

ESTUDIO DE HISTORIAS DE VIDA MEDIANTE ANÁLISIS DE CORRESPONDENCIAS. UNA APLICACIÓN A LA ENCUESTA DE TRANSICIÓN EDUCATIVO-FORMATIVA E INSERCIÓN LABORAL

Julio Abad González
Pilar Blanco Alonso
Ana García Gallego
Universidad de León

RESUMEN

En las últimas décadas, muchos son los trabajos que han puesto de manifiesto la importancia de adoptar una perspectiva dinámica a la hora de abordar el estudio de los procesos sociales. Por esta razón, de forma creciente la información que se recoge para realizar este tipo de estudios, se refiere a duraciones en distintos estados y/o tiempos transcurridos entre distintos eventos.

Este tipo de información dinámica, que se suele denominar "datos de historias de vida" (*event history data*), suele ser analizada utilizando procedimientos relacionados con el análisis de supervivencia o los modelos de regresión para datos de duración. Sin embargo, en esta comunicación se adopta un enfoque distinto al utilizar un método multivariante exploratorio como es el Análisis de Correspondencias y aplicarlo a la *super-matriz indicador* que recoge las historias de vida de los individuos como secuencias de eventos.

Para ilustrar la utilización de esta metodología, presentaremos una aplicación de la misma al análisis de datos procedentes de la *Encuesta de Transición Educativo-Formativa e Inserción Laboral* elaborada por el Instituto Nacional de Estadística en 2005. Dicha encuesta tiene por objetivo conocer las diferentes formas de transición desde la educación y la formación al mercado laboral mediante el estudio de las trayectorias formativas y laborales de 45.000 jóvenes entre los años 2001 y 2005.

Palabras clave: análisis de correspondencias, modelos de duración, historias de vida, transición al empleo.

1. INTRODUCCIÓN

Aunque desarrollado inicialmente para el estudio de tablas de contingencia bidimensionales, el Análisis de Correspondencias es una técnica estadística cuya aplicación puede ser útil en aquellas situaciones en que sea posible construir una matriz de datos positivos donde sea de interés comparar perfiles-fila, perfiles-columna o ambas cosas. Este hecho legitima su uso en otros ámbitos, y en particular, en el marco del estudio de datos de historias de acontecimientos (*event history data*).

Por ello, con el fin de mostrar la utilización de la técnica en este ámbito, presentamos una aplicación de la misma al análisis de datos procedentes de la *Encuesta de Transición Educativo-Formativa e Inserción Laboral* (EteFIL) elaborada por el Instituto Nacional de Estadística en 2005.

De acuerdo con este planteamiento, la estructura de esta comunicación es la siguiente:

- En el primer apartado se presentan brevemente las principales características teóricas del Análisis de Correspondencias para, a continuación, mostrar los pasos a seguir en su aplicación cuando se analizan datos de historias de acontecimientos.
- El segundo apartado se refiere a los datos: en primer lugar, se hace una descripción general de los objetivos y características de la Encuesta de Transición Educativo-

Formativa e Inserción Laboral (ETEFIL) y, seguidamente, se hace referencia a los datos que de ella se han extraído y que serán objeto de análisis.

- El tercer apartado muestra las distintas aplicaciones de esta técnica a los datos extraídos de la ETEFIL y describe los principales resultados obtenidos.
- Por último, se exponen las conclusiones más relevantes del trabajo realizado hasta la fecha ya que este estudio se inscribe en un proyecto más amplio cuyo desarrollo está aún en curso.

2. METODOLOGÍA

Como ya se ha señalado en la introducción, en este apartado describiremos brevemente los fundamentos teóricos del Análisis de Correspondencias para, seguidamente, presentar diferentes procedimientos que permiten su aplicación al estudio de historias de vida.

2.1. Análisis de Correspondencias

Comenzaremos este apartado abordando de forma breve en qué consiste esta técnica de análisis estadístico y cuáles son sus objetivos y características. No obstante, un desarrollo más exhaustivo, se puede encontrar en Benzécri (1992), Greenacre (1993) o Jambu (1989, pp. 193-328).

Como ya se ha señalado, el Análisis de Correspondencias es un método inicialmente adaptado para el tratamiento de tablas de contingencia –tablas de frecuencias bidimensionales– que permite estudiar las eventuales relaciones existentes entre las filas y columnas de dicha tabla a través de la representación gráfica simultánea de las mismas. No obstante, tal y como señala Jambu (1989, p. 252), su aplicación se puede extender (y de hecho, se ha extendido) al estudio de otros tipos de tablas de datos siempre que las nociones de perfiles-fila y perfiles-columna, a los que nos referiremos a continuación, tengan una interpretación natural.

