

TECNICAS FACTORIALES DE ANALISIS DE TABLAS MÚLTIPLES: NUEVOS DESARROLLOS EMPIRICOS¹

Abascal Fernández, Elena
Universidad Pública de Navarra

Fernández Aguirre, Karmele
Modroño Herrán, J. Ignacio
Universidad del País Vasco

Landaluce Calvo, M. Isabel
Universidad de Burgos

1. INTRODUCCION

El objetivo de este trabajo es presentar una recopilación de las diversas posibilidades que tienen las técnicas factoriales de análisis de tablas múltiples en las investigaciones de carácter empírico. Entre las técnicas factoriales múltiples se han seleccionado, por su versatilidad, la metodología STATIS y el Análisis Factorial Múltiple (AFM), ambas desarrolladas en el seno de la Escuela Francesa de Análisis de Datos.

Las dos técnicas seleccionadas, dada su naturaleza múltiple, están diseñadas para el análisis de tablas de tres dimensiones. Estas tablas tienen una gran riqueza, desde el punto de vista exploratorio, ya que, además de proporcionar información sobre las relaciones existentes entre las tres dimensiones o conjuntos, se pueden estructurar de forma bidimensional de diferentes maneras según la combinación seleccionada. La diferente estructuración de las tablas corresponde a diversos objetivos de estudio, por lo que los resultados que se desprendan del análisis de estos conjuntos de datos serán seguramente diferentes, pero también complementarios.

Estas metodologías permiten el análisis de diferentes tipos de tablas múltiples. Además, a lo largo de numerosos estudios realizados en los últimos años, hemos descubierto otras posibles aplicaciones, de gran interés en el campo del análisis de datos:

- *La posibilidad de analizar simultáneamente grupos de variables cualitativas, asociados a distintos colectivos de individuos.*
- *Estudiar la estabilidad de los resultados obtenidos en un ACP.*

¹ Este trabajo ha sido financiado por los Proyectos de Investigación PB98-0149 de la Dirección General de Enseñanza Superior del Ministerio de Educación y Cultura y 038.321- HA041/99 de la Universidad del País Vasco.

2. METODOLOGIA

2.1 AFM

El AFM, desarrollado por B. Escofier y J. Pagès , Escofier (1992), es un método que permite el análisis simultáneo de varios grupos de variables medidas sobre el mismo conjunto de individuos equilibrando la influencia de cada grupo.

Los grupos de variables pueden ser diferentes, incluso en la naturaleza y número de variables que los componen. La única condición es que las variables que integran un grupo sean de la misma naturaleza, cuantitativa o cualitativa.

Esta metodología general puede ser muy útil para el estudio de encuestas. Los métodos tradicionales, Análisis en Componentes Principales, ACP, Análisis Factorial de Correspondencias, AFC, y Análisis de Correspondencias Múltiples, ACM, sólo permiten estudiar las relaciones entre las variables de la misma naturaleza, todas cuantitativas o todas cualitativas, lo que restringe su utilización. Sin embargo, en ocasiones, es posible recoger gran cantidad de información que es interesante relacionar, aunque mida aspectos heterogéneos. Esta información puede ser estructurada en diferentes grupos y su análisis mediante el AFM proporciona una visión más rica del fenómeno a estudiar.

Así, por ejemplo, en una misma encuesta, un grupo puede estar definido por las variables que surgen de las preguntas relativas a un aspecto del objeto de estudio, y otro grupo de variables recoger las preguntas que estudian otro aspecto del mismo, sin necesidad de que el número de preguntas sea el mismo, ni que se formulen en el mismo tipo de escala. Un grupo puede estudiarse con una escala tipo Likert o diferencial semántico, mientras que las preguntas del otro pueden ser exclusivamente cualitativas, como la profesión, la zona de residencia, etc.

Otra forma usual de definir los grupos es a través de la variable tiempo, cuando se dispone de un conjunto de variables recogidas en diferentes momentos de tiempo, un grupo puede estar formado por la misma variable observada en diferentes períodos de tiempo o bien por todas las variables observadas en el mismo momento. Las variables de un grupo medidas sobre un conjunto de individuos o encuestados definen una tabla de dos dimensiones. El conjunto de grupos constituye una tabla de tres dimensiones (fig. 1). La única restricción es que el conjunto de individuos (filas de las tablas) sea el mismo en cada grupo.

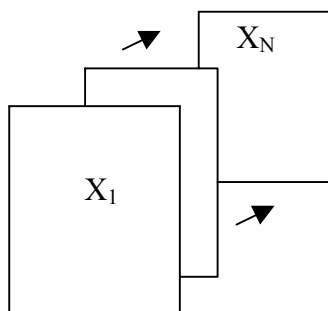


Figura 1

⇒

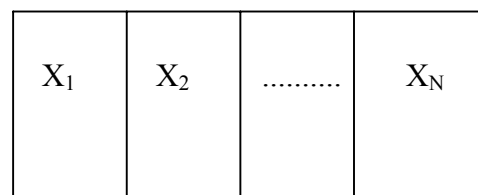


Figura 2

El objetivo del AFM es, como en cualquier análisis factorial, poner de manifiesto los principales factores de variabilidad de los individuos, pero en este caso, estando estos últimos descritos, de manera equilibrada, por los diversos grupos de variables.

El AFM proporciona indicadores, semejantes a los del resto de análisis factoriales, y gráficos que permiten estudiar:

- relaciones entre los grupos, además de medir su grado de semejanza
- relaciones entre las variables de un grupo y las del resto de los grupos
- semejanzas entre los individuos vistos a través de los diferentes grupos de variables.

La metodología del AFM se basa en el ACP. En primer lugar analiza cada grupo de variables o subtabla por separado, mediante un ACP. En la segunda fase analiza la tabla yuxtapuesta, donde cada subtabla ha sido ponderada por el inverso del primer valor propio del ACP. Esta ponderación equilibra el peso de los grupos en la obtención del primer factor. En este sentido, hay que señalar que para las variables continuas, el AFM se comporta como un Análisis en Componentes Principales (ponderando las variables); para las variables nominales se comporta como un Análisis de Correspondencias Múltiples (ponderando las modalidades).

El análisis proporciona las representaciones gráficas siguientes:

- los individuos caracterizados por un grupo de variables o punto de vista: nube parcial
- los individuos caracterizados por el conjunto de variables: nube global
- representación, en un mismo espacio, de los individuos caracterizados por el conjunto de variables y por cada uno de los grupos: nubes parciales y nube media o global
- las variables
- los grupos de variables.

Estos gráficos, junto con las ayudas a la interpretación clásicas del ACP, permiten examinar la existencia de estructuras comunes a todas, o a parte, de las nubes parciales, lo que se reflejaría en una inercia intra mínima o pequeña, o, lo que es lo mismo, en una inercia inter máxima o elevada. Aquéllos individuos cuyos puntos parciales (puntos que representan a cada individuo desde los diferentes grupos) se sitúen próximos (inercia intra débil) ilustran la estructura común de las distintas tablas analizadas. Por el contrario, aquellos individuos con puntos parciales asociados alejados (inercia intra alta) unos de otros, constituyen las excepciones a la estructura común.

Hay que señalar, que este método pone en evidencia factores comunes a todos los grupos, factores comunes a algunos grupos y factores específicos de algunos grupos.

Además, el AFM proporciona medidas globales de relación entre los grupos, basadas en el coeficiente RV de Y. Escoufier. Este coeficiente se obtiene a partir de los coeficientes de correlación lineal entre dos variables cualesquiera. Su valor está comprendido entre 0 (no existe relación entre las variables de los dos grupos considerados) y 1 (las nubes que representan a los grupos son homotéticas). Esta medida es completada con los coeficientes Lg que miden, además, la dimensionalidad

(número de factores de inercia considerable) de cada grupo. Estos coeficientes toman el valor 0 cuando no existe relación entre los grupos y no tienen límite superior.

La presentación completa del método supera los límites de este artículo y el lector interesado puede consultar Lebart (1995) y Escoufier (1992).

2.2 STATIS

El método STATIS, introducido por Y. Escoufier y L' Hermier Des Plantes en 1976 y desarrollado por C. Lavit, 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.

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), como por ejemplo series temporales. 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.
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, al igual que el AFM, proporciona una medida de relación entre tablas que coincide con el coeficiente RV de Y. Escoufier.
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 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, como ocurre en AFM, sino que asigna mayor peso a aquellas que presentan una estructura similar a la estructura común, penalizando, en cierto sentido, al resto. Una presentación completa de la metodología se puede estudiar en Lavit (1988).

3. NUEVOS DESARROLLOS EMPIRICOS

3.1 Análisis simultáneo de grupos de variables cualitativas, asociados a distintos colectivos de individuos

Ninguna de las dos técnicas descritas en la sección anterior permite el análisis simultáneo de un conjunto de grupos de variables de naturaleza cualitativa medidas en distintos colectivos de individuos. A continuación, se presenta una extensión de estas metodologías para solucionar este problema.

