

EXPLOTACIÓN DE LA ENCUESTA SOBRE LA SOCIEDAD DE LA INFORMACIÓN –ESI– MEDIANTE EL ANÁLISIS SIMULTÁNEO

Amaya Zárraga
Beatriz Goitisoló

Universidad del País Vasco

RESUMEN

En este trabajo se presenta el Análisis Simultáneo (Zárraga y Goitisoló, 2002 y 2003a) como una nueva metodología útil para el tratamiento de encuestas.

El Análisis Simultáneo permite analizar conjuntamente varias tablas de contingencia o de frecuencias, como por ejemplo las resultantes de una misma encuesta efectuada en diferentes países o en diferentes épocas o de encuestas diferentes efectuadas sobre una misma población. El método se puede extender al análisis de un conjunto de encuestas cuando los datos de los diferentes grupos de individuos son codificados en tablas disyuntivas completas yuxtapuestas para las mismas modalidades. El Análisis Simultáneo puede también utilizarse para el tratamiento estadístico de las preguntas abiertas contenidas en una encuesta (Zárraga y Goitisoló, 2003b).

Las ventajas de este método sobre otros ya existentes son resaltadas especialmente cuando las tablas a analizar poseen diferentes totales y diferentes marginales.

En el trabajo se aplica el Análisis Simultáneo a los datos procedentes de la Encuesta sobre la Sociedad de la Información elaborada por el Instituto Vasco de Estadística (EUSTAT) en el cuarto trimestre del año 2000, con el objeto de analizar la conducta de los individuos encuestados en relación con las nuevas tecnologías de la información y la comunicación y su comportamiento frente a Internet.

Palabras clave: *análisis simultáneo, análisis conjunto de tablas, análisis conjunto de encuestas, sociedad de la información.*

Introducción

El objetivo de este trabajo es doble. Por una parte, presentar un nuevo método de análisis factorial, denominado Análisis Simultáneo, que basado en las ya conocidas técnicas de Análisis de Correspondencias y Análisis Factorial Múltiple (Escofier y Pagès, 1988), permite el tratamiento y estudio simultáneo de varias tablas de datos, dando solución a los problemas existentes en el tratamiento con las técnicas clásicas.

El Análisis Simultáneo (Zárraga y Goitisoló, 2002 y 2003a) es especialmente apropiado para el estudio de varias tablas de contingencia, tablas disyuntivas completas, tablas disyuntivas incompletas y tablas de Burt o pseudo-tablas de Burt (Zárraga y Goitisoló, 1999).

Una situación interesante para el Análisis Simultáneo se presenta cuando se posee una misma encuesta realizada en distintos países o en diferentes períodos de tiempo. Las filas de cada tabla podrían ser grupos de individuos según el sexo, el intervalo de edad al que pertenecen y su situación laboral, etc. 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 ó 1990. Con el Análisis Simultáneo propuesto se pretende responder a preguntas del tipo, ¿las características de los individuos 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.

Otra circunstancia 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, etc.) 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 Simultáneo 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)?

El nuevo método de análisis que se presenta permite responder a estas preguntas estudiando las tendencias generales de las tablas, sin permitir que una de las tablas tenga una preponderancia sobre el resto.

El segundo objetivo del trabajo consiste en aplicar el Análisis Simultáneo a los datos procedentes de la Encuesta sobre la Sociedad de la Información elaborada por el Instituto Vasco de Estadística (EUSTAT) en el cuarto trimestre del año 2000.

Se analizará la conducta de los individuos encuestados en relación con las nuevas tecnologías y su comportamiento frente a Internet. Concretamente, se pondrán de manifiesto, según el nivel de estudios de los encuestados, las características comunes y/o diferentes que motivan su interés o desinterés, en su acceso a Internet, por los diversos temas –museos virtuales, servicios administrativos y ciudadanos, gestión bancaria de la bolsa y financiera, juegos y hobbies, etc– incluidos en la encuesta ESI.

