ESTUDIO COMPARATIVO DE ANALISIS ALTERNATIVOS DE TABLAS DISYUNTIVAS INCOMPLETAS

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

Resumen

El análisis de correspondencias múltiples (A.C.M.) estudia la relación entre varias variables cualitativas definidas sobre una misma población. Sin embargo, una de las principales fuentes de información son las encuestas donde es frecuente encontrar cierto número de datos ausentes y de preguntas condicionadas (a las que no toda la población ha de responder). En estos casos la codificación de los datos en una tabla disyuntiva completa supone la inclusión de modalidades de no respuesta que pueden alterar los resultados.

Keywords: Análisis de Correspondencias Múltiples, Tabla Disyuntiva Incompleta, Independencia entre Variables Cualitativas, Valores Propios, Tasas de Inercia

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

Las técnicas de análisis de datos surgen del deseo de obtener la máxima cantidad de información posible sobre un conjunto de variables y conocer las relaciones existentes entre ellas.

Con este objetivo surgen distintas técnicas que se adaptan a situaciones concretas dependiendo del número de variables que se posea, la naturaleza de las mismas (cualitativas, cuantitativas, ordinales, etc), el tipo de relación que se crea que las vincula (de interdependencia, causa-efecto, ...) y los objetivos del estudio (cuantificar la dependencia entre las variables, clasificarlas en función de algún criterio de semejanza, etc).

La información recogida en una encuesta es resumida en una tabla de datos que será de distinta naturaleza según el tipo de datos que se posea. En todos los casos se considera que está formada por un conjunto $\mathcal{I} = \{1, \ldots, i, \ldots, n\}$ de filas y un conjunto $\mathcal{J} = \{1, \ldots, j, \ldots, J\}$ de columnas. Estas tablas suelen ser de gran tamaño, resultando muy difícil interpretar la información que poseen sin la ayuda de las técnicas factoriales. El análisis factorial (A.F.) busca reducir el tamaño de dicha tabla poniendo de manifiesto las relaciones entre las filas y columnas y facilitar su interpretación mediante representaciones gráficas.

Aunque con el mismo objetivo, los métodos de análisis factorial han ido surgiendo en el tiempo para ser aplicados a tipos de datos distintos. Este hecho exige distintas transformaciones de los datos iniciales y diferentes definiciones de semejanza entre filas y columnas, lo que lleva a estudiar los distintos tipos de análisis (componentes principales y correspondencias simples y múltiples) por separado. Así, por ejemplo, el análisis en componentes principales (A.C.P.) trata tablas que cruzan individuos y variables cuantitativas; el análisis factorial de correspondencias fue creado para analizar la relación existente entre dos variables cualitativas definidas sobre un conjunto de modalidades a través de las tablas de contingencia (recibiendo el nombre de análisis de correspondencias simples) y posteriormente aplicado a otro tipo de tablas, y en concreto, a las tablas disyuntivas completas (T.D.C.) que cruzan individuos y modalidades de dos o más variables cualitativas, recibiendo en este caso el nombre de análisis de correspondencias múltiples (A.C.M.). Sin embargo, todos estos métodos se engloban dentro de un marco general.

Nuestro interés se centra en el análisis de encuestas, que habitualmente son estudiadas mediante el análisis de correspondencias múltiples de una tabla disyuntiva completa o mediante el análisis de correspondencias de la tabla de Burt asociada. Sin embargo, en las encuestas es frecuente encontrarse con una cantidad de respuestas ausentes; en estos casos la codificación de los datos en una T.D.C. supone la inclusión de modalidades de no respuesta que pueden alterar los resultados. Escofier (1981) propone una variante del análisis de correspondencias múltiples para este tipo de tablas disyuntivas incompletas que estudiaremos en detalle (análisis (A) de la sección § 6) En las secciones § 7, § 10 y § 11 proponemos otros tres análisis alternativos (denominados (B), (C) y (D)) y estudiamos su relación con el primero. En la sección § 16 se propone una corrección a los valores propios y tasas de inercia del análisis (A) para lo que será necesario estudiar la independencia entre las cuestiones (§ 15) y la relación con el análisis (C) (§ 13).

2 Datos Ausentes

Las razones por las que existen datos ausentes en las encuestas pueden ser diversas, revelando problemas que deberán ser tratados también de diferentes formas según se indica en § 5.

- Las no respuestas pueden estar causadas por un olvido involuntario del individuo y no tener un significado especial. Suelen representar una proporción muy pequeña de los datos y afectar por igual a todas las cuestiones e individuos.
- Una segunda razón para la existencia de la no respuesta corresponde a una actitud particular del entrevistado, deseo de no revelar cierta información (por ejemplo, los ingresos, la ideología política, etc). Este tipo de no respuesta no se reparte de forma aleatoria en la tabla de datos sino que afecta más a determinadas cuestiones y grupos de individuos.
- Otra razón por la que aparecen las tablas disyuntivas incompletas -definidas en § 3-, muy frecuente en las encuestas, se debe a la existencia de preguntas condicionadas, es decir, aquéllas a las cuales un individuo debe contestar o no dependiendo de cuál haya sido su respuesta a una cuestión anterior. Por ejemplo, se le pregunta si sabe o no inglés; a continuación, y sólo si ha respondido saber inglés, se le pregunta su nivel de inglés en determinados aspectos. En una encuesta pueden existir varios grupos de preguntas condicionadas (los que saben inglés, francés, sólo los que tienen familiares y se relacionan con ellos contestarán la frecuencia con que lo hacen, etc). En este caso, la no respuesta se agrupa en un determinado número de cuestiones (aquéllas cuya respuesta está condicionada por una pregunta anterior) y caracteriza a determinados grupos de individuos.

3 Definición de las tablas disyuntivas incompletas y notación

Se considera la tabla de datos que recoge en forma lógica y disyuntiva las respuestas de un conjunto de individuos a un conjunto de preguntas o cuestiones, poseyendo cada una de ellas un conjunto finito de modalidades de respuesta. En el análisis de correspondencias múltiples clásico se impone a todos los individuos la obligación de pertenecer a alguna de las modalidades de cada cuestión. Se denomina a la tabla así obtenida tabla disyuntiva completa y se denota por Z. Se dirá que tal tabla de datos es disyuntiva incompleta, la denotaremos por Z^*, cuando los individuos no dan respuesta a una o más de las cuestiones preguntadas.

	$q=1$		$\dots$	q	$\dots$	$q=Q$
	$j=1$ $\dots$	$j=J_1$		j		
1	1	0				
2	0	0				
3	0	1				
$\vdots$						
i				z_{ij}		
$\vdots$						
n						

Tabla disyuntiva completa (Z) o Tabla disyuntiva incompleta (Z^*)

donde:

$$
\begin{aligned}
\mathcal{Q} &= \{1,\dots,q,\dots,Q\} \quad \text{es el conjunto de variables o cuestiones a las cuales} \\
& \qquad\qquad\qquad\qquad\quad \text{debe responder el individuo} \\
\mathcal{J}_q &= \{1,\dots,j,\dots,J_q\} \quad \text{es el conjunto de modalidades de la variable } q \in \mathcal{Q} \\
\mathcal{J} &= \{1,\dots,j,\dots,J\} \quad \text{es el conjunto de modalidades de todas las variables} \\
&= \cup_{q=1}^{Q} \mathcal{J}_q \\
\mathcal{I} &= \{1,\dots,i,\dots,n\} \quad \text{es el conjunto de individuos} \\
z_{ij} &= \begin{cases} 1 & \text{si el individuo } i \in \mathcal{I} \text{ responde la modalidad } j \in \mathcal{J} \\ 0 & \text{en otro caso} \end{cases} \\
z_{i.} &= \sum_{j \in \mathcal{J}} z_{ij} \quad \text{es el número de cuestiones a las que responde el individuo } i \in \mathcal{I} \\
z_{.j} &= \sum_{i \in \mathcal{I}} z_{ij} \quad \text{es el número de individuos que eligen la modalidad } j \in \mathcal{J}
\end{aligned}
$$

Se denotará $z_{.j}^{q}$ cuando interese dejar constancia de la variable $q \in \mathcal{Q}$ a la que pertenece dicha modalidad

$$
z = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} z_{ij} \quad \text{es el total de la tabla}
$$

En las tablas disyuntivas incompletas -al igual que en las completas analizadas mediante el ACM clásico- las variables siguen estando definidas a través de un conjunto de modalidades a las que el individuo debe responder sobre su pertenencia ($z_{ij} = 1$) o no ($z_{ij} = 0$).

