INTRODUCCIÓN A LOS MODELOS LOGARÍTMICOS LINEALES[1]

Margarita Latiesa
Profesora de Métodos y Técnicas de Investigación Social en la Facultad de Ciencias Políticas y Sociología de la Universidad de Granada

Resumen

Se expone en este artículo la técnica multivariable de análisis de datos denominada Modelo logarítmico lineal, que se basa en niveles de medición nominal u ordinal. Se subdivide su contenido en tres partes: en la primera se examina la lógica y supuestos básicos del modelo, en la segunda se pone el énfasis en las operaciones y cálculos estadísticos necesarios para su utilización y, en la tercera, se expone la utilidad interpretativa y las ventajas frente a los análisis tradicionales de tablas de contingencia y frente a los Sistemas de ecuaciones y grafos.

Resum

Un dels diversos problemes que es presenten a la Sociologia és el de no disposar de mètodes d'anàlisi estadística-algebraica tan poderoses en el tractament de les variables qualitatives com existeixen en el cas de les quantitatives. En els darrers anys, s'han popularitzat de manera considerable les tècniques vinculades al tractament línial logarítmic de les dites variables. L'article que es presenta és un exemple d'aplicació del seu desenvolupament i càlcul en un cas particular de fàcil accés a un públic no especialitzat en matemàtiques.

Abstract

One of the problems in Sociology is the lack of the algebraic and statistical analysis methods as powerful in qualitative variables as in quantitative variables. During recent years, the popularity of techniques linked to the lineal logarithm processes has increased. This article, is an example of the processing and calculating applied to a particular case, which should be easily understood by non-mathematicians.

1. Este artículo es una parte reducida y reformada de la lección magistral leída en mayo de 1987.

He de agradecer la valiosa colaboración que he tenido en la confección de este trabajo de los siguientes profesores: Dr. Francisco Alvira Martín, Dr. Julio Carabaña Morales, Dr. Modesto Escobar Mercado y D. José Torres Mora.

INTRODUCCIÓN

En el contexto de una investigación social por encuesta la mayoría de las variables tienen un nivel de medición nominal u ordinal. Para analizar este tipo de variables se utilizan normalmente las tablas de contingencia. El análisis puede ser bivariable (cruce de dos variables), aunque también se introducen variables de control; y multivariable (cruce de n variables). Las técnicas multivariables en el análisis tabular pueden ser de dos tipos:

—Sistema de ecuaciones y grafos
—Modelos logarítmico lineales

Ambas técnicas permiten estudiar el cruce de varias variables simultáneamente, pero mientras en la primera es necesario determinar las variables dependientes e independientes, en la segunda, todas las variables son consideradas como independientes. El objetivo en el primer caso, es determinar el impacto causal que tienen unas variables sobre otras, y se hallan unos coeficientes similares a los coeficientes de los análisis de regresión; el objetivo en el segundo caso, es determinar las asociaciones e interacciones que puedan existir entre las variables.

En este artículo limitaremos la exposición a la segunda técnica mencionada: *los modelos logarítmicos lineales.* Esta técnica de análisis multivariable se ha desarrollado a partir de los trabajos de Goodman (1972, 1973, 1979); Bishop et al. (1975); Upton (1978); y en España ha sido expuesta por Alvira (1985, 1986), Sánchez Carrión (1984) y Ruiz-Maya (1990).

Frente al análisis tradicional de tablas de contingencia, donde se determina si existe asociación entre dos variables (Gi cuadrado)y la fuerza y dirección de la misma (coeficientes de asociación), el modelo logarítmico lineal presenta la notable ventaja de examinar varias variables simultáneamente y determina las asociaciones e interacciones que existen entre las mismas.

Frente al análisis multivariable denominado Sistema de ecuaciones y grafos, donde hay que definir previamente un modelo causal que especifique el tipo de relación que existe entre las variables, el modelo logarítmico lineal presenta la ventaja de que no es necesario establecer el modelo de relación entre las variables, ya que su objetivo es precisamente encontrar el mejor modelo que explique las casillas de la tabla de contingencia.

La exposición que vamos a hacer a continuación de los modelos lineales logarítmicos se desglosa en tres apartados:

—Primero: lógica y supuestos básicos
—Segundo: operaciones y cálculos estadísticos
—Tercero: ventajas y utilidad interpretativa

No se trata de una exposición minuciosa y profunda sino más bien de una panóramica general e introductoria en los tres niveles: lógico, operatorio e interpretativo.

