

# ANÁLISIS FACTORIAL CONJUNTO DE TABLAS DE CONTINGENCIA

Zárraga, A. y Goitisoló, B.  
Departamento de Economía Aplicada III,  
*Universidad del País Vasco-Euskal Herriko Unibertsitatea*

## Resumen

El objetivo de este trabajo consiste en desarrollar un método factorial para el análisis conjunto de varias tablas de contingencia que permita, de forma semejante al análisis de correspondencias, el estudio de la semejanza entre el conjunto de las filas, de las columnas y las relaciones entre ambos conjuntos. En este análisis conjunto se considera que dos filas son semejantes si lo son en los análisis de las tablas separadas.

Con el método propuesto es posible analizar conjuntamente varias tablas de contingencia, propiamente dichas, así como aquéllas que corresponden, por ejemplo, a una misma encuesta realizada en distintos países o en diferentes periodos de tiempo. Otro caso interesante para el análisis simultáneo de tablas se presenta cuando se poseen encuestas sobre temas diferentes realizadas sobre la misma población.

El método busca un análisis de las tendencias generales de las tablas, sin permitir que una de ellas tenga una preponderancia sobre el resto.

El método propuesto es extensible al análisis de un conjunto de encuestas cuando los datos de los distintos grupos de individuos son codificados en tablas disyuntivas completas y yuxtapuestas estas para las mismas modalidades y al análisis de encuestas en las que existen preguntas condicionadas.

**Keywords:** Análisis Factorial, Conjunto de Tablas de Contingencia, Análisis de Correspondencias

**Clasificación AMS:** 62H25

**Agradecimientos:** Este trabajo ha sido financiado por el Proyecto de Investigación UPV 038.321-HA041/99 de la Universidad del País Vasco (UPV/EHU) y el proyecto PB98-0149 de la Dirección General de Enseñanza Superior e Investigación Científica del Ministerio Español de Educación y Cultura.

# 1 Introducción

La necesidad de analizar simultáneamente varias tablas de datos ha dado lugar a la existencia de numerosos métodos factoriales. Sin embargo, la mayoría de ellos requieren tablas de variables cuantitativas y/o exigen que las diversas tablas de datos hagan referencia al mismo conjunto de individuos. Algunos de estos métodos serán comentados en § 3.

El objetivo de este trabajo consiste en desarrollar un método factorial para el análisis conjunto de varias tablas de contingencia que permita, de forma semejante al análisis de correspondencias, el estudio de la semejanza entre el conjunto de las filas, de las columnas y las relaciones entre ambos conjuntos. En este análisis conjunto se considera que dos filas son semejantes si lo son en los análisis de las tablas separadas.

Para tratar de analizar conjuntamente tablas de contingencia, algunos de los métodos que se han propuesto consisten en analizar la tabla obtenida como suma de las tablas de contingencia separadas, lo que exige que las modalidades en filas y columnas de las tablas de contingencia sean las mismas, analizar la tabla obtenida como yuxtaposición de las iniciales, el análisis intra (Escofier 1983) y más recientemente el análisis factorial múltiple intra-tablas (Bécue-Bertaut & Pagès 2000). Sin embargo, en § 2 y § 3 se muestra como existen situaciones donde ninguno de estos métodos permite un análisis de las semejanzas entre filas que mantenga la similitud existente en los análisis de las tablas separadas.

Con el método propuesto es posible analizar conjuntamente varias tablas de contingencia propiamente dichas, así como aquéllas que corresponden, por ejemplo, a una misma encuesta realizada en distintos países o en diferentes periodos de tiempo. Las filas de cada tabla pueden ser grupos de individuos según el sexo, el intervalo de edad al que pertenecen y su situación laboral y en columnas las respuestas a preguntas sobre la familia (¿Está a favor del divorcio?, ¿Cree que una mujer con hijos debería trabajar?, etc). Los resultados del análisis de una tabla de este tipo pueden ser distintos según los datos pertenezcan, por ejemplo, a España, a Italia o a Suecia y según hayan sido obtenidos en 1970, 1980 o 1990. El análisis de la yuxtaposición o suma puede revelar únicamente las relaciones dentro de una tabla o incluso unas relaciones que no existen en el análisis de cualquiera de las tablas por separado (§ 2). Con el análisis conjunto propuesto se pretende responder a preguntas del tipo, ¿Las características de los individuos (fila de las tablas) que responden estar a favor del divorcio son las mismas en los diferentes países?, ¿La opinión sobre si una madre debería trabajar ha variado en el tiempo?, ¿en qué grupos de individuos?, etc.

Otro caso interesante para el análisis simultáneo de tablas se presenta cuando se poseen encuestas diferentes realizadas sobre la misma población. Así, por ejemplo, se pueden tener las tablas con las mismas filas que en el ejemplo anterior (edad, sexo, situación laboral, ...) y columnas referentes a la familia en la primera tabla y sobre bienestar social en la segunda. El tipo de información que se obtiene es claramente diferente en los análisis separados. En el análisis conjunto se buscaría responder a cuestiones tales como, ¿Qué grupos de personas opinan que una madre debería cobrar un subsidio (obtenida de la encuesta sobre familia) y que debería tenerse en cuenta el número de hijos de un trabajador en la determinación de su salario (obtenida de la encuesta sobre bienestar social)?

Para poder responder a estas preguntas es necesario un análisis de las tendencias generales de las tablas, sin permitir que una de las tablas tenga una preponderancia sobre el resto.

El método propuesto es extensible al análisis de un conjunto de encuestas cuando los datos de los distintos grupos de individuos son codificados en tablas disyuntivas completas y yuxtapuestas estas para las mismas modalidades y al análisis de encuestas en las que existen preguntas condicionadas.

Dado que tenemos tablas de contingencia, al hablar de un grupo estamos haciendo referencia al conjunto de columnas pertenecientes a la misma tabla.

## 2 Problemas

Para mostrar los diversos problemas que existen al tratar varias tablas de contingencia simultáneamente, a los que tratamos de dar solución con el método propuesto en § 5, comenzaremos con un ejemplo sencillo obtenido de (Wardrop 1995) como ilustración de la Paradoja de Simpson (Simpson 1951). Esta se basa en el hecho de que la suma de tablas multivariantes puede no reflejar la misma información que el análisis de las tablas individuales.

Los datos corresponden a los resultados durante los años 1980 a 1982 de Larry Bird y Rick Robey, dos jugadores de los Boston Celtics de Baloncesto. Cada una de las tablas recoge los resultados de uno de los jugadores en dos tiros libres, el primero en filas y el segundo en columnas según que el resultado sea canasta (C) o un fallo (F).

Ejemplo 1: Paradoja de Simpson

	Larry Bird			Rick Robey		
	C2	F2		C2	F2	
C1	251	34	285	54	37	91
F1	48	5	53	49	31	80
	299	39	338	103	68	171

El artículo contrasta si es más probable obtener una canasta en el segundo intento después de encestar que después de fallar en el primero (fenómeno conocido por "hot hand").

Los contrastes indican que no existe dicho fenómeno cuando se analizan los datos correspondientes a Larry Bird y tampoco cuando se analizan los pertenecientes a Rick Robey. Sin embargo, si se suman ambas tablas los datos muestran el fenómeno.

Sobre estas tablas hemos realizado un análisis de correspondencias simples y encontramos la misma situación paradójica si se analizan las tablas individualmente respecto a su suma y también respecto a la yuxtaposición de las mismas.

Este ejemplo muestra claramente que el análisis factorial de un conjunto de tablas de contingencia no debe limitarse a una simple suma o yuxtaposición de las mismas.

El análisis de correspondencias simple de cada tabla por separado muestra una asociación entre C1 y F2 frente a F1 y C2 con unos valores propios de 0.0008069 y 0.0003789 para Larry Bird -Figura 1 (a)- y Rick Robey -Figura 1 (b)-, respectivamente.

Si analizamos conjuntamente ambas tablas esperamos obtener la misma asociación puesto que coincide en ambas tablas.

El análisis de la suma de ambas tablas, además de no permitir analizar la semejanza o diferencia entre los dos jugadores muestra la relación entre resultados iguales en los dos tiros -Figura 1 (c)-.

El análisis de la yuxtaposición de las tablas -Figura 1 (d)- muestra que Larry encesta el primer tiro (C1) con mayor frecuencia y enceste o falle el segundo tiro, frente a los primeros fallos de Rick (F1).

Este ejemplo sirve para ilustrar que el análisis factorial de un conjunto de tablas de contingencia basado en el análisis de la tabla suma de las iniciales y de la tabla obtenida como yuxtaposición de las mismas no es adecuado porque distorsiona la información que contienen las tablas iniciales.

El ejemplo pone de manifiesto además algunos de los problemas del análisis factorial de un conjunto de tablas de contingencia; como son el diferente efectivo de las tablas (338 frente a 171) y totales diferentes de la misma fila en las distintas tablas que suponen pesos de las filas diferentes (0.843 y 0.156 frente a 0.532 y 0.4678 respectivamente). Además cuando las tablas son de mayor dimensión y/o la información de cada una es diferente el análisis conjunto se verá influido por las inercias de las distintas tablas en cada eje. Con el propósito de poder ilustrar esto último utilizaremos el siguiente ejemplo:

### Ejemplo 2

	Tabla 1				Tabla 2			
	a	b	c		a	b	c	
A	66	7	7	80	10	0	0	10
B	7	2	1	10	0	44	1	45
C	7	1	2	10	0	1	44	45
	80	10	10	100	10	45	45	100

Las representaciones sobre un plano de los análisis separados de la tabla 1 -Figura 2 (a)-<sup>1</sup> y de la tabla 2 -Figura 2 (b)- son semejantes, el primer eje opone A frente a B y C mientras el segundo eje opone las modalidades B y C. El primer valor propio de cada uno de los análisis es 0.015625 y 1 de una inercia total de 0.025625 y 1.9130, respectivamente.

Dado que cada una de las tablas es simétrica, el análisis de las filas (A,B,C) coincide con el de las columnas (a,b,c).

