

La complementariedad del log-lineal y del análisis de correspondencias en la elaboración y el análisis de tipologías*

Carlos Lozares
Pedro López
Vicent Borràs

Universitat Autònoma de Barcelona. Departament de Sociologia
08193 Bellaterra (Barcelona). Spain

Resumen

Los análisis que combinan diversos métodos y técnicas según los diferentes objetivos planteados son susceptibles de obtener subproductos que, teniendo un interés sólo indirecto o relativo dentro del planteamiento global, permiten también mejorar los métodos de análisis y encontrar formas combinadas para construir objetos que sean pertinentes para la investigación. Es el caso de este artículo. En la construcción de tipologías se han empleado, aparte de otros métodos, dos concretamente que se han mostrado eficaces para dicha finalidad: el log-lineal y el análisis de correspondencias. Como resultado de dicho trabajo se ha analizado su complementariedad y sobre todo dónde radica analíticamente dicha complementariedad.

Este artículo es parte de una investigación cuyos resultados aparecen en este mismo número de PAPERS. El artículo tiene autonomía y entidad por sí mismo pero la lectura del resto de los artículos de este número facilitará su contextualización.

Palabras clave: análisis log-lineal, análisis de correspondencias, metodología, tipología, métodos estadísticos.

Abstract

Studies that combine different methods and techniques according to different objectives may give rise to by-products that only have an indirect or relative interest within the general approach, but that allow us to improve the methods of analysis and to find combined ways of building objects that are relevant to the research. This is the case of this article. In the construction of typologies, apart from other methods, two specific methods were used that proved effective for this purpose: the log-linear method and correspondence analysis. As a result of this work the complementary nature of these methods was studied, with particular reference to the analytical position of this complementariness.

This paper is part of a research work whose results are in this PAPERS volume. The paper is autonomous but the reading of the other papers of this volum can help to contextualize it.

Key words: log-linear analysis, correspondence analysis, methodology, typology, statistic method.

* Este artículo fue presentado en forma de comunicación en el V Congreso Español de Sociología, en el grupo de trabajo de metodología.

Sumario

El análisis log-lineal (AL) y el análisis de correspondencias (AC) son dos de los métodos más potentes y más frecuentemente usados en el análisis de las variables cualitativas en el ámbito de las ciencias sociales y en particular de la sociología. Ambos además suponen un toque de ingenio e intuición en la lógica y tratamiento de dichas variables además de provenir de dos tradiciones de análisis estadístico distintas —la anglosajona para el AL y la escuela francesa para el AC—. Durante años han seguido procesos diferentes dados sus orígenes y su diferente concepción. Con todo sólo recientemente se ha tratado de ver sus posibilidades complementarias¹. Estas reflexiones y la aplicación que aquí se hace se encaminan a dicho objetivo.

No entramos en este artículo en la exposición desarrollada de ninguno de los dos métodos, salvo consideraciones necesarias para los propósitos *ad hoc*, suponiendo que el lector interesado en esta temática posee de ambos un conocimiento suficiente.

1. El análisis de correspondencias, núcleo de diferentes métodos

El AC es popular y antiguo dentro de la literatura francesa dedicada al análisis de datos². El método no es tan popular en la literatura inglesa, aunque comienza a dársele una atención creciente³.

El AC descompone el X^2 de una tabla de contingencia usando los valores propios. El objetivo de esta descomposición consiste en llegar a un modelo de independencia de tal forma que las filas y columnas de la tabla de contingencia queden reflejados como puntos en un espacio de más baja dimensión, cada uno de los ejes está relacionado con los elementos de la descomposición del X^2 de Pearson. El AC

1. Por ejemplo en J.J. Daudin y P. Trecourt (1980), N.C. Lauro y A. Decarli (1982), M. Tenenhaus y F.W. Young (1985), L.A. Goodman (1986), T.P. Novak y D.L. Hoffman (1990), P.G.M. Van del Heijden y J. de Leeuw (1985), P.G.M. Van del Heijden, A. de Falguerolles y J. de Leeuw (1989).
2. La literatura en torno al análisis de correspondencias ha tenido un desarrollo no exento de polémica al confluir diferentes denominaciones, avances y perspectivas según países y tradiciones. Términos como *Optimal Scaling*, *Homogeneity Analysis*, *Conjoint Analysis* o *Analyse des Correspondances* han servido para identificar distintos enfoques y aportaciones de un problema común. Nuestro enfoque de esta técnica sigue las aportaciones de la tradición francesa derivadas de las obra de J.P. Benzécri (1973), L. Lebart (1977, 1985, 1987), B. Escoufier (1983) y M. Volle (1985), J. M. Cornejo (1988), entre otros, y de su implementación, por ejemplo, en el paquete informático de estadística SPAD-N.
3. S. Nishisato (1980), A. Gifi (1981), M.J. Greenacre (1984), L. Lebart, A. Morineau y K.M. Warwick (1984).

comprende dos modalidades de aplicación: el análisis factorial de correspondencias simples (AFC) y el análisis de correspondencias múltiples (ACM). Las reglas que rigen ambas modalidades conservan un mismo principio que se formula a partir del procedimiento de cálculo más simple del AFC, donde se analiza la relación entre dos variables consideradas de naturaleza nominal dispuestas en una tabla de contingencia o, de forma más general, sobre cualquier tabla bidimensional de números positivos. El ACM es la aplicación del AFC para matrices con un número cualquiera de variables o múltiples tablas de contingencia.