SUPUESTOS BÁSICOS Y LÓGICA DEL MODELO

La principal característica de este modelo es que no es necesario establecer previamente las relaciones causales que existen entre las variables, porque todas ellas son consideradas independientes. La variable dependiente, lo que queremos explicar, son las frecuencias de las casillas observadas en una tabla de contingencia. Se trata de explicar la probabilidad de que una persona presente una determinada combinación de categorías.

El objetivo de este análisis es precisamente hallar el mejor modelo, que sea capaz de explicar las casillas observadas en la tabla de contingencia. Se trata de detectar qué asociaciones y qué interacciones existen entre las variables y conforman el modelo.

La lógica del proceso que se sigue es la siguiente:

1) En una tabla de contingencia, existen varios efectos, que pueden influir para que aparezcan las casillas observadas:

—El efecto total, debido al número medio de casos en cada casilla, es decir, el tamaño de la muestra. Así, el tamaño de una casilla depende, en primer lugar, del número total de casos.
—El efecto de cada una de las variables que componen la tabla, debido a la distribución de los marginales de cada variable.
—El efecto de asociación de cada par de variables.
—El efecto de la interacción entre más de dos variables.

2) Nosotros no sabemos a priori qué efectos son los que realmente influyen en la tabla. Pero nuestro objetivo es encontrar un modelo que los determine y sirva para explicar las casillas originales.

Existen varios modelos posibles, según el número de variables que componen la tabla de contingencia. Así, en una tabla de dos variables existen cinco modelos posibles, y en una tabla de tres variables, dieciocho.

Normalmente, en el análisis de las tablas se contrasta el modelo de independencia (ausencia de relación conjunta entre las variables), por medio del X^2, pero el sistema logarítmico lineal incluye muchos más modelos, y el de independencia es uno más.

En un extremo, el modelo puede incluir sólo el efecto total (tamaño de la muestra), en el otro, puede incluir todos los efectos posibles.

3) A partir de cada uno de estos modelos, que incluye unos efectos y excluye otros, se puede construir una tabla de contingencia distinta para cada modelo. Esta es la tabla de las frecuencias esperadas bajo los supuestos del modelo.

Comparando cada una de estas tablas con las frecuencias observadas, unas estarán más cerca que otras o, dicho de otro modo, obtendremos diferentes bondades de ajuste para cada modelo.

En la práctica, no se prueba la bondad del ajuste de todos los modelos posibles, sino que se siguen otras estrategias que acortan el proceso y que explicaremos más adelante.

El modelo que seleccionaremos es aquel que cumpla dos condiciones:

a) incluya menos efectos.
b) la tabla de frecuencias esperadas reproduzca la tabla de frecuencias observadas (el modelo ajuste).

4)Una vez seleccionado el modelo y los efectos que influyen, es decir, las asociaciones e interacciones que existen entre las variables, cuantificamos los efectos, mediante el cálculo de los parámetros e interpretamos los resultados.

OPERACIONES Y CÁLCULOS ESTADÍSTICOS

Con objeto de exponer las operaciones y cálculos estadísticos que hay que realizar vamos a usar los datos procedentes de una investigación sobre la influencia de diversas variables en el abandono de los estudios universitarios

TABLA 1

Cruce de Actividad laboral y Abandono

Actividad laboral	*Abandono* *no*	*sí*	*Total*
no	245	45	290
sí	122	62	184
Total	367	107	474

de Ciencias Políticas y Sociología. Dicha investigación fue presentada como tesina en octubre de 1982 (Latiesa 1982, 1983).

Para ilustrar el modelo lineal logarítmico vamos a emplear dos ejemplos. *En el primero,* nos centraremos más en las operaciones y cálculos que hay que realizar. Emplearemos por tanto, el caso más sencillo de una tabla de dos variables. *En el segundo,* además de los cálculos, trataremos de ver la utilidad interpretativa del modelo lineal logarítmico, en un ejemplo de tres variables.

En la Tabla 1 aparece el cruce de las variables Actividad laboral y Abandono de los estudios.