Tanto en el análisis de la suma de ambas tablas -Figura 2 (d)- como en el análisis de la

---

<sup>1</sup>Las modalidades de la primera tabla (respectivamente la segunda) se denotarán con la letra seguida de un 1 (respectivamente de un 2).

yuxtaposición de ambas -Figura 2 (c)- el orden de los ejes cambia respecto a los análisis separados.

Para comprender la importancia de cada uno de los problemas mencionados, se analizará cada uno de ellos de forma separada modificando sobre el ejemplo anterior la segunda tabla y el efectivo de la primera.

Aunque el ejemplo 2 y los presentados a continuación, a diferencia del primero obtenido de datos reales, corresponden a situaciones extremas que difícilmente se dan en la práctica ilustran claramente los problemas de las situaciones reales (donde generalmente se presentan con menor magnitud pero simultáneamente). Sin embargo, los resultados generales obtenidos con ellos han sido verificados en gran cantidad de ejemplos analizados.

Puesto que la agregación de los datos supone una pérdida de información que incluso puede conllevar una interpretación errónea centraremos nuestro interés en el tratamiento de la juxtaposición.

## 2.1 Diferente efectivo de las tablas

Cuando el total de cada tabla (número de individuos) es distinto, el análisis de la juxtaposición de las tablas de contingencia se verá influenciado por este hecho; la tabla con mayor número de individuos tendrá una mayor contribución en la creación de los ejes, llegando incluso a crear el primero si su inercia coincide (o no es mucho menor) con la de la segunda tabla. Además hay que tener en cuenta que el efectivo relativo de las tablas no siempre tiene un significado. Si las tablas corresponden a encuestas diferentes de la misma población o a la misma encuesta realizada en diferentes poblaciones el efectivo de las mismas dependerá, por ejemplo, del presupuesto asignado a la realización del estudio.

Para ver la influencia del efectivo de cada tabla consideremos el siguiente ejemplo:

Ejemplo 3

	Tabla 1				Tabla 2			
	a	b	c		a	b	c	
A	66000	7000	7000	80000	65600	7200	7200	80000
B	7000	2000	1000	10000	7200	2025	775	10000
C	7000	1000	2000	10000	7200	775	2025	10000
	80000	10000	10000	100000	80000	10000	10000	100000

En estas tablas coinciden las marginales fila y columna y los valores propios (0.015625 y 0.01 para el primer y segundo eje) en ambas tablas. En la primera tabla, que coincide con la del ejemplo 2 pero con un efectivo mayor, -Figura 3 (a)- el primer eje enfrenta A con B y C mientras el segundo eje opone las modalidades B y C. El primer eje de la tabla 2 (respectivamente el segundo) corresponde al segundo eje (primero) de la tabla 1 -Figura 3 (b)-.

A diferencia de los ejemplos 1 y 2 existe una oposición entre la información de ambas tablas.

Al realizar el análisis de la yuxtaposición de las dos tablas -Figura 3 (e)- se comprueba que los dos ejes coinciden en interpretación (A frente a B, C) y en inercia de los ejes(0.0128125), sin permitir observar la oposición entre las modalidades B y C.

Cuando una cualquiera de las dos tablas es multiplicada por un valor, por ejemplo por 2, el primer eje del análisis de la tabla obtenida de la yuxtaposición corresponde a la tabla con mayor efectivo (-Figura 3 (c)- al multiplicar la primera tabla por 2 y -Figura 3 (d)- al multiplicar por 2 la segunda tabla).

Por tanto, el primer paso para equilibrar la influencia de las tablas, cara a un posterior análisis conjunto, será transformar cada una de ellas en frecuencias.

## 2.2 Pesos distintos de filas

Otro de los aspectos determinantes en el análisis conjunto de tablas de contingencia y posible responsable de una situación paradójica se presenta cuando ambas tablas tienen marginales diferentes.

En algunas situaciones, el análisis de diferentes encuestas (sobre hábitos de consumo y sobre medio ambiente, por ejemplo) en la misma población y en el mismo instante de tiempo, es aceptable considerar que las marginales sobre las filas (el peso de cada perfil fila) es aproximadamente el mismo. Sin embargo, cuando los grupos pertenecen a poblaciones diferentes (caso de la misma encuesta realizada en diferentes países) o la misma en diferentes periodos de tiempo, esas marginales no tienen por qué coincidir.

Las relaciones existentes dentro de cada tabla se verán alteradas si en el análisis no se mantienen los pesos y la métrica de la misma.

Por ejemplo si retomamos el ejemplo 1, transformamos cada tabla en frecuencias para evitar la influencia del distinto efectivo de cada una, y analizamos la yuxtaposición de las tablas transformadas, se comprueba que, en este caso, el análisis de la yuxtaposición -Figura 1 (f)- es similar al obtenido con las tablas de frecuencias absolutas. Es decir, no se mantiene la estructura interna de cada una de las tablas. En realidad esto sucede porque el análisis separado de cada tabla cambia al modificar los pesos propios de las filas por otros pesos (en la yuxtaposición, por la media de los pesos de las dos tablas). En el análisis de la tabla 1 del ejemplo 2 con el peso medio -Figura 2 (e)- se modifican las inercias de los ejes pero no el significado de los mismos. En cuanto a la segunda tabla, cambia el orden de los ejes -Figura 2 (f).

Este inconveniente nos induce a buscar un método de análisis factorial conjunto de tablas de contingencia donde se mantenga para cada tabla los pesos y también la métrica.

## 2.3 Inercias de las distintas tablas

El distinto orden de magnitud de la inercia proyectada sobre el primer eje en cada una de las tablas puede llevar a un análisis conjunto dominado por las diferencias de una de las tablas.

## Ejemplo 4

	Tabla 1				Tabla 2			
	a	b	c		a	b	c	
A	660	70	70	800	644	78	78	800
B	70	20	10	100	78	19	3	100
C	70	10	20	100	78	3	19	100
	800	100	100	1000	800	100	100	1000

La tabla 1 es de nuevo la primera del ejemplo 2 -Figura 4 (a)-. La tabla 2 -Figura 4 (b)- tiene un primer eje con un valor propio de 0.0256 que enfrenta las modalidades B y C (eje 2 de la tabla 1) y un segundo eje con un valor propio de 0.000625 que enfrenta la modalidad A a B y C. En este caso coinciden los efectivos y marginales de ambas tablas, siendo la inercia total de cada una muy próxima (0.025625 para la primera tabla y 0.026225 para la segunda).

Lo que distingue los análisis separados es la diferente inercia de cada uno de los ejes, fundamentalmente del primero. El análisis de las tablas yuxtapuestas -Figura 4 (c)- evidencia en primer lugar la oposición entre las modalidades B y C, puesto que este eje (segundo de la primera tabla y primero de la segunda) tiene una inercia mayor en los análisis separados.

Con el fin de equilibrar la influencia que cada tabla ejerza en el análisis conjunto y de no permitir que este último esté dominado por una tabla en concreto se introducirá una ponderación sobre las tablas.

### 3 Tratamientos que existen

Cuando la tabla de datos es de objetos (en general individuos) por variables **cuantitativas** existen varios métodos que analizan dos o más grupos de variables conjuntamente. Algunos de ellos exigen no sólo que el conjunto de individuos sea el mismo sino también que las variables coincidan y otros se limitan a la comparación de dos tablas.

El **Análisis Canónico** (Hotelling 1936) se limita a la comparación de dos tablas de datos individuos por variables (las mismas o distintas). Fue posteriormente generalizado al análisis de más tablas (Carroll 1968). El análisis canónico generalizado incluye como casos particulares el análisis discriminante, la regresión múltiple y el análisis de correspondencias múltiples (si las tablas analizadas son tablas disyuntivas completas) y el análisis en componentes principales normado.

El **Análisis Procrusteano** generalizado (Gower 1975), (Gower 1984) tiene su origen (Tucker 1958) en el estudio de dos conjuntos de variables (las mismas o distintas) medidas sobre el mismo conjunto de individuos. Busca una transformación de la nube de individuos de una de las tablas mediante translación y rotación para que se asemeje lo más posible a la segunda nube y posibilita de esta forma la comparación entre los grupos.

El objetivo del modelo **INDSCAL** (Individual Differences in Scaling) (Carroll & Chang 1970) y su caso particular el modelo **IDIOSCAL** (Individual Differences in Orientation

Scaling) (Tucker 1972) es la comparación de  $J$  matrices de disimilaridades. Exige no sólo que los individuos coincidan sino también las mismas variables en cada uno de los grupos.

Cuando las tablas corresponden a las mismas variables medidas para los mismos individuos en diferentes periodos de tiempo se pueden mencionar el **Doble Análisis en Componentes Principales** (Douroche & Dussaix 1972) y el **Análisis Factorial Evolutivo** (Chevalier 1977) que pondera cada una de las tablas asignando mayor importancia a aquellas más próximas en el tiempo.

Dado que el método que vamos a proponer permite variables diferentes o la misma variable codificada de forma distinta (por ejemplo el nivel de estudios en los países de la Unión Europea) y que consideramos importante la introducción de una ponderación sobre las tablas para equilibrar la influencia de estas en el análisis conjunto mencionaremos brevemente el **Análisis Factorial Múltiple** y el método **Statis**.

En todos los métodos anteriores cada objeto tiene el mismo peso en cada una de las tablas por separado y en el análisis conjunto. Este peso es en la mayoría de los métodos el inverso del número de objetos (generalmente individuos).

Cuando las tablas son de contingencia además de la suma y yuxtaposición de las mismas descartadas en el apartado anterior hay que mencionar el **Análisis Intra** sin ponderación de las tablas o con la ponderación propuesta en el **Análisis Factorial Múltiple Intra-tablas**.

Tanto el Análisis Factorial Múltiple como el método Statis buscan la intraestructura del conjunto de tablas representando los individuos compromiso calculados a partir de la representación de cada uno de los individuos en las distintas tablas (los individuos parciales). Además analizan una nube donde cada una de las tablas es representada por un punto para estudiar las diferencias entre las distintas tablas (Interestructura).

