son bajas, dando una idea pesimista del análisis efectuado. Se pueden recalcular las inercias proyectadas iniciales aplicando una determinada corrección (propuesta por Benzécri), pero en este estudio empírico no se ha considerado necesaria su aplicación.

10.02% de variabilidad total, un 7.19% el segundo eje y un 5.90% el tercero. La interpretación de estos ejes se realiza en base a las coordenadas, contribuciones absolutas (o participación de cada una de las modalidades en la creación de los distintos factores) y contribuciones relativas (o fidelidad de representación de las modalidades en cada eje), destacando en cada uno de ellos únicamente aquellas modalidades que tengan los mayores valores en estos indicadores señalados.

Figura 2
PLANO FACTORIAL 1-2 DEL ACM



En el gráfico construido con los dos primeros ejes factoriales (figura 2) se puede observar el protagonismo que adquieren las modalidades extremas (1 y 5) en la totalidad de los aspectos considerados sobre la cultura de la universidad, modalidades que se reparten entre el segundo y tercer cuadrante. Hay que señalar que son las modalidades menos frecuentes, esto es, seleccionadas por el menor número de individuos. Indicar, asimismo, que son respuestas asociadas a encuestados que pertenecen a las áreas de Ciencias Médicas y Ciencias Sociales, respectivamente, siendo estas últimas las únicas modalidades ilustrativas que presentan una cierta significatividad en el plano analizado.

se trabaja en equipo, es innovadora, imaginativa, hay cooperación, los resultados son a largo plazo y se valora la investigación.

Para finalizar la interpretación de los resultados obtenidos en esta aplicación del ACM señalar que no se han mencionado las variables ilustrativas, relativas a la caracterización de los encuestados, dado que ninguna de ellas resultaba significativa en los ejes analizados (sus valores test no superan en valor absoluto el 2, para un nivel de significación del 5%).

3.2 Resultados del STATIS Dual

El análisis de la tabla disyuntiva completa debidamente ponderada (ver 2.1.1), considerada como una tabla múltiple de naturaleza cuantitativa, formada por la yuxtaposición de las subtablas correspondientes a las cinco áreas de la Ciencia y la Tecnología a la que pertenecen los docentes que han colaborado en la encuesta analizada, a través de la metodología STATIS Dual ha proporcionado, entre otros, los siguientes resultados:

PRODUCTOS ESCALARES ENTRE MATRICES DE CORRELACION

	1	2	3	4	5
1	92.091				
2	66.141	92.015			
3	69.432	68.454	96.940		
4	73.704	68.951	71.924	99.404	
5	73.170	70.113	77.015	76.552	100.291

De la matriz de productos escalares entre las matrices de correlación asociadas a cada una de las subtablas consideradas destacar los elementos de la diagonal principal, normas de cada una de dichas matrices. Señalar que cuanto mayor es la norma, menor es el número de ejes factoriales con un peso importante en la matriz de correlación correspondiente, esto es, más fuerte es la estructura interna de la tabla analizada. En este caso las mayores normas corresponden a las tablas asociadas a las áreas de Humanidades y Ciencias Sociales (Tablas 5 y 4), lo que indica que son los grupos de encuestados con opiniones más homogéneas sobre la cultura de la universidad. Mientras que los grupos 1 y 2, de Ciencias exactas y naturales e Ingeniería y tecnología respectivamente, presentan las menores nor-

mas, esto es, son grupos con un comportamiento interno más heterogéneo, con mayor diversidad en sus opiniones sobre la misma realidad.

DISTANCIAS EUCLIDEAS ENTRE MATRICES DE CORRELACION

	1	2	3	4	5
1	0.000				
2	7.199	0.000			
3	7.083	7.214	0.000		
4	6.640	7.316	7.245	0.000	
5	6.785	7.217	6.573	6.826	0.000