Para explicar las frecuencias observadas que aparecen en las casillas de la Tabla existen cuatro efectos posibles:

—Efecto medio total. Efecto debido al número medio de casos en cada casilla (μ).
—Efecto de fila. Efecto debido a la distribución de los marginales de la variable Actividad laboral ($\lambda^{L}i$).
—Efecto de columna. Efecto debido a la distribución de los marginales de la variable Abandono ($\lambda^{A}j$).
—Efecto conjunto. Efecto debido a la asociación de las Variables Actividad laboral y Abandono (λ^{LA}_{ij})

Podemos construir cinco modelos según el número de efectos que incluyan:

1)Equiprobabilidad. Incluye exclusivamente el efecto total.
2)De columna. Tan solo incluye el efecto columna y total.
3)De fila. Tan solo incluye el efecto de fila y total.
4)Independencia. Incluye el efecto total, de fila y de columna. No incluye la asociación de las variables. Este es el modelo habitual de contraste de la hipótesis nula (ausencia de relación entre las variables) mediante el X^2.
5)Saturado. Incluye los cuatro efectos, es decir, todos.

A continuación vamos a exponer primero, los modelos; segundo, las tablas de frencuencias esperadas, bajo los supuestos de los modelos; y tercero, la prueba de significación X^2, para calcular la bondad del ajuste.

1)Equiprobabilidad. Se expresa del siguiente modo:

$$\text{Log}_e \text{ Fij} = \mu$$

Bajo el supuesto de que sólo exista el efecto total, las frecuencias esperadas son las siguientes:

Actividad laboral	*Abandono* *no*	*sí*	
no	119	119	X2 = 206*
sí	119	119	

Si comparamos las frecuencias observadas con las esperadas, vemos que difieren ostensiblemente, y que X^2 es significativo, luego hay que rechazar este modelo. La tabla original no queda explicada si incluimos tan solo el efecto total.

Si el modelo ajustara satisfactoriamente con los datos empíricos, procederíamos a calcular los parámetros, es decir μ, que es igual al $\log_e$ 119. No vamos a calcular los parámetros hasta que no encontremos un modelo satisfactorio, que ajuste bien.

2)Efecto fila. Se expresa del siguiente modo:

$$\mathrm{Log}_e\ Fij = \mu + \lambda^{L}i$$

Bajo el supuesto de que sólo exista el efecto de la variable Actividad laboral y el efecto total, las frecuencias esperadas son las siguientes:

Actividad laboral	*Abandono* *no*	*sí*	
no	145	145	
sí	92	92	X2 = 156*

De nuevo rechazamos el modelo que incluye tan sólo los efectos de fila y total (Gi-cuadrado significativo)

3)Efecto columna. Se expresa del siguiente modo:

$$\mathrm{Log}_e\ Fij = \mu + \lambda^{A}j$$

Las frecuencias esperadas son las siguientes:

Actividad laboral	*Abandono* *no*	*sí*	
no	184	54	X2 = 44*
sí	184	54	

También rechazamos este modelo.

4)Independencia. Se expresa del siguiente modo:

$$\text{Log}_e \text{ Fij} = \mu + \lambda^L i + \lambda^A j$$

Las frecuencias esperadas bajo el supuesto de que exista efecto total, de fila, de columna, pero no conjunto, son las siguientes:

	Abandono		
Actividad laboral	*no*	*sí*	
no	225	65	
sí	142	42	X2 = 20*

El Gi-cuadrado se va haciendo más pequeño a medida que incluimos más efectos en el modelo, pero sigue siendo significativo, luego tenemos que rechazar el modelo que no incluye el efecto conjunto de las dos variables.

5)Saturado. Se expresa del siguiente modo:

$$\text{Log}_e \text{ Fij} = \mu + \lambda^L i + \lambda^A j + \lambda^{LA}_{ij}$$

Los valores esperados bajo el supuesto del modelo saturado, coinciden con los valores observados, ya que incluye todos los efectos posibles. Gi-cuadrado, por tanto, es igual a 0 y no es significativo.

Luego éste es el único modelo que sirve para explicar las casillas de la tabla original. Implica que existe asociación entre las variables Actividad laboral y Abandono; que existen los efectos de cada una de las variables y el efecto total.

Una vez seleccionado el modelo, se procede al cálculo de cada uno de estos efectos (parámetros) para cuantificar su incidencia.