3.1.1 Propuesta Metodológica

Se parte de un conjunto de N tablas de naturaleza cualitativa, es decir compuestas por variables cualitativas, de forma que cada tabla tiene, en columnas las mismas K variables, pero, sin embargo, difieren en el número de filas (individuos).

La metodología que se propone consiste, en primer lugar, en efectuar sobre cada una de las N tablas un Análisis de Correspondencias Múltiple (ACM), seleccionando un número determinado de factores principales que, según el porcentaje de inercia que recojan, reflejarán en mejor ó peor medida el comportamiento de las tablas originales. Lógicamente el número de factores a seleccionar deberá representar una buena aproximación de los datos de partida. Las coordenadas de las modalidades sobre estos factores dan una idea de la relación que hay entre las distintas modalidades en los principales ejes ó fuentes de variabilidad. Para cada factor esta relación tendrá una interpretación diferente.

Posteriormente, y teniendo en cuenta que en este tipo de información los individuos no interesan en sí mismos, se construye una tabla que contenga en filas las modalidades y en columnas los factores principales del ACM de una sola de las N tablas (es decir, una tabla formada por las coordenadas de las modalidades sobre los ejes principales).

Esta nueva tabla recoge la información más relevante contenida en las variables originales. Dicha tabla tiene tantas filas como la suma total de modalidades de las variables, y tantas columnas como factores seleccionemos en el ACM, además de ser de naturaleza *cuantitativa*.

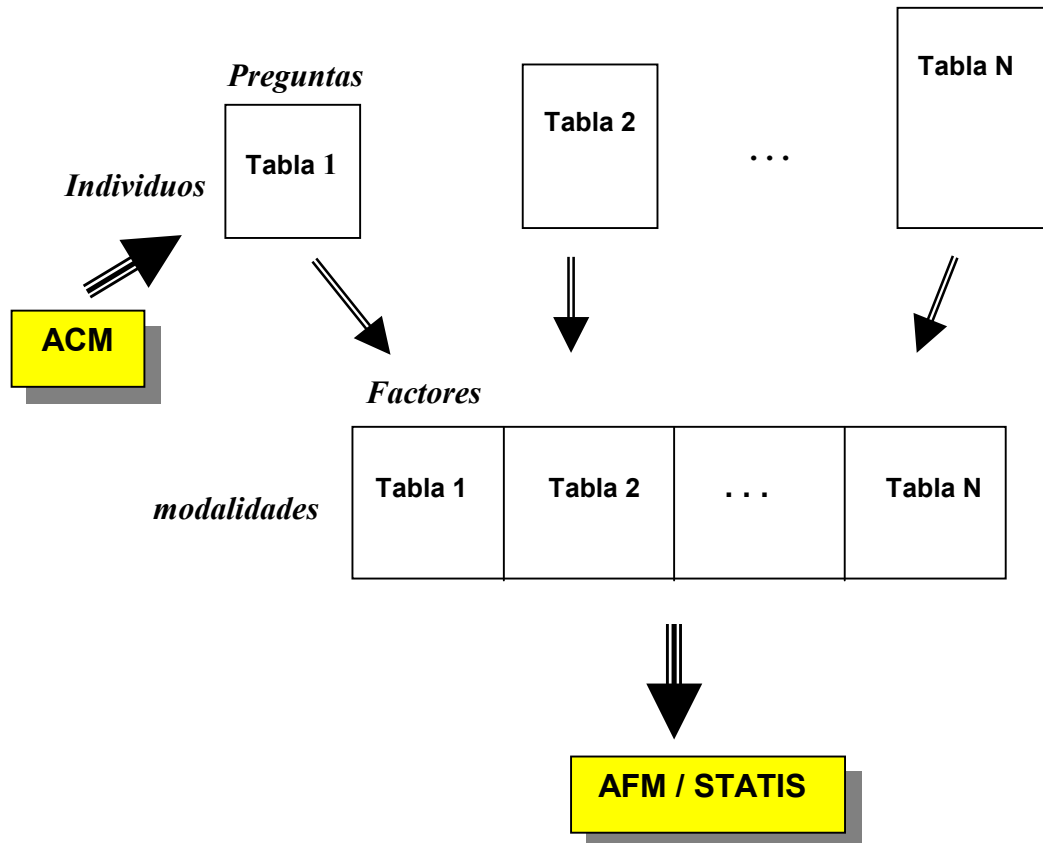
A continuación se forma una gran tabla, yuxtaponiendo, en lugar de las tablas originales, cada una de las tablas formadas por modalidades y factores principales de cada uno de los ACM realizados sobre las N tablas. Procediendo de esta manera, se obtiene un conjunto de N tablas, con el mismo número de filas, no necesariamente de columnas (i.e., en los ACM previos no tenemos por qué seleccionar el mismo número de factores principales para todas las tablas) y de naturaleza cuantitativa.

Estas dos características permiten solucionar el problema de análisis de las dos metodologías anteriormente mencionadas. Por un lado, al tener el mismo número de filas correspondientes a las modalidades de las mismas variables para las N tablas, se pueden comparar estas tablas artificiales mediante el Análisis Factorial Múltiple. Por otro, al ser de naturaleza cuantitativa, es posible utilizar, de la metodología Statis, el método STATIS, propiamente dicho, que utiliza como medida de distancia la obtenida a partir de los productos escalares entre individuos (filas).

El hecho de poder utilizar estas dos metodologías posibilita, por una parte, el análisis de tablas cualitativas de diferentes dimensiones por filas, obteniendo los mismos resultados que en los casos habituales donde ambos métodos pueden usarse sin ninguna transformación y, por otra parte, realizar una comparación entre ambos (ver Pagès [1996], Landaluce *et al* [1999]).

El AFM es más exhaustivo que el STATIS, y da una descripción más detallada de las relaciones entre las variables de las diferentes tablas, no obstante, el STATIS es más clarificador en cuanto a las relaciones entre las tablas diferentes, dando menos importancia a la intra estructura de las N tablas analizadas.

De forma esquemática, el proceso antes descrito sería el siguiente:



3.1.2 Aplicación empírica

Los datos elegidos son los de la Encuesta de Desigualdad Social del International Social Survey Program (ISSP), Zentralarchiv für Empirische Sozialforschung, Universidad de Colonia. Dicha encuesta fue realizada en 19 países en el año 1992 (excepto en el caso español, que fue a principios de 1993) y fueron recogidas un total de 176 variables, iguales para todos los países. La mayoría eran preguntas, por tanto, variables cualitativas y estaban medidas en la misma escala para todos los países. En cambio, el número de entrevistados fue radicalmente diferente para los distintos países; oscilaba entre 749 y 2502.

Como muestra se han elegido un total de 22 variables para 10 países. Estas variables eran preguntas realizadas a las personas encuestadas de dos tipos:

- ¿Cuánto cree que **debería de ganar** un ...<profesión>?
- ¿Cuánto cree que **realmente gana** un ...<profesión>?

Dicho grupo de variables eran, en su mayoría, continuas en la muestra original, pero fueron recodificadas a discretas en un paso previo, en una escala de 1 a 5. El motivo de la discretización no fue otro que ser introducidas en el análisis como elementos activos en la determinación de los factores del ACM, ya que de haber permanecido como continuas, serían tratadas como variables ilustrativas. Por otro lado, el hecho de incluir precisamente estas variables en el análisis se debe a que, en análisis de correspondencias previos realizados sobre éstas y el resto de variables incluidas en la encuesta, estas variables siempre aparecían como las más determinantes de los

principales factores. Por esta razón se decidió no incluir el resto de variables y mantener únicamente éstas, que además tienen mayor sentido económico.

Las profesiones consideradas abarcan un amplio espectro social, desde trabajadores no cualificados a ministros del gobierno. Estas profesiones pueden ser divididas, de forma no muy rigurosa, en profesiones de cuello blanco y de cuello azul, según su estatus socioeconómico sea considerado, respectivamente, alto ó bajo. En la tabla siguiente, aparecen las profesiones analizadas:

Profesiones (cuello blanco)
Médico
Gerente de gran empresa
Abogado
Propietario-gerente de gran fábrica
Juez
Ministro de gabinete

Profesiones (cuello azul)
Trabajador de fábrica cualificado
Dependiente Grandes Almacenes
Propietario de tienda pequeña
Granjero
Trabajador de fábrica no cualificado

Se analizan un total de 22 variables: se considera a las 11 profesiones citadas más arriba, sobre las que se hacen dos preguntas, la primera de cómo debería de estarlo y la segunda, si está bien pagada, a juicio del encuestado.

En la encuesta existía disponible un elevado número de variables socioeconómicas susceptibles de ser utilizadas como variables ilustrativas en los ACM de cara a la interpretación de los factores. Se han seleccionado un cierto número de ellas entre las cuales están sexo, edad, religión, tipo de afiliación política, años de escolarización, etc.