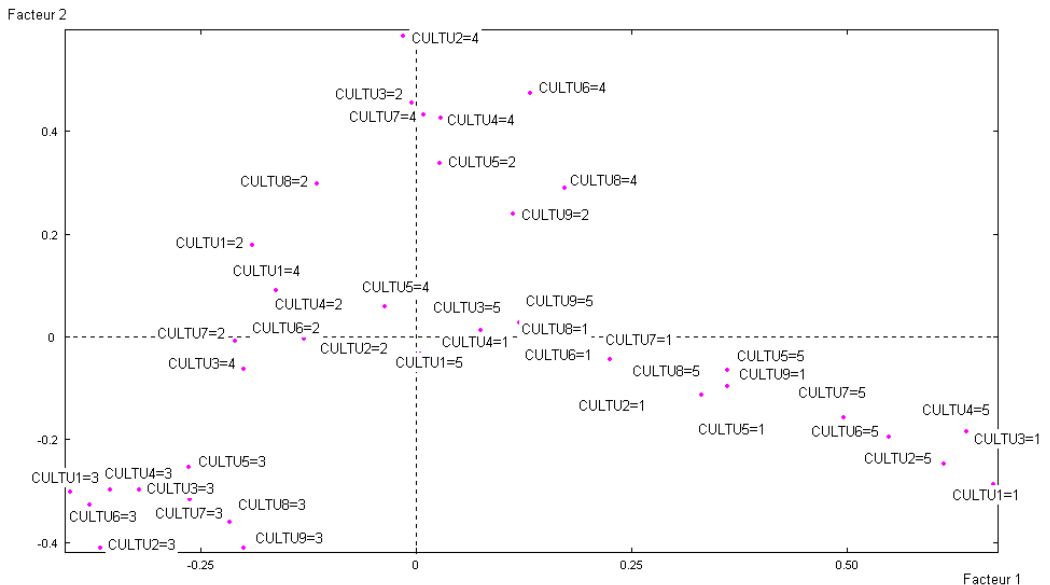
En lo que se refiere a la información que proporciona la matriz de distancias entre las matrices de correlación, destacar que existen ciertas diferencias entre las mismas, por lo que habrá que tenerlo en cuenta y buscar los motivos de tales distancias, esto es, las modalidades (distintas valoraciones en los nueve aspectos considerados) responsables. Destacar que la tabla que presenta las mayores distancias es la asociada al grupo 2, seguida de la tabla 3, áreas de Ingeniería y Tecnología y Ciencias Médicas, respectivamente. Mientras que la tabla asociada al área de Humanidades (tabla 5) es la que presenta las menores distancias. Para analizar los motivos de estas distancias resulta de gran ayuda el estudio de la tabla que recoge una medida, en porcentaje, de la descomposición de la distancia entre las matrices según las diferentes modalidades consideradas (ver tabla en anexo). Se puede comprobar que las nueve características sobre la cultura en la universidad consideradas contribuyen de manera similar a la distancia entre las tablas analizadas, con contribuciones totales (suma de las contribuciones de cada una de las modalidades) que oscilan entre el 10.31% de la primera característica, que hace referencia a la flexibilidad o rigidez, y el 11.88% de la correspondiente a la cooperación o rivalidad. Sin embargo, entre las propias modalidades las diferencias entre sus contribuciones son más notables, así con una contribución superior al 3% destacan: cultura abierta hacia el exterior (1), trabajo en equipo (5), innovadora (1), cooperación (1), resultados a largo plazo (1), valoración de la investigación (1) e investigación aplicada (5), como puede observarse todas ellas modalidades extremas.

A continuación, y centrándonos en el análisis de la intraestructura, se presenta el plano factorial (1,2) (figura 4) correspondiente al análisis de la matriz de correla-

ción compromiso (media ponderada de las matrices de correlación asociadas a cada uno de los grupos estudiados), en el que queda proyectado un 16% de la inercia total. En este plano aparecen proyectadas las 45 variables (indicadoras ponderadas de las modalidades de respuesta) consideradas en esta aplicación. La interpretación de estos ejes se realiza a través del examen de las correlaciones (coordenadas) con las variables iniciales.

Figura 4

PLANO FACTORIAL 1-2 DE STATIS DUAL



Previamente, señalar que el estudio de las ponderaciones asignadas a cada matriz de correlación inicial para la obtención de la matriz compromiso pone de manifiesto un resultado ya obtenido anteriormente, que los grupos 4 y 5 correspondientes a las áreas de Ciencias Sociales y Humanidades son los que más contribuyen a la definición de la matriz de correlación compromiso, esto es, son los grupos con opiniones más homogéneas sobre la cultura de la universidad, los que presentan una estructura interna más fuerte. No hay que olvidar, que esta metodología penaliza, en cierto sentido, a aquellos grupos con estructura interna más débil

(mayor variabilidad), y, por tanto, con un comportamiento menos cercano al comportamiento medio (compromiso).