Pero, en ocasiones, esas modalidades correspondientes a una misma variable no están definidas en forma completa, es decir, el individuo puede no pertenecer a ninguna de ellas; en otras ocasiones, a pesar de estar definidas en forma completa, el individuo puede no revelar a que modalidad pertenece; en ambos casos:

$$
z_{ij} = 0 \qquad i \in \mathcal{I} \quad \forall j \in \mathcal{J}_q \quad q \in \mathcal{Q}
$$

Por ello será necesario definir también un indicador que toma el valor 1 si el individuo i responde a la cuestión q y 0 en caso contrario:

$$
z_{i.}^{q} = \sum_{j \in \mathcal{J}_q} z_{ij} \tag{1}
$$

Y se denotará por z_q el número de individuos que han respondido a la cuestión q:

$$z_q = \sum_{j \in \mathcal{J}_q} z_{.j} \tag{2}$$

En resumen, las tablas disyuntivas incompletas se caracterizan porque en ellas dejan de cumplirse algunas de las relaciones que se dan en las completas:

$$\begin{aligned} z_{i.}^q &= 1 & \forall q \in \mathcal{Q} \quad & \forall i \in \mathcal{I} \\ z_q &= n & \forall q \in \mathcal{Q} \quad & \\ z_{i.} &= Q & \forall i \in \mathcal{I} \quad & \\ z &= nQ & & \end{aligned}$$

Como ya es sabido, en todo análisis de correspondencias se definen ((Escofier & Pagès 1992), (Lebart, Morineau & Tabard 1977) y (Abascal & Grande 1989) entre otros) las frecuencias relativas conjuntas y marginales como:

$$\begin{aligned} f_{ij} &= \frac{z_{ij}}{z} & \forall i \in \mathcal{I} \quad & \forall j \in \mathcal{J} \\ f_{i.} &= \frac{z_{i.}}{z} = \sum_{j \in \mathcal{J}} f_{ij} & \forall i \in \mathcal{I} & \\ f_{.j} &= \frac{z_{.j}}{z} = \sum_{i \in \mathcal{I}} f_{ij} & \forall j \in \mathcal{J} & \end{aligned}$$

y los perfiles fila i: $z_{ij}/z_{i.}$ y perfiles columna j: $z_{ij}/z_{.j}$ que forman las nubes $\mathcal{N}(\mathcal{I}) \in R^J$ y $\mathcal{N}(\mathcal{J}) \in R^n$ respectivamente.

4 Problema que plantea la aplicación del análisis de correspondencias múltiples habitual a una tabla disyuntiva incompleta

El problema que plantea este tipo de tablas es el mismo independientemente de la razón para esa ausencia de datos: la marginal sobre $\mathcal{I}$ ya no es constante.

Podría pensarse en aplicar directamente el análisis de correspondencias simples a esta tabla, (notar que originalmente se creó para el estudio de tablas de contingencia donde las marginales no son constantes); sin embargo, cuando se posee una tabla disyuntiva incompleta, la distancia χ^2 y los pesos definidos en el análisis clásico no se ajustan a los objetivos, ya conocidos de un análisis de correspondencias.

La aplicación de la distancia χ^2 entre dos perfiles fila i e i' sería:

$$d^2(i,i') = \sum_{j \in \mathcal{J}} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 = \sum_{j \in \mathcal{J}} \frac{z}{z_{.j}} \left(\frac{z_{ij}}{z_{i.}} - \frac{z_{i'j}}{z_{i'.}} \right)^2$$

Si los individuos i e i' no contestan al mismo número de preguntas, entonces $z_{i.}$ difiere de $z_{i'.}$ y por tanto la distancia χ^2 aumenta también con las respuestas comunes. Este es, por tanto, un concepto de distancia no deseable puesto que no reflejaría la similaridad entre individuos -en términos de modalidades comunes elegidas- buscada en un análisis de correspondencias.

La distancia χ^2 entre dos perfiles columna j y j' sería:

$$d^2(j,j') \;=\; \sum_{i\in\mathcal{I}} \frac{1}{f_{i.}}\left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}}\right)^2 = \sum_{i\in\mathcal{I}} \frac{z}{z_{i.}}\left(\frac{z_{ij}}{z_{.j}} - \frac{z_{ij'}}{z_{.j'}}\right)^2$$

En esta distancia χ^2 cada individuo tendría una ponderación distinta dependiendo del número de respuestas elegidas. No parece lógico asignar menos importancia a aquellos individuos que responden a la totalidad de las preguntas frente a quienes no lo hacen.

Por tanto, la aplicación directa del análisis de correspondencias clásico no es adecuada al estudio de las tablas disyuntivas incompletas.

5 Posibles alternativas al problema de las tablas disyuntivas incompletas: análisis factorial con marginal modificada

Una vez descartada la aplicación del análisis clásico, se debe buscar otra alternativa para el estudio de las tablas disyuntivas incompletas que se adapte mejor. Se busca un método que minimice la influencia de la no respuesta sobre el análisis.

Una solución evidente desde el punto de vista analítico sería eliminar aquellos individuos que no responden a todas las cuestiones, obteniendo de esta forma una tabla disyuntiva completa. Con esta alternativa perderíamos la información referente a esos individuos en el resto de las cuestiones; información que puede no ser importante cuando la no respuesta se debe a un descuido y por tanto representa una pequeña proporción sobre el total, pero que alteraría los resultados cuando revela una actitud particular o implica a un gran número de individuos (caso de las preguntas condicionadas).

Una práctica habitual es crear para cada variable con datos ausentes una modalidad de no respuesta, obteniendo de esta forma una tabla disyuntiva completa a la que se puede aplicar el análisis clásico. Esta solución puede ser adecuada cuando la no respuesta se debe a una actitud particular del individuo (deseo de no revelar determinada información por ejemplo); sin embargo, cuando la no respuesta se debe a un descuido involuntario la modalidad de no respuesta no tendría una interpretación adecuada y perturbaría los resultados.

En los casos en los que la no respuesta se debe a la existencia de preguntas condicionadas ya se ha indicado en § 2 que caracteriza a un grupo de individuos. A pesar de ello, la inclusión de una modalidad de no respuesta en cada cuestión no sería adecuada, puesto que se estaría creando una serie de modalidades todas ellas con el mismo perfil e idéntico a una de las modalidades de la pregunta condicionante (no saben inglés) o a una combinación lineal de ellas ("no tienen familiares" y "tienen pero no se relacionan"). Esto podría perturbar los resultados hasta el punto de llegar a crear uno de los primeros ejes del análisis, como así ocurre en la aplicación a la Encuesta de Condiciones de Vida de 1989 de la Comunidad Autónoma de Euskadi presentada en (Goitisolo & Zárraga 1998*a*).

Escofier (Escofier 1981) propone sustituir la marginal real de la tabla ($f_{i.} = z_{i.}/z$, que no es constante), por una marginal constante $g_{i.} = 1/n$, en todo el análisis.

Posteriormente (Benali 1985), (Benali 1988), (Benali & Escofier 1987) y (Escofier 1990) utilizan también esta técnica, pero siempre para el caso de tablas disyuntivas incompletas donde la no respuesta se debe a una omisión involuntaria y el caso de modalidades de efectivo débil. En el cálculo de la distancia entre dos individuos (y por tanto, en la inercia de la nube), las modalidades tienen una ponderación inversa a su efectivo. En consecuencia, las modalidades muy raras pueden influir demasiado; Benali y Escofier proponen eliminarlas y tratar a esos individuos como si no hubieran dado respuesta a la cuestión.

6 Análisis (A): $\quad Z^*/z$

En esta sección se analizará en detalle lo adecuado de esta sustitución y las consecuencias sobre el análisis, tanto para los casos de omisión involuntaria y de modalidades raras como para el caso de preguntas condicionantes, en el que centraremos nuestro interés.

Para ello, utilizaremos las frecuencias absolutas y marginales calculadas a partir de la tabla disyuntiva imcompleta, así como la marginal impuesta $g_{i.}$:

$$\begin{array}{ll} f_{ij}^{(A)} = z_{ij}/z & \\ f_{i.}^{(A)} = z_{i.}/z & \sum_{i\in\mathcal{I}} f_{i.}^{(A)} = 1 \\ f_{.j}^{(A)} = z_{.j}/z & \sum_{j\in\mathcal{J}} f_{.j}^{(A)} = 1 \\ g_{i.} = 1/n & \sum_{i\in\mathcal{I}} g_{i.} = 1 \end{array}$$

6.1 Nube de individuos: $\mathcal{N}(\mathcal{I})$

El punto i se representa en R^J por el perfil $\dfrac{f_{ij}^{(A)}}{g_{i.}} = n\dfrac{z_{ij}}{z}$. Este perfil es diferente del perfil obtenido en el análisis de correspondencias múltiples clásico (z_{ij}/Q) al ser el efectivo total de la tabla (z) distinto de nQ.

La distancia cuadrática propuesta entre dos individuos i e $i' \in \mathcal{I}$ es:

$$d^2(i,i')^{(A)} = \sum_{j\in\mathcal{J}} \frac{1}{f_{.j}^{(A)}} \left(\frac{f_{ij}^{(A)}}{g_{i.}} - \frac{f_{i'j}^{(A)}}{g_{i'.}} \right)^2 = \frac{n}{z}\left(n \sum_{j\in\mathcal{J}} \frac{1}{z_{.j}} \left(z_{ij} - z_{i'j}\right)^2 \right)$$

Se comprueba que únicamente las respuestas diferentes hacen aumentar la distancia, sin tener en cuenta si ambos individuos responden al mismo número de cuestiones o no.

La ponderación de cada modalidad es, al igual que en correspondencias múltiples con datos completos, el inverso de su efectivo (la distancia aumenta en mayor proporción cuando la modalidad poseída por sólo uno de los individuos es rara).

Considerar la distancia anterior entre los puntos i e i' es equivalente a buscar la distancia euclídea habitual en un espacio dotado de métrica $1/f_{.j}^{(A)}$.

Cada punto i está dotado de un peso $g_{i.} = 1/n$, que a pesar de no venir representado por la marginal $f_{i.}^{(A)}$ coincide con el peso que se asigna a los individuos en correspondencias múltiples habitual. Este peso constante significa que todos los individuos tienen la misma importancia, independientemente del número de cuestiones que han respondido. Es por tanto, más adecuado que $f_{i.}^{(A)}$.