### 3.1 Análisis Factorial Múltiple

El Análisis Factorial Múltiple (Escofier & Pagès 1984) estudia tablas de los mismos objetos (individuos) por variables (cuantitativas o cualitativas) estructuradas en grupos. Con el propósito de que el análisis conjunto no esté influenciado de manera preponderante por uno sólo de los grupos, cada uno se pondera por el inverso del primer valor propio del análisis separado del grupo.

Esta ponderación contrae la nube sin alterar la estructura interna del grupo. Iguala a 1 la inercia del primer eje de cada tabla impidiendo que cualquiera de los grupos pueda determinar por sí sólo el primer eje del análisis conjunto.

Desde el punto de vista de los individuos la ponderación elegida supone la elección de un individuo compromiso igual a la suma de los individuos representativos de cada tabla.

El análisis de la intraestructura se realiza llevando a cabo un análisis en componentes principales de los individuos compromiso, lo que permite utilizar las reglas habituales de interpretación entre individuos y variables.



El estudio de la interestructura (en el que cada grupo de variables es representado por un punto) se realiza en referencia a los ejes de intraestructura por lo que la interpretación es sencilla aunque la calidad de representación no es óptima.

Sobre los ejes factoriales del análisis conjunto se proyectan, además de las variables, los individuos vistos a través de cada uno de los grupos y a través del conjunto.

### 3.2 STATIS

El método STATIS (L'Hermier des Plantes 1976) analiza conjuntamente varias tablas de datos de individuos (los mismos) por variables cuantitativas que pueden ser distintas en cada tabla y posteriormente se generaliza el método a tablas de distintos individuos por las mismas variables.

La interestructura estudia los grupos de variables proyectándolos sobre los ejes que mejor los representan aunque estos ejes son difícilmente interpretables.

La nube compromiso elegida por STATIS (Structuration des Tableaux À Trois Indices de la Statistique) es una media ponderada de las nubes normalizadas (la inercia total de cada uno de los grupos es la unidad). Con la ponderación se pretende penalizar aquellas nubes que son más diferentes.

El individuo compromiso es una combinación lineal de objetos representativos de tal forma que esté lo más correlacionada posible a los objetos en el sentido del producto escalar de Hilbert-Schmidt.

### 3.3 Análisis Intra

El análisis intra desarrollado en (Escofier & Pagès 1992) realiza un análisis conjunto de tablas de contingencia que pueden tener diferente marginal fila y columna y distinto efectivo. Busca la comparación de perfiles de columnas homólogas. Para ello se centra cada subnube de perfiles homólogos respecto a su perfil medio manteniendo las distancias entre perfiles. Posteriormente se realiza un análisis de correspondencias simples sobre la yuxtaposición de tablas modificadas cuyo término general se puede expresar de la siguiente forma:

$$f_{ijt} - \frac{f_{ij.}}{f_{.j.}} f_{.jt} + f_{i..} f_{.jt}$$

donde los subíndices  $i$ ,  $j$  y  $t$  representan una fila, una columna y una tabla respectivamente, y siendo  $f_{ijt} = k_{ijt}/k_{...}$ , es decir, donde las frecuencias relativas se calculan en relación al total de todas las tablas yuxtapuestas.

La yuxtaposición de estas tablas modificadas tiene las mismas marginales fila y columna ( $f_{i..}$  y  $f_{.jt}$ ) que la yuxtaposición de las tablas originales.

Este análisis se plantea como complementario a la yuxtaposición de las tablas de contingencia cuya inercia puede descomponerse en una inercia inter y una inercia intra.

Es también posible realizar un análisis intra tabla en lugar de intra columnas homólogas, es decir, centrando los perfiles columna de la misma tabla respecto a su baricentro (marginal columna). Esto supondría realizar un análisis de correspondencias simple sobre la tabla yuxtapuesta de término general:

$$f_{ijt} - \frac{f_{i,t}}{f_{..t}} f_{.jt} + f_{i..} f_{.jt}$$

Sin embargo, este análisis no mantiene la estructura interna de cada tabla analizada por separado. Por una parte cuando las tablas tienen diferente efectivo las frecuencias relativas del análisis conjunto ( $k_{ijt}/k_{...}$ ) y de los análisis separados ( $f_{ijt} = k_{ijt}/k_{.t}$ ) son diferentes. Por otra parte, aunque no existan problemas de distinto efectivo de las tablas estamos modificando los pesos asignados a los perfiles fila y la métrica de los perfiles columna respecto a los análisis separados ( $f_{i..}$  respecto a  $f_{i.t}$ ).

Si en el ejemplo 2 aplicamos el análisis intra obtenemos -Figura 2 (g)- un primer valor propio de 0.748 asociado al segundo eje de los análisis separados de cada una de las tablas (modalidad B frente a C)

### 3.4 AFM Intra-Tablas

En (Bécue-Bertaut & Pagès 2000) se modifica el análisis intra tablas anterior ponderando éstas por el primer valor propio de los análisis separados de las tablas. En concreto se propone analizar la tabla de término general:

$$\frac{f_{ijt} - \frac{f_{i,t}}{f_{..t}} f_{.jt}}{f_{i..} f_{.jt}}$$

con pesos  $f_{i..}$  para las filas y pesos  $f_{.jt}$  para las columnas. De esta forma las filas tienen el mismo peso, el peso medio calculado sobre el conjunto de las tablas, en todos los análisis.

Al igual que en el análisis intra-tablas la imposición de unos pesos medios modifica la estructura interna de cada tabla distorsionando el análisis. La inclusión de la ponderación a cada tabla es incapaz de corregir este efecto. La aplicación de este análisis a los datos del ejemplo 2 -Figura 2(h)- muestra de nuevo el segundo eje de los análisis separados.

## 4 Transformación de los datos de una tabla de contingencia

Sea una tabla de contingencia donde  $\mathcal{I} = \{1, \dots, i, \dots, I\}$  y  $\mathcal{J} = \{1, \dots, j, \dots, J\}$  representen el conjunto de filas y de columnas de la tabla que corresponden a las modalidades de dos variables cualitativas. Sea  $k_{ij}$  y  $f_{ij}$  respectivamente las frecuencias absoluta y relativa de pertenecer a las modalidades  $i$  y  $j$  y sea el total de la tabla  $k$ .

- Existen varias formas de realizar un análisis de correspondencias simples. La más frecuente es definir los perfiles (fila o columna) centrados, en la matriz  $X$ , con sus

pesos , en la matriz diagonal  $P$ , y la métrica asociada a la distancia  $\chi^2$ , recogida en una matriz diagonal  $M$ . En el análisis de las filas los elementos genéricos de estas matrices son:<sup>2</sup>

$$x_{ij} = \frac{f_{ij}}{f_i} - f_{.j} \quad p_{ii} = f_i. \quad m_{jj} = \frac{1}{f_{.j}} \quad (1)$$

Se sabe que en análisis factorial (componentes, correspondencias simples y múltiples) se maximiza la inercia proyectada sobre los vectores  $u_s$  que forman las columnas de la matriz  $u$ :

$$u^T M X^T P X M u \quad (2)$$

De la diagonalización de la matriz simétrica:

$$A = M^{1/2} X^T P X M^{1/2} \quad (3)$$

se obtienen los valores propios  $\lambda_s$ , inercia proyectada sobre el eje  $s$ , y unos vectores propios  $u_s$ <sup>3</sup> normalizados para la métrica  $M$ :

$$u^T M u = I \quad (4)$$

donde  $I$  es la matriz identidad.

Las proyecciones del conjunto de perfiles fila sobre el eje  $s$ ,  $F_s$  son:

$$F_s = X M u_s$$

Dado que en una tabla de contingencia filas y columnas juegan papeles simétricos el estudio de los perfiles columna puede llevarse a cabo mediante el análisis de la matriz  $X$  con pesos  $P$  y métrica  $M$  de términos generales:

$$x_{ji} = \frac{f_{ij}}{f_{.j}} - f_i. \quad p_{jj} = f_j \quad m_{ii} = \frac{1}{f_i} \quad (5)$$

Existe además una relación entre los análisis de filas y columnas. Denotando  $F_s(i)$  la proyección de la fila  $i$  sobre el eje  $s$  y por  $G_s(j)$  la proyección de la columna  $j$  sobre el mismo eje, se tiene:

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{j \in \mathcal{J}} \frac{f_{ij}}{f_i} G_s(j)$$

$$G_s(j) = \frac{1}{\sqrt{\lambda_s}} \sum_{i \in \mathcal{I}} \frac{f_{ij}}{f_{.j}} F_s(i)$$

donde  $\lambda_s$  es el valor propio en el análisis tanto de filas como de columnas.

---

<sup>2</sup>En lo que sigue las matrices  $P$  y  $M$  serán siempre diagonales y se definirán, al igual que la matriz  $X$ , a través de un elemento general de las mismas. El orden de las matrices corresponde, excepto en las matrices en que se especifique lo contrario, a los cardinales de los conjuntos a los que pertenecen los subíndices del término general correspondiente.

<sup>3</sup> $\mathcal{S} = \{1, \dots, s, \dots, S\}$

- Sin embargo, esta no es la única forma en que puede llevarse a cabo el análisis.

En (Escofier & Pagès 1992) el análisis se realiza utilizando unas matrices  $X$ ,  $P$  y  $M$  de tal forma que la matriz  $X$  considerada en el análisis de las columnas coincide con la transpuesta de la utilizada en el análisis de las filas y donde las matrices de pesos y métrica en el análisis de las filas son las matrices correspondientes a la métrica y pesos respectivamente en el análisis de las columnas. Es decir, se consideran las matrices de elementos:

$$x_{ij} = \frac{f_{ij}}{f_i \cdot f_j} - 1 \quad p_{ii} = f_i \quad m_{jj} = f_j \quad (6)$$

La nube de los perfiles fila (respectivamente columna) definida por la matriz  $X$  ( $X^T$ ) con los pesos  $P$  ( $M$ ) y la métrica  $M$  ( $P$ ) es isomorfa de la obtenida en (1) (respectivamente en (5)).

