

## ANÁLISIS FACTORIAL MÚLTIPLE COMO TÉCNICA DE ESTUDIO DE LA ESTABILIDAD DE LOS RESULTADOS DE UN ANÁLISIS DE COMPONENTES PRINCIPALES

E. ABASCAL FERNÁNDEZ\*

M<sup>a</sup>. I. LANDALUCE CALVO\*\*

*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 preciso analizar la naturaleza de esta estructura y estudiar la estabilidad de los resultados. Consideramos que el mejor criterio es el análisis de la estabilidad de los mapas obtenidos en el análisis factorial.*

*El Análisis Factorial Múltiple (AFM), desarrollado por B. Escofier and J. Pagès (1992), permite el análisis simultáneo de tablas en las que un mismo conjunto de individuos se describe a través de varios grupos de variables. En este trabajo se muestra la eficacia de este método para el análisis de la estabilidad de los resultados obtenidos mediante Análisis de Componentes Principales.*

**Multiple Factor Analysis as a statistical technique to study the stability of the results obtained by Principal Component Analysis**

**Palabras clave:** Análisis factorial múltiple, análisis de componentes principales y estabilidad

**Clasificación AMS (MSC 2000):** 62-07, 62H25

---

\* Profesora Titular de Universidad. Departamento de Estadística e Investigación Operativa. Universidad Pública de Navarra.

\*\* Profesora Titular de Universidad. Departamento de Economía Aplicada. Universidad de Burgos.

– Recibido en septiembre de 2000.

– Aceptado en enero de 2002.

## 1. INTRODUCCIÓN

Una característica de los métodos factoriales es que siempre producen resultados más o menos interpretables 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 preciso 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 consistirá en verificar la estabilidad de las formas obtenidas en un análisis factorial. El objetivo es mostrar la eficacia del método de Análisis Factorial Múltiple (AFM) para verificar la estabilidad de los resultados de un Análisis de Componentes Principales (ACP).

## 2. LOS MÉTODOS DE 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.

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 de 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 sólo 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, Lebart (1995). 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 la definición y número de variables, o bien en perturbaciones de los datos, de diferentes formas, entre otras:

- a) Por omisión de una línea para estudiar si la influencia de ésta es decisiva.
- b) Agrupamiento de varias columnas.
- c) En la elección de las variables que definen el problema.

- d) En la codificación o definición de las variables.
- e) Alterando los datos mediante la suma de un error aleatorio. Es decir, afectando los datos con perturbaciones aleatorias, de esta forma se simulan errores de medida en las variables. Así, cada observación se modifica de la forma  $x_{ij}^{(r)} = x_{ij} + z_{ijr}$  siendo  $z_{ijr}$  un valor obtenido aleatoriamente de la distribución normal de media 0 y desviación típica  $\frac{1}{k}s_j$ . Para diferentes valores de  $k$  se obtienen nuevos valores.

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

Los métodos empíricos utilizados habitualmente para determinar el número de factores estables, realizan los análisis por separado de cada tabla y miden la correlación entre los factores obtenidos en cada uno de ellos y los generados en el análisis de la tabla original. Si los ejes son estables los factores obtenidos en la diversas tablas estarán altamente correlacionados. En caso contrario, si las correlaciones son semejantes con factores de distinto orden, el eje es arbitrario resultado del azar, y no se considera estable.

En este trabajo se propone un método empírico para el análisis de la estabilidad de los resultados proporcionados por la técnica exploratoria Análisis de Componentes Principales (ACP). El procedimiento consiste en generar diferentes tablas mediante alguna de las alteraciones indicadas y analizarlas simultáneamente utilizando una metodología de análisis de tablas múltiples, el Análisis Factorial Múltiple (AFM). Este análisis permitirá la verificación de la estabilidad de los factores. La forma de alteración de la tabla inicial seleccionada para ilustrar la bondad del método ha sido la última de las anteriormente citadas, esto es, mediante la suma de errores aleatorios. A continuación se presenta una breve exposición de los principios del AFM.