Al ACM pueden asimilarse otros muchos procedimientos de los que el ACM puede considerarse o interpretarse como su síntesis. M. Tenenhaus y F.W. Young (1985) desarrollan esta asimilación mutua y síntesis para el método de medias recíprocas, el análisis de varianza, el análisis de componentes principales y el análisis canónico generalizado. Estudian el ACM como una análisis de componentes principales explorando la relación entre los diversos criterios entre ambos métodos. El análisis de comparación y síntesis lo consiguen usando el diagrama de dualidad y el análisis de componentes principales del triplete Z, M, N como instrumentos útiles para sintetizar los métodos de análisis precedentes.

Z es la tabla de datos $n \times p$, las filas son los sujetos y las columnas son las variables. A estos datos se asocian cuatro espacios: el espacio de los sujetos, el de las variables, el espacio de los coeficientes de los sujetos y el espacio de los coeficientes de las variables.

M y N son matrices de métricas euclídeas definidas positivas vinculadas respectivamente al espacio de los sujetos y al de las variables. Se supone que la matriz N es diagonal y que sus elementos representan pesos positivos asociados a los sujetos sumando 1. Idem con M .

Se definen también las matrices de varianza-covarianza $V = Z'NZ$ y la matriz $W = ZMZ'$. Todo ello configura el triplete Z, M, N a partir del cual se pueden sintetizar, y es lo que hacen M. Tenenhaus y F.W. Young (1985), los diferentes métodos. Según estos autores estos instrumentos dan un vocabulario apropiado para una versión generalizada de la descomposición de los valores singulares de las matrices V y W , establecen y explicitan varios espacios y métricas asociados a las tablas de datos sujetos/variables, hacen posible la visualización de los diversos criterios del análisis de componentes principales y muestran cómo las soluciones de los diversos criterios de dicho análisis se relacionan explícitamente unos con otros por un diagrama de dualidades. Ello nos permite una visión unificada de una serie de métodos que poseen entre sí una manifiesta analogía.

2. El análisis de log-lineal

El AL es un método cuyo objetivo consiste en estudiar las relaciones entre variables cualitativas. La literatura sobre este método de análisis es abundante⁴.

4. Las referencias a este método comprenden una extensa literatura entre la que se puede citar S.J. Haberman (1979), Y.M. Bishop, S.E. Fienberg y P.W. Holland (1975), B.S. Everitt (1977), S.E. Fienberg (1980), D. Knoke y P.J. Burke (1980), R.R. Christensen (1990).

El AL no descompone el X^2 sino el logaritmo de las frecuencias de cada casilla por medio de un modelo lineal. En el caso de una tabla de contingencia de triple entrada, el logaritmo de la frecuencia absoluta de cada casilla es expresable como suma de la contribución de los distintos parámetros o factores que intervienen en su formación:

- de un elemento constante, que lo es para todas las casillas de una misma tabla y que depende de la frecuencia total y del número de casillas;
- de la contribución de cada fila, columna y nivel, correspondientes a la casilla para cada variable;
- de las posibles asociaciones entre pares entre filas, columnas y niveles (parámetros de primer orden de interacción),
- y de la posible triple asociación (parámetro de segundo orden de interacción).

En total, en el caso de que se trate de un modelo completo (o saturado) hay ocho parámetros que componen el logaritmo de cada casilla. La relación entre parámetros está sujeta a una condición similar a la del análisis de varianza: la suma cero para cada uno ellos extendida a una fila, columna o nivel. La ecuación para un modelo saturado y la dependencia de los parámetros, con tres variables, está dada por una ecuación lineal:

$$\text{Log}n_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YK} + \lambda_{ijk}^{XYZ}$$

Esto supone un número de $I \times J \times K$ ecuaciones. Evidentemente, el modelo saturado precedente siempre se cumple, la igualdad pasa a ser una identidad, pero la cuestión interesante consiste, por un principio elemental de parsimonia, en suponer y encontrar modelos más simples que den cuenta de dichas frecuencias con un grado de precisión aceptable dentro de un nivel dado de confianza. Para el caso de una tabla de contingencia de tres variables hay cuatro conjuntos de modelos globales posibles: aquél en el que el logaritmo de cada casilla se reduce a una constante; un conjunto de modelos de independencia (la ecuación correspondiente se reduce a los elementos de cada fila, columna o nivel); un conjunto de modelos en los que se da la asociación entre dos variables (primer orden de interacción), y, por fin, el modelo saturado (segundo orden de interacción). El modelo es jerárquico en el sentido de que si un parámetro es nulo lo son todos aquellos términos de orden inferior que tengan los mismo índices.