En este caso la matriz a diagonalizar en el análisis de las filas tiene también la expresión (3) con los valores propios  $\lambda_s$  y unos vectores propios  $u_s$  de los que se obtienen unos ejes de inercia normalizados para la métrica  $M$  (4).

El análisis de las columnas lleva a diagonalizar la matriz:

$$D = P^{1/2} X M X^T P^{1/2}$$

obteniéndose los mismos valores propios que en el análisis de las filas y unos vectores propios  $v_s$  normalizados para la métrica  $P$ .

Las proyecciones del conjunto de filas y de columnas en este análisis son:

$$\begin{aligned} F_s &= X M u_s = X M G_s \frac{1}{\sqrt{\lambda_s}} \\ G_s &= X^T P v_s = X^T P F_s \frac{1}{\sqrt{\lambda_s}} \end{aligned}$$

coincidentes con las del punto anterior.

- Es también frecuente (Lebart, Morineau & Fenelon 1985) realizar el análisis de los perfiles centrados modificados de tal forma que la distancia euclídea entre dos perfiles transformados coincida con la distancia  $\chi^2$  de los perfiles originales. Para los perfiles columna esta transformación se expresa como:

$$x_{ji} = \frac{1}{\sqrt{f_i}} \left( \frac{f_{ij}}{f_j} - f_i \right) \quad p_{jj} = f_j \quad m_{ii} = 1 \quad (7)$$

La proyección de estos perfiles se realiza igual que en el análisis con métrica  $1/f_i$  porque el efecto de la métrica es absorbido por los vectores.

Este análisis es equivalente al realizado con las matrices (5).

Este análisis correspondería a una transformación de los perfiles fila donde se ha modificado, no la métrica, sino los pesos:

$$x_{ij} = \sqrt{f_i} \left( \frac{f_{ij}}{f_i} - f_j \right) \quad p_{ii} = 1 \quad m_{jj} = \frac{1}{f_j} \quad (8)$$

A diferencia de la transformación de los perfiles columna es necesario proyectar los perfiles sin transformar para evitar que esta proyección contenga el peso y conseguir que perfiles iguales no se proyecten en puntos diferentes.

Este análisis es equivalente al obtenido con las matrices (1).

- Al igual que en (6) donde se ha realizado una transformación en la cual la matriz a utilizar en el análisis de las columnas es la transpuesta de la utilizada en el análisis de las filas, es también posible transformar las matrices para introducir los pesos y la métrica en la matriz  $X_*$  de la siguiente forma:

$$x_{ij}^* = \sqrt{f_i} \left( \frac{f_{ij}}{f_i \cdot f_j} - 1 \right) \sqrt{f_j} \quad (9)$$

El análisis podría llevarse a cabo utilizando todas las expresiones del análisis (6) si se considera que en este caso las matrices de pesos y métrica son unitarias.

La matriz a diagonalizar, al igual que sus valores propios, en el análisis de los perfiles fila, en este caso, tiene la expresión  $X_*^T X_*$  y coincide con la matriz  $A$  obtenida en el análisis de (6). Los vectores propios ( $u_s^*$ ) son normalizados para una métrica diferente (ahora la euclídea) y por tanto son proporcionales a los anteriores.

Las proyecciones tanto de los perfiles fila como de los perfiles columna no deben contener los pesos. Recordando que en el análisis factorial se maximiza: <sup>4</sup>

$$u^T M X^T P X M u$$

y descomponiendo las matrices  $P$  y  $M$  podemos expresarlo como:

$$u^T M^{1/2} M^{1/2} X^T P^{1/2} P^{1/2} X M^{1/2} M^{1/2} u$$

La matriz transformada es entonces:

$$X_* = P^{1/2} X M^{1/2}$$

Sin embargo, dado que la proyección se calcula como

$$X M u_s$$

Será necesario proyectar

$$P^{-1/2} X_* u_s^*$$

para evitar que las proyecciones contengan los pesos.

En concreto, en la transformación (9) propuesta para el análisis (6):

$$\begin{aligned} F_s &= P^{-1/2} X_* u_s^* = \\ &= P^{-1/2} X_* M^{1/2} G_s \frac{1}{\sqrt{\lambda_s}} \\ G_s &= M^{-1/2} X_*^T P^{1/2} F_s \frac{1}{\sqrt{\lambda_s}} \end{aligned}$$

---

<sup>4</sup>Las matrices  $X$ ,  $P$  y  $M$  pueden ser las de cualquiera de las alternativas presentadas (ecuaciones (1) y (5) a (8)).

- A partir del análisis de correspondencias clásico realizado a través de las matrices (6) e introduciendo los pesos de las filas y de las columnas en la matriz  $X$ , se ha obtenido el análisis (9) donde las matrices de pesos y métrica son la unidad.

La introducción de los pesos en la matriz  $X$  abre un camino para el tratamiento simultáneo de tablas de contingencia donde los pesos de las filas son diferentes.

## 5 Tratamiento simultáneo de varias tablas de contingencia

En § 2 y § 3.3 hemos comprobado la importancia de mantener en un análisis factorial conjunto de tablas de contingencia tanto los pesos como la métrica de cada una de las tablas y la necesidad de equilibrar la influencia de las tablas igualando el efectivo de cada una e introduciendo una ponderación en cada una de ellas.

A continuación proponemos un análisis conjunto de tablas de contingencia bajo dichas condiciones.

Sea  $\mathcal{G} = \{1, \dots, g, \dots, G\}$  el conjunto de tablas de contingencia que se posee. Cada una de ellas clasifica las respuestas de  $k_{..}^g$  individuos a dos variables cualitativas codificadas en modalidades. Todas las tablas tienen una de las variables (cuyas modalidades,  $\mathcal{I} = \{1, \dots, i, \dots, I\}$ , quedan recogidas en las filas) común. La otra variable de cada tabla de contingencia puede ser diferente o la misma, codificada de igual o de diferente forma. Las modalidades de la segunda variable en cada tabla son  $\mathcal{J}_g = \{1, \dots, j, \dots, J_g\}$ . Al yuxtaponer todas estas tablas de contingencia se posee un conjunto  $\mathcal{J} = \{1, \dots, j, \dots, J\}$  de columnas.

El elemento  $k_{ij}^g$  corresponde al número total de individuos que eligen simultáneamente las modalidades  $i$  de la primera variable y  $j$  de la segunda variable (perteneciente a la tabla  $g$ ).

Un primer paso para equilibrar la influencia de las tablas hemos visto que es igualar el efectivo de las tablas (§ 2.1). Con este fin cada una de las tablas de contingencia es transformada, dividiéndola por  $k_{..}^g$ , en una tabla de frecuencias relativas.

Se denotará:

$$\begin{aligned} f_{ij}^g &= \frac{k_{ij}^g}{k_{..}^g} \\ f_{i.}^g &= \sum_{j \in \mathcal{J}_g} f_{ij}^g \\ f_{.j}^g &= \sum_{i \in \mathcal{I}} f_{ij}^g \\ f_{..}^g &= \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_g} f_{ij}^g = 1 \end{aligned}$$

En § 2.3 hemos concluido la necesidad de introducir una ponderación, que denotaremos  $\alpha^g$ , en cada uno de los grupos (aquí tablas de contingencia) con el fin de equilibrar la influencia de las tablas. Esta ponderación dependerá de las inercias que resultan de realizar un análisis de correspondencias simples de cada una de las tablas tratadas por separado. Estas inercias

para la tabla  $g$  serán denotadas por  $\lambda_s^g$  (la inercia proyectada sobre el eje  $s$ ) y por  $\lambda_{TOT}^g$  (la inercia total). La ponderación puede ser la equivalente al análisis factorial múltiple para tablas continuas ( $1/\lambda_1^g$ ), la equivalente al método STATIS,  $\lambda_{TOT}^g$ ,  $1/\lambda_{TOT}^g$  para privilegiar grupos de dispersión mínima, 1 si no se desea una ponderación de las tablas, etc.

En (Pagès 1996) se estudia la influencia de las ponderaciones utilizadas en análisis factorial múltiple y el método STATIS dependiendo de la estructura inicial de la tabla de datos (individuos por variables cuantitativas).

## 5.1 Análisis de las columnas

A cada una de las  $g$  tablas de contingencia está asociada una subnube,  $\mathcal{N}(\mathcal{J}_g)$ , de  $J_g$  perfiles columna centrados representados por las filas de la matriz:

$$x_{ji} = \frac{f_{ij}^g}{f_{.j}^g} - f_i^g \quad j \in \mathcal{J}_g$$

con pesos  $f_{.j}^g$  y métrica  $1/f_i^g$ .

Para analizar conjuntamente las tablas sobreponderamos cada subnube por  $\alpha^g$  y dado que todas ellas están situadas en el mismo espacio,  $R^I$ , consideramos la nube global,  $\mathcal{N}(\mathcal{J})$ , que engloba los  $J$  perfiles columna. En esta nube estamos considerando métricas diferentes para cada subnube de perfiles de la misma tabla lo cual puede parecer que imposibilita el análisis conjunto. Sin embargo, en § 4 hemos comprobado cómo se transforman los perfiles columna de cada una de las tablas para considerar sus distancias euclídeas. En el análisis conjunto esto significa considerar las matrices  $X$ ,  $M$  y  $P$  de datos, métrica y pesos cuyos términos generales son los siguientes:

$$\begin{aligned} x_{ji} &= \frac{\sqrt{\alpha^g}}{\sqrt{f_i^g}} \left( \frac{f_{ij}^g}{f_{.j}^g} - f_i^g \right) \\ p_{jj} &= f_{.j}^g \\ m_{ii} &= 1 \end{aligned}$$

En este análisis, las distancias euclídeas entre perfiles columna de la misma subnube respetan las distancias  $\chi^2$  en la subnube original.