COEFICIENTES DE PONDERACION

MATRIZ 1	=	0.196
MATRIZ 2	=	0.191
MATRIZ 3	=	0.201
MATRIZ 4	=	0.205
MATRIZ 5	=	0.208

Se puede observar que los coeficientes de ponderación de las matrices de correlación iniciales, a partir de los cuales se define la matriz compromiso, no difieren mucho unos de otros, lo que significa que se ha obtenido un “buen” compromiso.

Sin embargo, también hay que tener en cuenta que el análisis factorial de la matriz de correlación compromiso, dado que se obtiene como media ponderada de las cinco matrices de correlación iniciales y cada una de éstas resume el comportamiento del grupo al que se asocia, va a poner de manifiesto un comportamiento que corresponde a la “media de los comportamientos medios” de los cinco grupos considerados respecto de sus opiniones sobre la cultura de la universidad, teniendo un mayor peso el correspondiente a los docentes de las áreas de Ciencias Sociales y Humanidades. Este es el principal motivo de las posibles diferencias entre los resultados aquí obtenidos y los alcanzados en el anterior análisis a través del ACM.

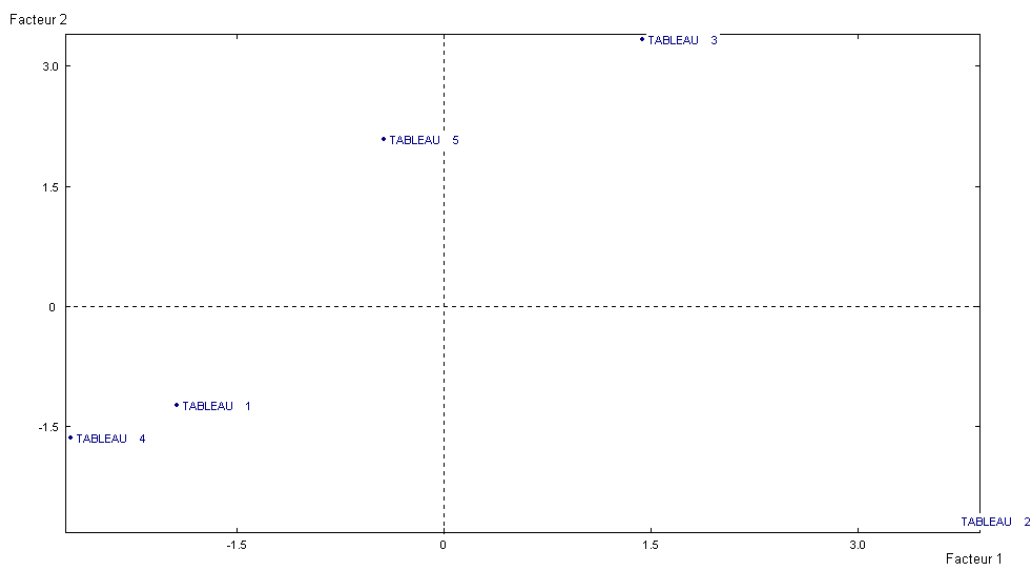
A pesar de todo lo comentado en el párrafo anterior, un estudio detallado de la figura 4 permite vislumbrar los mismos grupos de opinión definidos en el plano factorial 1-3 (figura 3) del ACM, aunque con los ejes rotados. Así, entre el primer y segundo cuadrantes se sitúan las valoraciones 2 y 4 de la totalidad de los 9 aspectos considerados sobre la cultura de la universidad. En el tercer cuadrante se proyectan todas las modalidades correspondientes a las valoraciones medias, 3, mientras que en el cuarto cuadrante encontramos las valoraciones extremas, 1 y 5. Hay que destacar que en STATIS Dual se han obtenido los mismos resultados que en ACM pero con menor número de ejes factoriales. Esto es debido a la pérdida de protagonismo de las modalidades con débil efectivo en la metodología STATIS, mientras que en ACM estas modalidades acaparan el primer eje factorial debido a su gran peso.

Para finalizar este análisis señalar que la metodología STATIS, al considerar la matriz de datos inicial como una yuxtaposición de cinco subtablas, también propor-

ciona un gráfico en el que cada punto representa a cada una de las matrices de correlación asociadas a dichas tablas (figura 5). Este gráfico resume, en cierto sentido, algunos de los resultados antes obtenidos.