Según el problema que se examine se puede validar el modelo deseado. Existen tests globales para ello, el de X^2 de Pearson y el de L^2 de máxima verosimilitud, que proporcionan pruebas de bondad de ajuste. Hay dos intereses fundamentales en el AL: el seleccionar, validar e interpretar un modelo presupuesto y, dado el primero, el validar e interpretar los parámetros, que dan lugar a la estimación de parámetros y de errores.

Para intepretar los parámetros, es decir, para ver la contribución de los distintos órdenes factoriales al modelo, es imprescindible haber validado el mode-

lo correspondiente. Pero el objetivo puede reducirse al primer interés, esto es, a examinar solamente cuál es el modelo más adecuado sin interpretar los parámetros. Es lo que se suele buscar en particular cuando el número de variables es elevado y/o el número de valores de cada una es grande. Hay varias razones para ello: la escala no es lineal sino logarítmica, muchas veces el objetivo consiste en saber simplemente si las variables están relacionadas y no en un análisis local y, sobre todo, la dificultad de interpretar parámetros de interacción de orden superior a dos y tanto más cuanto más elevado sea el orden. Así pues, si solamente se desea conocer la validez de los modelos el AL es un método ideal, superior a otros aunque sea elevado el número de variables en juego. En este sentido es más preciso que el AC y sus métodos aparejados precedentes. Si deseamos además interpretar la relación entre los distintos valores de las variables, es decir, los parámetros de diferentes órdenes de interacción, el AL no es aconsejable cuando el número de variables y el de categorías es elevado, exigiendo de forma paralela un número suficiente de efectivos totales.

3. La complementariedad entre el análisis de correspondencias y el análisis log-lineal

Según lo precedente, dados los objetivos de ambos métodos y sus posibilidades o condicionamientos, la situación ideal es aquella en que el AL nos oriente hacia el modelo más adecuado para las variables cualitativas en estudio (es decir, responde a la cuestión de las variables que están relacionadas) y el AC nos orientase sobre la relación entre los valores, lo que es equivalente a la interpretación de los parámetros de distintos órdenes (es decir, responde a la cuestión de cómo se relacionan las variables: qué categorías de ellas se asocian más o menos de lo esperado). Uno se refiere al nivel de las variables, otro, al de los valores o categorías de las mismas. Por tanto, lo primero es asegurarse del modelo, por medio del AL con el L^2 de máxima verosimilitud y, posteriormente, ver las relaciones entre categorías, interpretación de los parámetros, por el AC.

P.G.M. Van del Heijden y J. de Leeuw (1985: 435-446) estudian y aplican esta relación. Su análisis y ejemplificación los hacen solamente para dos y tres variables. Nosotros trataremos de introducir más variables y mostrar un primer ejercicio de análisis complementario. No reproduciremos sus demostraciones sino solamente las conclusiones necesarias para la comprensión y para nuestros propósitos de aplicación. Tampoco entraremos en la generalización del AC, que ellos mismos exponen y utilizan de la aportación de Escofier (1983), con el objetivo de analizar la diferencia entre distintos modelos.

Como anunciábamos precedentemente, el AC descompone la diferencia $(n_{ij} - e_{ij})$, donde los e_{ij} —valores esperados bajo un modelo determinado— se calculan a partir de los marginales de n_{ij} . El contenido de la relación establecida entre el AL y el AC y la lógica de su demostración (P.G.M. Van del Heijden y J. de Leeuw, 1985) consiste en mostrar que los valores n_{ij} y e_{ij} son iguales a la estimación de máxima verosimilitud de las frecuencias esperadas bajo un modelo log-lineal especificado. El resultado hace posible inter-

pretar la solución del análisis de correspondencias en términos de modelos de log-lineal y salvar los problemas de interpretación de los parámetros de log-lineal que surgen frecuentemente. Es decir, que tanto para análisis de tablas de contingencia de dos variables, de tres o de más alto nivel, la solución del AC se toma como la diferencia entre dos modelos log-lineales. Por ejemplo, para dos variables la diferencia entre el modelo saturado y el modelo de independencia, es decir, el parámetro de interacción de primer orden se interpreta como la solución del primer eje del AC. En el caso de tres variables, si uno de los primeros órdenes de interacción (asociación de dos variables) no parece interesante se puede construir una sola variable interactivamente entre las dos. Cuando se está especialmente interesado en la relación de una variable con las otras dos, ésta no debe codificarse interactivamente con ninguna de las otras dos. La solución del AC en este caso muestra los dos primeros órdenes de interacción, de esta variable con las otras dos, y el segundo orden de interacción.