La coordenada j-ésima del centro de gravedad de la nube es:

$$G_I^{(A)}(j) \;=\; f_{.j}^{(A)} \;=\; \frac{z_{.j}}{z} \qquad \forall j \in \mathcal{J}$$

que coincide en expresión con la correspondiente al centro de gravedad de la nube de individuos en correspondencias múltiples habitual.

Este será también el origen de los ejes de máxima inercia que se han de buscar.

La distancia de un punto al origen es:

$$d^2(i, O_I)^{(A)} \;=\; \frac{n^2}{z} \sum_{j \in \mathcal{J}} \frac{z_{ij}}{z_{.j}} + 1 - 2\,\frac{n\, z_{i.}}{z} \tag{3}$$

y multiplicando por el peso se obtiene su inercia:

$$\text{Inercia de } (i)^{(A)} \;=\; \frac{n}{z} \sum_{j \in \mathcal{J}} \frac{z_{ij}}{z_{.j}} + \frac{1}{n} - \frac{2}{z}\, z_{i.}$$

La suma para todos los puntos de la nube permite comprobar que su inercia total es función del número de respuestas ausentes (a través de z) pero no de su distribución en la tabla.

$$\text{Inercia total}^{(A)} \;=\; \frac{n\,J}{z} - 1$$

6.2 Nube de modalidades: $\mathcal{N}(\mathcal{J})$

El punto j se representa en R^n por el perfil $\dfrac{f_{ij}^{(A)}}{f_{.j}^{(A)}} = \dfrac{z_{ij}}{z_{.j}}$, es decir, el mismo que en el A.C.M. clásico.

La distancia cuadrática propuesta entre dos modalidades j y j' es:

$$d^2(j, j')^{(A)} \;=\; \sum_{i \in \mathcal{I}} \frac{1}{g_{i.}} \left(\frac{f_{ij}^{(A)}}{f_{.j}^{(A)}} - \frac{f_{ij'}^{(A)}}{f_{.j'}^{(A)}} \right)^2 = n \sum_{i \in \mathcal{I}} \left(\frac{z_{ij}}{z_{.j}} - \frac{z_{ij'}}{z_{.j'}} \right)^2$$

Esta distancia es semejante a la utilizada cuando la tabla es completa y apropiada también para el caso de preguntas condicionadas. Equivale a considerar la distancia euclídea en un espacio dotado de métrica $1/g_{i.}$.

Cada punto j está dotado de un peso $f_{.j}^{(A)} = z_{.j}/z$, coincide con el asignado en correspondencias múltiples de tablas completas y supone que cada modalidad tiene una importancia proporcional a la población que representa. Las modalidades tienen un peso en la construcción de los ejes tanto menor cuanto menor sea su efectivo.

6.2.1 Centros de gravedad

La coordenada i-ésima del centro de gravedad de la nube es:

$$G_J^{(A)}(i) \quad = \quad f_{i.}^{(A)} \quad = \quad \frac{z_{i.}}{z} \qquad\qquad \forall i \in \mathcal{I}$$

Este baricentro coincide en expresión, pero no en coordenadas, con el obtenido en el análisis de correspondencias múltiples cuando se posee una tabla disyuntiva completa, puesto que si existe algún dato ausente, esta marginal no es constante:

$$f_{i.}^{(A)} = \frac{z_{i.}}{z} \neq \frac{1}{n}$$

Cuando la tabla es disyuntiva incompleta, la coordenada i-ésima del centro de gravedad de cada cuestión es:

$$G_q^{(A)}(i) \quad = \quad \frac{z_{i.}^q}{z_q} \qquad\qquad \forall i \in \mathcal{I} \quad y \quad \forall\, q \in \mathcal{Q}$$

donde $z_{i.}^q$ y z_q son las variables definidas en (1) y (2) respectivamente.

Si no todos los individuos contestan a esa pregunta entonces:

$$G_q^{(A)}(i) \quad \neq \quad \frac{1}{n} \qquad\qquad \forall i \in \mathcal{I}$$

puesto que $z_{i.}^q$ es diferente de 1 para aquellos individuos que no responden a la cuestión y z_q difiere de n. Este centro de gravedad dependerá del número de individuos que hayan respondido a la cuestión y de quiénes sean esos individuos. En consecuencia, no es el mismo para todas las variables, ni coincide con el de la nube de todas las modalidades como ocurría en correspondencias de una tabla disyuntiva completa. En el caso de las preguntas condicionadas sí será el mismo para todas aquellas variables que dependan de la misma pregunta condicionante (y en las que por tanto, coincidan todos los individuos que han de responderlas).

Si todos los individuos contestan a esa cuestión entonces:

$$G_q^{(A)}(i) \quad = \quad \frac{1}{n} \qquad\qquad \forall i \in \mathcal{I}$$

puesto que $z_{i.}^q$ es 1 para todos los individuos y z_q es n. Pero este punto sólo coincide con el centro de gravedad de la nube si todos los individuos responden a todas las cuestiones (es decir, cuando se tiene una tabla completa), en caso contrario es un punto que representa la marginal modificada $g_{i.}$.

6.2.2 Elección del origen

Se quiere maximizar la inercia proyectada de la nube; pero esa inercia a proyectar interesa que esté basada en las modalidades en las que existe más diferencia entre los individuos. Es decir, si una modalidad ha sido elegida por todos los individuos, no interesa que dicha modalidad incremente la inercia por el hecho de que exista un individuo que no ha respondido a otra cuestión. Eso es lo que ocurre si se toma como origen $G_J^{(A)}$ (de $i-$ésima coordenada $f_{i.}^{(A)} = z_{i.}/z$), mientras que si se traslada la nube, tomando como origen O_J (cuya coordenada $i-$ésima es $g_{i.}$), las modalidades elegidas por todos los individuos no contribuirán a aumentar la inercia. Por ello el análisis se hará tomando como origen O_J, es decir, no tomando como origen el centro de gravedad de todas las variables sino el de aquéllas que son elegidas por todos los individuos. De esta forma la influencia de la no respuesta es minimizada.

6.2.3 Inercias

La distancia de un punto al origen:

$$d^2(j, O_J)^{(A)} = \frac{n}{z_{.j}} - 1 \tag{4}$$

permite obtener la inercia de un perfil columna:

$$\text{Inercia de } (j)^{(A)} = \frac{1}{z}(n - z_{.j}) \tag{5}$$

La inercia de una modalidad es mayor cuanto menor sea el número de individuos que la poseen. Es nula si todos los individuos la eligen y es máxima (n/z) si ningún individuo declara pertenecer a ella. Por tanto, al igual que en correspondencias múltiples sin datos ausentes, las modalidades raras pueden perturbar los resultados.

La inercia de una variable:

$$\text{Inercia de } (q)^{(A)} = \frac{1}{z}(nJ_q - z_q) \tag{6}$$

aumenta al disminuir el número de individuos que responden (z_q), y al incrementar el número de modalidades en que es clasificada (J_q). No es, por tanto, recomendable que este número de modalidades sea elevado (lo mismo que ocurre en el análisis clásico). En los cuestionarios con preguntas condicionadas habrá que tener especial cuidado con las cuestiones que no han de responder todos los individuos.

La suma para todos los puntos de la nube coincide con la obtenida en la nube de individuos.

6.3 Análisis con las nubes no centradas

En correspondencias múltiples clásico el análisis se puede hacer bien a través de las nubes centradas o bien a través de la nubes no centradas eliminando el primer eje (asociado a un valor propio igual a 1), que corresponde a la dirección de unión entre el origen del espacio y el baricentro de las nubes.

En el análisis con la marginal modificada esta relación no se mantiene y se deberá considerar siempre la nube trasladada al origen.

6.4 Obtención de los factores de ambas nubes

Calcular la sucesión de ejes $(u_{s\,(A)}, s \in \mathcal{S} = \{1, \ldots, s, \ldots, S\}, S \leq J)$ que maximizan la inercia proyectada de la nube $\mathcal{N}(\mathcal{I})$ equivale a:

$$\begin{array}{ll} \text{maximizar:} & u^T_{s\,(A)}\, M_{\,(A)}\, X^T_{(A)}\, P_{(A)}\, X_{\,(A)}\, M_{\,(A)}\, u_{s\,(A)} \\ \text{sujeto a:} & u^T_{s\,(A)}\, M_{\,(A)}\, u_{s\,(A)} = 1 \\ & u^T_{s\,(A)}\, M_{\,(A)}\, u_t = 0 \qquad\qquad \forall t < s \end{array}$$

donde:

- $X_{\,(A)}$ es una matriz $(n \times J)$ de término general:

$$x_{ij}^{(A)} = \frac{f_{ij}^{(A)}}{g_{i.} f_{.j}^{\,(A)}} - 1 = n\,\frac{z_{ij}}{z_{.j}} - 1$$

  Al igual que en análisis de correspondencias clásico cada elemento de esta matriz contiene las desviaciones entre la tabla de datos $f_{ij}^{(A)}$ y una tabla de término general que corresponde a la hipótesis de independencia. La diferencia con el análisis clásico radica en que la frecuencia relativa marginal correspondiente a las filas es impuesta en función del número de filas en lugar de obtenida a partir de los datos.

- $M_{\,(A)}$ es una matriz diagonal correspondiente a la métrica del espacio:

$$m_j^{(A)} = f_{.j}^{\,(A)} = \frac{z_{.j}}{z}$$

- $P_{(A)}$ es la matriz (también diagonal) de pesos:

$$p_i^{(A)} = g_{i.} = \frac{1}{n}$$

Se puede demostrar que la nube definida por las filas de la matriz $X_{\,(A)}$, con la métrica y los pesos considerados, es isomorfa de la definida en § 6.1 como objetivo del estudio. Ambas nubes mantienen las mismas distancias entre dos puntos cualesquiera.