### 3. EL ANÁLISIS FACTORIAL MÚLTIPLE COMO TÉCNICA DE VALIDACIÓN

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

El AFM se basa en la metodología de ACP y actúa en 2 etapas:

- 1) A cada grupo de variables se asocia una configuración euclídea o nube de individuos denominada *nube parcial*, que será analizada por separado, obteniendo los factores parciales. Es decir, realiza un ACP de cada una de las tablas  $X_j$ , donde  $X_j$  recoge la valoración de las variables del grupo  $j, j = 1, 2, \dots, J$ , en el conjunto de individuos.
- 2) Realiza un ACP de la tabla global resultado de yuxtaponer las  $J$  tablas  $X_j$ . En este análisis cada tabla  $X_j$  es ponderada por el inverso del primer valor propio obtenido en el ACP de la propia tabla,  $1/\lambda_j$ . Esta ponderación mantiene la estructura de cada tabla, ya que todas las variables han recibido la misma ponderación, pero consigue equilibrar la influencia de los grupos, ya que la inercia máxima de cada una de las nubes de individuos, definida por los diferentes grupos, vale 1 en cualquier dirección.

El objetivo global del análisis es examinar la existencia de estructuras comunes a todas, o a parte, de las nubes parciales.

#### 4. ANÁLISIS DE LA ESTABILIDAD MEDIANTE AFM

El AFM es un valioso instrumento para el estudio de la estabilidad de los resultados del ACP. Se considera que la tabla inicial, la analizada mediante ACP, constituye un grupo de variables medido sobre un conjunto de individuos, y cada una de las alteraciones generadas para medir la estabilidad constituyen un nuevo grupo.

El AFM permite el análisis simultáneo, equilibrando la influencia, de las tablas generadas con las distintas alteraciones y proporciona, como veremos, numerosos indicadores que permiten analizar la estabilidad. Genera además representaciones gráficas de gran poder ilustrativo.

- a) El estudio de la intra-estructura o compromiso permite detectar el número de factores comunes a las tablas. Si las tablas tienen la misma estructura, es decir, si el análisis es estable, se obtendrán factores comunes a todas las tablas y, en consecuencia, el número de éstos indicará el de factores estables.

Para este objetivo el AFM proporciona como principal indicador las inercias intra de los puntos individuo. Así, aquellos 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 estén alejados unos de otros (inercia intra alta), constituyen las excepciones a la estructura común.

- b) El estudio de la interestructura analiza la proximidad entre las diferentes tablas. Se realiza a través de tres acciones:

- La representación gráfica de los grupos. La coordenada de un grupo sobre un factor es la inercia acumulada del grupo sobre el eje del AFM. Indica los grupos que han determinado en mayor medida los factores y, en consecuencia, el número de factores estables. Además, el distanciamiento de un punto de los demás indica que el grupo correspondiente es el más alejado del compromiso, lo que permite detectar qué alteraciones de la tabla han producido una estructura más alejada de la común.
- Medidas globales de relación entre los grupos, basadas en el coeficiente RV de Escoufier (Dazy, 1996). 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). Para dos grupos cualesquiera  $j$  y  $h$  se tiene:

$$RV = \frac{\langle W_j D, W_h D \rangle}{\|W_j D\| \|W_h D\|} = \frac{\text{traza}(W_j D W_h D)}{\|W_j D\| \|W_h D\|}$$

Siendo  $W_j D = X_j X_j' D$  donde  $X_j$  es la matriz asociada al grupo  $j$  de variables,  $D$  una matriz diagonal que recoge los pesos de los individuos. La norma de  $W_j D$  se expresa como  $\|W_j D\| = \sqrt{\sum_s (\lambda_s^j)^2}$  donde  $\lambda_s^j$  es el  $j$ -ésimo valor propio del análisis de la tabla  $j$ . El numerador representa los productos escalares entre los representantes de los grupos.

Esta medida es completada con los coeficientes Lg (Escoufier y Pagès, 1992).

$$L_g(W_j D, W_h D) = \langle W_j D, W_h D \rangle$$

El producto escalar se interpreta como una medida de la relación entre los grupos. Del estudio de estos coeficientes se puede deducir o rechazar la estabilidad de los resultados obtenidos.

