



Comunicaciones XIV Reunión

MÉTODOS FACTORIALES DE ANÁLISIS DE TABLAS MÚLTIPLES COMO TÉCNICAS DE VALIDACIÓN DE LOS RESULTADOS DE UN ANÁLISIS DE COMPONENTES PRINCIPALES

Elena Abascal Fernández¹ - <u>eabascal@unavarra.es</u>

Mª Isabel Landaluce Calvo² - <u>iland@ubu.es</u>

¹Universidad Pública de Navarra

²Universidad de Burgos

Anales de Economía Aplicada

Oviedo 23
Junio 2000 4



Reservados todos los derechos.

Este documento ha sido extraído del CD Rom "Anales de Economía Aplicada. XIV Reunión ASEPELT-España. Oviedo, 22 y 23 de Junio de 2000".

ISBN: 84-699-2357-9

METODOS FACTORIALES DE ANALISIS DE TABLAS MULTIPLES COMO TECNICAS DE VALIDACION DE LOS RESULTADOS DE UN ANALISIS DE COMPONENTES PRINCIPALES¹

Abascal Fernández, Elena

eabascal@unavarra.es

Dpto. de Estadística e Investigación Operativa. Universidad Pública de Navarra

Landaluce Calvo, Ma Isabel

iland@ubu.es

Dpto. de Economía Aplicada. Universidad de Burgos

Palabras clave: Análisis Factorial Múltiple, Metodología STATIS, Componentes Principales, Estabilidad.

RESUMEN

Una característica de los métodos factoriales es que siempre producen resultados y éstos no son una simple descripción, sino que ponen de manifiesto la estructura existente entre los datos, de ahí la necesidad de estudiar la validez de los resultados. Es necesario analizar si representan una estructura existente entre ellos o simplemente es debida a las fluctuaciones de los datos o a la definición y codificación de las variables.

Existen diversas formas de verificar el significado de un análisis. En este trabajo, siguiendo a Lebart, se considera que el mejor criterio de validación será verificar la estabilidad de las formas obtenidas en un análisis factorial. Los estudios de esta estabilidad se realizan mediante métodos empíricos que trabajan sobre modificaciones de la tabla inicial y permiten verificar su estabilidad a través del mantenimiento de la configuración obtenida en el análisis.

El objetivo de este estudio es mostrar la eficacia de los métodos factoriales de análisis de tablas múltiples, en concreto Análisis Factorial Múltiple (AFM) y Metodología STATIS, para verificar la estabilidad de los resultados de un Análisis de Componentes Principales (ACP).

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.

1 INTRODUCCION

Una característica de los métodos factoriales es que siempre producen resultados y éstos no son una simple descripción, sino que ponen de manifiesto la estructura existente entre los datos, de ahí la necesidad de estudiar la validez de los resultados. Es necesario analizar si representan una estructura existente entre ellos o simplemente es debida a las fluctuaciones de los datos o a la definición y codificación de las variables.

Existen diversas formas de verificar el significado de un análisis. En este trabajo, siguiendo a Lebart (1995), se considera que el mejor criterio de validación será verificar la estabilidad de las formas obtenidas en un análisis factorial.

Greenacre (1993) considera dos tipos de estabilidad, interna y externa. La calidad o estabilidad interna puede verse afectada por la elección de las variables, la unidad medida, la codificación o el peso, así como por los errores de medida. En cuanto a la estabilidad externa, estudia si los datos son válidos como representativos de una población. Se considera estable si se obtiene ésta al considerar nuevas muestras. Esta última forma de estabilidad solo tiene sentido estudiarla cuando los datos proceden de un muestreo.

Los estudios de la estabilidad interna se realizan mediante métodos empíricos. Estos métodos trabajan sobre modificaciones de la tabla inicial y permiten verificar su estabilidad a través del mantenimiento de la configuración obtenida en el análisis. Las alteraciones de la tabla inicial se pueden producir en la definición y número de variables, o bien en perturbaciones de los datos. Los métodos tradicionales realizan los análisis por separado de cada tabla y miden la correlación entre los factores obtenidos en los análisis de las diferentes tablas y los factores de la original.