Partimos, por tanto, de una tabla de contingencia K que se obtiene clasificando una población o muestra según dos variables: la variable I , con n modalidades: $I = \{1, 2, \dots, i, \dots, n\}$, y la variable J , con p modalidades: $J = \{1, 2, \dots, j, \dots, p\}$, de modo que cada elemento de la tabla k_{ij} indica la frecuencia conjunta de la fila i y de la columna j . Las frecuencias marginales son designadas por $k_i = \sum_j k_{ij}$ para las filas y por $k_j = \sum_i k_{ij}$ para las columnas, siendo $k = \sum_i \sum_j k_{ij}$ la frecuencia total de la tabla. Todas estas frecuencias, que son absolutas, pueden ser transformadas en frecuencias relativas f_{ij} a través de la relación $f_{ij} = k_{ij}/k$.

Tablas de perfiles-fila y de perfiles-columna.

Dado que las distancias entre filas (o entre columnas) están condicionadas por sus masas, es necesario hacerlas comparables mediante su transformación en perfiles-fila y en perfiles-columna.

La tabla de perfiles-fila recoge las probabilidades condicionadas de presentar una determinada modalidad j de la variable J sabiendo que se presenta la modalidad i de la variable I (f_{ij}/f_i). El conjunto de perfiles-fila forma una nube de n puntos en el espacio de las p columnas. Cada punto i tiene por coordenadas en \mathbb{R}^p : $\{f_{ij}/f_i; j = 1, 2, \dots, p\}$, y su masa f_i es igual a su frecuencia relativa. El centro de gravedad G_I de esta nube es la media de los perfiles-fila ponderados por sus masas, y se corresponde con el perfil-fila medio, es decir, con las frecuencias marginales de las columnas (f_j).

Análogamente, la tabla de perfiles-columna recoge las probabilidades condicionadas de presentar una determinada modalidad i de la variable I sabiendo que se presenta la modalidad j de la variable J (f_{ij}/f_j). El conjunto de perfiles-columna constituye una nube de p puntos en el espacio de las n filas. Las coordenadas en \mathbb{R}^n del punto j son: $\{f_{ij}/f_j; i = 1, 2, \dots, n\}$, y cada punto está afectado de una masa f_j . El centro de gravedad G_J de la nube de perfiles-columna coincide

con el perfil-columna medio, y sus componentes son las frecuencias marginales de las filas (f_i).

Distancia ji-cuadrado

Se define la distancia χ^2 (ji-cuadrado) entre dos puntos-fila y entre dos puntos-columna como:

$$d_{\chi^2}^2(i, i') = \sum_{j=1}^p \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_i} - \frac{f_{i'j}}{f_i} \right)^2 \quad \text{y} \quad d_{\chi^2}^2(j, j') = \sum_{i=1}^n \frac{1}{f_i} \left(\frac{f_{ij}}{f_j} - \frac{f_{ij'}}{f_j} \right)^2$$

Ésta es una distancia euclídea ponderada por el inverso de la masa de la columna para no favorecer a aquellas columnas que tienen más masa. La distancia χ^2 va a otorgar una misma importancia a las columnas en el cálculo de la distancia entre dos perfiles-fila, y a las filas, en el cálculo de la distancia entre perfiles-columna, independientemente de cuáles sean sus frecuencias relativas.

Inercia de la nube de puntos

La inercia de la nube de puntos-fila respecto a su centro de gravedad es una medida de la dispersión de la nube, y se calcula como suma ponderada de las distancias entre los puntos-fila y su centro de gravedad usando como ponderación la masa de cada punto-fila y como métrica la de la distancia χ^2 :

$$H(I, G_I) = \sum_{i=1}^n f_i d_{\chi^2}^2(i, G_I) = \sum_{i=1}^n f_i \sum_{j=1}^p \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_i} - f_{.j} \right)^2 = \sum_{i=1}^n \sum_{j=1}^p \frac{(f_{ij} - f_i f_{.j})^2}{f_i f_{.j}}$$

Análogamente se obtiene la inercia de la nube de puntos-columna respecto a su centro de gravedad:

$$H(J, G_J) = \sum_{j=1}^p f_{.j} d_{\chi^2}^2(j, G_J) = \sum_{j=1}^p f_{.j} \sum_{i=1}^n \frac{1}{f_i} \left(\frac{f_{ij}}{f_j} - f_i \right)^2 = \sum_{i=1}^n \sum_{j=1}^p \frac{(f_{ij} - f_i f_{.j})^2}{f_i f_{.j}}$$

Como vemos, la inercia de la nube de puntos-fila coincide con la de la nube de puntos columna, y resulta ser igual al estadístico χ^2 de Pearson dividido entre la frecuencia total de la tabla k .