4. Ejemplo de aplicación

El ejemplo de aplicación consiste en una de las partes de la investigación que tratamos recurriendo en la primera fase y nivel de análisis a los datos de una encuesta⁵ con los que se ha tratado la problemática de la segmentación del mercado de trabajo o la construcción de una tipología del trabajo productivo. La explicación del funcionamiento del mercado laboral desde la segmentación, en contraposición con la visión neoclásica, se traduce en la estructuración de un mercado no homogéneo y dividido en segmentos diferenciados e interrelacionados jerárquicamente según procesos de internalización-externalización y en donde confluyen diversos factores de caracterización⁶. En base a estas consideraciones se plantea una hipótesis central y genérica: la existencia de una tipología de segmentos o situaciones de empleo que divide al empleo asalariado entre un segmento primario y otro secundario, el primero de los cuales comprende, a su vez, la distinción entre primario dependiente e independiente. La estructuración de estos segmentos se realiza en base a una serie de dimensiones de caracterización de los puestos de trabajo en términos de empleo donde se incluyen: la estabilidad laboral, la cualificación efectiva, el sala-

5. *Enquesta de la Regió Metropolitana de Barcelona. Condicions de vida i hàbits de la població. 1990.* Institut d'Estudis Metropolitans de Barcelona.
6. El modelo teórico de segmentación se elabora a partir de distintas contribuciones teóricas que han aportado diversos autores y autoras, entre ellos: institucionalistas (M.J. Piore, P.B. Doeringer), radicales (D.M. Gordon, R.C. Edwards, M. Reich), de la escuela de Cambridge (F. Wilkinson, J. Rubery, P. Villa) o de la literatura francesa (M. Maruani, F. Michom, M. Maurice, P. Desmarez). No obstante, la perspectiva adoptada parte sobre todo de la tradición de pensamiento marxista que se expresa en los trabajos de la escuela radical americana. El modelo teórico se elabora desde una conceptualización basada en el empleo y tiene su foco de atención en los procesos derivados del lado de la demanda. Para un desarrollo más extenso de la problemática, ver F. Miguélez (1989), P. López (1994), o el artículo que aparece en este mismo número de la revista.

rio y una caracterización de las empresas dada por el tamaño de los centros de trabajo, el sector de actividad, la titularidad y la presencia de representación laboral.

A efectos expositivos y de ilustración emplearemos un modelo de relación entre variables sencillo. A partir de él se tratará de mostrar un análisis parcial que evidencie la complementariedad entre el AC y el AL. Por otra parte, dentro de la problemática de la segmentación, esta tarea cumple un papel fundamental en el diseño de análisis, pues se encamina, como una primera etapa, al análisis del espacio de atributos o campo de variables en base al cual es posible derivar, en una segunda etapa, una tipología de segmentos de empleo.

Hemos considerado un modelo sencillo a partir de la interrelación de cinco variables, con una codificación en pocas categorías, que caracterizan el empleo asalariado (tabla 1).

En primer lugar presentaremos los resultados de la aplicación del AL a este conjunto de variables⁷. No nos extenderemos en los distintas pruebas ni estimaciones a las que da lugar el análisis logarítmico. Tan sólo destacaremos que el modelo final tiene como clase generadora: [P E R][R C][R I][C I][E I][P I].

Es decir, se obtiene un modelo con una interacción de segundo orden y cinco de primer orden. Las variables *Profesión*, *Empresa* y *Representación* presentan una gran asociación de base (segundo orden de interacción), constituyendo de alguna forma las condiciones contextuales de la segmentación laboral en el sentido de ser estructurales y generales con relación a los trabajadores y las

Tabla 1. Variables utilizadas en el análisis.

Variables	Categorías	Frecuencias	
P Profesión:	N (No técnico)	1671	(81,5%)
	T (Técnico)	380	(18,5%)
E Tipo de empresa:	PR (Privada)	1647	(80,3%)
	PU (Pública)	404	(19,7%)
R Representación laboral:	S (Sin representación)	953	(46,4%)
	C (Con representación)	1098	(53,6%)
I Ingresos mensuales:	INB (Bajos)	686	(33,4%)
	INM (Medios)	857	(41,8%)
	INA (Altos)	403	(19,7%)
	NSNC (Ns/Nc)	105	(5,1%)
C Tipo de contrato:	EVE (Eventual)	809	(39,4%)
	IND (Indefinido)	1242	(60,6%)
Total		2051	(100,0%)

7. Los cálculos se han realizado con el paquete estadístico SPSS.

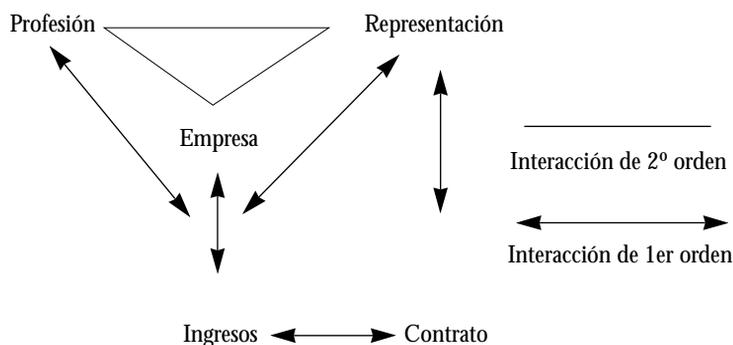


Gráfico 1

trabajadoras de las empresa⁸. Las otras dos variables, *Contrato* e *Ingresos*, presentan entre ambas una asociación de primer orden de interacción y guardan una relación con el resto también del mismo orden: ingresos con el triángulo de base y contrato con representación. El contrato actúa como condición de la relación laboral en la empresa y los ingresos expresan otra de las condiciones y resultados en los que se traduce la segmentación. Por tanto, el modelo global previo propuesto se puede esquematizar en el gráfico 1.