- Los coeficientes de correlación entre los factores parciales (obtenidos en el análisis parcial de cada nube) y los factores comunes o 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, obteniéndose una garantía de la estabilidad de los factores.
- c) Las representaciones gráficas simultáneas. El AFM obtiene, por un lado, las representaciones gráficas de la nube de individuos caracterizados por las variables de cada grupo (individuos parciales) y por el conjunto de variables (individuo medio) sobre el mismo plano factorial, que permiten observar las proximidades entre los puntos individuo de las diferentes tablas y detectar las posibles deformaciones. Por otro

lado, obtiene la representación de las variables que permite observar las proximidades entre los puntos variables de las diferentes tablas que corresponden a un mismo concepto. El análisis de éstas permitirá estudiar la estabilidad de cada una.

## 5. ANÁLISIS COMPARATIVO

El procedimiento propuesto es una aproximación no paramétrica en la misma línea que la propuesta por Greenacre en 1984, en la que se analiza el concepto de estabilidad en lugar de confianza. En este sentido, el procedimiento propuesto no es alternativo, sino complementario, a otros métodos paramétricos clásicos de estudio de la estabilidad como el análisis de los intervalos de confianza para los valores propios.

Cuando se trata de analizar la estabilidad interna el procedimiento clásico consiste en eliminar un elemento, cambiarlo, o simular errores en los datos para comprobar la estabilidad. En este caso, el AFM es una metodología que además de los indicadores habituales (correlaciones entre los factores) proporciona nuevas medidas que permiten estudiar la estabilidad de los factores (coeficientes RV y Lg) y detectar los elementos perturbadores, proporcionando, además, medidas de la importancia de éstos (inercias intra).

Una característica importante del AFM son las representaciones gráficas en las que se pueden estudiar las relaciones entre las variables de las diferentes tablas, lo que permite analizar la estabilidad de las variables y estudiar cómo se ven afectadas por las diferentes codificaciones, errores, etc. Si las tablas han sido generadas por remuestreo, esta representación es una forma de obtener las zonas de confianza empíricas para cada variable, semejantes a las zonas de confianza (Convex Hulls) obtenidas para los resultados del análisis de correspondencias múltiples por Greenacre (1984). La nube que contiene todas las representaciones de la misma variable correspondiente a las diferentes réplicas constituye una zona de confianza empírica para la variable.

En el caso de ACP, se han utilizado anteriormente dos procedimientos para obtener las representaciones: analizar la tabla inicial y adjuntar las réplicas como tablas ilustrativas (Lebart, 1995) o bien, analizarlas simultáneamente mediante el método STATIS (Holmes, 1985). Las diferencias entre éstos y a su vez con el AFM radican en el criterio de obtención del espacio común y en la interpretación de las representaciones. En particular, en el procedimiento de proyectar las réplicas como ilustrativas sobre el análisis de una de ellas, significa que solamente una de las tablas ha participado en la creación de los ejes y en consecuencia, no se tiene en cuenta la estructura de relaciones entre las variables de las otras tablas.

En cuanto al método STATIS, que presenta grandes similitudes con el AFM, los ejes factoriales también se obtienen de un compromiso de todas las tablas, sin embargo

su participación no está equilibrada, ya que tienen mayor contribución aquéllas que presentan una estructura interna más fuerte, esto es, con mayores similitudes.

De este breve estudio comparativo con otros métodos utilizados para el estudio de la estabilidad, se deduce que el AFM es un valioso instrumento para la determinación del número de factores y para estudiar la estabilidad de las formas. Únicamente presenta limitaciones para el estudio cuando el elemento sospechoso de producir inestabilidad (por fuerte contribución a la obtención del factor) es un único individuo.