Los países considerados son 10: República Federal de Alemania (WG), Gran Bretaña (GB), Estados Unidos (USA), Hungría (HUN), Noruega (NW), Checoslovaquia (CZ), URSS (RU), Nueva Zelanda (NZ), Filipinas (FIL) y España (SP). Del total de 19 países disponibles fueron desechados 9 principalmente por problemas en el proceso de discretización. Aún cuando dichas variables eran teóricamente continuas en origen, observamos cómo algunas de ellas eran realmente discretas; la gran mayoría de individuos tenía asignada como respuesta dos ó tres cantidades de las comúnmente denominadas redondas, lo cual indica posiblemente restricciones en las respuestas posibles en el cuestionario empleado en estos países. Es decir, que en el cuestionario dichas preguntas se habrían efectuado de forma cerrada. Dichos países no han sido considerados puesto que el método propuesto requiere la igualdad de categorías entre los distintos países, imposible de lograr para estos últimos con estas restricciones.

Se realizaron los ACM correspondientes a las 22 preguntas citadas antes para los 10 países, y como resultado se obtuvo que los factores tenían una interpretación bastante similar en todos los casos. Como es habitual en este tipo de encuestas, el primer factor refleja la no respuesta, asociada principalmente a personas de edad avanzada. El segundo factor es, básicamente, un factor talla que ordena, de arriba abajo, respuestas a las preguntas con valores altos que van descendiendo conforme nos vamos acercando a la parte negativa del factor. Tiene alguna relación con la edad, tendiendo a estar los jóvenes en la parte inferior y las personas de mediana edad en la superior. El tercer factor muestra un claro efecto Guttman cuando es representado frente al segundo factor; opone claramente respuestas extremas (1 y 5) a respuestas medias (alrededor del 3). El cuarto factor es ya menos interpretable; parece reflejar las opiniones de personas que creen que ciertas profesiones de cuello blanco deben (y realmente creen que lo hacen) ganar mucho y que ciertas de cuello azul deben (y también realmente creen que lo hacen) ganar poco. Las variables socioeconómicas ilustrativas no son de mucha ayuda, porque aparte de las de sexo y edad, no hay muchos efectivos de ellas ó no son muy significativas. En algunos países, los factores 3 y 4 parecen estar invertidos.

Se decidió mantener 4 factores de cada uno de los ACM de los 10 países, no obstante, para ilustrar las diferencias se ha incluido también el factor 5, a pesar de no haber encontrado ninguna explicación objetiva para él. Entonces, la tabla formada por las coordenadas de las modalidades sobre los factores es una tabla de $10 \times 5 = 50$ columnas y 132 filas (5 modalidades + modalidad de no respuesta * 22 variables = $6 \times 22 = 132$ modalidades totales). Sobre esta tabla se aplicó tanto un AFM como un STATIS, cuyos resultados se exponen a continuación.

i. Resultados del AFM

El AFM sobre una serie de tablas de naturaleza cuantitativa consiste, de forma resumida, en un ACP global de todas las tablas yuxtapuestas donde cada una de ellas ha sido previamente ponderada por el inverso de la raíz del primer valor propio procedente de un ACP realizado sobre la tabla parcial con el objetivo de equilibrar la influencia entre las tablas. Es decir, en la diagonalización de la tabla $X'X$, la matriz de correlación/covarianzas entre todas las variables de todas las tablas donde:

$$X = \left[\begin{array}{cccc} \frac{1}{\sqrt{\lambda_1^1}} X_1 & \frac{1}{\sqrt{\lambda_1^2}} X_2 & \cdots & \frac{1}{\sqrt{\lambda_1^j}} X_j \end{array} \right]$$

donde λ_1^j es el primer valor propio proveniente de un ACP (ó un ACM, según sea la naturaleza de las variables de partida) realizado sobre la j -ésima tabla. Posteriormente, se calculan las coordenadas de las variables de cada uno de los grupos respecto a los factores y se obtiene una representación global. En este caso, las matrices X_j que se emplean son las compuestas por los factores, cuantitativos, procedentes de los ACM realizados sobre cada una de las tablas cualitativas originales.

La matriz de correlación (triángulo inferior) entre los factores de los 10 ACM (antes de ponderar por los inversos de las raíces de los primeros valores propios) se recoge en la siguiente tabla :

	HUN3	HUN4	HUN5	NW1	NW2	NW3	NW4	NW5	CZ1	CZ2	CZ3	CZ4	CZ5	RU1	RU2	RU3	RU4
	RU5	NZ1	NZ2	NZ3	NZ4	NZ5	FIL1	FIL2	FIL3	FIL4	FIL5	SP1	SP2	SP3	SP4	SP5	
RU5	1.00																
NZ1	0.04	1.00															
NZ2	-0.03	-0.11	1.00														
NZ3	-0.11	0.08	-0.14	1.00													
NZ4	0.12	0.10	-0.19	0.14	1.00												
NZ5	-0.05	-0.20	-0.07	0.14	-0.05	1.00											
FIL1	-0.04	-0.96	0.20	-0.06	-0.09	0.10	1.00										
FIL2	-0.08	0.07	0.89	-0.10	-0.25	-0.11	0.01	1.00									
FIL3	-0.03	-0.03	-0.11	-0.54	-0.56	-0.25	0.01	-0.03	1.00								
FIL4	0.05	-0.13	0.02	-0.52	0.27	0.15	0.08	0.00	-0.04	1.00							
FIL5	-0.31	0.12	-0.04	-0.07	-0.19	0.05	-0.10	-0.04	0.01	0.05	1.00						
SP1	0.08	0.91	-0.25	0.04	0.07	-0.09	-0.96	-0.06	0.04	-0.07	0.02	1.00					
SP2	0.08	0.00	-0.89	0.05	0.15	0.05	-0.12	-0.81	0.16	0.07	-0.11	0.22	1.00				
SP3	-0.15	0.23	-0.28	0.83	0.22	0.11	-0.20	-0.19	-0.45	-0.46	0.02	0.16	0.11	1.00			
SP4	0.03	0.06	0.12	0.22	0.35	0.05	-0.08	0.11	-0.37	0.12	-0.28	0.11	-0.14	0.26	1.00		
SP5	0.22	0.00	0.42	-0.15	-0.57	-0.29	0.06	0.36	0.33	-0.26	-0.03	-0.06	-0.34	-0.25	0.00	1.00	

Tabla 1: Matriz de correlación

Las correlaciones entre factores del mismo orden entre países diferentes están en negrita, para facilitar su visualización. Se puede observar, en primer lugar, cómo todos los primeros factores están muy correlacionados entre sí, los segundos también, y así sucesivamente. Esto es porque los factores del ACM tienen básicamente la misma interpretación para todos los países.

En segundo lugar, es también notable el hecho de que la correlación entre factores homogéneos de distintos países es muy fuerte para los primeros factores, pero esa correlación va decayendo a medida que se consideran factores de orden superior. Esto es debido a que los factores de mayor orden están afectados, cada vez en mayor medida, de las fluctuaciones del muestreo y recogen menos características comunes de las variables originales.

Si se quiere obtener una medida de ligazón, ó de similaridad entre las tablas, se pueden observar los coeficientes RV, que pueden ser interpretados como coeficientes de correlación entre las tablas. Para las 10 tablas de factores, estos vienen dados en la tabla 2:

	WG	GB	USA	HUN	NW	CZ	RU	NZ	FIL	SP
WG	1.00									
GB	0.73	1.00								
USA	0.75	0.61	1.00							
HUN	0.54	0.55	0.59	1.00						
NW	0.64	0.58	0.61	0.54	1.00					
CZ	0.44	0.45	0.45	0.52	0.55	1.00				
RU	0.45	0.50	0.48	0.68	0.44	0.51	1.00			
NZ	0.58	0.59	0.68	0.51	0.75	0.63	0.42	1.00		
FIL	0.61	0.57	0.66	0.55	0.51	0.49	0.49	0.58	1.00	
SP	0.50	0.53	0.59	0.57	0.56	0.45	0.42	0.61	0.50	1.00

Tabla 2: Los coeficientes RV

Esta tabla proporciona bastante información, en el sentido de especificar qué pares de tablas están muy ó poco correlacionadas; por ejemplo Estados Unidos (USA) y Alemania Occidental (WG) tienen una correlación alta, de 0.75, mientras que entre Rusia (RU) y Nueva Zelanda (NZ) es bastante mediocre (0.42), indicando que son países que tienen poco que ver, al menos en cuanto a las relaciones entre las variables utilizadas en el estudio se refiere. De todas formas, es difícil hacerse una idea clara de

las relaciones, debido al número de países; evidentemente esto se agudiza a medida que crece el número de tablas a considerar. Aquí es donde el AFM es de especial interés, al permitir visualizar en un plano de factores principales, las posiciones relativas, no sólo de las tablas, sino de los factores que componen dichas tablas, para identificar cuáles de ellos son los que reflejan estas similitudes ó diferencias.