Metodología

Sea $G = \{1, \dots, g, \dots, G\}$ el conjunto de tablas de contingencia a analizar (figura 1). 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, $I = \{1, \dots, i, \dots, I\}$, quedan recogidas en las filas) en 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 g son $J_g = \{1, \dots, j, \dots, J_g\}$. Al yuxtaponer todas estas tablas de contingencia, se posee un conjunto $J = \{1, \dots, j, \dots, J\}$ de columnas.

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

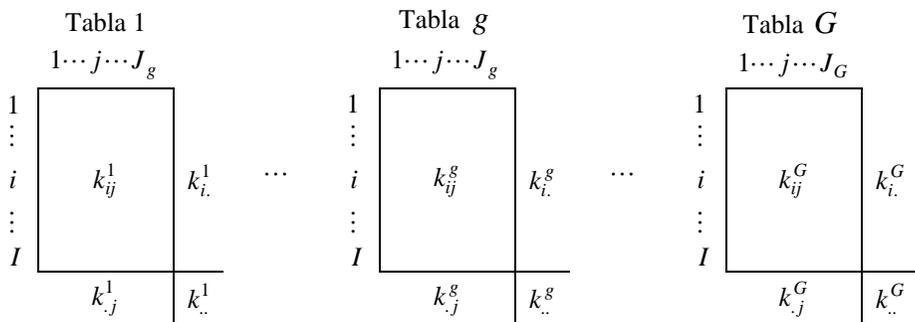


Figura 1: Sucesión de tablas

Si la sucesión de tablas de contingencia procede de una tabla ternaria, las modalidades J_g , $g = 1, \dots, G$, son comunes a las G tablas. El Análisis Simultáneo es, sin embargo, extensible al estudio de tablas con diferentes modalidades en columnas.

La metodología clásicamente utilizada para analizar estas tablas consiste en realizar los análisis factoriales de correspondencias separados de las diferentes tablas de contingencia y/o el análisis factorial de correspondencias de la yuxtaposición de las tablas (en el caso de que las filas de las tablas sean comunes). Sin embargo, los resultados derivados de esta manera de proceder pueden verse afectados, como se señala en Bécue-Bertaut y Pagès (2000) por:

- Las diferencias entre los perfiles de las marginales fila de las diferentes tablas.
- La importancia relativa de las tablas en el análisis, medida a través de las contribuciones de columnas, a su vez debidas a:

- las diferencias entre los totales de las tablas: manteniendo el resto de las cosas constantes, una tabla influye tanto más cuanto mayor es su efectivo total.
- diferencias de intensidad de estructura entre las tablas: manteniendo el resto de las cosas constantes, una tabla influye tanto más en el análisis global cuanto más fuerte es su estructura.

Una metodología para analizar varias tablas de naturaleza cualitativa ha sido propuesta por (Abascal y otros, 2001) considerando las diferencias de intensidad de estructura entre las tablas.

Más recientemente ha sido propuesto el Análisis Factorial Múltiple para Tablas de Contingencia (Becué-Bertaut y Pagès, 2003) con el mismo objetivo de analizar conjuntamente varias tablas de contingencia. Este último método, más próximo al Análisis Simultáneo, no tiene en cuenta los diferentes totales de las tablas e impone una marginal común para todas ellas. En el análisis global no se mantiene, entonces, la estructura interna de cada tabla.

El Análisis Simultáneo (Zárraga y Goitisoló, 2002 y 2003a) permite dar solución a los inconvenientes citados proporcionando una descripción conjunta de las diferentes estructuras contenidas dentro de cada una de las tablas así como una comparación de las mismas.

Para solucionar los problemas mencionados en el apartado 1, el método de análisis de la yuxtaposición de tablas ha de permitir:

- Equilibrar la influencia de las tablas transformando el efectivo de cada una de las tablas.
- Equilibrar la influencia de las tablas según las diferencias de intensidad de estructura entre las tablas.
- Conservar en un análisis factorial global tanto los pesos como la métrica de cada una de las tablas.

El Análisis Simultáneo ajusta los efectivos de las tablas como primer paso para equilibrar la influencia de las tablas. Para ello, cada una de las tablas de contingencia es transformada, dividiéndola por $k_{..}^g$, total de la g -ésima tabla, $g \in G$, en una tabla de frecuencias relativas.