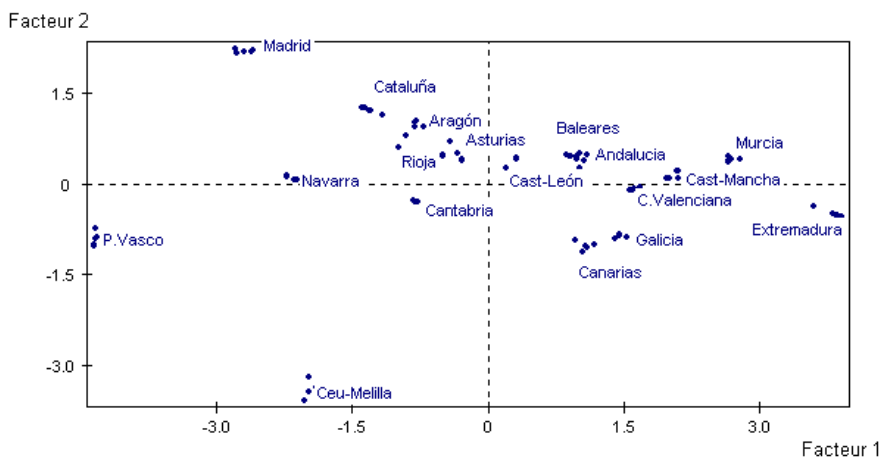
## **6. ANÁLISIS DE LA ESTABILIDAD DE LA ESTRUCTURA DE LOS SALARIOS MEDIOS DE LAS COMUNIDADES AUTÓNOMAS ESPAÑOLAS**

El objetivo de esta aplicación empírica es estudiar la estabilidad de los resultados obtenidos en un ACP de la tabla que recoge la estructura de los salarios medios de las Comunidades Autónomas españolas (CCAA) en el año 1995, a través del AFM. La tabla objeto del análisis recoge los salarios medios por ramas de actividad en las Comunidades Autónomas. Las ramas de actividad consideradas son: agricultura, ganadería y pesca (AGRICULTURA), Energía (ENERGIA), industria (INDUST), construcción (CONST), servicios de mercado (SER.MERCADO), servicios de no mercado (SER.N-MER).

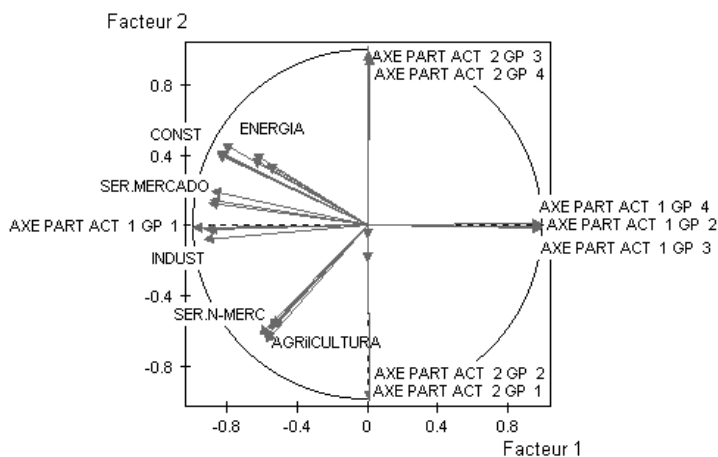
El análisis de componentes principales de esta tabla proporciona 2 factores cuyos valores propios son superiores a 1,  $\lambda_1 = 3,42$  y  $\lambda_2 = 1,15$ . Sin embargo, los intervalos de confianza de estos valores propios tiene intersección no vacía. Para estudiar su estabilidad se generan nuevas tablas con perturbaciones aleatorias, es decir, cada valor de la tabla original 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, siendo la tabla inicial el grupo 1.

### **6.1. 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, ramas de actividad, (gráfico 2).



**Figura 1.** Plano Factorial 1-2: Comunidades Autónomas. Puntos medios y Puntos Parciales.



**Figura 2.** Plano Factorial 1-2: Variables (activas) y Ejes Parciales (suplementarios) de las 4 tablas.

En ambos gráficos se puede observar una gran proximidad entre todos los puntos que representan al mismo individuo (gráf. 1) y a la misma variable (gráf. 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.