La resolución de este problema lleva a la diagonalización de la matriz $X_{(A)}^T P_{(A)} X_{(A)} M_{(A)}$ de orden $(J \times J)$ cuyo término general es:

$$a_{jj'}^{(A)} = \sum_{i \in \mathcal{I}} \frac{f_{ij}^{(A)} \, f_{ij'}^{(A)}}{g_{i.} \, f_{.j}^{(A)}} - f_{.j'}^{(A)} \tag{7}$$

Las proyecciones de la nube de individuos sobre los ejes de máxima inercia resultantes son:

$$F_s^{(A)} = X_{(A)} M_{(A)} u_{s\,(A)}$$

Su i-ésima coordenada adopta la expresión:

$$\begin{aligned} F_s^{(A)}(i) &= \sum_{j \in \mathcal{J}} \left(\frac{f_{ij}^{(A)}}{g_{i.} f_{.j}^{(A)}} - 1 \right) f_{.j}^{(A)} \, u_{sj}^{(A)} = \\ &= \sum_{j \in \mathcal{J}} \frac{f_{ij}^{(A)}}{g_{i.}} u_{sj}^{(A)} - \sum_{j \in \mathcal{J}} f_{.j}^{(A)} \, u_{sj}^{(A)} \end{aligned}$$

Al diagonalizar la matriz $X_{(A)} M_{(A)} X_{(A)}^T P_{(A)}$ cuyo término general es:

$$d_{ii'}^{(A)} = \sum_{j \in \mathcal{J}} \frac{f_{ij}^{(A)} \, f_{i'j}^{(A)}}{g_{i.} \, f_{.j}^{(A)}} - f_{i'.}^{(A)} - f_{i.}^{(A)} + g_{i.}$$

se obtienen los ejes $v_{s\,(A)}, s \in \mathcal{S}^{(*)}$ que maximizan la inercia proyectada de la nube $\mathcal{N}(\mathcal{J})$ y tras premultiplicar dicha matriz por $X_{(A)}^T P_{(A)}$ las proyecciones $G_s^{(A)}$ de dicha nube cuya j-ésima coordenada puede expresarse:

$$\begin{aligned} G_s^{(A)}(j) &= \sum_{i \in \mathcal{I}} \left(\frac{f_{ij}^{(A)}}{g_{i.} \, f_{.j}^{(A)}} - 1 \right) g_{i.} \, v_{si}^{(A)} = \\ &= \sum_{i \in \mathcal{I}} \frac{f_{ij}^{(A)}}{f_{.j}^{(A)}} v_{si}^{(A)} - \sum_{i \in \mathcal{I}} g_{i.} v_{si}^{(A)} \end{aligned}$$

6.5 Relaciones entre los factores

Los factores de ambas nubes se relacionan a través de las expresiones:

$$F_s^{(A)} = \frac{1}{\sqrt{\lambda_s^{(A)}}} X_{(A)} M_{(A)} G_s^{(A)} \tag{8}$$

$$G_s^{(A)} = \frac{1}{\sqrt{\lambda_s^{(A)}}} X_{(A)}^T P_{(A)} F_s^{(A)} \tag{9}$$

(*) Las relaciones de dualidad entre ambos espacios, que se verifican en todo análisis de correspondencias, permiten establecer que los subespacios de ajuste, asociados a valores propios no nulos, son de idéntica dimensión.

En el análisis de correspondencias con marginal modificada, igual que ocurre en el análisis clásico para la cantidad $f_{i.}^{(A)}$ los factores $F_s^{(A)}$ están centrados para la cantidad $g_{i.}$

$$\sum_{i \in \mathcal{I}} g_{i.} F_s^{(A)}(i) = 0$$

Aplicando la fórmula de transición (9):

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

que coincide con la relación baricéntrica del análisis de correspondencias.

Sin embargo, a diferencia del análisis de correspondencias clásico los factores $G_s^{(A)}$ no están centrados por la cantidad $f_{.j}^{(A)}$ porque el análisis se hace tomando como origen un punto diferente al centro de gravedad.

$$\sum_{j \in \mathcal{J}} f_{.j}^{(A)} G_s^{(A)}(j) = \sum_{i \in \mathcal{I}} f_{i.}^{(A)} v_{si}^{(A)} - \sum_{i \in \mathcal{I}} g_{i.} v_{si}^{(A)}$$

Si la marginal impuesta difiere de la marginal propia de la tabla esta cantidad es distinta de cero.

Por ello los factores $F_s^{(A)}$ no pueden interpretarse como el baricentro de los $G_s^{(A)}$ como en análisis clásico. Según la fórmula de transición (8):

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

El segundo sumatorio corresponde a la proyección del centro de gravedad de $\mathcal{N}(\mathcal{J})$, que al no haber sido tomado como origen de los ejes es diferente de 0.

Benali y Escofier (Benali & Escofier 1987) afirman que “Este término, en la práctica es casi nulo, lo que permite interpretar como en correspondencias múltiples clásico la abscisa de un individuo como el baricentro de las modalidades que ha elegido”. Hacen referencia a tablas disyuntivas incompletas en las que el efectivo de datos ausentes representa una proporción reducida en relación al total (caso de datos ausentes por olvido distribuidos de forma aleatoria a lo largo de la tabla). Sin embargo, puede alterar los resultados y su interpretación cuando se considera nulo en una tabla de datos en la cual la proporción de no respuesta es grande o corresponde a ciertos grupos de individuos (caso de tablas de datos con preguntas condicionadas) como se ha podido comprobar en la aplicación a la Encuesta de Condiciones de Vida de 1989 de la Comunidad Autónoma de Euskadi presentada en (Goitisolo & Zárraga 1998a).

Lo cierto es que este segundo sumatorio es el mismo para todos los individuos, aunque difiere para los distintos ejes, por lo que se podría trasladar los factores $F_s^{(A)}(i)$ de tal forma que en la representación superpuesta de ambas nubes un individuo siga estando representado en el baricentro de las modalidades que posee:

$$F_s^{(A)}(i)^* = \frac{1}{\sqrt{\lambda_s^{(A)}}} n \sum_{j \in \mathcal{J}} f_{ij}^{(A)} \; G_s^{(A)}(j)$$

6.6 Número de ejes

En el análisis de correspondencias de una tabla disyuntiva completa el número de ejes S es igual al número de modalidades activas menos el número de variables, porque todas las modalidades correspondientes a cada una de las variables se encuentran restringidas al mismo hiperplano. En el análisis de correspondencias con marginal modificada de una tabla disyuntiva incompleta, las modalidades de una misma cuestión no cumplen ningún tipo de restricción por lo que pueden existir tantos ejes como número de modalidades activas exista en el análisis. Si existen cuestiones con datos completos, sus modalidades mantendrán la misma restricción que en el análisis clásico por lo que la cantidad de ejes disminuirá en ese número de cuestiones completas.

La existencia de preguntas condicionadas en el análisis también reduce la cantidad de ejes, puesto que los individuos que han de responder a una pregunta condicionada vienen determinados por la respuesta a una modalidad (o combinación de ellas) anterior.

6.7 Ayudas a la interpretación

La contribución de los puntos a la formación del eje y su calidad de representación sobre el mismo se miden a través de las contribuciones absolutas y relativas:

$$\mathrm{CTA}_s^{(A)}(i) = \frac{\frac{1}{n}\left[F_s^{(A)}(i)\right]^2}{\frac{1}{n}\sum_{i \in \mathcal{I}}\left[F_s^{(A)}(i)\right]^2} \qquad \mathrm{CTR}_s^{(A)}(i) = \frac{\left[F_s^{(A)}(i)\right]^2}{d^2(i, O_I)^{(A)}}$$

$$\mathrm{CTA}_s^{(A)}(j) = \frac{f_{.j}^{(A)}\left[G_s^{(A)}(j)\right]^2}{\sum_{j \in \mathcal{J}} f_{.j}^{(A)}\left[G_s^{(A)}(j)\right]^2} \qquad \mathrm{CTR}_s^{(A)}(j) = \frac{\left[G_s^{(A)}(j)\right]^2}{d^2(j, O_J)^{(A)}}$$

Las distancia de los puntos al origen se han obtenido en (3) y (4).

Valores test y preguntas condicionadas

Morineau (Morineau 1984) desarrolla un test para buscar las modalidades de las variables (asociadas a grupos de individuos) que mejor caracterizan a un factor en análisis de correspondencias clásico. Una modalidad será más característica de un factor cuanto menos probable sea la hipótesis de aleatoriedad de las coordenadas de los individuos que lo forman sobre el total de los individuos del análisis.

Si n y $z_{.j}$ no son muy pequeños la variable:

$$G_s^{(A)}(j)\sqrt{z_{.j}\left(\frac{n-1}{n-z_{.j}}\right)}$$

sigue aproximadamente una distribución normal centrada y reducida. Las modalidades más características son las asociadas a los mayores valores test.

El cálculo se basa en la hipótesis de extracción aleatoria de las $z_{.j}$ coordenadas. Este supuesto no se mantiene para las variables activas que contribuyen a la formación del eje.

Estos valores tests son también válidos en el marco de la variante presentada del análisis de correspondencias aplicable a tablas disyuntivas incompletas, para aquellas variables que son respondidas por todos los individuos o si el número de no respuestas es pequeño y generado aleatoriamente entre los individuos. Pero cuando la no respuesta se debe a la existencia de preguntas condicionadas, los individuos que han de responder a la cuestión y que por tanto son clasificados en las modalidades, no son seleccionados de forma aleatoria entre el total de los individuos.