La estabilidad externa se puede estudiar a través de técnicas de remuestreo para obtener "zonas de confianza" representando las diferentes tablas obtenidas con las distintas muestras, sobre los mismos planos factoriales, proyectándolas como ilustrativas. El espacio común se puede obtener del análisis de la yuxtaposición de las tablas, de una tabla suma o bien de la tabla original.

El objetivo de este trabajo es mostrar la eficacia de los métodos factoriales de análisis de tablas múltiples, en concreto Análisis Factorial Múltiple (AFM) y Metodología STATIS, para verificar la estabilidad interna de los resultados de un Análisis de Componentes Principales (ACP), comparando las aportaciones de ambos métodos. La estabilidad externa ya ha sido tratada mediante STATIS, Holmes (1985,1989).

2 LAS TÉCNICAS FACTORIALES PARA LA VALIDACIÓN

Un mapa se considera estable si su forma permanece aproximadamente igual cuando se producen pequeñas alteraciones en los datos, es decir, si la orientación definida en el mismo no está determinada por aspectos aislados de los datos. Las modificaciones que se generan van destinadas a estudiar aquellos elementos que pueden incidir sobre la calidad y estabilidad de los resultados del análisis. Estas alteraciones de la tabla inicial se pueden producir en:

- Las dimensiones de la tabla, alterando el conjunto de individuos o el de variables, mediante supresión o fusión de líneas.
- En la definición o codificación de las variables.
- Simulando errores de medida en las variables mediante la adicción de perturbaciones aleatorias.

Al producir estas alteraciones se generan nuevas tablas. En todos éstos casos, se dispone de una tabla original y de otras, que pueden diferenciarse en el conjunto de individuos, en el valor de los datos o en bien en la definición o número de variables.

En todos los casos el objetivo será estudiar si la configuración de las representaciones gráficas de las diferentes tablas es la misma o si por el contrario, se producen alteraciones considerables.

2.1. El Análisis Factorial Múltiple (AFM) como técnica de validación

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 surgir de la utilización conjunta de variables de diferente naturaleza, cuantitativas y cualitativas, del empleo de tablas que provienen de otras de tres dimensiones o del manejo de un mismo conjunto de variables medidas en distintos periodos de tiempo. La posible estructuración en grupos de variables de los datos originales enriquece su estudio. En este caso los objetivos perseguidos no se limitan a la obtención de una tipología de los individuos definida a través del conjunto de variables, sino que se amplían a la búsqueda de posibles relaciones entre las estructuras obtenidas en el seno de cada uno de los grupos.

El objetivo es poner de manifiesto los principales factores de variabilidad de los individuos, estando estos últimos descritos, de manera equilibrada, por los diversos grupos de variables. Desde este punto de vista, el AFM proporciona los resultados clásicos de los análisis factoriales clásicos (Análisis en Componentes Principales, ACP, Análisis Factorial de Correspondencias, AFC, y Análisis de Correspondencias Múltiples, ACM). Esto es, eje por eje, se obtienen las coordenadas, contribuciones y

cosenos cuadrado de los individuos, los coeficientes de correlación entre las variables continuas y los factores y, para cada modalidad de las variables nominales, la coordenada, y el valor test asociado, del centro de gravedad de los individuos que presentan dicha modalidad. 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).

A cada grupo de variables se asocia una nube de individuos denominada nube parcial, que será analizada por separado, obteniendo los factores parciales. En este método se descompone, de forma aditiva, la inercia de la nube global (unión de todas las nubes parciales), inercia total, en inercia intra (inercia de las nubes correspondientes a la imagen de cada individuo desde los diferentes puntos de vista, grupos, estudiados respecto a los centros de gravedad de cada individuo) e inercia inter (inercia de la nube correspondiente a los centros de gravedad de cada individuo). Se desea 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. El AFM proporciona una representación superpuesta de estas nubes parciales proyectándolas, sobre los ejes del análisis global de todos los grupos, como elementos ilustrativos. Aquéllos individuos cuyos puntos parciales (puntos que representan a cada individuo desde bs diferentes grupos) se sitúen próximos (inercia intra débil) ilustran la estructura común de las distintas tablas analizadas. Por el contrario, aquéllos 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.