Se denotará:

$$f_{ij}^g = \frac{k_{ij}^g}{k_{..}^g} \quad f_{.j}^g = \sum_{i \in I} f_{ij}^g \quad i \in I$$

$$f_{i.}^g = \sum_{j \in J} f_{ij}^g \quad f_{..}^g = \sum_{i \in I} \sum_{j \in J_g} f_{ij}^g = 1 \quad j \in J$$

$$g \in G$$

A fin de controlar la influencia en el análisis global de aquellas tablas con fuerte estructura, el Análisis Simultáneo introduce una ponderación, que denotaremos α^g , $g \in G$, en cada uno de los grupos (aquí tablas de contingencia). Esta ponderación dependerá de las inercias que resultan de realizar un análisis de correspondencias simple de

cada una de las tablas tratadas por separado. Estas inercias para la tabla g , $g \in G$, serán denotadas por λ_s^g (la inercia proyectada sobre el eje s , $s \in S$) y por λ_{TOT}^g (la inercia total). La ponderación puede ser la equivalente al Análisis Factorial Múltiple (Escofier y Pagès, 1984) para tablas continuas ($1/\lambda_i^g$), la equivalente al método STATIS (L'Hermier des Plantes, 1976 y García y Simon, 2002), ($1/\lambda_{TOT}^g$) para privilegiar grupos de dispersión mínima, 1 si no se desea una ponderación de las tablas, etc.

Si se considera $\alpha^g = 1/\lambda_1^g$, $g \in G$, ponderación utilizada en la aplicación, el primer valor propio del análisis factorial del conjunto de tablas de contingencia estará comprendido entre 0 y G . Si se desea permanecer en el marco del análisis de correspondencias y que el primer valor propio del análisis del conjunto de tablas esté comprendido entre 0 y 1, bastará utilizar α^g / α como ponderación con $\alpha = \sum_{g \in G} \alpha_g$.

Por último, para que cada tabla conserve sus pesos y métrica en el análisis global, el método que presentamos requiere transformaciones en las nubes de perfiles columna y perfiles fila que se detallan a continuación.

Análisis de las columnas

A cada una de las g , $g \in G$, tablas de contingencia está asociada una subnube, $N(J_g)$, de J_g perfiles columna centrados:

$$\left\{ \begin{array}{l} \frac{f_{ij}^g}{f_{.j}^g} - f_i^g \mid i \in I \end{array} \right\} \quad \begin{array}{l} j \in J_g \\ g \in G \end{array}$$

con los pesos $f_{.j}^g$, $j \in J_g$, y la métrica de matriz asociada la matriz diagonal de $1/f_{.j}^g$, $i \in I, g \in G$.

Para analizar conjuntamente las tablas sobreponderamos cada subnube por α^g , $g \in G$, y dado que todas ellas están situadas en el mismo espacio, R^I , consideramos la nube global, $N(J)$, 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.

Se pueden transformar los perfiles columna de cada una de las tablas para considerar sus distancias euclídeas. En el análisis conjunto esto significa considerar los perfiles columna:

$$\left\{ \frac{\sqrt{\alpha^g}}{\sqrt{f_{.j}^g}} \left(\frac{f_{ij}^g}{f_{.j}^g} - f_i^g \right) \mid i \in I \right\} \quad \begin{array}{l} j \in J_g \subset J \\ g \in G \end{array}$$

con los pesos $f_{.j}^g$, $j \in J$, y la métrica euclídea usual.

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

Análisis de las filas

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

$$\left\{ \begin{array}{l} \frac{f_{ij}^g}{f_{i.}^g} - f_{.j}^g \mid j \in J_g \end{array} \right\} \quad \begin{array}{l} i \in I \\ g \in G \end{array} \quad (1)$$

con pesos $f_{i.}^g$, $i \in I$, y la métrica de matriz asociada la matriz diagonal de $1/f_{.j}^g$, $j \in J_g$, $g \in G$.