Si la pregunta condicionante no caracteriza al eje, entonces la no respuesta se distribuye de forma aleatoria sobre el eje y los valores tests siguen indicando las modalidades que caracterizan a un eje a pesar de corresponder a preguntas condicionadas.

Por el contrario, si la variable condicionante contribuye a la formación del eje como activa o lo caracteriza bien -en el caso de una variable suplementaria- entonces la no respuesta de las preguntas que condiciona no se repartirá de forma aleatoria sobre el eje. En ese caso unos valores tests altos no indicarán las modalidades características del eje sino aquellas respondidas por un conjunto determinado de individuos. Para corregir estos valores tests correspondientes a preguntas condicionadas será necesario calcular la media y varianza de las coordenadas sobre el eje únicamente en función de los individuos que han de responder a la cuestión.

6.8 Rango de los valores propios: $(0, nQ/z)$

En este análisis la inercia de una modalidad es $\frac{n-z_{.j}}{z}$ -ecuación 5- Por tanto, esta inercia se encuentra entre 0 (si todos los individuos eligen dicha modalidad) y $\frac{n}{z}$ si no la elige ninguno.

La inercia máxima de una variable proyectada sobre un determinado eje es la inercia máxima de una modalidad, $\frac{n}{z}$, debido a la ortogonalidad de las modalidades de una misma variable.

La inercia máxima de la nube de modalidades proyectada sobre un determinado eje es la inercia máxima de una variable por el número de variables, $\frac{nQ}{z}$, situación que corresponde a una dependencia total entre todas las variables.

7 Análisis (B): Z^*/nQ

7.1 Notación

Incluyendo las modalidades de no respuesta en la tabla disyuntiva incompleta se convierte en una tabla disyuntiva completa y por tanto su total es nQ

A partir de esta tabla se obtiene la de frecuencias. En ella:

$$\begin{aligned} f_{ij}^{(B)} &= \frac{z_{ij}}{nQ} \\ \sum_{i\in\mathcal{I}} f_{ij}^{(B)} &= \sum_{i\in\mathcal{I}} \frac{z_{ij}}{nQ} = \frac{z_{.j}}{nQ} \\ \sum_{j\in\mathcal{JA}} f_{ij}^{(B)} &= \sum_{j\in\mathcal{JA}} \frac{z_{ij}}{nQ} = \frac{1}{n} \\ \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{JA}} f_{ij}^{(B)} &= \sum_{i\in\mathcal{I}} f_{i.}^{(B)} = 1 \end{aligned}$$

siendo $\mathcal{JA}$ el conjunto $\mathcal{J}$ ampliado con las modalidades de no respuesta.

Una vez calculadas las frecuencias relativas se utilizan éstas para el análisis eliminando las modalidades de no respuesta.

A la hora de realizar el análisis hay que tener en cuenta que el total de la tabla y de las marginales ya no es 1 sino que:

$$\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} f_{ij}^{(B)} = \sum_{i\in\mathcal{I}} f_{i.}^{(B)} = \sum_{j\in\mathcal{J}} f_{.j}^{(B)} = \frac{z}{nQ}$$

7.2 Problema que plantea la aplicación del análisis de correspondencias múltiples habitual a una tabla disyuntiva incompleta

Calculando las distancias entre dos perfiles fila y entre dos perfiles columna obtenemos las mismas expresiones que en el análisis (A) multiplicadas por $\frac{z}{nQ}$.

Al igual que en el análisis (A) la distancia entre dos individuos incrementa con respuestas comunes si los individuos no responden el mismo números de preguntas y en el cálculo de la distancia entre dos modalidades cada individuo tendría una ponderación distinta dependiendo del número de respuestas elegidas.

La aplicación del análisis de correspondencias clásico con las frecuencias $f_{ij}^{(B)}$ tampoco es válida.

Lo mismo que en el análisis (A) estos problemas desaparecen si se impone la marginal $g_{i.} = 1/n$.

7.3 Nube de individuos: $\mathcal{N}(\mathcal{I})$

El punto i se representa en R^J por el perfil z_{ij}/Q. Esta nube es homotética de la analizada en (A). El factor de escala $\frac{z}{nQ}$ aparece siempre que se calculen coordenadas en el espacio R^J, por ejemplo la coordenada j-ésima del centro de gravedad de la nube (y origen de los ejes de máxima inercia) es $\frac{z_{.j}}{nQ}$ la distancia entre dos individuos i e i' es:

$$d^2(i,i')^{(B)} = \frac{n}{Q}\sum_{j\in\mathcal{J}}\frac{1}{z_{.j}}\left(z_{ij}-z_{i'j}\right)^2 = \frac{z}{nQ}\, d^2(i,i')^{(A)}$$

y la distancia de un individuo al origen es:

$$d^2(i,O_I)^{(B)} = \frac{z}{nQ}d^2(i,O_I)^{(A)} \tag{10}$$

El peso de los puntos i es el asignado en el análisis (A), es decir, $g_{i.} = 1/n$.

La inercia de un perfil fila es:

$$\text{Inercia de } (i)^{(B)} = \frac{z}{nQ}\text{ Inercia de } (i)^{(A)} \tag{11}$$

7.4 Nube de modalidades: $\mathcal{N}(\mathcal{J})$

El punto j se representa en R^n por el perfil $z_{ij}/z_{.j}$. La nube de modalidades es la misma que la analizada en (A). Por ello coinciden en los análisis (A) y (B) el centro de gravedad $(z_{i.}/z)$, el origen $g_{i.} = 1/n$ y las distancias entre dos perfiles columna y de un perfil columna al origen:

$$d^2(j,j')^{(B)} = n\sum_{i\in\mathcal{I}}\left(\frac{z_{ij}}{z_{.j}}-\frac{z_{ij'}}{z_{.j'}}\right)^2$$
$$d^2(j,O_J)^{(B)} = \frac{n}{z_{.j}}-1$$

La diferencia entre las nubes de modalidades se refleja en los pesos de los puntos (en el análisis (B) es $z_{.j}/nQ$ en lugar de $z_{.j}/z$).

Al mantenerse las distancias y variar los pesos las inercias de cada modalidad, cuestión y de la nube global son:

$$\text{Inercia de } (j)^{(B)} = \frac{n-z_{.j}}{nQ} = \frac{z}{nQ}\text{ Inercia de } (j)^{(A)} \tag{12}$$

$$\text{Inercia de } (q)^{(B)} = \frac{z}{nQ} \text{ Inercia de } (q)^{(A)} \tag{13}$$
$$\text{Inercia total}^{(B)} = \frac{z}{nQ} \text{ Inercia total}^{(A)} \tag{14}$$

7.5 Obtención de los factores

Calcular la sucesión de ejes $(u_{s\,(B)}, s \in \mathcal{S} = \{1, \ldots, s, \ldots, S\}, S \leq J)$ que maximizan la inercia proyectada de la nube $\mathcal{N}(\mathcal{I})$ equivale a:

$$\begin{array}{lll} \text{maximizar:} & u_{s\,(B)}^T\, M_{(B)}\, X_{(B)}^T\, P_{(B)}\, X_{(B)}\, M_{(B)}\, u_{s\,(B)} & \\ \text{sujeto a:} & u_{s\,(B)}^T\, M_{(B)}\, u_{s\,(B)} = 1 & \\ & u_{s\,(B)}^T\, M_{(B)}\, u_t = 0 & \forall t < s \end{array}$$

donde los términos generales de las matrices $X_{(B)}, M_{(B)}$ y $P_{(B)}$ son:

$$\begin{aligned} x_{ij}^{(B)} &= n\,\frac{z_{ij}}{z_{.j}} - 1 \\ m_j^{(B)} &= \frac{z_{.j}}{nQ} \\ p_i^{(B)} &= \frac{1}{n} \end{aligned}$$

8 Relación entre los planos factoriales de los análisis (A): Z^*/z y (B): Z^*/nQ

Comparando las matrices X, P y M en los dos análisis puede comprobarse que las dos primeras coinciden mientras que las matrices M mantienen la siguiente relación:

$$M_{(B)} = \frac{z}{nQ}\, M_{(A)}$$

Por ello, las matrices que han de ser diagonalizadas en el segundo análisis (tanto en la nube de filas como de columnas) y los valores propios asociados coinciden con los del primer análisis multiplicados por el término z/nQ.

Los vectores directores de los ejes de máxima inercia proyectada de la nube de individuos (u_s) son normalizados para métricas diferentes ($M_{(A)}$ y $M_{(B)}$) guardando la relación:

$$u_s^{(B)} = \sqrt{\frac{nQ}{z}}\; u_s^{(A)}$$

Las proyecciones de esta nube son:

$$\begin{aligned} F_s^{(B)} &= X_{(B)}\, M_{(B)}\, u_s^{(B)} = \\ &= \sqrt{\frac{z}{nQ}}\, F_s^{(A)} \end{aligned}$$

Los vectores directores de la nube de columnas (v_s) y las proyecciones de esta nube coinciden en ambos análisis porque la métrica en el espacio R^n coincide $(P_{(B)} = P_{(A)})$.