Por lo tanto, la inercia de la nube será tanto mayor cuanto mayor sea la distancia entre los puntos y su centro de gravedad o, planteado de otro modo, cuanto más se alejen las frecuencias de la condición de independencencia.

Obtención de las dimensiones y coordenadas factoriales

El objetivo del Análisis de Correspondencias será determinar sendos subespacios que maximicen las inercias de las proyecciones de las respectivas nubes de puntos-fila y puntos-columna.

Para el ajuste de la nube de puntos-fila, si denotamos por $\psi_{\alpha i}$ la proyección del punto-fila i sobre el eje factorial o dimensión α , la dimensión 1 será aquella que maximice la inercia o varianza ponderada de las proyecciones de los puntos-fila sobre dicha dimensión ($\lambda_1 = \sum_i f_i \psi_{1i}^2$). La dimensión 2 también será aquella que maximice la inercia de las proyecciones de los puntos-fila sobre dicha dimensión ($\lambda_2 = \sum_i f_i \psi_{2i}^2$) pero con la restricción de que las coordenadas de los proyecciones de los puntos-fila sobre la dimensión 2 deben ser ortogonales a las obtenidas para la dimensión 1 ($\sum_i f_i \psi_{1i} \psi_{2i} = 0$). Las dimensiones sucesivas serán también aquellas que maximicen la inercia de las proyecciones con la restricción de ser ortogonales a las dimensiones anteriores.

El ajuste de la nube de puntos-columna se realizará de forma análoga. Si se denota por $\varphi_{\alpha j}$ la proyección del punto-columna j sobre el eje factorial o dimensión α , la dimensión 1 será aquella que maximice la inercia ponderada de las proyecciones de los puntos-columna sobre dicha dimensión ($\lambda_1 = \sum_j f_j \varphi_{1j}^2$) las dimensiones siguientes serán esas que maximicen la inercia de las proyecciones pero con la restricción de ser ortogonales a todas las dimensiones obtenidas anteriormente.

Relaciones de transición o cuasi-baricéntricas.

Aunque, como se acaba de señalar, el Análisis de Correspondencias da lugar a sendas soluciones para las nubes de los puntos-fila y de los puntos-columna, una característica importante de esta técnica es que ambas soluciones están estrechamente ligadas a través de las denominadas relaciones de transición o cuasi-baricéntricas, que vienen dadas por las expresiones siguientes:

$$\Psi_{\alpha i} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^p \frac{f_{ij}}{f_i} \varphi_{\alpha j} \quad \text{y} \quad \varphi_{\alpha j} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^n \frac{f_{ij}}{f_j} \Psi_{\alpha i}$$

Estas relaciones permiten obtener, para cada dimensión, las coordenadas de los puntos-fila a partir de las coordenadas de los puntos-columna, y viceversa. Las relaciones de transición prueban también la igualdad de las dimensiones del ajuste de la nube de puntos-fila y de la nube de puntos-columna, siendo el número total de dimensiones extraídas igual al mínimo de $n-1$ y $p-1$.

Por último, hay que destacar que la existencia de estas relaciones cuasi-baricéntricas, justifican la representación gráfica simultánea de las filas y de las columnas de la tabla, y constituyen una de las propiedades más destacables del Análisis de Correspondencias.

Reglas de interpretación.