Por otro lado, el AFM puede ser interpretado como un análisis multicanónico, ya que tiene como uno de sus objetivos la búsqueda de los factores comunes a todos los grupos estudiados. En este sentido, los factores globales del AFM pueden ser considerados como las variables generales de un análisis multicanónico (variables relacionadas con el conjunto de los grupos de variables). Y los factores de las nubes parciales pueden ser interpretados como las variables canónicas (combinación lineal de las variables de un grupo más relacionada con la variable general).

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.

2.2. El método STATIS como técnica de validación

El método STATIS, introducido por Y. Escoufier y L' Hermier Des Plantes en 1976 y desarrollado por C. Lavit, permite el análisis exploratorio simultáneo de un conjunto de matrices de datos cuantitativos. Estas matrices pueden estar referidas a las mismas o a distintas variables (columnas) y a un determinado conjunto de individuos (filas) o pueden recoger información sobre las mismas variables (columnas), medidas en diferentes conjuntos de individuos (filas). A cada una de estas configuraciones le corresponde una estrategia: la primera da prioridad a las posiciones relativas de los individuos (es el método STATIS) y la segunda estudia las relaciones entre las variables (es el método STATIS Dual). Para analizar tablas que cruzan los mismos individuos y las mismas variables es posible utilizar cualquiera de las dos métodos.

El objetivo básico de este método es la búsqueda de una estructura común a las tablas consideradas, denominada *intraestuctura*. Esta búsqueda se puede formular de la siguiente manera: ¿las distancias son estables en todas las tablas analizadas?.

En este trabajo, el conjunto de tablas analizado corresponde a la segunda de las configuraciones antes presentadas, por lo que la versión de la metodología que se va a utilizar es la denominada STATIS Dual, cuya presentación general se realiza a continuación.

El método se descompone en las siguientes etapas sucesivas:

- 1. Estudio de la interestructura. Es el estudio de las diferentes tablas. Consiste en una comparación global de la estructura de todas las matrices de datos. Para ello el método STATIS utiliza la matriz de correlación como elemento representativo de cada tabla y el producto escalar de Hilbert-Schmidt para definir una distancia entre estos elementos representativos. Señalar, en este punto, que los individuos pierden total protagonismo en esta técnica. El coeficiente de asociación entre grupos es, al igual que en AFM, el coeficiente RV.
- 2. Búsqueda de un compromiso. En esta etapa se resumen todas las tablas en una sola denominada "compromiso", que será representativa de todas las consideradas en el análisis y, por tanto, de la misma naturaleza que los elementos representativos de cada una de ellas. Esta matriz compromiso se obtiene como media ponderada de las matrices de correlación correspondientes a cada grupo.
- 3. **Estudio de la** *intraestructura*. El "compromiso" determinado en la etapa anterior va a permitir representar las posiciones-compromiso de los puntos (variables), que corresponden

al conjunto de las tablas. A través de un ACP de la matriz compromiso se obtiene la imagen euclídea compromiso. Cuando las distancias entre tablas determinadas en la interestructura son pequeñas, se puede afirmar que existe una estructura común a éstas. Esta estructura queda descrita por las distancias compromiso obtenidas en esta etapa. Hay que señalar que, como en un ACP clásico, los ejes del plano principal compromiso son interpretados estudiando las correlaciones de las variables con los ejes del compromiso.

Todas las distancias consideradas en cada una de las etapas 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 la contribución de cada una de las variables, dando estas contribuciones una explicación global de la dispersión y de los principales responsables de la misma.

La metodología proporciona, además, 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 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.

3 ESTUDIO DE LA ESTABILIDAD INTERNA DE UN ACP MEDIANTE LOS DOS METODOS

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.