Dado que se busca analizar conjuntamente las g , $g \in G$, tablas y que 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 (1) completando el resto de las coordenadas con 0. El perfil fila parcial i^g , $i \in I$, $g \in G$, tiene por tanto por coordenadas:

$$i^g = \begin{cases} \frac{f_{ij}^g}{f_{i.}^g} - f_{.j}^g & \text{si } j \in J_g \\ 0 & \text{en otro caso} \end{cases} \quad (2)$$

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

Con el fin de realizar el análisis conjunto, se busca para cada fila un representante, que llamaremos "compromiso" y denotaremos i^c , $i \in I$, que caracterice 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 , dotado de la métrica de matriz asociada la matriz diagonal de $\alpha^g / f_{.j}^g$, $j \in J_g \subset J$, forman la nube $N(I^c)$. 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:

$$In(i^c) = \sum_{g \in G} \alpha^g In(i^g) \quad (3)$$

y que, por tanto, la inercia de la nube compromiso se exprese como suma ponderada de las inercias de las nubes parciales. Para ello se define el compromiso i^c como una media ponderada de los perfiles fila parciales i^g con pesos p_{i^c} , $i \in I$,

$$i^c = \frac{\sum_{g \in G} \sqrt{f_{i.}^g}}{\sum_{g \in G} \sqrt{f_{i.}^g}} i^g \quad p_{i^c} = \left(\sum_{g \in G} \sqrt{f_{i.}^g} \right)^2$$

con $\sum_{i \in I} p_{i^c} = p \neq 1$.

Si se considera $p_i^* = p_{i^c} / p$ (con $\sum_{i \in I} p_i^* = 1$), los resultados factoriales sólo se verán alterados en la proporción $1/p$. Se considera p_{i^c} en la medida en que hace las fórmulas más simples.

Obtención de los factores

Con el fin de buscar la relación entre los análisis de filas y columnas en los desarrollos siguientes, el análisis factorial del conjunto de las tablas de contingencia se realiza obteniendo los valores propios (λ_s) y vectores propios (u_s), $s \in S$, resultantes de la diagonalización de la matriz $X^T X$ donde el término general de la matriz X es:

$$x_{ij} = \sqrt{\alpha^g} \sqrt{f_{i.}^g} \left(\frac{f_{ij}^g}{f_{i.}^g f_{.j}^g} - 1 \right) \sqrt{f_{.j}^g} \quad \begin{array}{l} i \in I \\ j \in J \\ g \in G \end{array} \quad (4)$$

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

$$r_{ii} = p_{i^c} \quad n_{jj} = f_{.j}^g \quad \begin{array}{l} i \in I \\ j \in J_g \subset J \\ g \in G \end{array}$$

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 , $g \in G$, y se define la matriz Q , también diagonal por bloques, donde cada bloque es, a su vez, una matriz diagonal de orden (I) y de término general $f_{i.}^g$, $i \in I$, $g \in G$.

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 ($\lambda_s, u_s \in R^J$) y columnas ($\mu_s, v_s \in R^I$) es:

$$\begin{aligned}\lambda_s &= \mu_s & s \in S \\ u_s &= \lambda_s^{-1/2} X^T v_s & s \in S\end{aligned}\quad (5)$$

$$v_s = \lambda_s^{-1/2} X u_s \quad s \in S \quad (6)$$

Proyecciones de los perfiles

Se calculan las proyecciones sobre el eje s , $s \in S$, de los perfiles fila parciales $F_s(i^g)$, de los compromisos $F_s(i^c)$, $i \in I$, y de los perfiles columna $G_s(j)$, $j \in J$. Las proyecciones de todos los perfiles fila parciales y compromisos son respectivamente $F_s(I^G)$ con $I^G = \{I^g / g \in G\}$ y $F_s(I^c)$ y de todas las columnas $G_s(J)$. Las proyecciones sobre el eje s , $s \in S$, 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:

$$F_s(I^c) = R^{-1/2} X u_s = \lambda_s^{1/2} R^{-1/2} v_s \quad (7)$$

$$G_s(J) = N^{-1/2} X^T v_s = \lambda_s^{1/2} N^{-1/2} u_s \quad (8)$$

$$F_s(I^G) = Q^{-1/2} Y u_s \quad (9)$$

Ayudas a la interpretación

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

Las contribuciones de los puntos a la formación del eje s , $s \in S$, son calculadas de la manera 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:

$$cta_s(i^c) = \frac{P_{i^c} F_s^2(i^c)}{\lambda_s} \quad cta_s(j) = \frac{f_{j^g} G_s^2(j)}{\lambda_a} \quad \begin{array}{l} i \in I \\ j \in J_g \subset J \\ g \in G \end{array}$$