Las nubes de puntos-fila y de puntos-columna se representan en los planos formados por los primeros ejes factoriales tomados de dos en dos, y para su interpretación es necesario tener en cuenta los siguientes elementos:

- **Tasas de inercia.** Indican la importancia de cada dimensión dado que miden la proporción de la inercia total de la nube ($H = \chi^2/k = \sum_\alpha \lambda_\alpha$) que explica la dimensión α ($\tau_\alpha = \lambda_\alpha / \sum_\alpha \lambda_\alpha$).
- **Contribuciones absolutas.** Expresan en qué medida contribuye una fila i o una columna j a la inercia proyectada sobre cada dimensión. Se define como la proporción de la inercia explicada por el factor α que corresponde a la fila i ($C_\alpha(i) = f_i \Psi_{\alpha i}^2 / \lambda_\alpha$) o a la columna j ($C_\alpha(j) = f_j \varphi_{\alpha j}^2 / \lambda_\alpha$).
- **Contribuciones relativas o cosenos cuadrados.** Informan de si un punto está bien representado sobre un eje factorial. La contribución relativa de la dimensión α al punto-fila i (o al punto-columna j) es el porcentaje de la inercia de a la fila i (o a la columna j) que queda explicado por la dimensión α ($\text{Cos}^2_\alpha(i) = \Psi_{\alpha i}^2 / \sum_\alpha \Psi_{\alpha i}^2$ y $\text{Cos}^2_\alpha(j) = \varphi_{\alpha j}^2 / \sum_\alpha \varphi_{\alpha j}^2$, respectivamente).

Elementos suplementarios.

Una última característica interesante del Análisis de Correspondencias es la posibilidad de representar elementos suplementarios en los planos factoriales obtenidos del análisis. Se trata de añadir filas o columnas a la tabla base para, una vez obtenidos los ejes factoriales, obtener sus coordenadas factoriales y representarlas junto al resto de elementos. A estos elementos se les denomina también ilustrativos, dado que, a diferencia del resto de elementos – denominados activos –, no intervienen en la determinación de los ejes factoriales, sino que se

introducen a posteriori con el fin de "ilustrar" o facilitar la interpretación de las representaciones gráficas.

2.2. Extensiones del Análisis de Correspondencias: aplicación al estudio de historias de vida

Como ya hemos apuntado, el Análisis de Correspondencias puede extenderse a otros tipos de tablas que no son *strictu sensu* de contingencia, siempre y cuando siga teniendo sentido estudiar las diferencias entre los perfiles-fila, entre los perfiles-columna, o entre los dos. En este sentido, una de las generalizaciones más simples y, probablemente la más utilizada, es el **Análisis de Correspondencias Múltiples** (a la versión a la que nos hemos referido anteriormente se le suele denominar Análisis de Correspondencias Simples). Esta técnica consiste en aplicar el Análisis de Correspondencias a una *tabla disyuntiva completa*, es decir, una tabla cuyas filas corresponden a individuos, y que tiene por columnas las distintas modalidades –codificadas como indicadores de presencia/ausencia (unos y ceros)– de un conjunto de variables cualitativas.

Asimismo, también tiene sentido la utilización del Análisis de Correspondencias en el estudio de datos de historias de vida. Los datos de historias de vida o datos de historias de acontecimientos (*event history data*) son registros biográficos en los que, para cada individuo de la muestra, se recoge no sólo una secuencia con los estados en que se encuentra éste a lo largo del periodo de estudio, sino también cuánto tiempo se encuentra en cada uno de ellos. Cada una de estas historias individuales puede ser codificada en términos de frecuencias del siguiente modo: cada frecuencia indicaría el tiempo que cada individuo ha estado en cada uno de los distintos estados mutuamente excluyentes considerados durante el periodo de estudio (por ejemplo, cuántos meses ha estado un individuo estudiando, trabajando o desempleado durante un año). Si no sólo es de interés el tiempo transcurrido en cada estado, sino también el orden en que los diferentes estados se van sucediendo en el tiempo, se construye una nueva tabla yuxtaponiendo las *m* subtablas obtenidas de la división del periodo de estudio en *m* subperiodos. En último término, si la unidad de medida del tiempo coincide con el subperiodo definido, se llegaría a una "supermatriz de indicadores" de presencia/ausencia (compuesta de unos y ceros) que indicaría si el individuo estaba en ese estado durante ese subperiodo o no.

La Figura 1 recoge sendos ejemplos de ambos tipos de codificación: la tabla "agregada" muestra como el individuo 1 permaneció 6 meses estudiando (*E*), 3 trabajando (*T*) y 3 desempleado (*D*); esa misma información aparece desglosada por meses de modo que cada patrón 100, 010 ó 001 indica en qué situación está el individuo ese mes (*E*, *T* o *D*, respectivamente). Así, entre enero y junio estudiaba, de julio a septiembre estaba desempleado, y entre octubre y diciembre tenía trabajo.