En este modelo se observa una fuerte vinculación entre *Contrato* e *Ingresos*, seguida de la relación entre *Profesión* e *Ingresos*. El resto quedan muy por debajo de estas interacciones.

Una vez aceptado el modelo general interesa conocer cómo se realiza esta asociación. Ahora se trata de estudiar con detalle la forma y el sentido de las distintas interacciones a partir de la estimación de los parámetros del modelo. Nos detendremos solamente en la relación que las tres variables de base (P E R) mantienen con la variable de *Ingresos* (*I*). Nos interesa tratar esta relación que implica más variables para que a través del AC se evidencie la relación entre los valores de las variables, es decir, podamos tener un método para esclarecer la complejidad que da el análisis de los parámetros del AL en cuanto superan el primer orden de interacción.

Los *Ingresos* tienen asociación con cada una de las variables de base —*Empresa*, *Representación* y *Profesión*—, además de que éstas entre sí mantie-

8. De hecho, también hemos ensayado la introducción de la variable *Tamaño de la empresa*, el resultado abunda en la idea de que constituyen un conjunto de cuatro variables —las tres precedentes y esta nueva— con una asociación —ahora triple orden de interacción— representando la misma idea de condiciones de base o de contexto ecológico de la segmentación laboral. Simplemente se ha eliminado esta variable por su efecto redundante y en aras de la simplificación. Alternativamente se podía haber eliminado la variable de representación laboral y conservar la de *Tamaño de la empresa*, pues ambas mantienen un muy alto grado de asociación.

nen un doble efecto aunque débil. Por ello el ejemplo que tomaremos para la interpretación de los parámetros del AL a partir del AC es el de la asociación entre los *Ingresos* y una variable única que denominaremos *Contexto de la segmentación (S)* configurada a partir de los valores de las tres variables, es decir, se obtiene una variable con ocho valores (ver tabla 2).

Según lo visto anteriormente, la solución del AC entre las variables *Ingresos* (I) y la nueva variable construida *S*, es la diferencia entre el modelo saturado y el modelo de independencia, esto es, la solución del AC nos sirve para interpretar el primer efecto de la interacción entre las variables precedentes.

Pero antes de aplicar este análisis veamos la interpretación de los parámetros del AL que resulta entre *I* y *S*, lo que nos permitirá mostrar la semejanza y complementariedad del AL con el AC. El resultado esperado es que el único modelo aceptable es el de saturación, es decir, el de asociación entre las dos variables. En la tabla 3 aparece la prueba que muestra este resultado, así como la estimación de los parámetros de primer orden (aparecen en negrita los valores significativos y se reproducen a continuación en la tabla 4).

Los ingresos mensuales totales bajos están sobre todo asociados a la «no cualificación de trabajador y a la ausencia de representación laboral». La situación significativa de mayor vinculación a los bajos salarios corresponde a la de «la empresa privada, sin representación laboral y sin cualificación del trabajador». También la vinculación es grande en «la empresa pública con idéntica situación de representación laboral y cualificación» pero no es significativo.

Tabla 2. Tabla de contingencia entre los ingresos y la variable contextual construida.

	S	NSPR	NCPR	NSPU	NCPU	TSPR	TCPR	TSPU	TCPU	Total
I										
Bajos	370 48,6	180 27,0	19 46,8	56 27,3	30 29,7	15 12,6	12 23,1	5 4,9	686 33,4	
Medios	265 34,8	312 46,9	17 42,9	113 55,2	33 32,8	38 31,9	26 51,5	53 48,6	857 41,8	
Altos	80 10,5	136 20,4	3 8,8	32 15,7	30 30,2	59 49,5	13 25,4	50 45,6	403 19,7	
NS/NC	47 6,2	38 5,7	1 1,5	4 1,8	7 7,3	7 6,0	0 0,0	1 1,0	105 5,1	
	761	667	40	204	101	119	51	109	2051	
Total	37,1	32,5	1,9	9,9	4,9	5,8	2,5	5,3	100,0	

NSPR: No técnico, Sin representación laboral, Empresa privada; **NCPR:** No técnico, Con representación laboral, Empresa privada; **NSPU:** No técnico, Sin representación laboral, Empresa pública; **NCPU:** No técnico, Con representación laboral, Empresa pública; **TSPR:** Técnico, Sin representación laboral, Empresa privada; **TCPR:** Técnico, Con representación laboral, Empresa privada; **TSPU:** Técnico, Sin representación laboral, Empresa pública; **TCPU:** Técnico, Con representación laboral, Empresa pública.

Tabla 3. Resultados del análisis log-lineal entre ingresos y la variable contextual.