Figura 5

PLANO FACTORIAL (1-2) DE LA NUBE DE MATRICES DE CORRELACIÓN



El estudio de este gráfico pone de manifiesto que la tabla 2 asociada al área de Ingeniería y Tecnología es la que se distancia más del comportamiento medio (es la más alejada del origen de coordenadas), esto es, las opiniones de los docentes de dicho área presentan mayor variabilidad. Las tablas 1 y 5 asociadas a los grupos de Ciencias exactas y naturales y Humanidades son las más próximas al centro de gravedad, siendo las áreas con opiniones menos heterogéneas en el seno de su correspondiente grupo de docentes. Además, existe una cierta asociación (comportamiento similar) entre las tablas 1 (Ciencias exactas y naturales) y 4 (Ciencias Sociales).

Hay que señalar que los ejes de este plano factorial (ejes de la interestructura) no son interpretables, en el sentido de que la posición de las tablas en el mismo no se puede explicar en función de las variables iniciales.

4. COMENTARIOS FINALES

En este trabajo se ha puesto de manifiesto la existencia de dos métodos complementarios para el análisis factorial de tablas de datos de naturaleza cualitativa, considerando la posibilidad que habitualmente presentan estas tablas de ser estructuradas en diferentes subtablas (asociadas a las distintas modalidades de una variable nominal, generalmente de caracterización). Los métodos son Análisis de Correspondencias Múltiples (ACM) y una novedosa extensión de la Metodología STATIS Dual para tablas cualitativas.

Los resultados aquí obtenidos permiten deducir las siguientes ventajas y desventajas de una técnica frente a otra, así como las similitudes que ambas presentan:

- En ACM y en STATIS Dual los individuos pierden total protagonismo, lo cual no supone inconveniente alguno ya que en los estudios de datos procedentes de encuestas (especialmente de opinión) lo que realmente interesa son las clases de individuos, clases que están recogidas en las distintas variables indicadoras (modalidades de respuesta). Asimismo, resulta de gran interés el estudio de la relación entre las variables, a través de la relación entre las modalidades, estudio que ambos métodos permiten.

- En STATIS Dual todas las tablas consideradas actúan como activas en el análisis, mientras que en ACM la variable seleccionada para definir las distintas subtablas a considerar debe actuar como ilustrativa (el área de los encuestados en este estudio). Este último método proporciona otra alternativa para el análisis de las diferencias entre los cinco grupos que consiste en analizar uno de ellos como activo y el resto de los grupos como ilustrativos. Sin embargo, esta alternativa no se ha considerado en este trabajo ya que la realidad analizada es distinta a la que se estudia considerando la totalidad de la muestra, de ahí que se presenten los dos estudios con mayor homogeneidad desde el punto de vista de las dimensiones de la tabla activa.

- En la metodología STATIS Dual las denominadas modalidades “raras” o con pocos efectivos pierden el protagonismo que tienen en ACM, por ello los primeros ejes factoriales ponen de relieve una realidad más global que los obtenidos en ACM.

- La metodología STATIS proporciona una medida de las diferencias entre las distintas tablas, lo cual es una gran ayuda en caso de tener un número elevado de colectivos a estudiar, ya que permite analizar si alguno de los colectivos tiene un comportamiento tan particular que merece ser estudiado por separado. El plano factorial global es una media ponderada de las tablas, por lo que la realidad de algunos de los colectivos puede predominar sobre la del resto, aspecto que hay que tener en cuenta, ya que puede ocurrir que el comportamiento de alguno de los colectivos quede totalmente eclipsado en los resultados globales.

- Para finalizar este trabajo, poner especial énfasis en que las dos metodologías aquí presentadas, ACM y STATIS Dual de la TDC ponderada, apropiadas para el estudio exploratorio del mismo conjunto de datos cualitativos, no deben ser consideradas como técnicas alternativas, sino como complementarias, ya que representan dos puntos de vista distintos, aunque con grandes similitudes, para analizar la misma información. La aplicación de ambas técnicas sobre el mismo conjunto de datos proporciona un estudio más exhaustivo de la misma realidad.