La calidad de representación de un punto sobre el eje s , $s \in S$, se mide por las contribuciones relativas. Se calcula como el cociente de la inercia proyectada del punto sobre el eje s , $s \in S$, entre la inercia total del punto o como el cuadrado de la proyección sobre el eje s , $s \in S$, entre la distancia, al cuadrado, del punto al origen:

$$ctr_s(i^c) = \frac{F_s^2(i^c)}{d^2(i^c, 0)} \quad ctr_s(j) = \frac{G_s^2(j)}{d^2(j, 0)} \quad \begin{array}{l} i \in I \\ j \in J_g \subset J \\ g \in G \end{array}$$

siendo la distancia, al cuadrado, de la columna $j, j \in J$ al origen:

$$d^2(j,0) = \sum_{i \in I} \alpha^g \left(\frac{f_{ij}^g}{f_{i \cdot}^g} - f_{i \cdot}^g \right)^2 \frac{1}{f_{i \cdot}^g} = \sum_s G_s^2(j) \quad g \in G$$

y la distancia, al cuadrado, del compromiso al origen:

$$d^2(i^c,0) = \sum_{g \in G} \alpha^g \frac{f_{i \cdot}^g}{p_{i^c}} \sum_{j \in J_g} \frac{1}{f_{i \cdot}^g} \left(\frac{f_{ij}^g}{f_{i \cdot}^g} - f_{i \cdot}^g \right)^2 = \sum_s F_s^2(i^c) \quad i \in I$$

Las relaciones entre las proyecciones de filas (compromiso) y columnas así como sus propiedades se pueden consultar en (Zárraga y Goitisoló, 2002 y Goitisoló, 2002)

Aplicación del Análisis Simultáneo Propuesto

Se desea conocer en detalle la conducta de los individuos en relación con las nuevas tecnologías y en particular el comportamiento frente a Internet. A pesar de ser un tema de actualidad, existen aún escasos estudios encaminados a conocer la actitud de la población hacia estas tecnologías. Para responder a cuestiones, como por ejemplo, ¿qué temas interesan a los estudiantes?, ¿qué grupos de personas utilizan más Internet para chatear?, ¿los temas de interés en el acceso a Internet están relacionados con la situación profesional?, ¿es la misma relación en los diferentes niveles de estudios? se dispone de la Encuesta sobre la Sociedad de la Información elaborada por el Instituto Vasco de Estadística –EUSTAT– en el cuarto trimestre del año 2000. Esta encuesta pretende recabar información sobre el uso de las nuevas Tecnologías de la Información y de la Comunicación entre la población de 15 y más años de la Comunidad Autónoma de Euskadi –CAE–.

De los 4627 individuos encuestados en el cuarto trimestre del año 2000 se seleccionan los 1018 individuos que afirman conectarse a Internet.

Además de la información de la Encuesta sobre la Sociedad de la Información se dispone, para cada uno de los individuos, de variables sociodemográficas extraídas de la Encuesta de Población en Relación con la Actividad (P.R.A.) elaborada también por el Eustat. Las variables de caracterización utilizadas son presentadas en la tabla 1.

Las variables que hacen referencia a los temas que interesan al usuario en su acceso a Internet como son los museos virtuales –TMU–, los contactos y debates políticos –TPO–, la educación a distancia –TED–, la telemedicina y el telediagnóstico –TME–, los servicios administrativos y ciudadanos –TSA–, la preparación de viajes –TVI–, los diarios electrónicos y las noticias –TNO–, la búsqueda de empleo –TBE–, los derechos de los consumidores –TDC–, la gestión bancaria, de la bolsa y los servicios financieros –TFI–, las bibliotecas y documentación –TBI– y los juegos y hobbies –TJU– aparecen en la tabla 2.