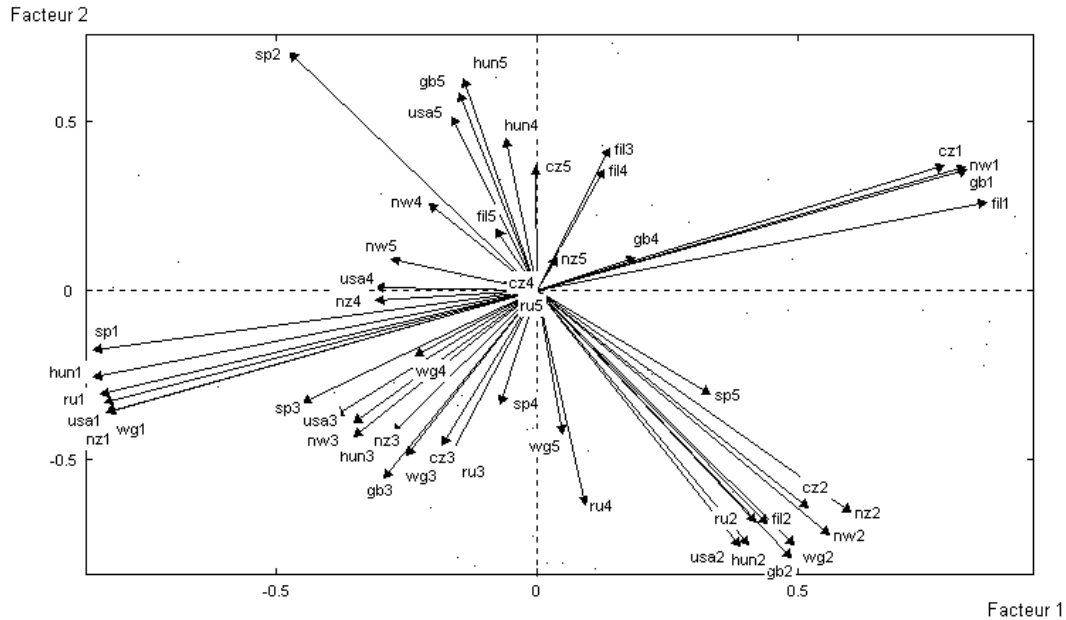


Gráfico 1: Plano factorial 1-2. Factores Parciales

En el gráfico anterior, se visualizan las posiciones de los factores de cada país sobre los dos primeros ejes del AFM. Es clara la elevada correlación entre los factores del mismo orden para los distintos países, especialmente para los dos primeros, y de éstos con respecto a los ejes del análisis global del AFM (especialmente claro si se efectúa una pequeña rotación en el sentido de las agujas del reloj). La correlación pierde intensidad a partir del tercer factor. En el caso del factor 5, se observa que las correlaciones entre los quintos factores de los 10 países son ya poco evidentes.

En el siguiente gráfico, se ven las posiciones de los grupos sobre los dos primeros factores del AFM:

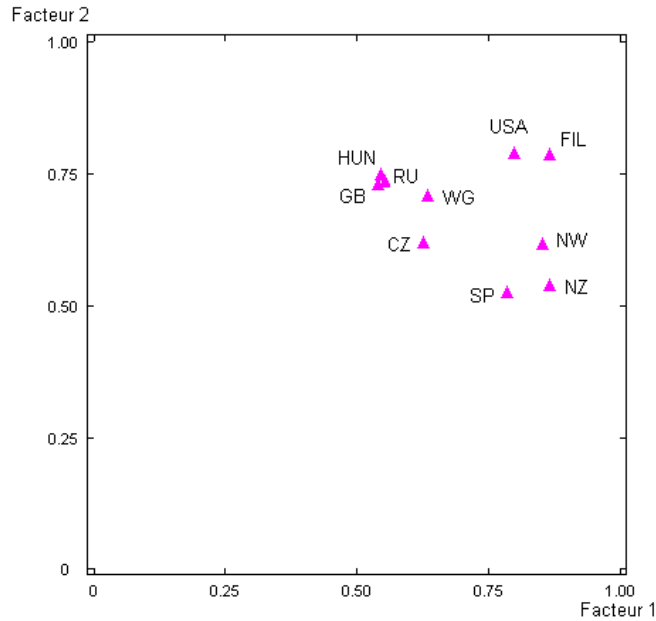


Gráfico 2: Plano Factorial 1-2. Grupos (Países)

En este gráfico es difícil destacar demasiadas diferencias entre los países, que aparecen todos bastante próximos.

ii. Resultados del STATIS

El procedimiento STATIS analiza las matrices $X_j X_j'$ de productos escalares centrados entre individuos, llamadas impropriadamente nubes de individuos, para las 10 tablas, en este caso. Son, por tanto, matrices de orden 132×132 . Posteriormente analiza las distancias y similitudes entre ellas, y obtiene una nube compromiso, ó nube media, lo que va a permitir mostrar las desviaciones (medidas como correlaciones ó como distancias euclídeas) de las diferentes nubes respecto de la nube compromiso.

En este caso, los productos escalares RV entre nubes se recogen en la matriz de la tabla 3 (similar a una matriz de correlación). En ésta se muestra con claridad que hay un país notablemente más alejado del resto, que es España, con correlaciones que sólo en un caso supera el 0.5. En general, no obstante, las nubes de individuos (modalidades de las variables originales) puede decirse que son bastante parecidas.

	WG	GB	USA	HUN	NW	CZ	RU	NZ	FIL	SP
WG	1.000									
GB	0.806	1.000								
USA	0.851	0.760	1.000							
HUN	0.695	0.672	0.718	1.000						
NW	0.823	0.767	0.932	0.710	1.000					
CZ	0.709	0.655	0.828	0.701	0.825	1.000				
RU	0.705	0.705	0.794	0.813	0.774	0.788	1.000			
NZ	0.792	0.748	0.943	0.681	0.952	0.844	0.772	1.000		
FIL	0.776	0.699	0.821	0.728	0.778	0.759	0.769	0.794	1.000	
SP	0.487	0.489	0.449	0.587	0.453	0.366	0.444	0.456	0.526	1.000

Tabla 3: Matriz de productos escalares RV

Si se procede al cálculo de la nube media y se obtienen sus ejes principales, se pueden calcular las posiciones relativas de las tablas respecto a los dos primeros ejes principales, que dan una idea más exacta de las posiciones relativas de los países:

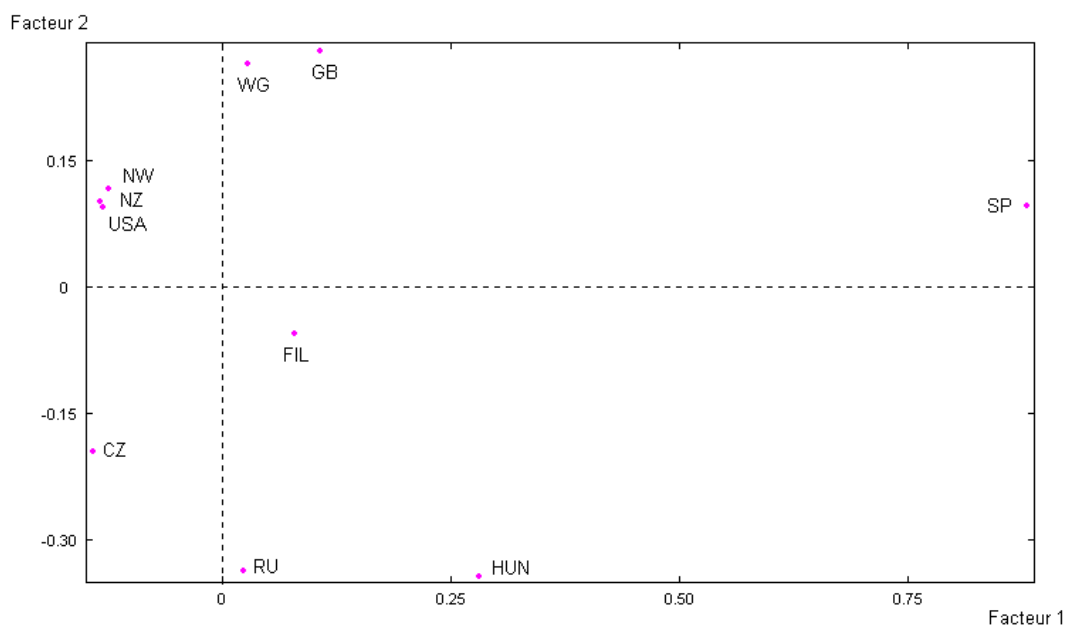


Gráfico 3: Plano Factorial 1-2. Grupos (Países)

Aquí se obtiene que el factor 1 separa claramente a España del resto de países, mientras que el factor 2, con menos claridad en términos de distancia, separa países occidentales desarrollados de países menos desarrollados, como Filipinas, ó países de economía y sociedad comunistas, al menos hasta hacía muy poco tiempo. La separación entre países es más clara que la obtenida en el AFM, lo que no es de sorprender, puesto que el método STATIS utiliza una ponderación que destaca las tablas más alejadas de la nube compromiso.