## 5.2 Análisis de las filas

En cada una de las  $g$  tablas de contingencia se definen los perfiles fila centrados,

$$\frac{f_{ij}^g}{f_i^g} - f_{.j}^g \quad j \in \mathcal{J}_g \tag{10}$$

con pesos  $f_i^g$  y métrica  $1/f_{.j}^g$ .

Dado que se busca analizar conjuntamente las  $g$  tablas y los perfiles fila de cada tabla se representan en espacios distintos, cada una en un espacio de dimensión  $J_g$ , es necesario buscar un espacio común en el que poder realizar el análisis. Este espacio común es  $R^J$ , en él se representan los perfiles fila de cada una de las tablas, llamados perfiles fila parciales. Las coordenadas de estos puntos corresponden a las definidas en (10) completando el resto de las coordenadas con 0, es decir, se define el perfil fila parcial:

$$i^g = \begin{cases} \frac{f_{ij}^g}{f_i^g} - f_j^g & \text{si } j \in \mathcal{J}_g \\ 0 & \text{en caso contrario} \end{cases} \quad (11)$$

El conjunto de perfiles fila parciales pertenecientes a la misma tabla forman la subnube de puntos que denotaremos  $\mathcal{N}(\mathcal{I}^g)$ .

Con el fin de realizar el análisis conjunto se busca para cada fila un compromiso, que denotaremos  $i^c$ , que represente a la subnube de perfiles fila formada por todos los puntos que representan la misma fila en las distintas tablas. El conjunto de todos los compromisos representados en  $R^J$  forman la nube  $\mathcal{N}(\mathcal{I}^c)$ . El compromiso es elegido con el objetivo de que su inercia se pueda expresar como una suma ponderada de las inercias de los perfiles fila parciales:

$$\text{In}(i^c) = \sum_{g \in \mathcal{G}} \alpha^g \text{In}(i^g)$$

y que, por tanto, la inercia de la nube compromiso se exprese en función de las inercias de las nubes parciales:

$$\text{In}(\mathcal{N}(\mathcal{I}^c)) = \sum_{g \in \mathcal{G}} \alpha^g \text{In}(\mathcal{N}(\mathcal{I}^g))$$

Para ello se define el compromiso como una media ponderada de los perfiles fila parciales, con ponderaciones

$$m_{ig} = \frac{\sqrt{f_i^g}}{\sum_{g \in \mathcal{G}} \sqrt{f_i^g}}$$

Es decir, el compromiso es:

$$\begin{aligned} i^c &= \sum_{g \in \mathcal{G}} m_{ig} i^g = \\ &= \sum_{g \in \mathcal{G}} \frac{\sqrt{f_i^g}}{\sum_{g \in \mathcal{G}} \sqrt{f_i^g}} i^g \end{aligned}$$

con peso:

$$p_{ic} = \left( \sum_{g \in \mathcal{G}} \sqrt{f_i^g} \right)^2$$



En efecto, para comprobarlo comencemos obteniendo la distancia del compromiso al origen:

$$\begin{aligned}
d^2(i^c, 0) &= \sum_{j \in \mathcal{J}} \alpha^g \frac{1}{f_{.j}^g} \left[ \frac{\sqrt{f_{i.}^g}}{\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g}} \left( \frac{f_{ij}^g}{f_{i.}^g} - f_{.j}^g \right) \right]^2 = \\
&= \sum_{g \in \mathcal{G}} \alpha^g \frac{f_{i.}^g}{\left( \sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} \right)^2} \sum_{j \in \mathcal{J}_g} \frac{1}{f_{.j}^g} \left( \frac{f_{ij}^g}{f_{i.}^g} - f_{.j}^g \right)^2
\end{aligned}$$

Recordando la definición del perfil fila parcial realizada en (11), el segundo sumatorio corresponde a la distancia de un perfil fila parcial al centro de gravedad de la subnube formada por todos los perfiles fila parciales correspondientes a la misma tabla que denotaremos  $d^2(i^g, 0)$ .

De esta forma, la distancia puede escribirse:

$$d^2(i^c, 0) = \sum_{g \in \mathcal{G}} \alpha^g \frac{f_{i.}^g}{\left( \sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} \right)^2} d^2(i^g, 0)$$

y multiplicando por el peso del compromiso, obtenemos su inercia en función de las inercias de los perfiles fila parciales:

$$\text{In}(i^c) = \sum_{g \in \mathcal{G}} \alpha^g \text{In}(i^g)$$

Es decir, el análisis de las filas se podría realizar utilizando la matriz  $X$  de compromisos, con la matriz de pesos  $P$  y la matriz diagonal que define la métrica  $M$ :

$$\begin{aligned}
x_{ij} &= \frac{\sqrt{f_{i.}^g}}{\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g}} \left( \frac{f_{ij}^g}{f_{i.}^g} - f_{.j}^g \right) \\
p_{ii} &= \left( \sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} \right)^2 \\
m_{jj} &= \alpha^g \frac{1}{f_{.j}^g}
\end{aligned}$$

### 5.3 Obtención de los factores

Con el fin de buscar la relación entre los análisis de filas y columnas en los desarrollos siguientes y teniendo en cuenta las equivalencias entre análisis obtenidas en § 4, el análisis

factorial del conjunto de las tablas de contingencia se realiza obteniendo los valores ( $\lambda_s$ ) y vectores propios ( $u_s$ ) resultantes de la diagonalización de la matriz:

$$A = X^T X \quad (12)$$

donde el término general de la matriz  $X$  es:

$$x_{ij} = \sqrt{\alpha^g} \sqrt{f_{i.}^g} \left( \frac{f_{ij}^g}{f_{i.}^g f_{.j}^g} - 1 \right) \sqrt{f_{.j}^g} \quad j \in \mathcal{J} \quad (13)$$

y se utilizarán además las matrices diagonales  $R$  de orden  $I \times I$  y  $N$  de orden  $J \times J$ , cuyos términos generales son:

$$\begin{aligned} R : \quad r_{ii} &= \left( \sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} \right)^2 \\ N : \quad n_{jj} &= f_{.j}^g \end{aligned}$$

Se define también la matriz  $Y$ , diagonal por bloques, donde cada bloque de la diagonal es la matriz  $X_g$  de orden  $(I \times J_g)$  que contiene las coordenadas de la matriz  $X$  para el conjunto de columnas  $J_g$  pertenecientes a la tabla  $g$  y se define la matriz  $Q$ , también diagonal por bloques, donde cada bloque es, a su vez, una matriz diagonal de orden  $(I \times I)$  y de término general  $f_{i.}^g$ .

Cuando la ponderación elegida para los grupos es el inverso del primer valor propio de los análisis separados, garantizando de esta forma que el primer eje del análisis conjunto no esté formado por un sólo grupo, la inercia total se puede expresar como una suma ponderada de las inercias de cada uno de los grupos igual que sucede en el análisis factorial múltiple de variables continuas. Es decir,

$$\text{si: } \alpha^g = \frac{1}{\lambda_1^g} \quad \Rightarrow \quad \lambda_{TOT} = \sum_{g \in \mathcal{G}} \frac{\lambda_{TOT}^g}{\lambda_1^g}$$

#### 5.4 Relación entre $u$ y $v$

Dado que en el análisis de filas se diagonaliza la matriz  $X^T X$  y en el de columnas la matriz  $X X^T$  la relación entre los valores propios y vectores directores de los análisis de filas ( $\lambda_s$  y  $u_s$ ) y columnas ( $\mu_s$  y  $v_s$ ) es:

$$\begin{aligned} \lambda_s &= \mu_s \\ u_s &= \frac{1}{\sqrt{\lambda_s}} X^T v_s \end{aligned} \quad (14)$$

$$v_s = \frac{1}{\sqrt{\lambda_s}} X u_s \quad (15)$$

## 5.5 Proyecciones

Se calculan las proyecciones sobre el eje  $s$  de los perfiles fila parciales  $F_s(i^g)$ , de los compromisos  $F_s(i^c)$  y de los perfiles columna  $G_s(j)$ . Las proyecciones de todos los perfiles fila parciales y compromisos son respectivamente  $F_s(\mathcal{I}^g)$  y  $F_s(\mathcal{I}^c)$  y de todas las columnas  $G_s(\mathcal{J})$ . Las proyecciones sobre el eje  $s$  de filas y columnas se calculan recordando que es necesario eliminar el efecto de la introducción de los pesos y los perfiles fila parciales se proyectan como suplementarios eliminando también el efecto de la introducción del peso:

$$\begin{aligned} F_s(\mathcal{I}^c) &= R^{-1/2} X u_s \\ G_s(\mathcal{J}) &= N^{-1/2} X^T v_s \\ F_s(\mathcal{I}^g) &= Q^{-1/2} Y u_s \end{aligned}$$

## 5.6 Propiedades de los factores

Tanto las proyecciones de los perfiles columna como de los perfiles fila parciales están centradas para sus pesos.

- Las proyecciones del conjunto de columnas son:

$$G_s(\mathcal{J}) = N^{-1/2} X^T v_s$$

Para una determinada columna:

$$G_s(j) = \sum_{i \in \mathcal{I}} \sqrt{\alpha^g} \sqrt{f_i^g} \left( \frac{f_{ij}^g}{f_i^g f_j^g} - 1 \right) v_{si}$$

donde  $v_{si}$  es la coordenada  $i$ -ésima del vector director  $v_s$ .

El baricentro de estas proyecciones es:

$$\begin{aligned} \sum_{j \in \mathcal{J}_g} f_j^g G_s(j) &= \sum_{j \in \mathcal{J}_g} \sqrt{\alpha^g} \left( f_j^g \sum_{i \in \mathcal{I}} \frac{1}{\sqrt{f_i^g}} \frac{f_{ij}^g}{f_j^g} v_{si} - f_j^g \sum_{i \in \mathcal{I}} \sqrt{f_i^g} v_{si} \right) = \\ &= \sum_{i \in \mathcal{I}} \frac{\sqrt{\alpha^g}}{\sqrt{f_i^g}} f_i^g v_{si} - \sum_{i \in \mathcal{I}} \sqrt{\alpha^g} \sqrt{f_i^g} v_{si} = \\ &= 0 \end{aligned} \tag{16}$$

es decir, las proyecciones de las columnas están centradas porque las columnas de cada tabla, y por tanto, del conjunto han sido centradas.