Tabla 1: *características Sociodemográficas.*

SEX		Sexo
	1	Varón
	2	Mujer
THI		Territorio Histórico de residencia
	1	Álava
	2	Bizkaia
	3	Guipúzcoa
ECI		Estado Civil
	1	Soltero/a
	2	Casado/a
	3	Otro
EDA		Edad cumplida durante semana de refer.
	1	15-24 años
	2	25-34 años
	3	35-44 años
	4	45 ó más
EST		Nivel de instrucción terminado
	1	Primarios o menos
	2	Secundarios
	3	Superiores
RAC		Relación con la actividad
	1	Estudiantes
	2	Ocupados
	3	Resto (inactivos y parados)

RAE		Rama de actividad econ. (sector)
	1	Industria
	2	Construcción
	3	Servicios
	4	Otros
PRO		Profesión, ocupación u oficio
	1	Profesionales y técnicos
	2	Personal directivo
	3	Empleados administrativos
	4	Comerciantes y vendedores
	5	Personal de servicios
	6	Obreros
	7	Otros (16, inact, par 1er, ayud fami)
SPR		Situación profesional
	1	Empleadores
	2	Autónomos
	3	Miembros de cooperativas
	4	Asalariados del sector público
	5	Asalariados del sector privado
	6	Parados buscan 1er empleo
	7	Otros (16, inact, ayud fami)
TCO		Tipo de contrato (para ocupados asalariados)
	1	Indefinido fijo (permanente o discontinuo)
	2	Temporal (formación, estacional, ocasional u otros)
	3	Otros

Tabla 2: temas de Interés en el Acceso a Internet.

TMU		Museos Virtuales
TPO		Contactos y debates políticos
TED		Educación a distancia
TME		Telemedicina, telediagnóstico
TSA		Servicios administrativos, ciudadanos
TVI		Preparación de viajes
TNO		Diarios electrónicos, noticias
TBE		Búsqueda de empleo
TDC		Derechos de los consumidores
TFI		Gestión bancaria, de la bolsa y financiera
TBI		Bibliotecas, documentación
TJU		Juegos, hobbies
	1	Interesado
	2	No interesado

En los gráficos, cada una de las modalidades será identificada por la etiqueta de la variable, los 3 caracteres en la primera columna de las tablas 1 y 2, seguida del número de la modalidad. En el caso de los perfiles fila parciales y de los perfiles columna de los análisis simultáneos, se precederá esta notación del número indicador del grupo al que pertenece dicho perfil.

Análisis Simultáneo según el Nivel de Estudios

Para estudiar las diferencias entre los individuos en función de su nivel de estudios, se realiza a continuación un análisis simultáneo de dos tablas donde se clasifica a los individuos según su nivel de estudios terminados (Figura 2). En la primera tabla se incluyen aquellos individuos con estudios secundarios o inferiores y en la segunda los encuestados que poseen estudios superiores. Cada una de las tablas clasifica, además, a los individuos en función de las características sociodemográficas (en filas) y los temas de interés en su acceso a Internet (en columnas).

	Estudios No Superiores	Estudios Superiores
	Temas	Temas
Variables de Caracterización		

Figura 2: tabla analizada.

El **primer factor** de este análisis, posee una inercia de 1.917, 67.27% de la inercia total. Este valor es muy próximo al máximo alcanzable en este caso (dos), indicando que es un eje común y de inercia importante en ambas tablas.

(Figura 4) y enfrentados a los mayores de 24 años, trabajadores interesados en temas financieros –TFI1– e indiferentes hacia los juegos y hobbies –TJU2–.

El **segundo factor** explica el 10.15% de la inercia total. Enfrenta, principalmente, el interés por la búsqueda de empleo –TBE1–, por la preparación de viajes –TVII–, por la telemedicina y el telediagnóstico –TME1– y por la educación a distancia –TED1– a las modalidades que indican indiferencia hacia estas materias. Aunque son los perfiles columna pertenecientes a la tabla correspondiente a los encuestados con estudios superiores los que más contribuyen a la creación de este segundo eje, un 72.47% (ver tabla 3), las proyecciones de los perfiles columna pertenecientes a la tabla de los encuestados con estudios no superiores permiten extender la interpretación a todos los individuos.