Y ¿qué significan los factores? Si se posicionan en un gráfico las coordenadas de las modalidades originales respecto a los factores, se ve que la interpretación es la misma que en los ACM correspondientes a los países por separado, y por ende, a los factores globales del AFM. Es decir, el factor 1 se refiere a la No Respuesta, el 2 recoge un efecto talla, el 3 opone respuestas medias a extremas (efecto Guttman en el plano (2,3)) y el 4 asocia determinadas modalidades de valor alto con otras de bajo, correspondientes a oficios de diferente consideración social y económica.

Para terminar este estudio empírico, señalar que la interpretación realizada de los resultados no es completa, ya que no era éste el objetivo del mismo, sino plantear una nueva posibilidad de las metodologías AFM y STATIS para el análisis simultáneo de tablas cualitativas referidas a distintos colectivos de individuos.

3.2 Estudio de la estabilidad de los resultados obtenidos en un ACP

3.2.1 Generalidades

Una característica de los métodos factoriales es que los resultados que proporcionan al aplicarlos a grandes tablas de datos ponen de manifiesto la estructura existente entre los mismos. Ahora bien, estos resultados pueden verse afectados por las fluctuaciones de los datos o por la definición y codificación de las variables, incluso por la representatividad de la muestra utilizada, de ahí, la necesidad de realizar estudios sobre la estabilidad de resultados, Lebart (1995) y Greenacre (1993).

Los estudios de estabilidad trabajan a menudo sobre modificaciones de la tabla inicial. Si la configuración obtenida en el análisis inicial se mantiene al introducir diversas modificaciones, los resultados del análisis pueden considerarse estables. Las alteraciones de la tabla inicial se pueden producir en la definición y número de variables, en el conjunto de individuos, o bien en perturbaciones de los datos.

En nuestros trabajos hemos utilizado los métodos factoriales de análisis de tablas múltiples, en concreto Análisis Factorial Múltiple (AFM) (Abascal y Landaluce 2000) y Metodología STATIS, para verificar la estabilidad interna de los resultados de un Análisis de Componentes Principales (ACP), comparando las aportaciones de ambos métodos. El método STATIS ha sido utilizado anteriormente por Holmes (1985,1989).

La idea básica consiste en estudiar simultáneamente todas las tablas, generadas en las diferentes alteraciones, mediante los métodos de análisis de tablas múltiples. Las características de éstos permiten la comparación de los resultados de los análisis de las diferentes tablas y proporcionan buenos indicadores de las diferencias obtenidas.

3.2.2 Aplicación empírica

El objetivo de esta aplicación empírica es estudiar la estabilidad de los resultados obtenidos mediante un ACP de la tabla que recoge la estructura porcentual de consumo alimenticio de las Comunidades Autónomas españolas.

i. A través del AFM

La tabla objeto de este análisis recoge la distribución del porcentaje de gasto en 10 categorías de alimentos (cuya descripción se presenta posteriormente) por Comunidades Autónomas (CCAA). Esta tabla constituye el grupo 1. Para estudiar su estabilidad se generan nuevas tablas con perturbaciones aleatorias, es decir, cada valor se altera mediante la adición de una perturbación generada por una distribución normal cuya varianza es una fracción de la varianza inicial de la variable. Se generan así tres nuevas tablas que corresponden a perturbaciones con varianzas $1\%S_j$, $10\%S_j$, $20\%S_j$ respectivamente y constituyen los grupos 2 a 4.

GASTOS			
C	X _{ij}	X _{ij}	X _{ij}
C		+	+
A		N(0;1%S _j)	N(0;10%S _j)
A			N(0;20%S _j)

G1: Pan y cereales

G2: Carne

G3: Pescado

G4: Leche, derivados y huevos

G5: Aceites y grasas

G6: Frutas y hortalizas

G7: Patatas

G8: Azúcar

G9: Café, té y chocolate

G10: Otros

La estabilidad de la nube se estudia a través del análisis de las representaciones gráficas de los puntos de las nubes parciales y de sus inercias con respecto a su centro de gravedad o nube media, las inercias intra. Para ello se exponen a continuación los dos primeros planos factoriales correspondientes a las nubes de individuos, CCAA (gráfico 4) y de variables, gastos, (gráfico 5).

En ambos gráficos se puede observar una gran proximidad entre todos los puntos que representan al mismo individuo (gráfico 4) y a la misma variable (gráfico 5), resultado que pone de manifiesto la existencia de una débil inercia intra. Esto es, las tablas analizadas tienen una estructura muy similar

Se producen algunas excepciones que merecen cierta atención. Así, en el plano de las variables, los vectores que representan al grupo de gasto en carne, presentan entre sí un ángulo mayor, que el resto de los gastos estudiados, lo cual nos indica que se trata de una variable con un comportamiento menos estable que las demás. Hay que destacar que es la que presenta los valores más altos en los estadísticos básicos: media y desviación típica. En lo que se refiere al plano de los individuos, en él se puede observar el comportamiento “menos homogéneo” de ciertas CCAA, entre las que podemos destacar: Cataluña, Navarra, Madrid, Andalucía, La Rioja y Galicia. El estudio minucioso de las tablas que recogen las inercias intra, tanto de los puntos medios como de los puntos parciales que representan a estas regiones, permite poner de manifiesto en qué eje factorial este comportamiento es más acusado y cuáles son los puntos responsables del mismo.