- Las proyecciones del conjunto de los perfiles fila parciales son:

$$F_s(\mathcal{I}^g) = Q^{-1/2} Y u_s$$

y de uno en particular:

$$F_s(i^g) = \sum_{j \in \mathcal{J}_g} \sqrt{\alpha^g} \left( \frac{f_{ij}^g}{f_i^g f_j^g} - 1 \right) \sqrt{f_j^g} u_{sj} \quad (17)$$

siendo  $u_{sj}$  la coordenada  $j$ -ésima del vector director  $u_s$ .

Por tanto, el baricentro de las proyecciones de los perfiles fila parciales correspondientes al grupo  $g$  es:

$$\begin{aligned} \sum_{i \in \mathcal{I}} f_i^g F_s(i^g) &= \sqrt{\alpha^g} \sum_{i \in \mathcal{I}} \left( f_i^g \frac{1}{f_i^g} \sum_{j \in \mathcal{J}_g} \frac{f_{ij}^g}{\sqrt{f_j^g}} u_{sj} \right) - \sqrt{\alpha^g} \sum_{i \in \mathcal{I}} \left( f_i^g \sum_{j \in \mathcal{J}_g} \sqrt{f_j^g} u_{sj} \right) = \\ &= \sqrt{\alpha^g} \sum_{j \in \mathcal{J}_g} \sqrt{f_j^g} u_{sj} - \sqrt{\alpha^g} \sum_{j \in \mathcal{J}_g} \sqrt{f_j^g} u_{sj} = \\ &= 0 \end{aligned}$$

- Para el conjunto de los compromisos sus proyecciones son:

$$F_s(\mathcal{I}^c) = R^{-1/2} X u_s \quad (18)$$

La proyección para un compromiso concreto:

$$\begin{aligned} F_s(i^c) &= \sum_{j \in \mathcal{J}} \frac{\sqrt{f_i^g}}{\sum_{g \in \mathcal{G}} \sqrt{f_i^g}} \sqrt{\alpha^g} \left( \frac{f_{ij}^g}{f_i^g f_j^g} - 1 \right) \sqrt{f_j^g} u_{sj} = \\ &= \frac{1}{\sum_{g \in \mathcal{G}} \sqrt{f_i^g}} \sum_{g \in \mathcal{G}} \left( \sqrt{\alpha^g} \sum_{j \in \mathcal{J}_g} \frac{f_{ij}^g}{\sqrt{f_i^g}} \frac{1}{\sqrt{f_j^g}} u_{sj} \right) - \\ &\quad - \frac{1}{\sum_{g \in \mathcal{G}} \sqrt{f_i^g}} \sum_{g \in \mathcal{G}} \left( \sqrt{\alpha^g} \sqrt{f_i^g} \sum_{j \in \mathcal{J}_g} \sqrt{f_j^g} u_{sj} \right) \end{aligned} \quad (19)$$

En consecuencia, las proyecciones de los compromisos no están centradas.

- Los perfiles de cada tabla (perfiles fila parciales y perfiles columna) han sido centrados para mantener las relaciones dentro de cada tabla y en consecuencia están también centradas sus proyecciones. Los compromisos son una media ponderada de los perfiles fila parciales con unas ponderaciones diferentes a los pesos de los perfiles fila parciales y por tanto no están centrados ni sus proyecciones tampoco.

## 5.7 Relaciones entre $F_s(i^g)$ , $F_s(i^c)$ y $G_s(j)$

Al igual que en análisis de correspondencias simples existen también relaciones de transición entre las proyecciones de filas (compromiso) y columnas. Además se pueden relacionar con las proyecciones de los perfiles fila parciales.

1. Relación entre  $F_s(i^c)$  y  $F_s(i^g)$

La proyección de un compromiso obtenida en (19):

$$F_s(i^c) = \sum_{g \in \mathcal{G}} \frac{\sqrt{f_{i.}^g}}{\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g}} \sum_{j \in \mathcal{J}_g} \sqrt{\alpha^g} \left( \frac{f_{ij}^g}{f_{i.}^g f_{.j}^g} - 1 \right) \sqrt{f_{.j}^g} u_{sj}$$

se puede expresar teniendo en cuenta la expresión (17):

$$\begin{aligned} F_s(i^c) &= \sum_{g \in \mathcal{G}} \frac{\sqrt{f_{i.}^g}}{\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g}} F_s(i^g) = \\ &= \sum_{g \in \mathcal{G}} m_{ig} F_s(i^g) \end{aligned}$$

En proyección, al igual que en el espacio, el compromiso es una media ponderada de los perfiles fila parciales.

2. Relación entre  $F_s(i^g)$  y  $G_s(j)$

Premultiplicando la relación (14) por  $Q^{-1/2} Y$ :

$$Q^{-1/2} Y u_s = Q^{-1/2} Y N^{1/2} N^{-1/2} X^T v_s \frac{1}{\sqrt{\lambda_s}}$$

y expresando la ecuación anterior en función de las proyecciones:

$$F_s(\mathcal{I}^g) = Q^{-1/2} Y N^{1/2} G_s(\mathcal{J}) \frac{1}{\sqrt{\lambda_s}}$$

La proyección para un perfil fila parcial es, en función de las proyecciones de las columnas:

$$\begin{aligned} F_s(i^g) &= \sum_{j \in \mathcal{J}_g} \sqrt{\alpha^g} f_{.j}^g \left( \frac{f_{ij}^g}{f_{i.}^g f_{.j}^g} - 1 \right) G_s(j) \frac{1}{\sqrt{\lambda_s}} = \\ &= \sqrt{\alpha^g} \frac{1}{\sqrt{\lambda_s}} \left( \sum_{j \in \mathcal{J}_g} \frac{f_{ij}^g}{f_{i.}^g} G_s(j) - \sum_{j \in \mathcal{J}_g} f_{.j}^g G_s(j) \right) \end{aligned}$$

como el segundo sumatorio es, según (16), nulo:

$$F_s(i^g) = \sqrt{\alpha^g} \frac{1}{\sqrt{\lambda_s}} \sum_{j \in \mathcal{J}_g} \frac{f_{ij}^g}{f_{i.}^g} G_s(j)$$

Excepto por la ponderación del grupo  $\sqrt{\alpha^g}$  y el coeficiente  $\frac{1}{\sqrt{\lambda_s}}$ , la proyección de un perfil fila parcial es, como en análisis de correspondencias, el baricentro de los perfiles columna correspondientes a la tabla a la que pertenece.

### 3. Relación entre $F_s(i^c)$ y $G_s(j)$

Como ya hemos visto:

$$F_s(i^c) = \sum_{g \in \mathcal{G}} m_{ig} F_s(i^g)$$

expresando la proyección del perfil fila parcial en función de las columnas:

$$F_s(i^c) = \sum_{g \in \mathcal{G}} \left( \sqrt{\alpha^g} m_{ig} \left( \frac{1}{\sqrt{\lambda_s}} \sum_{j \in \mathcal{J}_g} \frac{f_{ij}^g}{f_{i.}^g} G_s(j) \right) \right)$$

Es decir, la proyección del compromiso es, salvo por los coeficientes  $\sqrt{\alpha^g}$  y  $\frac{1}{\sqrt{\lambda_s}}$ , la media ponderada de los baricentros de los perfiles columna correspondientes a cada tabla.

### 4. Relación entre $G_s(j)$ y $F_s(i^c)$ o $F_s(i^g)$

Premultiplicando por  $N^{-1/2} X^T$  la relación (15) obtenemos las proyecciones de las columnas en función de los compromisos:

$$\begin{aligned} N^{-1/2} X^T v_s &= N^{-1/2} X^T R^{1/2} R^{-1/2} X u_s \frac{1}{\sqrt{\lambda_s}} \\ G_s(\mathcal{J}) &= N^{-1/2} X^T R^{1/2} F_s(\mathcal{I}^c) \frac{1}{\sqrt{\lambda_s}} \end{aligned}$$

La proyección del perfil columna  $j$  perteneciente al grupo  $g^*$  se puede expresar de la siguiente manera:

$$G_s(j \in g^*) = \sqrt{\alpha^g} \frac{1}{\sqrt{\lambda_s}} \left( \sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} \right) \left( \sum_{i \in \mathcal{I}} \sqrt{f_{i.}^{g^*}} \left( \frac{f_{ij}^{g^*} - f_{i.}^{g^*} f_{.j}^{g^*}}{f_{i.}^{g^*} f_{.j}^{g^*}} \right) F_s(i^c) \right)$$

Una columna se sitúa del lado de los compromisos a los que se asocia más que en la hipótesis de independencia y al lado contrario de aquéllos a los que se asocia menos.

Y la proyección en función de los perfiles fila parciales:

$$G_s(j \in g^*) = \sqrt{\alpha^g} \frac{1}{\sqrt{\lambda_s}} \left( \sum_{i \in \mathcal{I}} \sqrt{f_{i.}^{g^*}} \left( \frac{f_{ij}^{g^*} - f_{i.}^{g^*} f_{.j}^{g^*}}{f_{i.}^{g^*} f_{.j}^{g^*}} \right) \left( \sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} F_s(i^g) \right) \right)$$

## 5.8 Ayudas a la interpretación

En el análisis propuesto se pueden obtener las mismas ayudas a la interpretación de los análisis factoriales habituales.