9 Estudio de la asociación entre las cuestiones

El estudio de la independencia entre dos variables cualitativas q y $q' \in \mathcal{Q}$ con $\mathcal{J}_q$ y $\mathcal{J}_{q'}$ modalidades respectivamente se realiza habitualmente a través de su tabla de contingencia donde han de ser clasificados todos los individuos. Para ello, en el caso de que alguna de las cuestiones no haya sido respondida por todos los individuos, será necesario tener presente las modalidades de no respuesta. La tabla de contingencia tendrá la forma:

$$
\begin{array}{|c|ccccc|}
\hline
 & 1 \ \ldots & j' & \ldots & J_{q'} & a' \\
\hline
1 & & & & & \\
\vdots & & & & & \\
j & & b_{jj'}^{qq'} & & & \\
\vdots & & & & & \\
J_q & & & & & \\
a & & & & & \\
\hline
\end{array}
$$

donde a y a' representan las modalidades de no respuesta de las cuestiones q y q' respectivamente.

Un elemento $b_{jj'}^{qq'}; j \in \mathcal{J}_q; j' \in \mathcal{J}_{q'}$ de esta tabla indica el número de individuos que pertenecen simultáneamente a las modalidades j y j' de las cuestiones q y q' respectivamente. Es decir:

$$b_{jj'}^{qq'} = \sum_{i \in \mathcal{I}} z_{ij} z_{ij'} = \text{Card}\{i : z_{ij} = z_{ij'} = 1 | i \in \mathcal{I}\} \qquad j \in \mathcal{J}_q \quad j' \in \mathcal{J}_{q'} \qquad q, q' \in \mathcal{Q}$$

Analizar la independencia entre las cuestiones q y q' lleva a comparar los términos:

$$\frac{b_{jj'}^{qq'}}{n} \quad \text{y} \quad \frac{b_{jj}^{qq} \ b_{j'j'}^{q'q'}}{n^2} \qquad j \in \mathcal{J}_q \quad j' \in \mathcal{J}_{q'} \qquad q, q' \in \mathcal{Q}$$

que, en función de la tabla disyuntiva incompleta, equivale a comparar:

$$\frac{\sum_{i \in \mathcal{I}} z_{ij} z_{ij'}}{n} \quad \text{y} \quad \frac{\sum_{i \in \mathcal{I}} z_{ij} \sum_{i \in \mathcal{I}} z_{ij'}}{n^2} \qquad j \in \mathcal{J}_q \quad j' \in \mathcal{J}_{q'} \qquad q, q' \in \mathcal{Q}$$

Sin embargo, la no respuesta en el caso de cuestionarios con preguntas condicionadas es determinada por una cuestión diferente. Si, por ejemplo, las dos cuestiones han sido condicionadas por la misma pregunta -de tal forma que ambas tienen en común el total de las respuestas ausentes- no existe independencia. Puesto que el interés se centra en la asociación entre las modalidades respondidas, parece más apropiado no tener en consideración la relación entre las modalidades de no respuesta a la hora de analizar la independencia entre ambas cuestiones.

En ACM clásico el análisis simultáneo de la dependencia entre las Q cuestiones se realiza a través de tabla de Burt definida en función de la tabla disyuntiva completa (Z) de la siguiente forma:

$$B = Z^T Z$$

La tabla B de término general $b_{jj'}^{qq'}$ está formada por Q^2 bloques. El bloque $(q \times q')$, de orden $(J_q \times J_{q'})$ es la tabla de contingencia que cruza las respuestas a las cuestiones q y q'.

El elemento $b_{jj'}^{qq'}$, $j, j' \in \mathcal{J}$ es nulo si las dos modalidades son distintas y pertenecen a la misma cuestión y representa el número de individuos que eligen una determinada modalidad (denominado también por $z_{.j}$ o $z_{.j}^{q}$ -en § 3-) si las modalidades j y j' coinciden.

Por analogía se define la matriz B^* de orden $(J \times J)$ que será denominada pseudo-tabla de Burt y puede expresarse en función de la tabla disyuntiva incompleta:

$$B^* = Z^{*T} Z^*$$

La diferencia entre B y B^* radica en que en la matriz B^* el total de cada una de las subtablas que la forman no es constante, sino la cantidad de individuos que responden a las cuestiones q y $q' \in \mathcal{Q}$ (no necesariamente coincidente con el número total de individuos encuestados).

Se denotará como D_{B^*} la matriz cuadrada de orden J donde los términos de la diagonal principal coinciden con los de la diagonal de la matriz B^* y el resto de los elementos son nulos.

Se define la matriz L de orden $(J \times J)$, cuyo elemento $\quad l_{jj'} = 1 \;\; \forall j$ y $j' \in \mathcal{J}$

10 Análisis: (C): B^*/n

Analizar las relaciones, de la forma definida anteriormente, entre cada par de cuestiones equivale a realizar un análisis factorial mediante las matrices $X_{(C)}$, $M_{(C)}$ y $P_{(C)}$ cuyos términos generales son:

$$\begin{aligned}
x_{jj'}^{(C)} &= \left(\frac{b_{jj'}^{qq'}}{n} - \frac{b_{jj}^{qq}\, b_{j'j'}^{q'q'}}{n^2} \right) \left(\frac{b_{jj}^{qq}\, b_{j'j'}^{q'q'}}{n^2} \right)^{-1} \\
m_j^{(C)} &= \frac{b_{jj}^{qq}}{n} \\
p_j^{(C)} &= \frac{b_{jj}^{qq}}{n}
\end{aligned}$$

Estas matrices pueden ser expresadas en función de la matriz (B^*) y las matrices L y D_{B^*} de la siguiente forma:

$$X_{(C)} \;\; = \;\; n\, D_{B^*}^{-1} \left[B^* - \frac{1}{n}\, D_{B^*} L D_{B^*} \right] D_{B^*}^{-1}$$

$$
\begin{aligned}
M_{(C)} &= \frac{1}{n} D_{B^*} \\
P_{(C)} &= \frac{1}{n} D_{B^*}
\end{aligned}
$$

Aplicando la expresión $A_{(C)} = X_{(C)}^T \, P_{(C)} \, X_{(C)} \, M_{(C)}$ se obtiene la matriz que ha de ser diagonalizada para obtener los ejes de máxima inercia proyectada:

$$
A_{(C)} = D_{B^*}^{-1} \left[B^* - \frac{1}{n} D_{B^*} L D_{B^*} \right] D_{B^*}^{-1} \left[B^* - \frac{1}{n} D_{B^*} L D_{B^*} \right]
$$

11 Análisis (D): B^*/nQ^2

En lugar de eliminar de cada una de las tablas de contingencia las modalidades de no respuesta para crear la pseudo tabla de Burt se podría haber partido de una tabla de Burt (incluyendo las modalidades de no respuesta), haber calculado el total de la tabla (nQ^2) con el que obtener las frecuencias conjuntas $(\frac{b_{jj}^{qq}}{nQ})$ y marginales para finalmente eliminar las modalidades de no respuesta.

De esta forma se compararían los términos:

$$
\frac{b_{jj'}^{qq'}}{nQ^2} \quad \text{y} \quad \frac{b_{jj}^{qq}}{nQ} \, \frac{b_{j'j'}^{q'q'}}{nQ} \qquad j \in \mathcal{J}_q \quad j' \in \mathcal{J}_{q'} \qquad q, q' \in \mathcal{Q}
$$

es decir, se utilizaría la misma tranformación que en correspondencias múltiples clásico, dividir la tabla entre nQ^2 en lugar de entre n.

El análisis se realiza utilizando las matrices:

$$
\begin{aligned}
x_{jj'}^{(D)} &= \left(\frac{b_{jj'}^{qq'}}{nQ^2} - \frac{b_{jj}^{qq} \, b_{j'j'}^{q'q'}}{n^2 Q^2} \right) \left(\frac{b_{jj}^{qq} \, b_{j'j'}^{q'q'}}{n^2 Q^2} \right)^{-1} \\
m_j^{(D)} &= \frac{b_{jj}^{qq}}{nQ} \\
p_j^{(D)} &= \frac{b_{jj}^{qq}}{nQ}
\end{aligned}
$$

12 Relación entre los planos factoriales y los valores propios de los análisis (C): B^*/n y (D): B^*/nQ^2

Las matrices X de ambos análisis coinciden mientras que las matrices de pesos y métrica son proporcionales:

$$
X_{(D)} = X_{(C)}
$$

$$
\begin{aligned}
M_{(D)} &= \frac{1}{Q} M_{(C)} \\
P_{(D)} &= \frac{1}{Q} P_{(C)}
\end{aligned}
$$

Por ello, tanto la matriz a diagonalizar como los valores y vectores propios y las proyecciones estan relacionados:

$$
\begin{aligned}
A_{(D)} &= X_{(D)}^T P_{(D)} X_{(D)} M_{(D)} = \\
&= \frac{1}{Q^2} A_{(C)} \\
\lambda_s^{(D)} &= \frac{1}{Q^2}\, \lambda_s^{(C)} \\
u_{s\,(D)} &= \sqrt{Q}\; u_{s\,(C)} \\
F_s^{(D)} &= X_{(D)}\, M_{(D)}\, u_{s\,(D)} = \\
&= \frac{1}{Q}\, F_s^{(C)}
\end{aligned}
$$

13 Relación entre los análisis (A): Z^*/z y (C): B^*/n

Para poner de manifiesto la equivalencia entre ambos análisis comprobaremos que la matriz que ha de ser diagonalizada en el análisis de la tabla disyuntiva incompleta es el cuadrado, excepto por un factor de escala, de aquélla que ha de ser diagonalizada en el análisis de la pseudo-tabla de Burt.