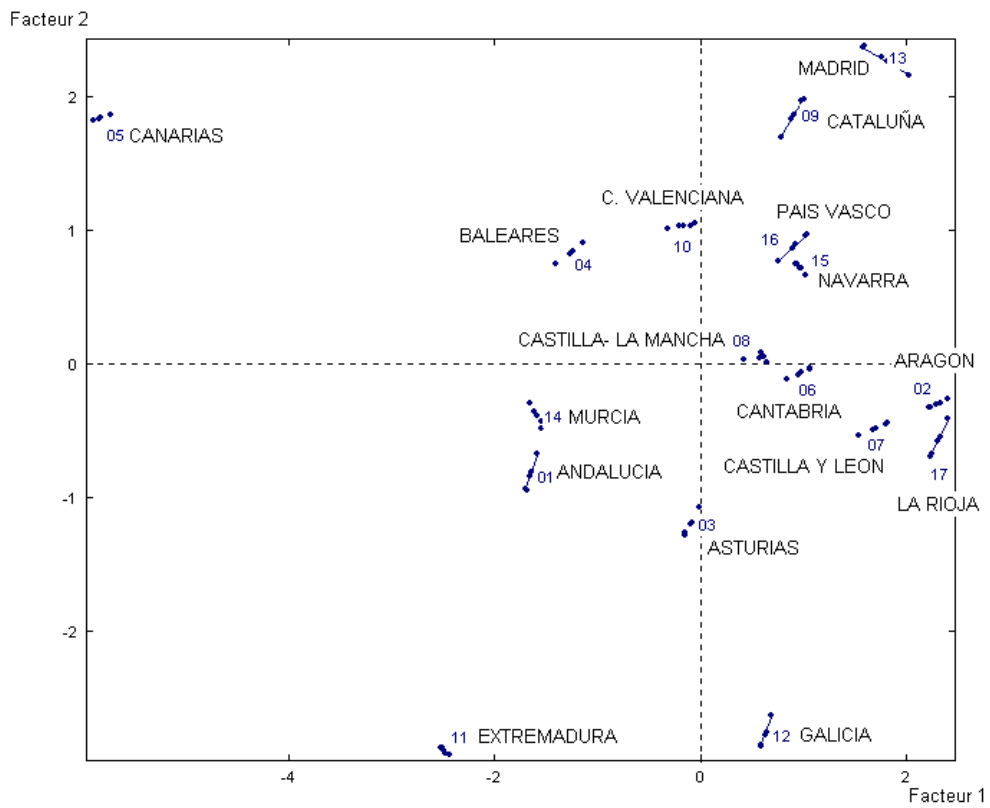


Gráfico 4: Plano Factorial 1-2

Comunidades Autónomas. Puntos medios y Puntos parciales

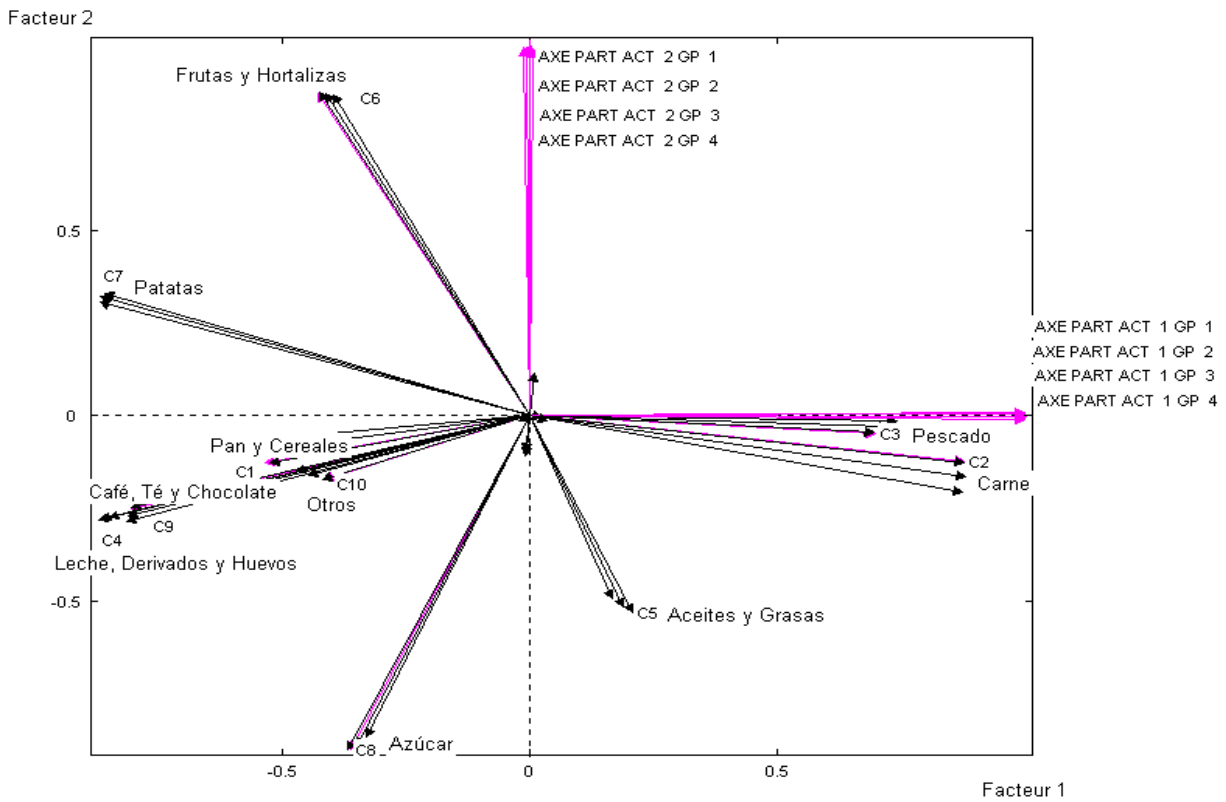


Gráfico 5.Plano Factorial 1-2:

Variables (activas) y Ejes Parciales (suplementarios) de las 4 tablas

**INDIVIDUOS CON MAYORES INERCIAS
INTRA**

EJE 1	INDIVIDUOS	INER	ACUM .
	Madrid	25.22	25.22
	País Vasco	10.03	35.25
	Castilla y León	9.52	44.77
	Baleares	8.87	53.64

EJE 2

INDIVIDUOS	INER	ACUM .
Cataluña	16.18	16.18
La Rioja	15.14	31.32
Andalucía	14.77	46.09
Galicia	10.51	56.60

**INDIVIDUOS PARCIALES CON MAYORES INERCIAS
INTRA**

EJE 1	INDIVIDUOS	INER	ACUM .
	Madrid	4	13.73
	Madrid	1	5.88
	País Vasco	4	5.55
	Castilla y León	4	5.31
	Baleares	4	4.95
	Madrid	2	4.79

EJE 2

INDIVIDUOS	INER	CUMUL
Cataluña	4	9.08
La Rioja	4	8.58
Andalucía	4	8.45
Galicia	4	6.14
Madrid	4	5.33
Asturias	4	4.72

El estudio *comparativo de la proximidad o semejanza entre las diferentes nubes*. Se realiza a través de diferentes indicadores destacando los siguientes resultados:

- La lectura de la matriz de correlaciones entre los factores parciales (tabla 5), pone de manifiesto la estabilidad de los resultados obtenidos en este estudio empírico. Ello se observa tanto en las fuertes correlaciones entre los factores del mismo orden, correspondientes a las diferentes tablas, como a las correlaciones prácticamente nulas entre los factores de distinto orden. Esto es, esta matriz nos proporciona una visión previa de las similitudes entre las cuatro tablas analizadas, indicando que las primeras direcciones de variabilidad de cada grupo (que recogen el 94% de la inercia total) manifiestan estructuras comunes a los mismos.

	101	102	103	104	105	205	0.00	0.00	0.00	0.01
105					0.99					
Perturbación 1%										
1.00	1.00	0.00	0.00	0.00	0.00	1.00	-0.01	-0.01	0.03	0.02
	0.00	1.00	0.00	0.00	0.00	0.01	0.99	-0.13	-0.01	0.01
	0.00	0.00	1.00	0.00	0.00	0.01	0.13	0.99	0.01	-0.01
	0.00	0.00	0.00	1.00	0.00	-0.03	0.01	-0.01	1.00	0.01
	0.00	0.00	0.00	0.00	1.00	-0.02	-0.01	0.01	-0.01	0.99
Perturbación 20%										
0.00	1.00	0.00	0.00	0.00	0.00	0.99	-0.02	-0.02	0.06	0.04
0.00	0.00	1.00	-0.01	0.00	0.00	0.01	0.97	-0.22	-0.03	0.03
0.00	0.00	0.01	1.00	-0.01	0.00	0.02	0.22	0.97	0.02	-0.02
0.00	0.00	0.00	-0.02	-1.00	0.00	0.06	-0.02	0.03	-0.98	-0.02
0.02	0.00	0.00	-0.02	-1.00	0.00	-0.04	-0.02	0.02	-0.02	0.96

Tabla 5: Matriz de correlaciones entre los factores parciales

- Las matrices L y RV permiten , nuevamente, comprobar la estabilidad de los resultados obtenidos. Son grupos con una dimensionalidad parecida (se observa en los coeficientes de la diagonal principal de la matriz L) y con una estructura interna prácticamente igual (se observa en los coeficientes de la matriz RV).

COEFICIENTES Lg DE RELACION ENTRE GRUPOS					COEFICIENTES RV DE RELACION ENTRE GRUPOS				
	1	2	3	4		1	2	3	4
1	1.51				1	1.00			
2	1.52	1.53			2	1.00	1.00		
3	1.49	1.49	1.47		3	1.00	1.00	1.00	
4	1.45	1.46	1.45	1.43	4	0.99	0.99	1.00	1.00

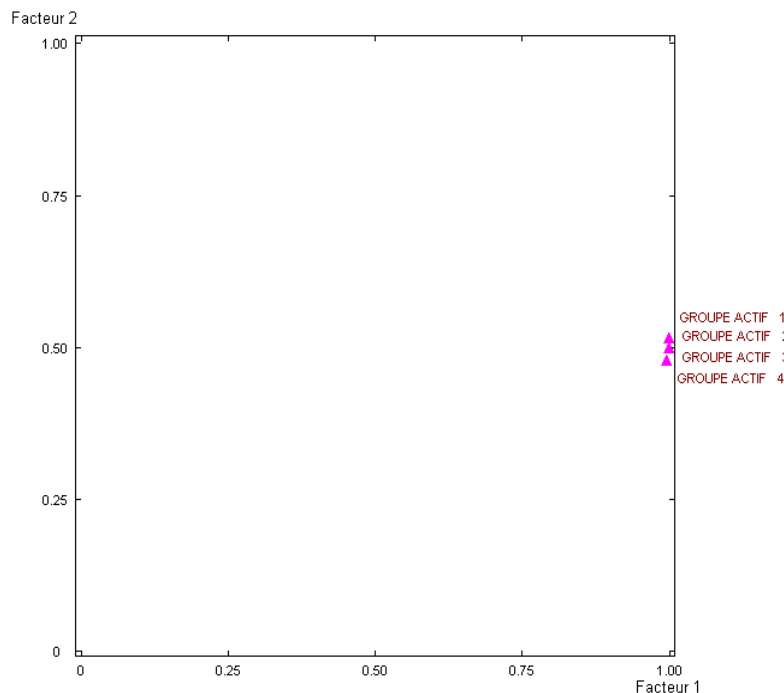
Tabla 6: Coeficientes L y RV de relación entre grupos

- La lectura del gráfico 6, plano factorial en el que cada punto representa a cada uno de los grupos analizados, pone de manifiesto que los ejes presentados recogen una realidad común a las tablas consideradas, ya que la contribución es la misma para todas y cada una de ellas.