Prueba de que los efectos de orden K o superiores son cero						
K	GL	L^2	Prob	X^2	Prob	Iteración
2	21	287,530	,0000	286,989	,0000	2
1	31	3006,870	,0000	4144,561	,0000	0
Prueba de que los efectos de orden K son cero						
K	GL	L^2	Prob	X^2	Prob	Iteración
1	10	2719,340	,0000	3857,572	,0000	0
2	21	287,530	,0000	286,989	,0000	0
Pruebas de asociación parcial						
Efecto	GL	X^2 Parcial	Prob	Iteración		
S	7	1967,498	,0000	2		
I	3	751,843	,0000	2		
Estimación de los parámetros						
S*I	Parámetro	Coef.	Err. Est.	Valor-Z	Intervalo al 95%	
1	,6310989092	,10620	5,94251	,42295	,83925	
2	-,3067422139	,09887	-3,10263	,50052	,11297	
3	-,8559420177	,12269	-6,97673	-1,09641	-,61548	
4	-,0311303262	,11150	,27919	-,24968	,18741	
5	-,0848210056	,09822	,86358	-,27733	,10769	
6	-,2696093370	,11499	-2,34460	-,49499	-,04423	
7	,7804230761	,29670	2,63032	,19889	1,36196	
8	,0917120091	,29640	,30941	-,48924	,67267	
9	-,7434464895	,40301	-1,84475	-1,53334	,04645	
10	,2603880617	,16985	1,53307	-,07251	,59329	
11	,3568520786	,15354	2,32421	,05592	,65778	
12	-,2475304093	,18396	-1,34555	-,60810	,11304	
13	-,0364111202	,17646	-,20634	,38227	,30945	
14	-,5442542502	,16762	-3,24690	-,87280	-,21571	
15	,0164865862	,17634	,09349	-,32914	,36211	
16	-,7326835605	,20942	-3,49869	-1,14314	-,32223	
17	-,4277736046	,16574	-2,58102	-,75262	-,10293	
18	,6509964840	,16141	4,03313	,33463	,96736	
19	,2349968087	,37650	,62415	-,50295	,97295	
20	,4111525239	,35113	1,17095	,27706	1,09936	
21	,3654706234	,37311	,97952	-,36583	1,09677	

Tabla 4. Parámetros de asociación de análisis log-lineal.

S	I			
	INB	INM	INB	NSNC
NSPR	0,63	-0,30	-0,85	0,52
NCPR	-0,03	-0,08	-0,26	0,37
NSPU	0,78	0,09	-0,74	-0,12
NCPU	0,26	0,35	-0,24	-0,37
TSPR	-0,03	-0,54	0,01	0,56
TCPR	-0,73	-0,42	0,65	0,50
TSPU	0,23	0,41	0,30	-1,00
TCPU	-1,11	0,50	1,07	-0,46

La situación significativa más opuesta a esta vinculación con los salarios bajos es la del «trabajador con cualificación, con representación laboral en la empresa donde trabaja, sea ésta pública o privada».

Los ingresos mensuales medios están más asociados a la «empresa pública». Si la «cualificación es alta» la vinculación es mayor, de igual forma si se da «representación laboral». La situación de la «empresa privada» es prácticamente opuesta.

Los ingresos mensuales totales altos están vinculados exclusivamente a la «alta cualificación del trabajador» en particular en las «empresas con representación laboral» y, en estas condiciones, es mayor en la «pública» que en la «privada». La situación opuesta es la menos vinculada con los salarios elevados.

El NS/NC sobre los ingresos mensuales está vinculado exclusivamente a la «empresa privada» y es mayor cuanto «más cualificado» se está y menor cuanto mayor es la «representación laboral».

Los resultados del AC de esta misma relación entre ingresos y la variable contextual de segmentación construida se presentan en la tabla 5 y el gráfico 2⁹. Del análisis de la tabla y de la observación del gráfico se extraen conclusiones similares a las precedentes del AL.

Primero, la proyección importante que todas las variables tienen sobre el primer eje, el 81%, que es el que da la solución del AL.

El eje está recorrido por la escala de los ingresos. Del lado izquierdo los ingresos altos, aproximadamente en el centro los salarios medios y a la derecha del gráfico los salarios altos. Los NS/NC están en la parte central con algo de proyección sobre el segundo eje.

Correlativamente se emplazan también *Profesión, Representación y Empresa*. Las altas cualificaciones de técnicos están de un lado, izquierda de la representación, más próximas a los salarios altos; las bajas cualificaciones en el otro lado de los salarios bajos, derecha, o próximos al centro. En igualdad

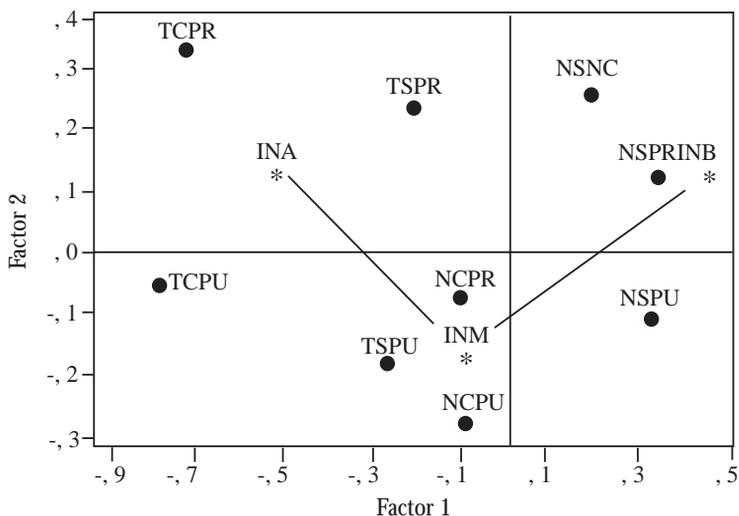
9. Los cálculos se han realizado con el paquete estadístico SPAD-N.