Matriz a diagonalizar en el análisis factorial (A)

La matriz a diagonalizar en el análisis factorial con marginal modificada de la nube $\mathcal{N}(\mathcal{I})$ cuyo término general aparecía recogido en la ecuación (7) se puede expresar matricialmente como:

$$
A_{(A)} \;=\; \frac{n}{z}\, D_{B^*}^{-1} \left[Z^{*T} Z^* - \frac{1}{n}\, D_{B^*} L D_{B^*} \right]
$$

Equivalencia entre ambos análisis

Multiplicando la matriz $A_{(A)}$ por sí misma se obtiene:

$$
A_{(A)}^2 \;=\; \frac{n^2}{z^2}\, A_{(C)}
$$

La diagonalización de la matriz $A_{(A)}$ produce los valores propios $\lambda_s^{(A)}$ y los ejes de máxima inercia proyectada $u_{s\,(A)}$. Diagonalizando su cuadrado se obtienen los mismos vectores propios y el cuadrado de sus valores propios:

$$\begin{aligned} A_{(A)}\, u_{s\,(A)} &= \lambda_s^{(A)}\, u_{s\,(A)} \\ A^2_{(A)}\, u_{s\,(A)} &= \left[\lambda_s^{(A)}\right]^2 u_{s\,(A)} \end{aligned}$$

Por tanto, la relación entre los valores propios de ambos análisis es:

$$\lambda_s^{(C)} = \left(\frac{z}{n}\lambda_s^{(A)}\right)^2$$

Los vectores propios (ejes de máxima inercia proyectada) son proporcionales porque están normalizados para métricas diferentes.

$$\begin{aligned} M_{(C)} &= \frac{1}{n} D_{B^*} = \frac{z}{n} M_{(A)} \\ 1 &= u^T_{s\,(C)}\, M_{(C)}\, u_{s\,(C)} = u^T_{s\,(A)}\, M_{(A)}\, u_{s\,(A)} \\ u_{s\,(C)} &= \sqrt{\frac{n}{z}}\, u_{s\,(A)} \end{aligned}$$

Ambos análisis son por tanto equivalentes, producen factores proporcionales aunque para diferentes valores propios y tasas de inercia asociadas. Recordando que:

$$X_{(C)} = z A_{(A)}\, D^{-1}_{B^*}$$

y que en el análisis de la pseudo-tabla de Burt los factores $G_s^{(C)}$ coinciden con los $F_s^{(C)}$

$$\begin{aligned} G_s^{(A)} &= \frac{1}{\sqrt{\lambda_s^{(A)}}} X^T_{(A)}\, P_{(A)}\, F_s^{(A)} = \\ &= \frac{1}{\sqrt{\lambda_s^{(A)}}} X^T_{(A)}\, P_{(A)}\, X_{(A)}\, M_{(A)}\, u_{s\,(A)} = \\ &= \frac{1}{\sqrt{\lambda_s^{(A)}}} A_{(A)}\, u_{s\,(A)} \\ G_s^{(C)} &= F_s^{(C)} = \\ &= X_{(C)}\, M_{(C)}\, u_{s\,(C)} = \\ &= z\, A_{(A)}\, D^{-1}_{B^*}\, \frac{1}{n} D_{B^*}\, \sqrt{\frac{n}{z}}\, u_{s\,(A)} = \\ &= \sqrt{\frac{z}{n}}\, A_{(A)}\, u_{s\,(A)} = \\ &= \sqrt{\lambda_s^{(A)}}\, \sqrt{\frac{z}{n}} G_s^{(A)} \end{aligned}$$

14 Relaciones entre los valores propios de los cuatro análisis

Resumiendo las relaciones entre los valores propios que hemos visto:

- $$\left(\frac{z}{nQ}\lambda_s^{(A)}\right)^2 = \left(\lambda_s^{(B)}\right)^2 = \lambda_s^{(D)} = \frac{1}{Q^2}\lambda_s^{(C)}$$
- Los valores propios del análisis (A) pertenecen al intervalo $(0, \frac{nQ}{z})$
- Los valores propios de los análisis (B) y (D) pertenecen al intervalo $(0, 1)$
- Los valores propios del análisis (C) pertenecen al intervalo $(0, Q^2)$

15 Independencia entre las cuestiones a través del análisis (A)

Al igual que en el análisis de correspondencias múltiples con datos completos, el análisis de la tabla disyuntiva incompleta está asociado a unas tasas de inercia pequeñas que no deben ser interpretadas como partes de información explicada por los ejes. A continuación se estudia el caso en el cual las cuestiones son independientes dos a dos y se propone una corrección a estas tasas de inercia en el análisis con marginal modificada de una tabla disyuntiva incompleta.

La matriz de diagonalización en el análisis con marginal modificada de la tabla disyuntiva incompleta, tiene un término general (recogido en la ecuación (7)) que puede ser expresado en función de los elementos de la tabla disyuntiva incompleta de la siguiente forma:

$$a_{jj'}^{(A)} = \frac{n}{z}\left(\frac{1}{z_{.j}}\sum_{i\in\mathcal{I}} z_{ij}\, z_{ij'} - \frac{1}{n}\, z_{.j'}\right) \qquad j, j' \in \mathcal{J}$$

Notar que:

- $\sum_{i\in\mathcal{I}} z_{ij}\, z_{ij'} = 0 \quad$ si $\quad j, j' \in \mathcal{J}_q \quad$ y $\quad j \neq j' \quad q \in \mathcal{Q}$
  porque un individuo no puede pertenecer a dos modalidades de la misma cuestión
- $\sum_{i\in\mathcal{I}} z_{ij}\, z_{ij'} = z_{.j} \quad$ si $\quad j, j' \in \mathcal{J}_q \quad$ y $\quad j = j' \quad q \in \mathcal{Q}$
  número de individuos que pertenecen a una determinada modalidad
- $\sum_{i\in\mathcal{I}} z_{ij}\, z_{ij'} = \frac{z_{.j}\, z_{.j'}}{n} \quad$ si $\quad j \in \mathcal{J}_q\,, \quad j' \in \mathcal{J}_{q'}\,, \quad q \neq q' \quad q, q' \in \mathcal{Q}$ y además existe independencia entre ambas cuestiones

y que por tanto:

- $a_{jj'}^{(A)} = -\frac{z_{.j'}}{z} \quad$ si $\quad j, j' \in \mathcal{J}_q \quad$ y $\quad j \neq j' \quad q \in \mathcal{Q}$

- $a_{jj'}^{(A)} = \dfrac{n - z_{.j'}}{z}$ si $j, j' \in \mathcal{J}_q$ y $j = j'$ $q \in \mathcal{Q}$
- $a_{jj'}^{(A)} = 0$ si $j \in \mathcal{J}_q$, $j' \in \mathcal{J}_{q'}$, $q \neq q'$ $q, q' \in \mathcal{Q}$ y existe independencia entre ambas cuestiones.

En consecuencia, la matriz $A_{(A)}$ que se ha de diagonalizar tiene la forma:

$$A_{(A)} = \begin{bmatrix} A^1_{(A)} & & & & 0 \\ & \ddots & & & \\ & & A^q_{(A)} & & \\ & & & \ddots & \\ 0 & & & & A^Q_{(A)} \end{bmatrix}$$

donde cada submatriz $A^q_{(A)}$ es:

$$A^q_{(A)} = \begin{bmatrix} \dfrac{n - z^q_{.1}}{z} & \dfrac{-z^q_{.2}}{z} & \dfrac{-z^q_{.3}}{z} & \dots & \dfrac{-z^q_{.J_q}}{z} \\ \dfrac{-z^q_{.1}}{z} & \dfrac{n - z^q_{.2}}{z} & \dfrac{-z^q_{.3}}{z} & \dots & \dfrac{-z^q_{.J_q}}{z} \\ \dfrac{-z^q_{.1}}{z} & \dfrac{-z^q_{.2}}{z} & \dfrac{n - z^q_{.3}}{z} & \dots & \dfrac{-z^q_{.J_q}}{z} \\ & & & \ddots & \\ \dfrac{-z^q_{.1}}{z} & \dfrac{-z^q_{.2}}{z} & \dfrac{-z^q_{.3}}{z} & \dots & \dfrac{n - z^q_{.J_q}}{z} \end{bmatrix}$$

Siendo $z^q_{.j}$ equivalente al $z_{.j}$ anterior que representa el número de individuos que eligen la modalidad j de la cuestión q.

Para diagonalizar la matriz $A_{(A)}$ se ha de resolver la ecuación:

$$|A_{(A)} - \lambda_s{}^{(A)} I| \;=\; 0$$

Debido a la estructura diagonal por bloques de la matriz $A_{(A)}$, los valores $\lambda_s{}^{(A)}$ (incluyendo los valores nulos) que solucionan esa ecuación son los resultantes de resolver el sistema de Q ecuaciones siguiente:

$$|A^q_{(A)} - (\lambda^q_s)^{(A)} I_{J_q}| \;=\; 0$$

siendo I_{J_q} la matriz identidad de orden J_q.

Las matrices $A^q_{(A)}$ pueden expresarse como:

$$A^q_{(A)} \;=\; \frac{n}{z} I_{J_q} - E^q$$

siendo:

$$E^q = \begin{bmatrix} \frac{z^q_{.1}}{z} & \frac{z^q_{.2}}{z} & \frac{z^q_{.3}}{z} & \dots & \frac{z^q_{.J_q}}{z} \\ \vdots & & & \ddots & \vdots \\ \frac{z^q_{.1}}{z} & \frac{z^q_{.2}}{z} & \frac{z^q_{.3}}{z} & \dots & \frac{z^q_{.J_q}}{z} \end{bmatrix}$$

Como se puede comprobar fácilmente, esta matriz (de orden J_q) es de rango 1 y su traza es z_q/z (número de individuos que responden la cuestión q entre el número total de respuestas de los n individuos a las Q cuestiones), por ello tiene un valor propio $\mu^q = z_q/z$ y $(J_q - 1)$ valores propios nulos.