Figura 1. Tabulación de datos de historias de vida para Análisis de Correspondencias

Tabla agregada				Supermatriz de Indicadores											
<i>i</i>	<i>E</i>	<i>T</i>	<i>D</i>	ene	feb	mar	abr	may	jun	jul	ago	sep	oct	nov	dic
<i>i</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>	<i>ETD</i>
1	6	3	3	100	100	100	100	100	100	001	001	001	010	010	010
2	9	0	3	100	100	100	100	100	100	100	100	100	001	001	001
3	0	6	6	001	001	001	001	001	001	010	010	010	010	010	010
...

El tratamiento de esta tabla nos permitirá determinar, mediante un Análisis de Correspondencias Múltiples de la supermatriz de indicadores, qué individuos y qué estados tienen perfiles semejantes (o diferentes) entre sí en función de qué estado han ocupado durante cada subperiodo. Además y previamente al análisis de esta tabla, van der Heijden y de Leeuw (1989) proponen que se aplique el Análisis de Correspondencias Simples para estudiar la matriz subperiodos×estados que recoge de forma agregada en qué estados se encuentran los individuos para cada uno de los subperiodos de tiempo definidos. El ACS de esta tabla definirá el

"perfil medio" de la nube mientras que el ACM de la supermatriz de indicadores revelará qué individuos y qué estados se alejan de este "perfil medio".

3. DATOS: ETEFIL-2005

La aplicación que presentamos se ha realizado a partir de micro-datos proporcionados por el Instituto Nacional de Estadística (INE) correspondientes a la *Encuesta de Transición Educativo-Formativa e Inserción Laboral* (ETEFIL-2005)¹. De acuerdo con su resumen metodológico, "es la primera investigación estadística de esta magnitud desarrollada en España" dedicada al "estudio del conjunto de itinerarios seguidos dentro del sistema educativo y las transiciones entre el estudio y el trabajo". Esta encuesta es fruto de la colaboración entre el Ministerio de Educación y Ciencia, el Ministerio de Trabajo y Asuntos Sociales y el propio INE, que fue el encargado de llevarla a cabo en el año 2005.

3.1. Población, muestra y periodo de referencia

La población objeto de la encuesta son todas aquellas personas que, siendo menores de 25 años a 31/12/2001, terminaron estudios del sistema educativo no universitario durante el curso 2000/01, o abandonaron la educación secundaria obligatoria sin obtener la titulación en ese mismo curso o completaron programas del sistema de formación ocupacional en 2001.

La muestra final está formada por 45.620 individuos y es el resultado de la agregación de las siete muestras correspondientes a cada uno de los siete colectivos siguientes: alumnos que abandonaron el sistema educativo sin el título de graduado en Educación Secundaria Obligatoria (AESO), graduados en Educación Secundaria Obligatoria (ESO), graduados en Bachillerato (BACH), graduados en Ciclos Formativos de Formación Profesional de Grado Medio (CFGM), y de Grado Superior (CFGS), alumnos que finalizaron un curso del Plan Nacional de Formación e Inserción Profesional (FIP), y alumnos que finalizaron un programa de Escuelas Taller y Casas de Oficios (ETCO).

Aunque el trabajo de campo se realizó entre abril y septiembre de 2005, el periodo de referencia es único para cada individuo, dado que se trata de un estudio longitudinal y retrospectivo acerca de lo realizado por el encuestado desde que finalizó o abandonó el sistema educativo hasta la fecha de la entrevista. No obstante, y con el fin de homogeneizar todas las observaciones, nos hemos centrado en los años 2002 a 2004, periodo durante el cual todos los individuos se encuentran en observación.

3.2. Variables estudiadas en el Análisis de Correspondencias

Como ya se ha indicado, uno de los resultados más interesantes del Análisis de Correspondencias son las representaciones gráficas a que da lugar. Sin embargo, este aspecto gráfico de la técnica no permite el que se incluyan una gran cantidad de variables en el análisis puesto que la profusión de puntos podría dar lugar a una mayor confusión en vez de a una clarificación de los resultados. Por ello, hemos optado inicialmente por restringir el estudio a las siguientes variables por ser las de mayor interés: el mes de observación (entre enero de 2002 y diciembre de 2004), el colectivo de pertenencia (los siete colectivos considerados a los que nos referimos al hablar de la muestra), y la situación básica del encuestado (ver Tabla 1), que recoge de forma exhaustiva los distintos estados mutuamente excluyentes en que un individuo puede estar en cada mes del periodo de referencia.