REFERENCIAS

- ABASCAL, E., FERNÁNDEZ, K., LANDALUCE, M.I., MODROÑO, J.I. (2001): «Diferentes aplicaciones de las técnicas factoriales de análisis de tablas múltiples en las investigaciones mediante encuestas». *Metodología de Encuestas*. Vol. 3, Núm. 2, pp. 251-279.
- DAZY, F., LE BARZIC J.F. (1996). «L'analyse des Données Evolutives. Méthodes et Applications». Editions Technip. Paris.
- ESCOFIER B., PAGÈS, J. (1992). «Análisis Factoriales Simples y Múltiples: objetivos, métodos e interpretación». Traducción de la 2ª edición. Servicio editorial de la Universidad del País Vasco.
- GOITISOLO, B. (2001). «El Análisis Simultáneo. Propuesta y Aplicación de un Nuevo Método de Análisis Factorial de Tablas de Contingencia». *Tesis Doctoral*. Universidad del País Vasco.
- GREENACRE M. (1984). «Theory and Applications of Correspondence Analysis». *Academic Press*, London
- LANDALUCE, M^a I. (1997). «Análisis factorial de tablas mixtas: nuevas equivalencias entre ACP normado y ACM». *Qüestió*, vol. 21, p. 99-108

- LANDALUCE, M. I., Fernández, K., Modroño, J. I. (1999). «Reflexiones sobre el Uso Comparativo del Análisis Factorial Múltiple (AFM) y de la Metodología Statis para el Análisis de Tablas Múltiples» *Methodologica*, Vol. 7, pp. 37-65.
- LAVIT, C. (1988). «Analyse conjointe de tableaux quantitatifs». Masson. Paris.
- SPAD v4.5 (2000). «Système Portable d'Analyse des Données Numeriques», CISIA, Montreuil, France.
- L'HERMIER DES PLANTES, H. (1976). «STATIS: Structuration de Tableaux À Trois Indices de la Statistique». *Thèse (3c)*. USTL, Montpellier.
- ZÁRRAGA, A. Y GOTISOLO, B. (2000). «Independencia Entre las Cuestiones en el análisis Factorial de Tablas Disyuntivas Incompletas con Preguntas Condicionadas». *Questiío*, 23 (3), pp. 465-488.

ANEXO**CONTRIBUCIONES DE LAS MODALIDADES A LA DISTANCIA
ENTRE MATRICES**

CULTU11	=	2.24 %
CULTU12	=	1.79 %
CULTU13	=	1.42 %
CULTU14	=	1.90 %
CULTU15	=	2.96 %
CULTU21	=	3.49 %
CULTU22	=	1.89 %
CULTU23	=	1.68 %
CULTU24	=	1.68 %
CULTU25	=	1.87 %
CULTU31	=	2.47 %
CULTU32	=	1.58 %
CULTU33	=	1.63 %
CULTU34	=	1.79 %
CULTU35	=	3.45 %
CULTU41	=	4.32 %
CULTU42	=	2.07 %
CULTU43	=	1.68 %
CULTU44	=	1.91 %
CULTU45	=	1.86 %
CULTU51	=	2.59 %
CULTU52	=	1.97 %
CULTU53	=	2.16 %
CULTU54	=	2.46 %
CULTU55	=	2.60 %
CULTU61	=	4.09 %
CULTU62	=	1.80 %
CULTU63	=	1.74 %
CULTU64	=	1.83 %
CULTU65	=	2.42 %
CULTU71	=	3.65 %
CULTU72	=	1.64 %
CULTU73	=	1.69 %
CULTU74	=	1.95 %
CULTU75	=	2.41 %
CULTU81	=	3.04 %

CULTU82	=	1.53 %
CULTU83	=	1.93 %
CULTU84	=	2.08 %
CULTU85	=	2.01 %
CULTU91	=	2.61 %
CULTU92	=	1.96 %
CULTU93	=	1.56 %
CULTU94	=	1.47 %
CULTU95	=	3.17 %

MCA AND WEIGHTED DUAL STATIS.TWO COMPLEMENTARY TECHNIQUES FOR THE ANALYSIS OF A VIEW OF THE UNIVERSITY CULTURE

ABSTRACT

In order to perform an exploratory analysis from a dataset of qualitative variables structured in different tables, several alternatives can be considered. The most traditional is the application of MCA, either letting one of the tables as active and the rest as illustrative or by means of the use of illustrative variables. In this paper it is proposed, as an alternative, an extension of the STATIS methodology which allows for the analysis of several active tables composed of qualitative data. In order to accomplish this, we start from the complete disjunctive tables related to the initial table and then the categories related to the qualitative variables considered are weighted in an appropriate way.

Key words: qualitative variables, multiple tables, MCA and STATIS.

AMS Classification: 62H25