GRP.	COORDENADAS					CONTRIBUCIONES				
	1	2	3	4	5	1	2	3	4	5
GR 1	1.00	0.52	0.44	0.18	0.10	25.0	25.6	26.2	24.7	25.1
GR 2	1.00	0.52	0.45	0.18	0.10	25.0	25.7	26.9	24.9	24.7
GR 3	1.00	0.50	0.41	0.18	0.10	25.0	24.8	24.3	25.0	24.9
GR 4	1.00	0.48	0.38	0.18	0.10	24.9	23.9	22.6	25.4	25.3

Tabla 7: Coordenadas y ayudas a la interpretación de los grupos

Gráfico 6: Plano factorial 1-2 Grupos de variables (tablas)



La coordenada de un grupo sobre un factor se puede considerar como una medida de la importancia de la dirección de dispersión en las nubes (de individuos y variables) asociadas a cada grupo. Esta coordenada se interpreta como la contribución absoluta de las variables del grupo al factor, es decir, una medida de relación entre ambos elementos. Esto es, indica los grupos que han determinado en mayor medida los factores. La lectura de esta tabla pone de manifiesto que los ejes presentados recogen una realidad común a las tablas consideradas, ya que la contribución es la misma para todas y cada una de ellas.

ii. A través del STATIS DUAL

Una segunda posibilidad de estudio de la estabilidad mediante análisis de tablas múltiples, se basa en la aplicación del método STATIS DUAL a la tabla múltiple formada por la original y aquella que surge al eliminar uno o varios individuos de la primera.

En este análisis de la estabilidad interna de los resultados del ACP que se presenta, el número de tablas consideradas son dos: la tabla original y aquella que surge al eliminar un individuo con gran contribución a la formación del primer eje, en el análisis de la tabla original, la Comunidad de Canarias.

Dentro de las técnicas factoriales de análisis de tablas múltiples la que permite el estudio comparativo de dos tablas que difieren en el número de individuos es la metodología STATIS DUAL. A continuación se exponen e interpretan los resultados más destacados obtenidos al utilizar este método:

- Los valores propios de las matrices de correlación asociadas a las distintas tablas analizadas, así como el porcentaje de inercia acumulado, son los siguientes:

NUBE1										NUBE 2										
MATRIZ DE CORRELACIONES 1										MATRIZ DE CORRELACIONES 2										
7	8	1	2	3	4	5	6	7	8	6	7	1	2	3	4	5	6	7	8	
8	1	2	3	4	5	6	7	1	2	7	1	2	3	4	5	6	7	1	2	
1	1.000									1	1.000									
2	-0.251	1.000								2	-0.476	1.000								
3	-0.725	0.299	1.000							3	-0.800	0.166	1.000							
4	0.342	-0.697	-0.556	1.000						4	0.460	-0.511	-0.484	1.000						
5	-0.187	-0.073	0.337	-0.115	1.000					5	-0.246	-0.332	0.320	0.016	1.000					
6	0.013	-0.518	-0.255	0.062	-0.365	1.000				6	0.124	-0.208	-0.163	-0.359	-0.315	1.000				
7	0.174	-0.907	-0.423	0.617	-0.080	0.646	1.000			7	0.545	-0.813	-0.408	0.318	0.202	0.407	1.000			
8	0.335	-0.153	-0.257	0.539	0.369	-0.677	0.006	1.000		8	0.343	-0.313	-0.306	0.698	0.338	-0.733	0.167	1.000		
9	0.286	-0.605	-0.437	0.765	-0.136	0.010	0.596	0.452	1.000	9	0.467	-0.270	-0.363	0.636	-0.003	-0.527	0.115	0.663	1.000	
10	-0.226	-0.548	0.142	0.296	0.256	0.128	0.424	0.291	0.351	10	-0.208	-0.316	0.332	0.033	0.433	-0.208	-0.011	0.374	0.049	
	0.351	1.									0.049	1.								

NUBE 1			
HISTOGRAMA DE LOS 10 PRIMEROS VALORES PROPIOS			
NUM	VALOR	PORC.	
	PROPRIO	ACUM.	
1	4.0378	40.38	*****
2	2.2067	62.45	*****
3	1.9130	81.58	*****
4	0.7755	89.33	*****
5	0.4334	93.66	*****
6	0.2860	96.52	*****
7	0.2345	98.87	*****
8	0.0978	99.85	**
9	0.0139	99.99	*

NUM	VALOR	PORC.	PROPIO	ACUM.
1	3.7372	37.37		
2	2.4969	62.34		
3	1.8711	81.05		
4	0.6249	87.30		
5	0.4998	92.30		
6	0.3808	96.11		
7	0.2042	98.15		
8	0.1630	99.78		
9	0.0197	99.98		
10	0.0023	100.00	*	

Tabla 8: Resultados sobre la forma de las diferentes nubes de columnas

Al analizar y comparar los resultados anteriores, se observa que existen mínimas diferencias entre las estructuras internas, de gasto en alimentación, que presentan las dos tablas consideradas. Únicamente merece destacar el mayor porcentaje de inercia que recoge el primer eje factorial del análisis de la tabla original, con respecto del recogido por el factor del mismo orden de la tabla modificada. Esta reducción se debe a que el individuo eliminado tenía un gran peso en la creación de la primera dirección de variabilidad.

- Los siguientes resultados, productos escalares entre las matrices de correlación y las distancias euclídeas entre las mismas, son dos medidas que corroboran los comentarios realizados en el párrafo anterior, sobre las semejanzas entre las tablas. En la primera matriz destacan las semejantes normas que presentan las tablas analizadas, valores que aparecen recogidos en la diagonal principal. 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 en cuestión. En lo que se refiere a la matriz de distancias, destacar que existe una cierta distancia entre ambas tablas. Estudiaremos cuáles son las variables (gastos) responsables en mayor o menor medida de la misma.

PRODUCTOS ESCALARES ENTRE MATRICES DE CORRELACIONES				DISTANCIAS EUCLIDAS ENTRE MATRICES DE CORRELACIONES				
	1	2		1	2		1	2
1	25.769		1	0.000				
2	22.854	24.556	2	2.149	0.000			

Tabla 9: Situación de las matrices de correlaciones

Este método proporciona, además, una medida, en porcentaje, de la descomposición de las distancias entre las matrices de correlación asociadas a las tablas de datos, según las variables. Esta medida permite vislumbrar, de forma general, cuáles son las variables responsables de las desviaciones existentes.

Tabla 10: Contribuciones de las variables en porcentaje a las distancias entre matrices de correlaciones

PAN Y CEREALES	=	5.55 %	LECHE, DERIVADOS Y HUEVOS	=	9.74 %
CARNE	=		ACEITES Y GRASAS	=	4.78 %
10.13 %					
PESCADO	=	1.76	FRUTAS Y HORTALIZAS	=	16.41 %
%					

PATATAS	=	17.75	%
AZUCAR	=	2.91	%
CAFÉ, TE Y CHOCOLATE	=	18.20	%
OTROS	=	12.77	%

Se puede comprobar que existen importantes diferencias, en lo que a las contribuciones de las variables a la distancia entre las dos tablas se refiere, así los grupos de alimentación relativos a Café, té y chocolate, Patatas y Frutas y hortalizas contribuyen en conjunto con más del 50%, siendo éstos claramente los causantes de las diferencias entre las estructuras internas de las dos tablas (señalar que la Comunidad de Canarias, eliminada en la tabla 2, es la que presenta los mayores porcentajes de gasto en estos grupos, en la tabla 1). Mientras que los grupos relativos a Pescado, Azúcar y Aceites y grasas, no superan cada uno de ellos el 5% de contribución.

A continuación se presenta la matriz de correlación compromiso (media ponderada de las matrices de correlación asociadas a cada una de las tablas estudiadas), así como las ponderaciones correspondientes a cada una de las 2 tablas.

Debido a la gran similitud que presentan las dos tablas consideradas, el peso de ambas en la definición de la matriz de correlación compromiso es muy similar. No obstante, hay que señalar el mayor peso de la tabla original, reflejo de que es la que presenta una estructura interna más fuerte. Esta metodología penaliza, en cierto sentido, a aquellos grupos con estructura interna más débil.

CONSTRUCCION DE LA MATRIZ DE CORRELACION COMPROMISO
= MEDIA PONDERADA DE LAS MATRICES DE CORRELACIONES
POR LOS COEFICIENTES

1 = 0.507
2 = 0.493

	1	2	3	4	5	6	7	8	9	10
1	1.000									
2	-0.362	1.000								
3	-0.762	0.233	1.000							
4	0.400	-0.605	-0.521	1.000						
5	-0.216	-0.201	0.329	-0.051	1.000					
6	0.068	-0.365	-0.210	-0.146	-0.341	1.000				
7	0.357	-0.861	-0.416	0.469	0.059	0.528	1.000			
8	0.339	-0.232	-0.281	0.617	0.354	-0.705	0.085	1.000		
9	0.375	-0.440	-0.401	0.701	-0.070	-0.255	0.359	0.556	1.000	
10	-0.217	-0.434	0.236	0.166	0.343	-0.038	0.210	0.332	0.202	1.000

Tabla 11: Matriz de correlación compromiso

El análisis de esta matriz va a poner de manifiesto una realidad que corresponde a la “media de las realidades medias” de los dos grupos considerados.