Asimismo, como ilustrativas se incluyen, además del sexo del individuo, otras variables que surgen al agrupar algunas de las situaciones descritas en la Tabla 1: si estudia en el

¹ http://www.ine.es/prodyser/micro_etefil.htm. Consultado el 17/10/2007.

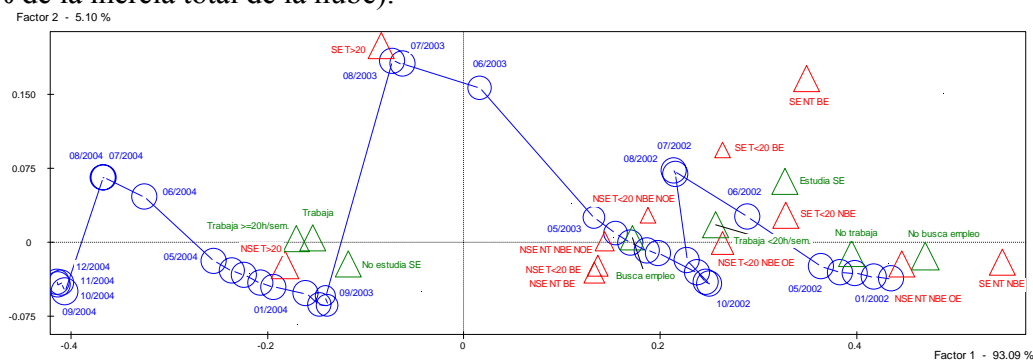
sistema educativo o no, si trabaja o no, y en su caso, si trabaja menos de 20 h./semana o no, y si busca empleo o no.

Tabla1. Categorías de la variable "situación básica del encuestado"

Etiqueta	Categoría
SE T>20	Estudia en el sistema educativo (SE) y trabaja más o igual de 20 h. semanales
SE T<20 BE	Estudia en el SE, trabaja menos de 20 h. semanales y busca empleo
SE T<20 NBE	Estudia en el SE, trabaja menos de 20 h. semanales y no busca empleo
SE NT BE	Estudia en el SE, no trabaja y busca empleo
SE NT NBE	Estudia en el SE, no trabaja y no busca empleo
NSE T>20	No estudia en el SE y trabaja más o igual de 20 h. semanales
NSE T<20 BE	No estudia en el SE, busca empleo y trabaja menos de 20 h. semanales.
NSE T<20 NBE OE	No estudia en el SE, trabaja menos de 20 h. semanales, no busca empleo y recibe algún tipo de educación/formación fuera del SE
NSE T<20 NBE NOE	No estudia en el SE, trabaja menos de 20 h. semanales, no busca empleo y no recibe ningún tipo de educación/formación fuera del SE
NSE NT BE	No estudia en el SE, no trabaja y busca empleo
NSE NT NBE OE	No estudia en el SE, no trabaja, no busca empleo y recibe algún tipo de educación/formación fuera del SE
NSE NT NBE NOE	No estudia en el SE, no trabaja, no busca empleo y no recibe ningún tipo de educación/formación fuera del SE