Tabla 5. Resultados del análisis de correspondencias entre ingresos y la variable contextual.

Eje	Valor propio	%	% acumulado
1	0,1138	81,35	81,35
2	0,0223	15,97	97,32
3	0,0037 (0,1399)	2,68	100,00

Coordenadas y contribuciones de las categorías sobre los ejes 1 a 3

	Coordenadas			Contribuciones			Coseno cuadrado		
	1	2	3	1	2	3	1	2	3
INB	0,39	0,08	-,04	44,3	9,9	12,4	0,95	0,04	0,01
INM	-,07	-,17	0,01	1,7	54,0	2,5	0,14	0,85	0,01
INM	-,55	0,16	-,03	53,0	22,2	5,1	0,92	0,08	0,00
NSNC	0,15	0,25	0,24	1,0	13,9	80,0	0,16	0,43	0,41
NSPR	0,34	0,08	-,02	38,5	11,1	4,4	0,94	0,05	0,00
NCPR	-,09	-,07	0,07	2,5	8,1	45,3	0,45	0,28	0,27
NSPU	0,31	-,12	-,17	1,7	1,2	14,4	0,70	0,10	0,20
NCPU	-,05	-,28	-,04	0,2	35,6	4,1	0,03	0,95	0,02
TSPR	-,19	0,23	0,03	1,5	11,6	1,5	0,39	0,59	0,01
TCPR	-,71	0,33	-,02	25,4	28,3	0,4	0,82	0,18	0,00
TSPU	-,26	-,19	-,14	1,4	4,1	13,9	0,53	0,30	0,17
TCPU	-,79	-,03	-,11	28,9	0,2	16,0	0,98	0,00	0,02

**Gráfico 2.** Representación de la «categoría» sobre los ejes 1 y 2.

de cualificación, es decir, tanto altas como bajas, los trabajadores en empresas con representación laboral están más del lado de las profesiones altas, a la izquierda en los ejes de ingresos superiores, y los que no tienen representación laboral más a la derecha, de ingresos inferiores. Lo público y lo privado no tiene demasiado peso en la determinación del eje, aunque sí se nota una ligera inclinación a ingresos superiores en la empresa pública en iguales condiciones de cualificación y representación laboral. Los que trabajan en empresas privadas están más cercanos a los «no saben/no contestan» con relación a los ingresos.

5. Conclusiones

La exposición que hemos realizado nos ha permitido mostrar la coincidencia de resultados entre dos métodos de análisis multivariable de gran utilidad para el sociólogo. El ejemplo que hemos ilustrado se considera un modelo en el que se involucran efectos de interacción de segundo orden, con un número elevado de parámetros estimados, cuyo análisis e interpretación en términos de una solución de AL resulta a menudo difícil. Más aún cuando el modelo es más complejo, como era también nuestro modelo general de partida. La complementariedad con el AC se plantea precisamente en términos de lectura e interpretación de las relaciones entre las variables del modelo, y más específicamente entre las categorías de éstas. Constatando que ambos procedimientos nos conducen a la resolución de un mismo problema de análisis, el AL se convierte en un instrumento valioso para concluir los modelos de relación al nivel de las variables, mientras que el AC lo es para especificar la naturaleza e interpretación de esos modelos al nivel de las categorías.

El ejemplo que aquí se ha expuesto no es más que una de las posibilidades de uso complementario de ambos métodos, esto es, aquélla en que se compara una asociación entre dos variables, una de las cuales es el resultado de la unión de otras. De esta forma se ha podido evidenciar una misma solución del AL en términos del AC y que resulta de la diferencia entre un modelo saturado y el de independencia.

A partir de aquí se abren nuevas posibilidades de análisis que nos facilitarían por ejemplo la comparación de efectos más complejos y de diferencias no necesariamente referidos a modelos de independencia. Para ello hay que considerar la generalización del AC que P.G.M. van der Heijden y J. de Leew (1985, 1989) utilizan siguiendo el trabajo de B. Escofier (1983) o bien tomando el análisis de residuos que proponen Th. P. Novak y D.L. Hoffman (1990).

La literatura de origen anglosajón que ha planteado la complementariedad entre el AL y el AC ha razonado desde la posición de cómo el AC puede ayudar a mejorar un AL. La visión contraria, cómo un AL puede ayudar a mejorar un AC, seguramente propiciaría nuevas posibilidades de análisis. Por ejemplo, cuando el número de categorías y de valores alcanza un número elevado el AL se convierte en un instrumento difícil de aplicar, mientras que el AC, en su versión multivariable, elude ese tipo de problemas. Siendo ésta una ventaja de

importancia, el AC se podría concebir como una primera fase del análisis que condujera a la reducción de la información multivariable y reutilizable posteriormente en un AL con menos variables que permitiera modelizar las relaciones significativas del campo de variables estudiado.