Tabla 3: *contribuciones de cada tabla al An. Simultáneo*

	Eje 1	Eje 2	Eje 3
Estudios No Superiores	0.5002	0.2753	0.3062
Estudios Superiores	0.4998	0.7247	0.6938

Puede comprobarse cómo la búsqueda de empleo –TBE1– está vinculada con contratos temporales –TCO2–. El interés por la telemedicina y el telediagnóstico –TME1–, la educación a distancia –TED1– y por la preparación de viajes –TVI– es más intenso entre las mujeres –SEX2–, los individuos entre 25 y 34 años y empleados administrativos –PRO3– y reducido entre los hombres –SEX1–, comerciantes y vendedores –PRO4–, empleadores –SPR1– e individuos mayores de 44 años.

Utilizando las proyecciones de los perfiles fila parciales de las modalidades de edad (Figura 5) puede apreciarse cómo la diferencia apuntada, en temas de interés, entre hombres –SEX1– y mujeres –SEX2– es más acentuada entre los encuestados con niveles de estudios superiores. Además, puede afirmarse que con estudios secundarios o inferiores no existe diferencia en la atracción por los juegos y hobbies –TJU– y la gestión bancaria y servicios financieros –TFI–. También es posible apreciar la gran homogeneidad entre los jóvenes de menos de 25 años independientemente de su nivel de estudios, mientras que entre los 25 y 34 años y entre los parados que buscan su primer empleo –SPR6– existe una mayor tendencia hacia la búsqueda de empleo –TBE1– entre los individuos con estudios superiores que entre aquéllos sin ellos. Lo contrario, es decir, una mayor indiferencia por la búsqueda de empleo –TBE2– con estudios superiores, se produce a partir de los 45 años y entre el personal directivo –PRO2– y comerciantes y vendedores –PRO4–.

Los empleados administrativos –PRO3– se interesan más por la educación a distancia –TED1–, por la telemedicina y el telediagnóstico –TME1– y por la preparación de viajes –TVII– cuando su nivel de estudios es mayor.

El **tercer factor**, eje horizontal de las figuras 6 y 7, con un porcentaje de inercia explicada del 7.14% enfrenta principalmente a asalariados del sector público –SPR4– con los asalariados del sector privado –SPR5–.