4. APLICACIONES Y RESULTADOS

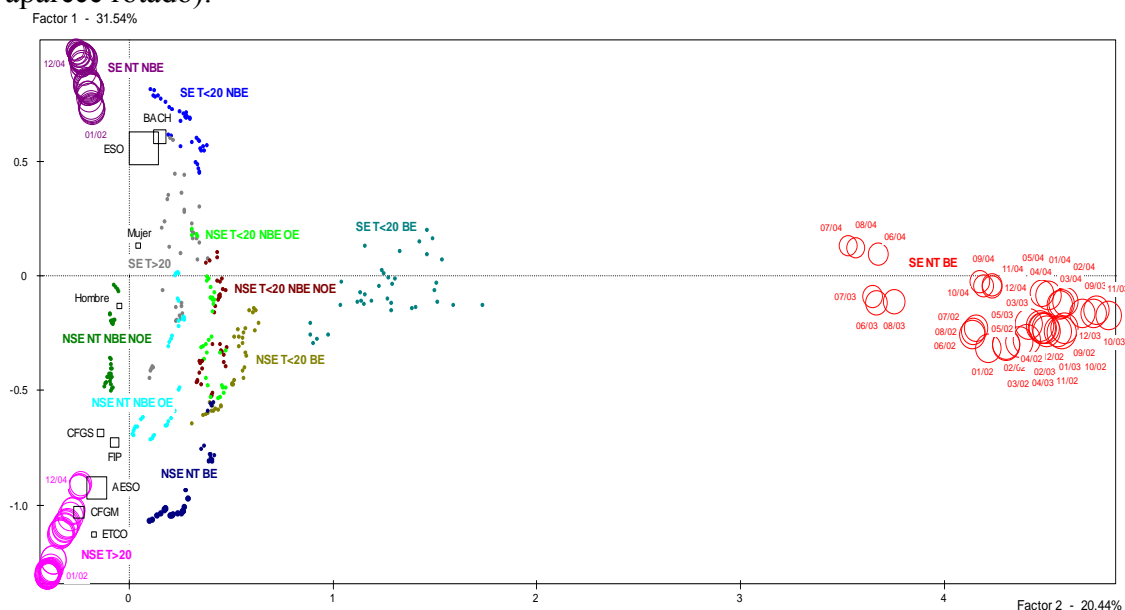
Tal y como se indicaba en el apartado 2.2, el primer análisis es el que se ha de realizar sobre la tabla que define el "perfil medio", es decir, aquella que recoge de forma agregada en qué situaciones se encuentran los individuos para cada uno de los subperiodos de tiempo definidos. A esta tabla, hemos añadido como columnas suplementarias las agrupaciones de las situaciones básicas. Los resultados del análisis se sintetizan en el gráfico siguiente que recoge la representación simultánea de filas y columnas sobre el primer plano factorial (que explica un 98,19% de la inercia total de la nube).



La primera dimensión, a la que corresponde la mayor parte de la inercia (93,09%), queda muy identificada con la secuencia temporal ya que los meses se ordenan prácticamente de forma cronológica en sentido siguiente: desde enero de 2002, en el extremo positivo, hasta diciembre de 2004, en el extremo negativo. Desde el punto de vista de las columnas, las situaciones con mayores contribuciones absolutas son "SE NT NBE", con un 59,8%, en el extremo positivo, y "NSE T>20", con un 26,5% en el negativo. Por lo tanto, este primer eje muestra la transición de una etapa de continuación de estudios (año 2002, sobre todo el primer semestre) a otra de trabajo a tiempo completo (año 2004, sobre todo el segundo semestre). El segundo factor es mucho menos importante que el primero pero aporta una información complementaria. Desde el punto de vista de las filas, separa en positivo los meses de junio, julio y agosto de los tres años (sobre todo de 2003) del resto de meses. Respecto de las columnas, los esta-

dos que más contribuyen son "SE T>20", con un 78,3% y, en menor medida, "SE NT BE", con un 10,2%, ambos en el extremo positivo, frente al resto de situaciones. Esta dimensión refleja la actividad laboral de los meses de verano en los que los encuestados conjugan estudio y trabajo (2003) o estudio y búsqueda de empleo (2002).

A continuación, se aplica un ACM a la supermatriz de indicadores para encontrar las situaciones que se alejan del "perfil medio" descrito por el análisis anterior. Los cuatro primeros factores explican² un 76,02% de la inercia total, correspondiendo un 51,98% al primer plano factorial, que es el que se muestra en el gráfico siguiente (por razones de espacio el plano aparece rotado):