Bibliografía

- BENZÉCRI, J.P. (1973). *L'Analyse des données*. París: Dunod.
- BISHOP, Y.M.; FIENBERG, S.E.; HOLLAND, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: Cambridge University Press.
- CORNEJO ÁLVAREZ, José Manuel (1988). *Técnicas de investigación social: el análisis de correspondencias (Teoría y Práctica)*. Barcelona: Promociones y Publicaciones Universitarias.
- CHRISTENSEN, Ronald R. (1990). *Log-Linear Models*. Nueva York: Springer-Verlag.
- CRIVISQUI, Eduardo M. (1993). *Análisis factorial de correspondencias. Un instrumento de investigación en ciencias sociales*. Asunción: Universidad Católica de Asunción.
- DAUDIN, J.J.; TRECCOURT, P. (1980). «Analyse factorielle de correspondances et modèle Log-lineal. Comparaison de deux méthodes sur un exemple». *Revue de Statistique Appliquée*, 28, p. 5-24.
- ETXEBERRIA, J. y otros (1995). *Análisis de datos y textos*. Madrid: RA-MA.
- EVERITT, Brian S. (1992). *The Analysis of Contingency Tables*. Londres: Chapman and Hall.
- FIENBERG, S.E. (1980). *The Analysis of Cross-Classified Categorical Data*. Cambridge: MIT Press.
- GIFI, A. (1981). *Non-linear multivariate analysis*. Leiden: DSWO Press.
- GOODMAN, L.A. (1986). «Some useful extensions of the usual Correspondence Analysis approach and usual Log. Linear Models Approach in Analysis of Contingency Tables». *International Statistical Review*, 54, p. 243-309.
- KNOKE, D.; BURKE, P.J. (1980). *Log Linear Models*. Beverly Hills: Sage Publications.
- GREENACRE, M.J. (1984). *Theory and applications of correspondence analysis*. Londres: Academic Press.
- GRUP D'ESTUDIS SOCIOLÒGICS SOBRE LA VIDA QUOTIDIANA I EL TREBALL (1994). *Transformaciones del trabajo y bienestar social*. Informe del proyecto de investigación financiado por la DGICYT.
- HABERMAN, S.J. (1979). *Analysis of Qualitative Data*. Nueva York: Academic Press.
- LAURO, N.C.; DECARLI, A. (1982). «Correspondence analysis and Log-lineal models in multiway contingency tables study: Some remarks on experimental data». *Metron*, 40, p. 213-234.
- LEBART, Ludovic (1977). «La validité des résultats en analyse des données». *Consumation*, 1, enero-marzo, p. 41-69.
- LEBART, L.; MORINEAU, A.; WARWICK, K.M. (1984). *Multivariate descriptive analysis: Correspondence analysis and related techniques for large matrices*. Nueva York: Wiley.
- LEBART, L.; MORINEAU, A.; FENELON, J.P. (1985). *Tratamiento estadístico de datos. Métodos y programas*. Barcelona: Marcombo.
- LEBART, L.; MORINEAU, A.; LAMBERT, T. (1987). *SPAD.N: Système Portable pour l'Analyse des Données. Version 1.0. Manuel de Référence*. Sèvres: CISIA.
- LÓPEZ ROLDÁN, Pedro (1994). *La construcción de tipologías en Sociología: propuesta metodológica de construcción, análisis y validación. Aplicación al estudio de la seg-*

- mentación del mercado de trabajo en la Región Metropolitana de Barcelona*. Tesis doctoral. Bellaterra: Universitat Autònoma de Barcelona.
- MIGUÉLEZ LOBO, Faustino (1989). *La segmentación del mercado de trabajo*. Mimeo.
- NISHIATO, S. (1980). *Analysis of Categorical Data: Dual Scaling and its Applied its Applications*. Toronto: University of Toronto Press.
- NOVAK, Th.P.; HOFFMAN, D.L. (1990). «Residual Scaling: An Alternative to Correspondence Analysis for the Graphical Representation of Residuals from Log-linear Models». *Multivariate Behavioral Research*, 25, 3, p. 351-370.
- SÁNCHEZ CARRIÓN, Juan Javier (1984). *Introducción a las técnicas de análisis multivariable aplicadas a las ciencias sociales*. Madrid: Centro de Investigaciones Sociológicas.
- TENENHAUS, M.; YOUNG, F.W. (1985). «An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data». *Psychometrika*, 50, 1, p. 91-119.
- VAN DEL HEIJDEN, P.G.M.; DE LEEUW, J. (1985). «Correspondence analysis used complementary to loglinear analysis». *Psychometrika*, 50, p. 429- 447.
- VAN DEL HEIJDEN, P.G.M.; DE FALGUEROLLES, A.; DE LEEUW, J. (1989). «A combined approach to contingency table analysis using correspondence analysis and Log-linear analysis». *Applied Statistics*, 38, p. 249- 292.
- VOLLE, Michel (1985). *Analyse des données*. París: Economica.