A través de la relación entre los valores propios de $A^q_{(A)}$ y de esta matriz E^q:

$$\begin{aligned} A^q_{(A)}\, u_{s\,(A)} &= (\lambda^q_s)^{(A)}\, u_{s\,(A)} \\ \tfrac{n}{z} I_{J_q}\, u_{s\,(A)} - E^q\, u_{s\,(A)} &= (\lambda^q_s)^{(A)}\, u_{s\,(A)} \\ E^q\, u_{s\,(A)} &= \left(\tfrac{n}{z} - (\lambda^q_s)^{(A)}\right) u_{s\,(A)} \\ E^q\, u_{s\,(A)} &= \mu^q_s\, u_{s\,(A)} \end{aligned}$$

donde $s = 1, \dots, J_q$, se obtienen los valores propios de cada matriz $A^q_{(A)}$:

$$(\lambda^q_s)^{(A)} = \begin{cases} \dfrac{n}{z} & \text{si } s = 1, \dots, J_q - 1 \\ \dfrac{n}{z} - \dfrac{z_q}{z} & \text{si } s = J_q \end{cases} \qquad \forall q \in \mathcal{Q}$$

y de la matriz $A_{(A)}$ que resultan ser:

$$\lambda_s^{(A)} = \begin{cases} \dfrac{n}{z} & \text{si } s = 1, \dots, J - Q \\ \dfrac{n}{z} - \dfrac{z_q}{z} & \text{si } s = (J - Q + 1), \dots, J \qquad \forall q \in \mathcal{Q} \end{cases}$$

Donde existen tantos valores propios nulos como cuestiones a las que responden todos los individuos ($z_q = n,\ q \in \mathcal{Q}$).

Al igual que en el caso de tablas disyuntivas completas, la independencia entre cuestiones no se refleja en una inercia nula y no se debe, obviamente, a asociaciones entre las cuestiones sino a un efecto de estructura o de construcción de la tabla disyuntiva incompleta. Cada uno de los J ejes estaría recogiendo una inercia trivial que evidentemente engorda los valores propios del análisis de la tabla disyuntiva incompleta cuando no existe la independencia.

16 Tasas de inercia

Cuando no existen datos ausentes, la aplicación del análisis de correspondencias simples a la tabla de Burt lleva a la obtención de los mismos factores del análisis de la tabla disyuntiva completa. En el caso particular de dos cuestiones, ambos análisis producen los mismos factores que el análisis de correspondencias simples de la tabla de contingencia. Sin embargo, los valores propios que se obtienen en los tres análisis son diferentes y dan lugar a distintas tasas de inercia proyectada sobre cada uno de los ejes.

Se pone de manifiesto de esta forma que el análisis de correspondencias múltiples (bien se realice a través de la tabla disyuntiva completa o bien a partir de la tabla de Burt) estudia las relaciones de dependencia entre cada par de cuestiones y se revela el escaso interés de los valores propios como medida de la información explicada por cada uno de los factores.

Benzécri (Benzécri 1979) propone por ello, basándose en la equivalencia entre el análisis de correspondencias de la tabla disyuntiva completa y de la tabla de contingencia cuando el número de cuestiones es dos, la siguiente corrección de los valores propios:

$$\lambda_s^* = \left(\frac{Q}{Q-1}\right)^2 \left(\lambda_s - \frac{1}{Q}\right)^2 \qquad\qquad s \in \mathcal{S}^*$$

donde $\mathcal{S}^*$ es el subconjunto de $\mathcal{S}$ con valores λ_s (valores propios resultantes del análisis de la tabla disyuntiva completa) superiores a $1/Q$. Al estar los valores propios λ_s comprendidos entre 0 y 1, el término $\left(\frac{Q}{Q-1}\right)^2$ permite obtener unos valores propios corregidos comprendidos también entre 0 y 1, que hacen posible su comparación con otros análisis. Esta modificación de los valores propios lleva a definir las tasas de inercia proyectada:

$$\tau_s^* = \frac{\lambda_s^*}{\sum\limits_{s \in \mathcal{S}^*} \lambda_s^*}$$

Esta corrección encuentra su justificación, para el caso de más de dos cuestiones, en la equivalencia entre los análisis de la tabla disyuntiva completa y de la tabla de Burt. En éste último análisis se introducen en la diagonal principal tablas que cruzan una cuestión consigo misma haciendo incrementar la inercia total y donde el resto de las tablas de contingencia aparecen dos veces (Greenacre 1993).

Una razón alternativa para calcular las tasas de inercia corregidas se encuentra en el estudio del caso particular donde las Q cuestiones son independientes dos a dos. A pesar del nulo interes del análisis de correspondencias clásico cuando se conoce (en ocasiones únicamente tras los resultados del análisis) la independencia de las cuestiones, su aplicación proporciona una inercia total no nula y $(J - Q)$ factores con valores propios asociados iguales a $1/Q$ (Zárraga 1989), demostrando que los valores propios del análisis (exista o no independencia) recogen una inercia trivial debida a un efecto de estructura o de construcción de la tabla disyuntiva completa.

Cuando existen datos ausentes, el análisis del caso en el cual las cuestiones son independientes, en la forma definida, revela que las tasas de inercia calculadas como los valores propios entre la inercia total, no son una buena medida de la asociación entre las cuestiones recogida por cada eje, por ello se propone, en el caso general en que las cuestiones no son independientes, calcular los valores propios y las tasas de inercia de la siguiente forma:

$$\begin{aligned} \lambda_s^* &= \left(\frac{z}{n(Q-1)}\right)^2 \left(\lambda_s{}^{(A)} - \frac{n}{z}\right)^2 \qquad\qquad s \in \mathcal{S}^* \\ \tau_s^* &= \frac{\lambda_s^*}{\sum\limits_{s \in \mathcal{S}^*} \lambda_s^*} \end{aligned}$$

donde $\lambda_s{}^{(A)}$ son los valores propios obtenidos en el análisis - (A) - con marginal modificada de la tabla disyuntiva incompleta cuando no existe independencia y $\mathcal{S}^*$ es el subcojnunto de $\mathcal{S}$ con valores $\lambda_s{}^{(A)}$ mayores que n/z.

Dado que los valores propios $\lambda_s{}^{(A)}$ pertenecen al intervalo $(0, \frac{nQ}{z})$, los valores propios corregidos están comprendidos entre 0 y 1.

Estos nuevos valores propios y tasas de inercia coinciden, si la tabla es disyuntiva completa, en la que $z = nQ$, con los propuestos por Benzécri (1979).

Referencias

Abascal, E. & Grande, I. (1989), *Métodos multivariantes para la investigación comercial. Teoría, aplicaciones y programación BASIC*, Ariel economía.

Benali, H. (1985), *Stabilité de l'analyse en composantes principales et de l'analyse des correspondances multiples en présence de certains types de perturbations. Méthodes de dépouillement d'enquêtes. Thèse de troisième cycle*, Université de Rennes I.

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.

Benali, H. & Escofier, B. (1987), 'Stabilité de l'analyse factorielle des correspondances multiples en cas de données manquantes et de modalités à faibles effectifs', *Revue de Statistique Appliquée* **XXXV**(1), 41–51.

Benzécri, J. (1979), 'Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire, addendum et erratum à', *Les Cahiers de l'Analyse des Données* **IV**(3), 377–378.

Escofier, B. (1981), 'Traitement des questionnaires avec non réponse, analyse des correspondances avec marge modifiée et analyse multicanonique avec contrainte', *INRIA*.

Escofier, B. (1990), 'Traitement des variables incomplètes en analyse des correspondances multiples', *Revue de Modulad* **5**, 13–27.

Escofier, B. & Pagès, J. (1992), *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación*, Servicio Editorial Universidad del País Vasco.

Goitisolo, B. & Zárraga, A. (1998*a*), Application of the incomplete disjunctive tables study to the C.A.V. living conditions survey., *in* 'Analyses Multidimensionnelles des donnes. IV$^{\text{éme}}$ Congrès International NGUS$'$97', Fernandez-Aguirre, K. and Morineau, A., pp. 301–313.

Goitisolo, B. & Zárraga, A. (1998*b*), Equivalence between the incomplete disjunctive table and the associated burt pseudo-table analysis., *in* 'Analyses Multidimensionnelles des donnes. IV$^{\text{éme}}$ Congrès International NGUS$'$97', Fernandez-Aguirre, K. and Morineau, A., pp. 227–238.

Greenacre, M. (1990), 'Some limitations of multiple correspondence analysis', *Computational Statistics Quartely* **3**, 249–256.

Greenacre, M. (1993), *Correspondence Analysis in Practice*, Academic Press.

Lebart, L., Morineau, A. & Tabard, N. (1977), *Techniques de la description statistique*, Dunod.

Morineau, A. (1984), 'Note sur la caractérisation statistique d'une classe et les valeurs-tests', *Bulletin Technique du CESIA* **2**(1-2), 20–27.

Zárraga, A. (1989), *Análisis de correspondencias múltiples por bandas. Aplicación al estudio de una gran encuesta. Tesis Doctoral*, Universidad del País Vasco.

Zárraga, A. & Goitisolo, B. (1999), 'Independencia entre las cuestiones en el análisis factorial de tablas disyuntivas incompletas con preguntas condicionadas', *Qüestiió* **23**(3), 465–488.