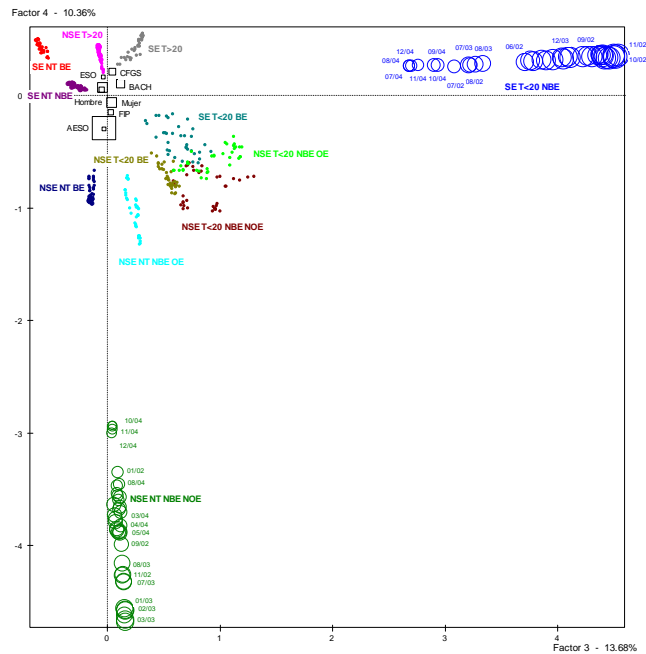


El plano con los factores 3 y 4, que explica un 24,04% adicional, se muestra en el gráfico adjunto. En ambos planos factoriales, se han representado de un mismo color los 36 meses correspondientes a cada una de las 12 situaciones básicas. Al igual que en los gráficos anteriores, el tamaño del icono de cada punto es proporcional a su contribución relativa al plano, por lo que las modalidades con iconos pequeños estarán mal representadas en el plano y no han de ser tenidas en cuenta en la interpretación del mismo. El primer plano tiene una disposición triangular: extremo positivo del primer factor, "se nt nbe"; extremo negativo, del primer factor "nse t>20"; y extremo positivo del segundo factor, "se nt be".

Por su parte, el segundo plano queda definido por las situaciones "se t<20 nbe" (extremo positivo del tercer factor) y "nse nt nbe noe" (extremo negativo del cuarto factor). Por lo tanto, estas 5 situaciones básicas son las que tienen un perfil más alejado del "perfil medio". El que aparezcan próximos entre sí los puntos correspondientes a los meses de una misma situación indica que una mayoría de los individuos que están en esa situación permanecen en ella durante todo el periodo (por ejemplo, "se nt nbe" y "nse t>20").

Sin embargo, en las otras tres situaciones se observa que los valores más próximos al origen (el "perfil medio") corresponden a los últimos meses de 2004 ("nse nt nbe noe"), a los meses de verano ("se nt be") o a ambos ("se t<20 nbe"), lo cual indica que durante esos meses esas situaciones se alejan menos del "perfil medio" antes descrito.

² Dado que los autovalores del ACM dan una idea muy pesimista de la inercia extraída en el análisis, las tasas de inercia se han recalculado de acuerdo con la fórmula propuesta por Benzécri (1992, pp. 411-412).



5. CONCLUSIONES

Con esta comunicación hemos pretendido presentar el Análisis de Correspondencias (AC) como una técnica multivariante de reducción de datos que puede ser útil en el estudio de historias de vida. A tal fin, se muestran dos formas complementarias de aplicar esta técnica utilizando para ello datos procedentes de la ETEFIL-2005. Los resultados obtenidos de cada análisis permiten identificar a través de las representaciones gráficas de los planos factoriales característicos del AC el "perfil medio" de la nube así como los estados e individuos que más se alejan de este "perfil medio".

6. REFERENCIAS BIBLIOGRÁFICAS

- BENZÉCRI, J.-P. (1992). *Correspondence analysis handbook*, Marcel Dekker, New York.
- DE LEEUW, J.; VAN DER HEIJDEN, P.G.M.; KREFT, I. (1985): "Homogeneity analysis of event history data", *Methods of Operations Research*, 50, pp. 299-316.
- GREENACRE, M. (1993). *Correspondence analysis in practice*, Academic Press, London.
- INSTITUTO NACIONAL DE ESTADÍSTICA (2007). *Encuesta de transición educativo-formativa e inserción laboral. Resumen Metodológico*. (<http://www.ine.es/daco/daco42/etefil/notaetefil.pdf>. Consultado el 17/10/2007).
- JAMBU, M. (1989). *Exploration informatique et statistique des données*, Dunod, Paris.
- MARTENS, B. (1994). "Analyzing event history data by cluster analysis and multiple correspondence analysis: an example using data about work and occupations of scientists and engineers" en M. Greenacre & J. Blasius (eds.), *Correspondence analysis in the social sciences*, Academic Press, San Diego.
- MOREAU, J.; DOUDIN, P.-A.; CAZES, P. (2000). *L'analyse des correspondances et les techniques connexes. Approches nouvelles pour l'analyse statistique des données*, Springer, Berlin.

VAN DER HEIJDEN, P.G.M. (1987). *Correspondence analysis of longitudinal categorical data*, DSWO, Leiden.

VAN DER HEIJDEN, P.G.M.; DE LEEUW, J. (1989). "Correspondence analysis with special attention to the analysis of panel data and event history data", en C.C. Clogg (ed.), *Sociological methodology*, Basil Blackwell, Oxford.