**Tabla 1.** Individuos con las mayores inercias intra.

Eje 1			Eje 2		
Individuos	INER	ACUM	Individuos	INER	ACUM
<b>RIOJA</b>	24.02	24.02	CANARIAS	23.50	23.50
<b>C. VALENCIANA</b>	16.48	40.51	CEU-MELILLA	14.69	38.19
GALICIA	10.95	51.45	<b>CAST-MANCHA</b>	13.23	51.42

Además, en lo que se refiere a los individuos el estudio minucioso de las inercias intra, tanto de los puntos medios (tabla 1) como de los puntos parciales que representan a las regiones, permite poner de manifiesto cuáles de ellas presentan un comportamiento más heterogéneo y en qué eje factorial este comportamiento es más acusado, además de señalar cuáles son los puntos parciales responsables del mismo. Así, en el primer eje factorial son las comunidades de La Rioja, de Valencia y Galicia las que tienen las mayores inercias intra. En el segundo eje factorial son Canarias, Ceuta y Melilla y la región de Castilla-La Mancha las que presentan un comportamiento menos homogéneo. Destacar que en todos los casos es el punto-parcial asociado al grupo 4 el mayor responsable de este resultado, resultado lógico si tenemos en cuenta que estos puntos corresponden a la tabla que surge de la mayor transformación aplicada a la tabla original.

**Tabla 2.** Matrices de correlaciones entre los factores parciales.

	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

**Tabla original**

	101	102	103	104	105
201	<b>1.00</b>	0.00	0.00	0.00	0.00
202	0.00	1.00	0.01	0.01	0.00
203	0.00	0.01	<b>1.00</b>	0.02	0.00
204	0.00	0.00	0.02	<b>-1.00</b>	0.00
205	0.00	0.00	0.00	0.00	<b>1.00</b>

**Perturbación 1%**

	101	102	103	104	105
301	<b>-1.00</b>	0.02	0.00	0.00	0.00
302	-0.02	<b>-1.00</b>	0.02	0.03	-0.01
303	0.00	0.01	<b>-0.99</b>	0.14	-0.01
304	0.00	-0.03	-0.14	<b>-0.98</b>	0.05
305	0.01	0.01	0.00	<b>-0.06</b>	-0.97

**Perturbación 10%**

	101	102	103	104	105
401	<b>-0.99</b>	0.02	-0.01	0.01	0.02
402	-0.02	<b>-0.95</b>	0.24	0.11	0.06
403	-0.01	0.26	<b>0.88</b>	0.31	0.01
404	0.02	0.04	-0.32	<b>0.92</b>	0.01
405	0.01	0.04	0.02	0.05	<b>0.95</b>

**Perturbación 20%**

## 6.2. Análisis de la Interestructura

Es el estudio comparativo de la proximidad entre las diferentes nubes. La lectura de la matriz de correlaciones entre los factores parciales (tabla 2), 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 manifiestan estructuras comunes a los mismos.

**Tabla 3.** Coeficientes Lg de relación entre grupos.

	1	2	3	4
1	<b>1.178</b>			
2	1.177	<b>1.176</b>		
3	1.169	1.169	<b>1.168</b>	
4	1.167	1.166	1.161	<b>1.197</b>

**Tabla 4.** Coeficientes RV de relación entre grupos.

	1	2	3	4
1	1.000			
2	<b>1.000</b>	1.000		
3	<b>0.997</b>	<b>0.997</b>	1.000	
4	<b>0.983</b>	<b>0.983</b>	<b>0.982</b>	1.000

Del estudio de las matrices L y RV (tablas 3 y 4) se deduce, nuevamente, 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).

**Tabla 5.** Correlaciones entre las variables canónicas y los factores del análisis global.

CORRELACIONES					
FAC.	1	2	3	4	5
GR 1	1.00	1.00	1.00	1.00	0.99
GR 2	1.00	1.00	1.00	1.00	0.99
GR 3	1.00	1.00	0.99	0.99	0.96
GR 4	1.00	0.99	0.97	0.98	0.90

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

este estudio empírico, las correlaciones son totales para los dos primeros factores y sólo a partir del tercer eje algunos coeficientes descienden levemente. Por tanto, podemos concluir que son ejes que traducen una tendencia presente en todos los grupos.