3.1 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. 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 adicción 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

$\begin{array}{ccc} C & & \\ C & & X_{ij} \\ A & & \\ A & & \end{array}$	$\begin{matrix}X_{ij}\\+\\N(0,1\%S_j)\end{matrix}$	$\begin{matrix}X_{ij}\\+\\N(0,10\%S_j)\end{matrix}$	$\begin{matrix}X_{ij}\\+\\N(0,20\%S_j)\end{matrix}$
--	--	---	---

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

Análisis de la intraestructura o compromiso:

Este análisis consiste en el estudio de las inercias de los puntos de las nubes parciales, con respecto a su centro de gravedad. Para ello se exponen a continuación los dos primeros planos factoriales correspondientes a las nubes de individuos, CCAA (gráfico 1) y de variables, gastos, (gráfico 2).

En ambos gráficos se puede observar una gran proximidad entre todos los puntos que representan al mismo individuo (graf.1) y a la misma variable (graf.2), resultado que pone de manifiesto la existencia de una débil inercia intra y, como consecuencia, una elevada inercia inter. Esto es, las tablas analizadas tienen una estructura muy similar.

Gráfico 1. Plano Factorial 1-2: Comunidades Autónomas. Puntos medios y Puntos parciales

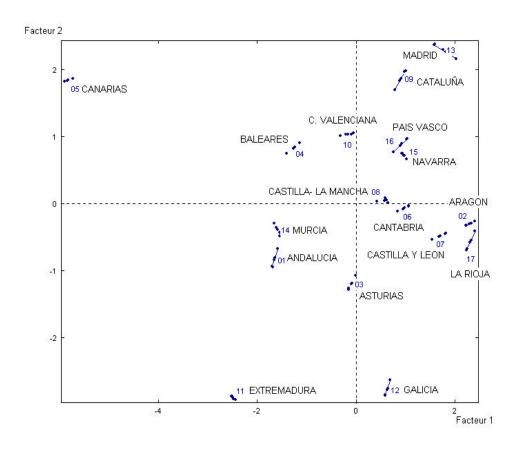
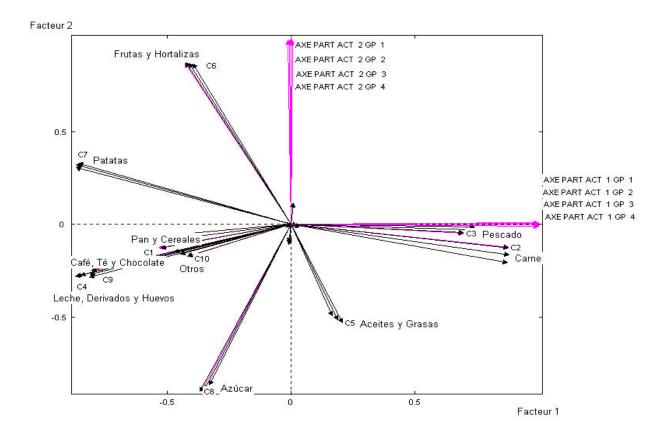


Gráfico2.Plano Factorial 1-2:

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



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.

INDIVIDUS AYANT LES PLUS FORTES INERTIES INTRA

AXE 1

+	+++
INDIVIDUS	INER CUMUL
+	++
13	25.22 25.22
16	10.03 35.25
07	9.52 44.77
04	8.87 53.64
i	

AXE 2

	INER CUMUL	INDIVIDUS
09	16.18 16.18 15.14 31.32 14.77 46.09	17 01

INDIVIDUS PARTIELS AYANT LES PLUS FORTES INERTIES INTRA

AXE 1

++ INDIVIDUS							
13 13 16 07 04	4 1 4 4	13.73 5.88 5.55 5.31 4.95	13.73 19.60 25.16 30.47 35.42				
13 +	2	4.79 +	40.21				

AXE 2

INDIVIDUS		+ INER	++ CUMUL
09	4	9.08	9.08
17	4	8.58	17.66
01	4	8.45	26.11
12	4	6.14	32.26
13	4	5.33	37.59
03	4	4.72	42.31
4			

Análisis de la Interestructura

Es el estudio comparativo de la proximidad entre las diferentes nubes. De este análisis se pueden destacar los siguientes resultados:

- La lectura de la matriz de correlaciones entre los factores parciales, 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.