### 5.8.1 Contribuciones de los puntos a la formación de los ejes

Las contribuciones de los puntos a la formación de los ejes se calculan como es habitual dividiendo la inercia proyectada de un punto (peso por coordenada al cuadrado) entre la suma de las inercias de todos los puntos de la nube sobre el eje  $s$ , es decir:

$$\begin{aligned} \text{cta}_s(i^c) &= \frac{\left(\sum_{g \in \mathcal{G}} \sqrt{f_i^g}\right)^2 F_s^2(i^c)}{\lambda_s} \\ \text{cta}_s(j) &= \frac{f_{.j}^g G_s^2(j)}{\lambda_s} \end{aligned}$$

Ya que se cumple:

$$\begin{aligned} \lambda_s &= \sum_{j \in \mathcal{J}} f_{.j}^g G_s^2(j) = \\ &= \sum_{i \in \mathcal{I}} \left(\sum_{g \in \mathcal{G}} \sqrt{f_i^g}\right)^2 F_s^2(i^c) \end{aligned}$$

### 5.8.2 Calidad de representación de los puntos sobre los ejes

Las contribuciones relativas de cada punto indican la calidad de representación de un punto sobre el eje  $s$ . Se calculan como el cociente de la inercia proyectada del punto sobre el eje  $s$  entre la inercia total del punto o como el cuadrado de la proyección sobre el eje  $s$  entre la distancia del punto al origen:

$$\begin{aligned} \text{ctr}_s(i^c) &= \frac{F_s^2(i^c)}{d^2(i^c, 0)} \\ \text{ctr}_s(j) &= \frac{G_s^2(j)}{d^2(j, 0)} \end{aligned}$$

siendo la distancia de la columna  $j$  al origen:

$$\begin{aligned} d^2(j, 0) &= \sum_{i \in \mathcal{I}} \alpha^g \left(\frac{f_{ij}^g}{f_{.j}^g} - f_i^g\right)^2 \frac{1}{f_i^g} = \\ &= \sum_s G_s^2(j) \end{aligned}$$

y la distancia del compromiso al origen la ya obtenida en función de las distancias de los

perfiles fila parciales:

$$\begin{aligned}
 d^2(i^c, 0) &= \sum_{g \in \mathcal{G}} \alpha^g \frac{f_{i.}^g}{\left( \sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} \right)^2} d^2(i^g, 0) = \\
 &= \sum_s F_s^2(i^c) \\
 d^2(i^g, 0) &= \sum_{j \in \mathcal{J}} \left( \frac{f_{ij}^g}{f_{i.}^g} - f_{.j}^g \right)^2 \frac{1}{f_{.j}^g}
 \end{aligned}$$

## 6 Ejemplo

La aplicación de este método propuesto al ejemplo 2 muestra cómo el análisis conjunto utilizando como ponderación para cada uno de los grupos el inverso del primer valor propio del análisis separado de cada tabla, presenta un plano factorial semejante al correspondiente a los análisis separados -Figura 2 (i)-. Dado lo particular del ejemplo, los planos de las dos tablas analizadas por separado son semejantes, el análisis sigue siendo igualmente válido sin la ponderación de los grupos ( $\alpha^g = 1$ ) -Figura 2 (j)-.

La aplicación al ejemplo 1 del análisis conjunto (con o sin ponderación -Figura 2 (g)- y -Figura 2 (h)- respectivamente) muestra la asociación obtenida en el análisis de cada tabla por separado (enestar la primera canasta y fallar la segunda frente a fallar en primer lugar y enestar el segundo tiro)

## 7 Conclusión

- El método propuesto permite el análisis simultáneo de varias tablas de contingencia manteniendo cada una de ellas su propia estructura y su propia especificidad.
- Se ha comprobado que el método propuesto, cuando se aplica a una tabla de contingencia coincide con el análisis de correspondencias simples. Al ser aplicado a una tabla disyuntiva completa se obtienen los mismos resultados del análisis de correspondencias múltiples clásico. Y cuando la tabla que se analiza es disyuntiva incompleta el método proporciona los resultados del análisis de correspondencias con marginal impuesta. (Zárraga & Goitisoló 2000)



## Referencias

- Bécue-Bertaut, M. & Pagès, J. (2000), Analyse factorielle multiple intra-tableaux. application à l'analyse simultanée de plusieurs questions ouvertes., in 'JADT 2000: 5<sup>es</sup> Journées Internationales d'Analyse Statistique des Données Textuelles'.
- Benali, H. (1988), 'Données manquantes et modalités à faible effectif en analyse des correspondances multiples et conditionnelle', *Data Analysis and Informatics V* pp. 311–318.
- Benzécri, J. (1983), 'Analyse de l'inertie intraclasse par l'analyse d'un tableau de correspondance', *Les Cahiers de l'Analyse des Données VIII*(3), 351–358.
- Carroll, J. (1968), 'Generalization of canonical correlation analysis to three or more sets of variables', *Proc. Am. Psychol. Ass.* pp. 227–228.
- Carroll, J. & Chang, J. (1970), 'Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition', *Psychometrika* **35**(3), 283–319.
- Chevalier, A. (1977), 'Contribution à l'analyse descriptive des chroniques multidimensionnelles: Application à l'étude évolutive du trafic fluvial du port de paris', *PhD thesis, Université Pierre et Marie Curie. Paris.*
- Dazy, F. & Le Barzic, J. (1996), *L'Analyse des Données Évolutives. Méthodes et applications.*, Technip.
- Douroche, J. & Dussaix, A. (1972), 'Double analyse en composantes principales', *Metra International.*
- Escofier, B. (1983), 'Généralisation de l'analyse des correspondances à la comparaison de tableaux de fréquence', *INRIA* **207**, 1–33.
- Escofier, B. (1988), 'Analyse des correspondances multiples conditionnelle', *Data Analysis and Informatics V* pp. 333–342.
- Escofier, B. (1990), 'Analyse des correspondances multiples conditionnelle', *Revue de Modulad* pp. 13–27.
- Escofier, B. & Drouet, D. (1983), 'Analyse des différences entre plusieurs tableaux de fréquence', *Les Cahiers de l'Analyse des Données VIII*(4), 491–499.
- Escofier, B. & Pagès, J. (1984), 'L'analyse factorielle multiple', *Cahiers du Bureau Universitaire de Recherche Opérationnelle* **42**, 1–68.
- Escofier, B. & Pagès, J. (1992), *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación*, Servicio Editorial Universidad del País Vasco.
- Gower, J. (1975), 'Generalized procrustes analysis', *Psychometrika* **40**(1), 33–51.
- Gower, J. (1984), *Handbook of Applicable Mathematics 6*, Lloyd E.H., chapter Procrustes analysis.
- Hotelling, H. (1936), 'Relation between two sets of variables', *Biometrika* **28**, 129–149.

- Kendall, M. (1977), Multivariate contingency tables and some further problems in multivariate analysis, in P. Krishnaiah, ed., 'Multivariate Analysis IV', North-Holland, pp. 483–494.
- Lavit, C. (1988), *Analyse Conjointe de Tableaux Quantitatifs*, Masson. Paris.
- Lebart, L., Morineau, A. & Fenelon, J. (1985), *Tratamiento estadístico de datos*, Marcombo, chapter IV.5: Análisis de Correspondencias (pp 307-335).
- Lebart, L., Morineau, A. & Piron, M. (1995), *Statistique exploratoire multidimensionnelle*, Dunod.
- L'Hermier des Plantes, H. (1976), *STATIS: Structuration de Tableaux À Trois Indices de la Statistique*, Thèse (3c), USTL, Montpellier.
- Pagès, J. (1996), 'Eléments de comparaison entre l'analyse factorielle multiple et la méthode statis', *Revue de Statistique Appliquée* **XLIV**(4), 81–95.
- Simpson, E. (1951), 'The interpretation of interaction in contingency tables', *Journal of the Royal Statistical Society, Ser B* **13**, 238–241.
- Tucker, L. (1958), 'An inter-battery method of factor analysis', *Psychometrika*.
- Tucker, L. (1972), 'Relations between multidimensional scaling and three-mode factor analysis', *Psychometrika* **37**(1), 3–27.
- Wardrop, R. (1995), 'Simpson's paradox and the hot hand in basketball', *The American Statistician* **49**(1), 24–28.
- Zárraga, A. & Goitisoló, B. (2000), 'Estudio comparativo de análisis alternativos de tablas disyuntivas incompletas', *Documentos de Trabajo. Biltoki* **8**, 1–29.