- La lectura del histograma de los valores propios de la matriz de correlación compromiso pone de manifiesto que la realidad ahora estudiada tiene mayor dimensionalidad que la analizada con la tabla única, esto es, existen más factores de variabilidad con peso destacable. En concreto, ahora son tres los factores que recogen algo más de la mitad de la inercia total.

DESCOMPOSICION DE LA INERCIA SEGUN LAS
MATRICES DE CORRELACION

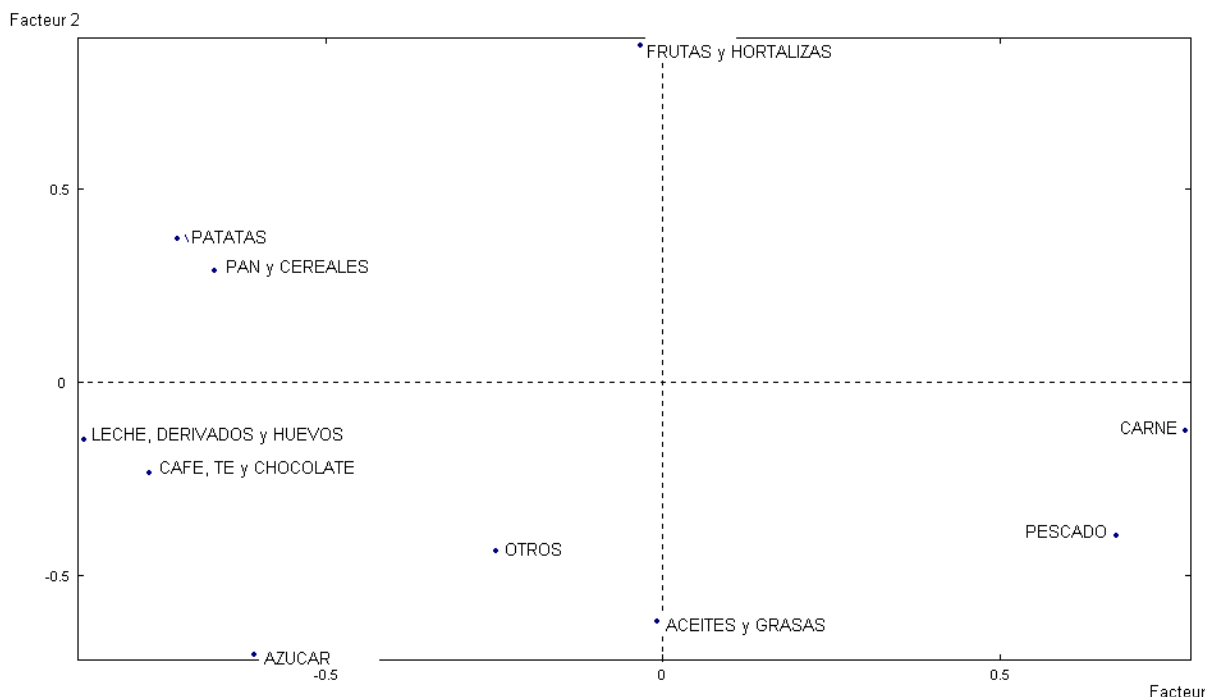
MATRIZ DE CORRELACION 1 = 48.67 %
MATRIZ DE CORRELACION 2 = 51.33 %

HISTOGRAMA DE LOS 10 PRIMEROS VALORES PROPIOS

NUM	VALOR	POR. PROPIO	ACUM
1	3.7552	37.55	*****
2	2.3199	60.75	*****
3	1.8590	79.34	*****
4	0.7893	87.23	*****
5	0.5008	92.24	*****
6	0.3271	95.51	*****
7	0.2690	98.20	*****
8	0.1542	99.74	****
9	0.0209	99.95	*
10	0.0046	100.00	*

- El plano factorial 1-2 correspondiente al análisis de la matriz de correlación compromiso (gráfico 7) pone de manifiesto las relaciones medias, para los dos colectivos de comunidades consideradas, entre los 10 grupos de gasto en alimentación estudiados. Este plano, como era de esperar por los resultados ya analizados con anterioridad, mantiene una equivalencia casi perfecta con el correspondiente al análisis de la sección anterior (gráfico 5 del AFM). Así, los gastos porcentuales en Carne y Pescado que en el primer análisis tenían un gran peso en el primer factor, han quedado relegados, en lo que a su asociación se refiere, a un segundo factor en el análisis de la tabla múltiple, apareciendo enfrentados a los gastos en leche, derivados y huevos. Sin embargo, en el primer factor de este análisis mantiene un peso relevante en la definición del mismo los gastos en carne, no ocurriendo lo mismo con los destinados a pescado. Los gastos con mayor contribución en este primer factor son los correspondientes a los grupos de aceites y grasas, fruta y hortalizas, azúcar, otros, pan y cereales y café, té y chocolate.

*Gráfico 7: Plano factorial 1-2
Análisis de la matriz compromiso (gastos)*



Se comprueba de nuevo, con todos estos indicadores del STATIS dual presentados y analizados, la estabilidad interna de los resultados obtenidos en el ACP de

los gastos porcentuales de las CCAA españolas en los grupos de alimentación considerados.

4. COMENTARIOS FINALES

Los análisis factoriales exploratorios multivariantes son de gran aplicabilidad a datos reales y la metodología existente sugiere diversas combinaciones que pueden ser objeto de estudio.

En este trabajo, se exponen dos nuevas posibilidades de aplicación de análisis de tablas múltiples a título meramente experimental.

En el caso de la primera, se estudia la posibilidad de analizar simultáneamente diversas tablas de datos de naturaleza cualitativa, referidas a distintos colectivos de individuos. Se presentan los primeros pasos, que se van a completar tanto con el estudio de relaciones, indicadores, ... del método experimentado, como con la aplicación a otras tablas de datos reales y simulados.

En el caso de la segunda, supone la propuesta de los métodos de análisis de tablas múltiples como técnicas empíricas para estudiar la estabilidad de los resultados obtenidos en un ACP. Se proponen como métodos complementarios a otras técnicas existentes (remuestreo, validación cruzada, ...), destacando su sencillez y fácil aplicabilidad.

5. REFERENCIAS

- Abascal, E. y Grande, I. (1989): "Métodos Multivariantes para la Investigación Comercial. Teoría, Aplicaciones y Programación BASIC". Ed. Ariel
- Abascal, E. y Landaluce, I. (2000): "Análisis Factorial Múltiple como técnica de estudio de la estabilidad interna de los resultados de un Análisis de Componentes Principales. Una aplicación empírica". *Questiio* (en revisión)
- Aluja, T. y Morineau, A. (1999): "*Aprender de los Datos: El Análisis de Componentes Principales*". EUB Barcelona.
- Dazy, F. y Le Barzic, J.F. (1996): "*L'Analyse des Données Evolutives*". Technip. Paris
- Greenacre, M.J. (1993). "*Correspondence analysis in practice*". Academic Press London.
- Escofier, B. y Pagès, J. (1992) "*Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación.*" Servicio editorial de la Universidad de País Vasco
- Holmes, S. (1985): "*Outils Informatiques pour l'Evaluation de la Pertinence d'un Résultat en Analyse des Données*". Thèse USTL, Montpellier
- Holmes, S. (1989): "Using the Bootstrap and the RV Coefficient in the Multivariate Context in *Data Analysis, Learning Symbolic and Numeric Knowledge*", E. Diday (ed.). Nova Science, New York, pp. 119-132
- Landaluce, M^a I. (1995). "*Estudio de la estructura de gasto medio de las Comunidades Autónomas españolas. Una aplicación del Análisis factorial múltiple*". Tesis doctoral. Universidad del País Vasco
- Landaluce, M.I., Fernández, K. y Modroño, J.I. (1999): "*Reflexiones sobre el uso comparativo del Análisis Factorial Múltiple y de la metodología STATIS para el análisis de tablas múltiples*". *Methodologica*, N^o 7 pp. 37-66

- Lavit C. [1988]. *Analyse conjointe de tableaux quantitatifs*. Masson. Paris.
- Lebart, L.; Morineau, A y Piron, M. (1995). “*Statistique exploratoire multidimensionnelle*” Dunod Paris..
- L’Hermier Des Plantes, H. [1976]. *STATIS, Structuration de Tableaux à Trois Indices de Statistique*. Thèse de Doctorat. Université de Montpellier.
- Pagès J. [1996]. “*Eléments de comparaison entre l’analyse factorielle multiple et la méthode STATIS*”. Rev. Statistique Appliquée (XLIV).
- SPAD-version 3.21 [1997]. *Logiciel diffusé par CISIA*. 1 av. Herbillon 94160 Saint-Mandé