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

## 7. CONCLUSIONES

El AFM ha permitido estudiar las relaciones entre los factores de los análisis por separado de cada tabla, de forma semejante a otras técnicas, pero con procedimientos más directos y sencillos. Este procedimiento propuesto no es alternativo, sino complementario, a otros métodos clásicos de estudio de la estabilidad como el análisis de los intervalos de confianza para los valores propios o los métodos de remuestreo (validación cruzada, bootstrap, etc). Sin embargo, como método de validación empírica presenta buenas propiedades ya que, no sólo proporciona medidas de la estabilidad de los ejes sino también, indicadores de cómo afectan las perturbaciones en las diferentes variables a la estabilidad, así como los individuos que más contribuyen a la inestabilidad, aspectos no tratados por los métodos más habituales. Además, proporciona representaciones gráficas de las diferencias entre las tablas y entre sus estructuras sobre un marco de referencia común.

## AGRADECIMIENTOS

Los autores desean agradecer los comentarios de dos evaluadores anónimos que sin duda han ayudado a la presentación de este artículo.

## REFERENCIAS

- Abascal, E. y Grande, I. (1989). *Métodos Multivariantes para la Investigación Comercial. Teoría, Aplicaciones y Programación BASIC*. Ed. Ariel.
- 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.
- Escofier, B. y Pagès, J. (1992). *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación*. Servicio editorial de la Universidad de País Vasco.

- Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. Academic Press London.
- Greenacre, M. J. (1993). *Correspondence analysis in practice*. Academic Press London.
- Landaluce, M<sup>a</sup>. 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.
- Lebart, L., Morineau, A. y Piron, M. (1995). *Statistique exploratoire multidimensionnelle*. Dunod, París.

## ENGLISH SUMMARY

### MULTIPLE FACTOR ANALYSIS AS A STATISTICAL TECHNIQUE TO STUDY THE STABILITY OF THE RESULTS OBTAINED BY PRINCIPAL COMPONENT ANALYSIS

E. ABASCAL FERNÁNDEZ\*

M<sup>a</sup>. I. LANDALUCE CALVO\*\*

*It is commonly accepted that Factor Analysis always get results and that these results are not only descriptive but also explanatory of the underlying structure existing in the data. There is a need of analysing the nature of this structure, studying the validity of these results.*

*We consider that the best validity criterion is the analysis of the stability of the maps obtained by Factor Analysis.*

*The Multiple Factor Analysis (MFA), developed by B. Escofier and J. Pagès (1992), allows a comparative analysis of a weighted group of tables referred to the same individuals (rows) and to the same or different variables (columns) per table. In this paper we point out, that this method gives the appropriate results and graphics to assess the stability of the results obtained by PCA.*

**Keywords:** Multiple Factor Analysis, Principal Component and Stability

**AMS Classification (MSC 2000):** 62-07, 62H25

---

\* Profesora Titular de Universidad. Departamento de Estadística e Investigación Operativa. Universidad Pública de Navarra.

\*\* Profesora Titular de Universidad. Departamento de Economía Aplicada. Universidad de Burgos.

– Received September 2000.

– Accepted January 2002.

It is commonly accepted that Factor Analysis always get results and that these results are not only descriptive but also explanatory of the underlying structure existing in the data. There is a need of analysing the nature of this structure, studying the validity of these results.

Lebart (1995) consider that the best validity criterion is the analysis of the stability of the maps obtained by Factor Analysis.

A map is considered «stable» in two different ways: firstly at the level of the data matrix itself (internal stability) and, secondly, if the data are a valid sample representative of a population, at the level of the wider population (external stability). In this paper we only consider the first case.

We say that a map is «internally stable» if the orientation of the plane is not determined by isolated aspects of the data. Different ways of investigating this internal stability are possible. We consider modifications of the original table by adding to the data a random error. So, an observation  $x_{ij}$  is modified in the following way:

$$x_{ij}^{(r)} = x_{ij} + z_{ijr}$$

where  $z_{ijr}$  is a normally distributed random variable with 0 mean and  $\frac{1}{k}s_j$  standard deviation. The different values of  $k$  provide different tables.

The tables thus obtained and the original table are joined together forming a broad table containing the same rows and columns but with different values.

MFA, developed by B. Escofier and J. Pagès (1992), allows a comparative analysis of a weighted group of tables referred to the same individuals (rows) and to the same or different variables (columns) per table. In this paper we point out, through an empirical application, that this method gives the appropriate results and graphics to assess the stability of the results obtained by PCA.