Cálculo del efecto medio: (μ)

$$\mu = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \text{Log}_e \text{ Fij}$$

$$\mu = \frac{1}{4} (\log_e 245 + \log_e 45 + \log_e 122 + \log_e 62) =$$

$$= \frac{5.5 + 3.8 + 4.8 + 4.1}{4} = 4.5$$

Cálculo del efecto de la variable Actividad laboral ($\lambda^{L}i$)

$$\lambda^{L}i = \frac{1}{IJ} \sum_{j=1}^{J} \log_e (Fij/Fnoi, j)$$

$$\lambda^{L}no = \frac{1}{4} (\log_e 245/122 + \log_e 45/62) = .10$$

$$\lambda^{L}sí = \frac{1}{4} (\log_e 122/245 + \log_e 62/45) = -.10$$

Cálculo del efecto de la variable Abandono ($\lambda^{A}j$)

$$\lambda^{A}j = \frac{1}{IJ} \sum_{i=1}^{I} \log_e (Fij/Fi,noj)$$

$$\lambda^{A}no = \frac{1}{4} (\log_e 245/45 + \log_e 122/62) = .57$$

$$\lambda^{A}sí = \frac{1}{4} (\log_e 45/245 + \log_e 62/122) = -.57$$

Cálculo del efecto conjunto de las dos variables

$$\lambda^{LA}ij = \frac{1}{IJ} \log_e (Fij (Fnoi, noj)/Fi, noj (noi,j))$$

$$\lambda^{LA}nono = \frac{1}{4} \log_e (245 (62)/122 (45)) = .25$$

$$\lambda^{LA}sísí = \frac{1}{4} \log_e (62 (245)/45 (122)) = .25$$

$$\lambda^{LA}nosí = \frac{1}{4} \log_e (45 (122)/245 (62)) = -.25$$

$$\lambda^{LA}síno = \frac{1}{4} \log_e (122 (45)/62 (245)) = -.25$$

Luego se cumple:

$$\lambda^{L}no = -\lambda^{L}sí \qquad \lambda^{A}no = -\lambda^{A}sí$$

$$\sum_{i=1}^{I} \lambda^{L}i = 0 \qquad\qquad \sum_{j=1}^{J} \lambda^{A}j = 0$$

Y también se cumple:

$$\lambda^{LA}nono = \lambda^{LA}sísí = -\lambda^{LA}nosí = -\lambda^{LA}síno$$

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \lambda^{LA}ij = 0$$

El modelo permite predecir los valores de las casillas. Para comprobarlo, transformamos la Tabla 1 en sus logaritmos naturales:

	Abandono	
Actividad laboral	*no*	*sí*
no	5.5	3.8
sí	4.8	4.1

Y procedemos al cálculo para cada una de las casillas:

— Casilla no trabaja, no abandona =

$$\mu + \lambda^{L}no + \lambda^{A}no + \lambda^{LA}nono = 4.5 + .10 + .57 + .25 = 5.5$$

—Casilla no trabaja, sí abandona =

$$\mu + \lambda^{L}no + \lambda^{A}sí + \lambda^{LA}nosí = 4.5 + .10 - .57 - .25 = 3.8$$

—Casilla sí trabaja, no abandona =

$$\mu + \lambda^{L}sí + \lambda^{A}no + \lambda^{LA}síno = 4.5 - .10 + .57 - .25 = 4.8$$

—Casilla sí trabaja, sí abandona =

$$\mu + \lambda^{L}sí + \lambda^{A}sí + \lambda^{LA}sísí = 4.5 - .10 - .57 + .25 = 4.1$$

Luego comprobamos que el modelo saturado siempre reproduce exactamente los valores observados en las casillas.

Los efectos que aparecen como más importantes son:

—Los individuos se distribuyen desigualmente en la variable Abandono (.57)
—Existe asociación entre las variables Actividad laboral y Abandono (.25)

VENTAJAS Y UTILIDAD INTERPRETATIVA

En el caso de las tablas de dos variables, el cálculo de las frecuencias esperadas es sencillo, pero cuando existen más variables, el cálculo se complica y necesitamos un algoritmo iterativo para la estimación. Es necesario utilizar el ordenador para efectuar estos cálculos. El único modelo en el que sí podríamos calcular las frecuencias esperadas sin esta ayuda es el de Independencia, multiplicando los marginales y dividiendo por el total.