MATRIZ DE CORRELACIONES ENTRE FACTEURES PARCIALES

- 1	101	102	103	104	105		
101	1.00						
102	0.00	1.00					
103	0.00	0.00	1.00				
104	0.00	0.00	0.00	1.00			
105	0.00	0.00	0.00	0.00	1.00		
	Pertu	rbació	n 1%				
201	1.00	0.00	0.00	0.00	0.00		
202	0.00	1.00	-0.01	0.00	0.00		
203	0.00	0.01	1.00	-0.01	0.00		
	0.00						
205	0.00	0.00	0.00	0.01	0.99		
Perturbación 10%							
301	1.00	-0.01	-0.01	0.03	0.02		
302	0.01	0.99	-0.13	-0.01	0.01		
303	0.01	0.13	0.99	0.01	-0.01		
304	-0.03	0.01	-0.01	1.00	0.01		
305	-0.02	-0.01	0.01	-0.01	0.99		
	Pert	urbaci	ón 20%	ó			
401	0.99	-0.02	-0.02	0.06	0.04		
402	0.01	0.97	-0.22	-0.03	0.03		
403 j	0.02	0.22	0.97	0.02	-0.02		
404	0.06	-0.02	0.03	-0.98	-0.02		
405	-0.04	-0.02	0.02	-0.02	0.96		

- Del estudio de las matrices L y RV se deduce, nuevamente, la estabilidad de los resultados obtenidos. Son grupos con una dimensionalidad parecida (se observa en bs 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).

COEFFICIENTES Lg DE RELACION ENTRE GRUPOS

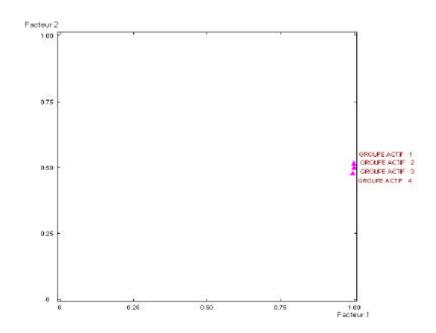
- [1	2	3 4	4
2 3	1.51 1.52 1.49 1.45	1.49		1.43

COEFFICIENTS RV DE LIAISON ENTRE GROUPES

	1	2	3	4
2	1.00 1.00 1.00 0.99	1.00		1.00

- La lectura del grafico3, 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.

Gráfico 3: Plano factorial 1-2 Grupos de variables (tablas)



COORDONNEES ET AIDES A L'INTERPRETATION DES GROUPES ACTIFS

	C	OORE	ONN	EES	- 1		CONT	RIBUT	IONS		·
GRP.	1	2	3	4	5	1	2	3 4	5		
GR 1 GR 2 GR 3 GR 4	1.00 1.00 1.00	0.52 0.52 0.50	0.44 0.45 0.41	0.18 0.18 0.18	0.10 0.10 0.10	25.0 25.0 25.0	25.6 25.7 24.8	6 26.2 7 26.9 3 24.3	24.7 24.9 25.0	25.1 24.7 24.9	

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.

- La existencia de factores comunes a todos los grupos también puede ser detectado a través del cálculo del coeficiente de correlación entre el factor global y el correspondiente a cada uno de los grupos analizados (esto es, entre las variables canónicas y las variables generales). Cuando la correlación es fuerte el factor global traduce una tendencia que está presente en todas las tablas, es decir, se trata de un factor común. En nuestro ejemplo, estas correlaciones son totales para los tres primeros factores y sólo a partir del cuarto eje algunos coeficientes descienden levemente. Por tanto, podemos concluir que son ejes que traducen una tendencia presente en todos los grupos.