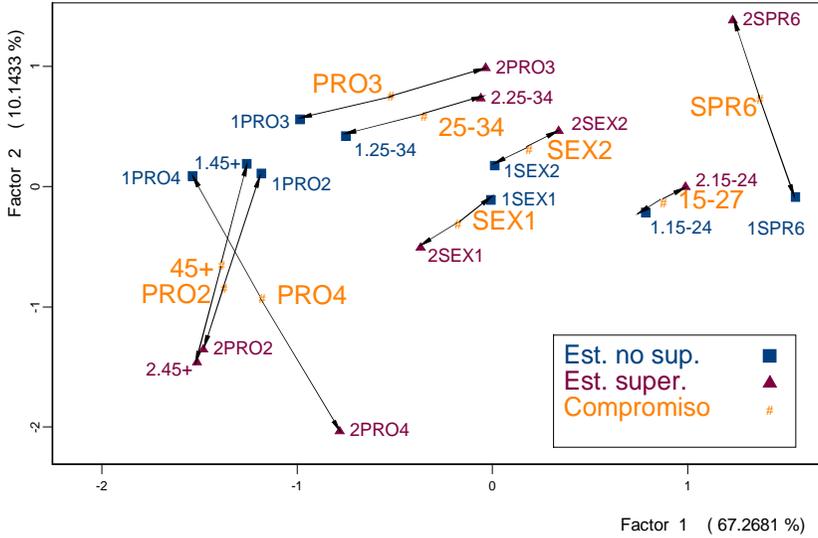


Figura 5: *perfiles fila parciales y compromisos .*

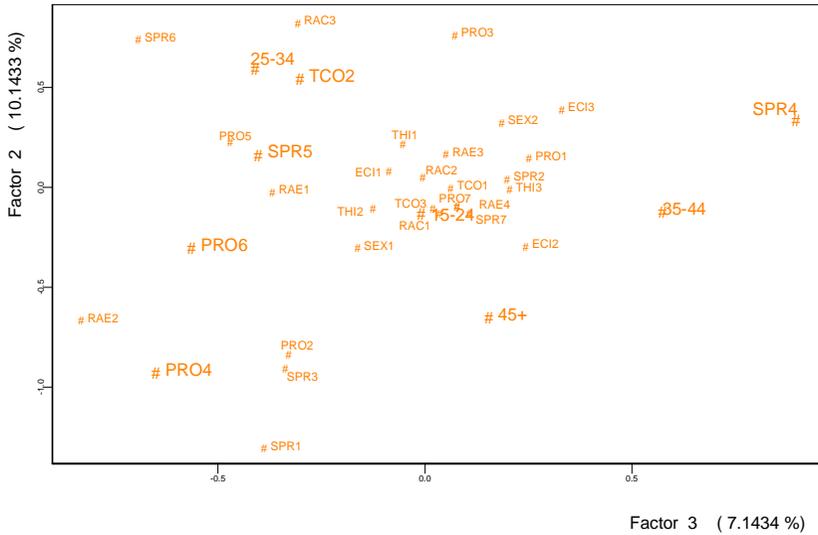


Figura 6: *proyección de los compromisos.*

Referencias

- Abascal, E. y otros (2001) Diferentes aplicaciones de las técnicas factoriales de análisis de tablas múltiples en las investigaciones mediante encuestas. *Metodología de Encuestas*, 3 (2) 251-279.
- Bécue-Bertaut, M. y Pagès, J. (2000) Analyse Factorielle Multiple Intra-tableaux. Application à l'Analyse Simultanée de Plusieurs Questions Ouvertes. JADT 2000: *5és Journées Internationales d'Analyse Statistique des Données Textuelles*.
- Bécue-Bertaut, M. y Pagès, J. (2003) A principal axes method for comparing contingency tables: MFACT. *Computational Statistics & Data Analysis*, available online 21 June 2003.
- Escofier, B. y Pagès, J. (1984) L'Analyse Factorielle Multiple. *Cahiers de Bureau Universitaire de Recherche Opérationnelle*, 42, 1-68.
- Escofier, B. y Pagès, J. (1988 (1998, 3ra edición). *Analyses Factorielles Simples et Multiples. Objectifs, méthodes et interprétation*, Dunod, París.
- EUSTAT- Instituto Vasco de Estadística (www.eustat.es).
- García, I. y Simón, E. (2002) Implementación del método STATIS mediante SPAD. *Metodología de Encuestas*, 4 (1) 129-139.
- Goitisoló, B. (2002) , *El Análisis Simultáneo. Propuesta y Aplicación de un Nuevo Método de Análisis Factorial de Tablas de Contingencia*, Tesis doctoral, Universidad del País Vasco. Servicio editorial.
- L'Hermier des Plantes, H. (1976) *STATIS: Structuration de Tableaux À Trois Indices de la Statistique*. Thèse (3c), USTL, Montpellier.
- Zárraga, A. y Goitisoló, B. (1999) Independencia entre las cuestiones en el análisis factorial de tablas disyuntivas incompletas con preguntas condicionadas. *Qüestió* 23 (3) 465-488.
- Zárraga, A. y Goitisoló, B. (2002) Méthode factorielle pour l'analyse simultanée de tableaux de contingence. *Revue de Statistique Appliquée*, L (2) 47-70.
- Zárraga, A. y Goitisoló, B. (2003a) Étude de la structure Inter-tableaux à travers l'Analyse Simultanée. *Revue de Statistique Appliquée*, LI (3) 39-60.
- Zárraga, A. y Goitisoló, B. (2003b) Analyse de Tableaux Ternaires de Données Textuelles. *Lexicometrica*, 4.