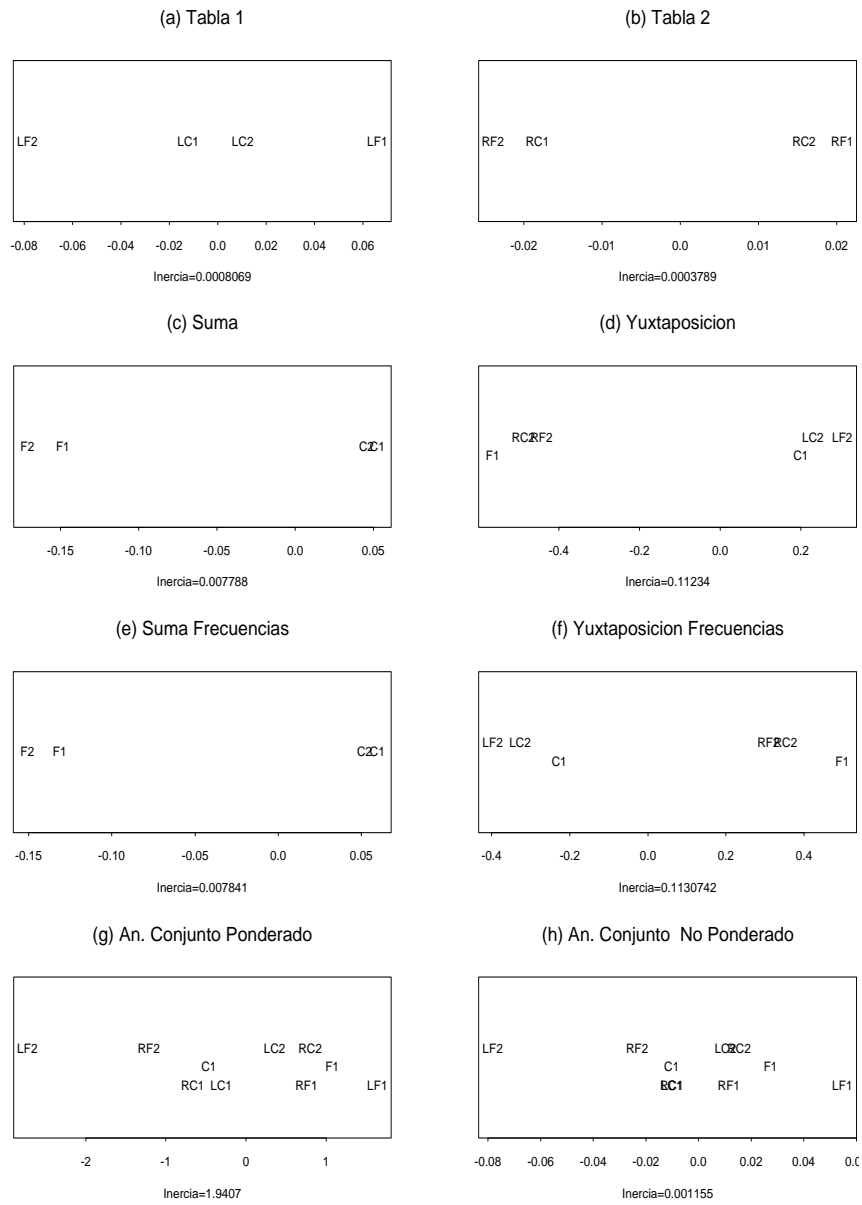


Figura 1: Ejemplo 1

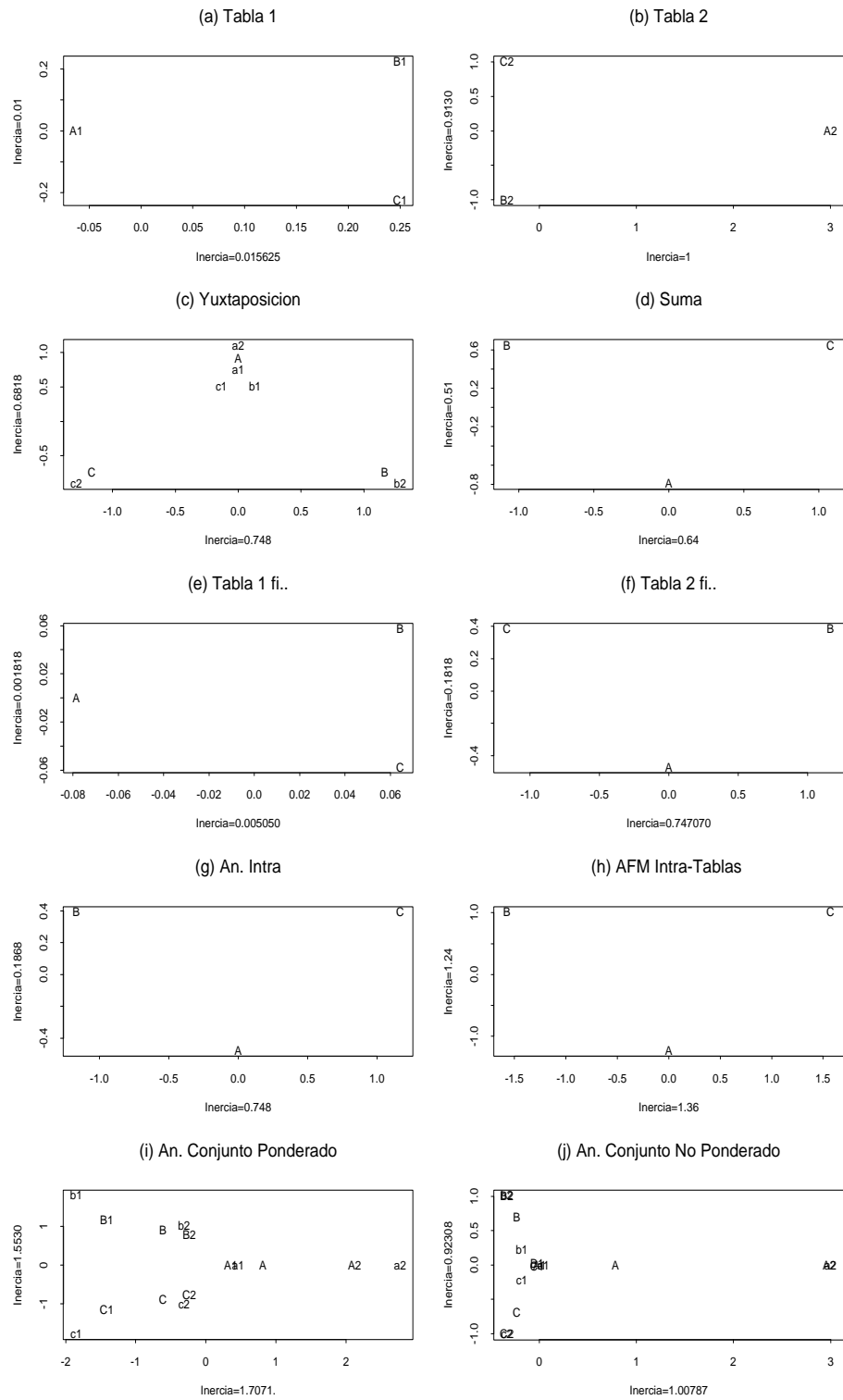


Figura 2: Ejemplo 2

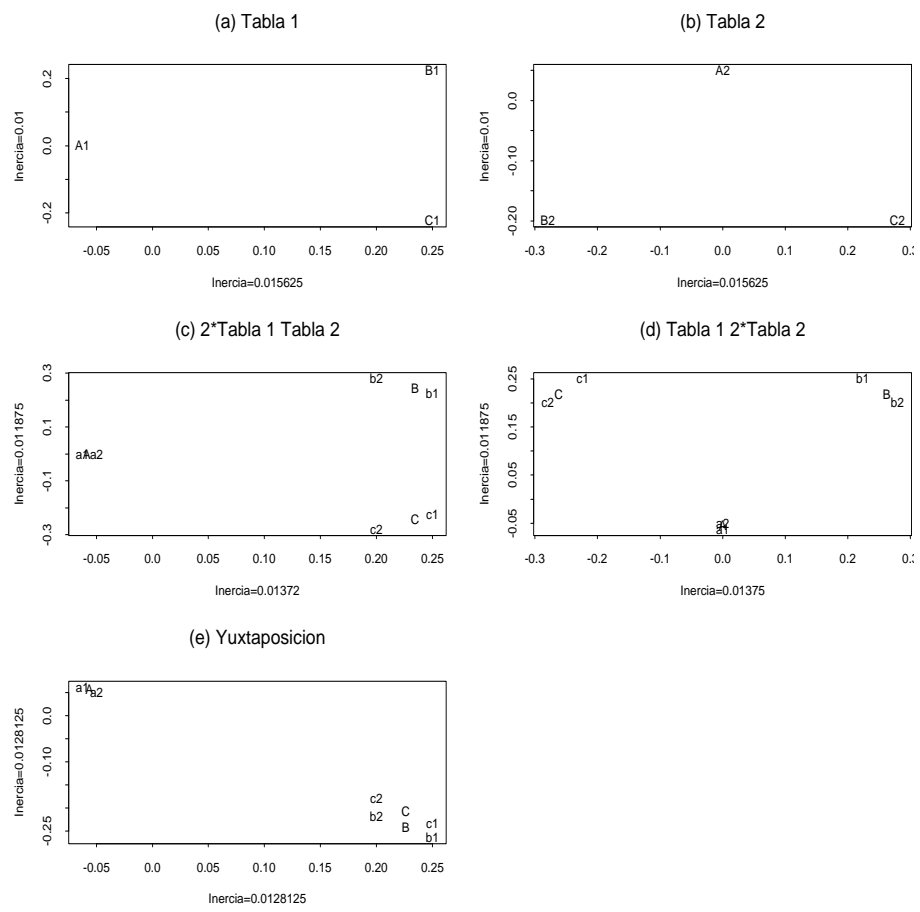


Figura 3: Ejemplo 3

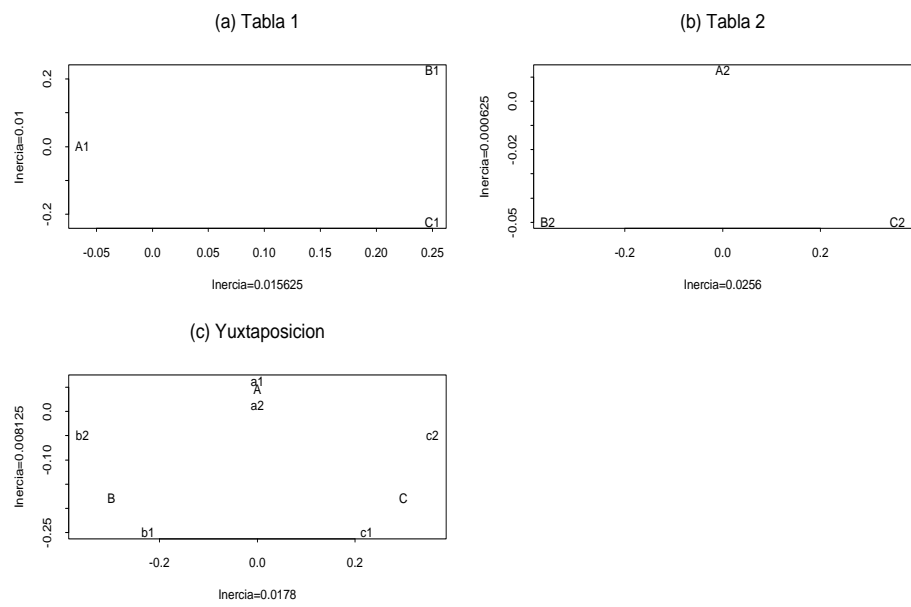


Figura 4: Ejemplo 4