CORRELATIONS ENTRE LES VARIABLES CANONIQUES ET LES FACTEURS DE L'ANALYSE GLOBALE

		DRREL		NS		+
FAC.	1	2	3	4 5		
GR 1 GR 2 GR 3 GR 4	1.00 1.00 1.00	1.00 1.00 1.00	1.00 1.00 1.00	0.99 0.99	0.99 0.98 1.00	

En este caso, todos estos resultados ponen de manifiesto la estabilidad de los resultados obtenidos en el ACP de la tabla original.

3.2 A través del STATIS DUAL

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

Análisis de la interestructura

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

1. RESULTATS SUR LA FORME DES DIFFERENTS NUAGES DE COLONNES NUAGE 1 MATRICE DE CORRELATIONS 1 | 1 2 3 4 5 6 7 8 9 10 1 | 1.000 2 | -0.251 1.000 3 | -0.725 0.299 1.000 4 | 0.342 - 0.697 - 0.556 1.000 8 | 0.335-0.153-0.257 0.539 0.369-0.677 0.006 1.000 9 | 0.286-0.605-0.437 0.765-0.136 0.010 0.596 0.452 1.000 10 | -0.226 -0.548 0.142 0.296 0.256 0.128 0.424 0.291 0.351 1.000 | 1 2 3 4 5 6 7 8 9 10 VALEURS PROPRES HISTOGRAMME DES 10 PREMIERES VALEURS PROPRES |NUM | VALEUR | PORC. | | PROPRE | CUMUL | 2.2067 1.9130 0.7755 0.4334 ***** 0.2860 | 96.52 | ***** 0.2345 | 98.87 | ***** 6 0.0978 | 99.85 | ** | 0.0139 | 99.99 | * 10 | 0.0014 |100.00 | * NUAGE 2 MATRICE DE CORRELATIONS 2 | 1 2 3 4 5 6 7 8 9 10 1 | 1.000 2 | -0.476 1.000 3 | -0.800 0.166 1.000 4 | 0.460 - 0.511 - 0.484 1.000 6 | 0.124 - 0.208 - 0.163 - 0.359 - 0.315 1.000 8 | 0.343-0.313-0.306 0.698 0.338-0.733 0.167 1.000 | 1 2 3 4 5 6 7 8 9 10

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. Unicamente 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, cuyos motivos serán estudiados, esto es, cuáles son las variables (gastos) responsables en mayor o menor medida de la misma.

```
2. SITUATION DES MATRICES DE CORRELATIONS LES UNES PAR RAPPORT AUX AUTRES PRODUITS SCALAIRES ENTRE MATRICES DE CORRELATIONS
```

DISTANCES EUCLIDIENNES ENTRE MATRICES DE CORRELATIONS

```
1 | 0.000
2 | 2.149 0.000
| 1 | 2
```

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.

INTERPRETATION DES DISTANCES ENTRE MATRICES DE CORRELATIONS CONTRIBUTIONS DES VARIABLES EN POURCENTAGE LA SOMME DES DISTANCES AU CARRE SE DECOMPOSE SELON LES VARIABLES :

PAN Y CEREALES = 5.55 % **CARNE** = 10.13 % **PESCADO** = 1.76 % LECHE, DERIVADOS Y HUEVOS = 9.74 % = 4.78 % ACEITES Y GRASAS FRUTAS Y HORTALIZAS = 16.41 % PATATAS = 17.75 % AZUCAR = 2.91 % = 18.20 % CAFÉ, TE Y CHOCOLATE OTRSO = 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.

Análisis del compromiso y de la intraestructura

- A continuación se presenta la matriz de correlación compromiso (media ponderada de las matrices de correlación asociadas a cada uno de los estados estudiados), así como las ponderaciones correspondientes a cada una de la s 2 tablas.

CONSTRUCTION DE LA MATRICE DE CORRELATIONS COMPROMIS = MOYENNE PONDEREE DES MATRICES DE CORRELATIONS PAR LES COEFFICIENTS

1 = 0.5072 = 0.493

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 aquéllos grupos con estructura interna más débil.

17

MATRICE DE CORRELATIONS COMPROMIS

```
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
  8 | 0.339-0.232-0.281 0.617 0.354-0.705 0.085 1.000
7 8
       3
            4
 | 1
                  6
                             10
```

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.

```
DECOMPOSITION DE L'INERTIE SELON LES MATRICES DE CORRELATIONS MATRICE DE CORRELATIONS 1=48.67\,\% MATRICE DE CORRELATIONS 2=51.33\,\%
```

REPRESENTATION GRAPHIQUE DE LA MATRICE DE CORRELATIONS COMPROMIS CHAQUE POINT DU NUAGE CORRESPOND A UNE DES 10 VARIABLES DECOMPOSITION DE L'INERTIE SELON LES AXES PRINCIPAUX VALEURS PROPRES TRACE DE LA MATRICE: 10.0000

```
HISTOGRAMME DES 10 PREMIERES VALEURS PROPRES
NUM | VALEUR | POUR. |
  |PROPRE |CUMUL|
 |
***********
 0.5008 | 92.24 |
       | 95.51 | ******
  10.3271
           *****
  | 0.2690 | 98.20 |
  0.1542
       99.74
 8
 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 4) 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 1 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 gatos 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.

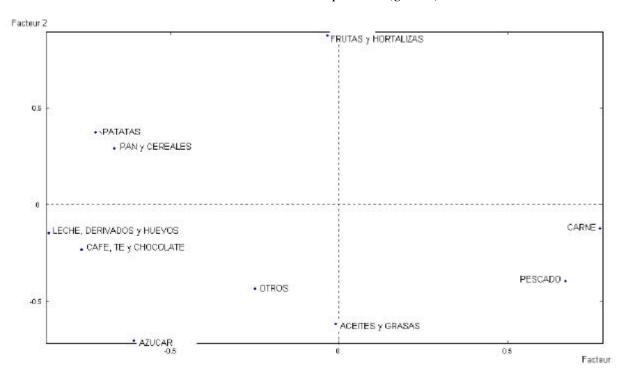


Gráfico 4: Plano factorial 1-2 Análisis de la matriz compromiso (gastos)

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 CONCLUSIONES

En este trabajo se ha estudiado la estabilidad interna de los resultados de un Análisis en Componentes Principales a través del uso de técnicas factoriales de Análisis de Tablas Múltiples: Análisis Factorial Múltiple (AFM) y metodología STATIS Dual.

Esta aplicación ha puesto de manifiesto la idoneidad de estos métodos para el objetivo perseguido dado que permiten el análisis comparativo de un conjunto de tablas definidas de diferentes formas. Asimismo, hay que destacar la riqueza de indicadores que ambas técnicas proporcionan para el estudio de las semejanzas y/o diferencias entre las tablas consideradas y, en consecuencia, de la estabilidad de las formas que definen.

Hay que señalar, para finalizar, que en este trabajo no se han examinado todas las posibilidades que estas metodologías ofrecen, ya que, por una parte, la técnica STATIS puede utilizarse también para el análisis de la estabilidad interna que se ha realizado a través del AFM (con alteraciones aleatorias de la tabla inicial). Y, además, al permitir el estudio comparativo de tablas referidas a las mismas variables y distintos individuos (alternativa usada en este trabajo), también se puede analizar con esta técnica la estabilidad externa de los resultados de un ACP. Por otra parte, el AFM analiza simultáneamente matrices de datos de naturaleza cuantitativa y cualitativa, por lo que se puede utilizar para el estudio de la estabilidad interna de los resultados proporcionados por otras técnicas factoriales como Análisis Factorial de Correspondencias (AFC) y Análisis de Correspondencias Múltiples (ACM).

5 BIBLIOGRAFÍA

- 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 practique". 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, Mª I. (1995). "Estudio de la estructura de gasto medio de las Comunidades Autónomas españolas. Una aplicación del Análisis factorial multiple". 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, Nº 7 (en imprenta)
- 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.
- SPAD-version 3.21 [1997]. Logiciel diffusé par CISIA. 1 av. Herbillon 94160 Saint-Mandé