También es sencillo, en el caso de dos variables, contrastar la bondad del ajuste de los 5 modelos. Pero con más variables el número de modelos aumenta y sería muy engorroso contrastarlos todos.

Nuestro interés consiste en encontrar el modelo que ajuste bien y que incluya el menor número de efectos posibles. El procedimiento que se suele utilizar para encontrar el mejor modelo es calcular, en primer lugar, los parámetros o efectos del modelo saturado (las lambdas) y elegir aquellos que sean significativos. A continuación se contrasta la bondad del ajuste del nuevo modelo seleccionado y la significación de los parámetros, con el fin de volver a seleccionar, si es posible, un modelo más sencillo, y así sucesivamente. Una vez elegido el modelo, se calculan las frecuencias esperadas y los valores de los efectos (parámetros o lambdas).

TABLA 2

Cruce de Actividad laboral, Horario y Abandono

		Abandono		
Actividad laboral	*Horario*	*no*	*sí*	*Total*
no	mañana	100	10	110
no	tarde	70	25	95
no	noche	75	10	85
sí	mañana	17	7	24
sí	tarde	50	20	70
sí	noche	55	35	90
Total		367	107	474

TABLA 3

Contraste de la bondad del ajuste
de los modelos y los parámetros

Modelo $\mu + \lambda^{L} + \lambda^{H} + \lambda^{A} + \lambda^{LH} + \lambda^{HA} + \lambda^{LA} + \lambda^{LHA}$		
$X^2 = 0$		
Modelo $\mu + \lambda^{L} + \lambda^{H} + \lambda^{A} + \lambda^{HA} + \lambda^{LA} + \lambda^{LH}$		
$X^2 = 9$		p= .012
parámetros		
	y^2	P
λ^{LH}	34	.00
λ^{LA}	15	.00
λ^{HA}	6	.06
Modelo $\mu + \lambda^{L} + \lambda^{H} + \lambda^{A} + \lambda^{LA} + \lambda^{LH}$		
$X^2 = 14$		p= .006
parámetros		
	y^2	P
λ^{LH}	39	.00
λ^{LA}	21	.00

En el ejemplo que vamos a exponer a continuación de tres variables (Tabla 2) nos detendremos más en estos aspectos de selección del modelo y, sobre todo, pondremos el énfasis en la utilidad del modelo logarítmico lineal para interpretar las tablas de contingencia multidimensionales.

En la Tabla 3 se contrasta la bondad del ajuste de varios modelos y de los parámetros, con el fin de encontrar el mejor modelo que explique la tabla de contingencia.

Para comentar los resultados obtenidos en la Tabla 3, comenzaremos por el modelo saturado. Este modelo incluye todos los efectos (parámetros), Gicuadrado es igual a 0 (no significativo) y ajusta bien.

Como el modelo logarítmico lineal es jerárquico, a continuación contrastamos el modelo que suprime la interacción de las tres variables.

El contraste entre las casillas esperadas bajo el nuevo modelo y las observadas da un X^2 de 9, con una significación de .012. Este modelo lo podemos aceptar o rechazar, según el nivel de confianza que estemos dispuestos a exigir. Si el nivel de confianza es del 99%, aceptaríamos el modelo como bueno y esto implica que no existe interacción entre las tres variables. Si por el contrario, exigimos que el nivel de confianza sea del 95%, rechazaríamos el modelo porque las diferencias entre las frecuencias esperadas y observadas son significativas.

Veamos qué ocurre en cada uno de los dos supuestos:

1) Aceptamos el modelo. Esto implica que pasamos a contrastar las lambdas de las tres asociaciones que incluye y, a continuación probaríamos con otro modelo que incluya menos efectos.

Los parámetros λ^{LH} y λ^{LA} son significativos con un nivel de confianza del 99%, sin embargo λ^{HA} no lo es. En el siguiente modelo que vamos a contrastar se suprime, por tanto, esta asociación (λ^{HA}).

El modelo que resulta incluye tan solo las asociaciones de Actividad laboral y Horario (λ^{LH}) y Actividad laboral y Abandono (λ^{LA}). Gi-cuadrado es igual a 14 y es significativo, luego rechazamos este modelo.

El modelo más sencillo que reproduce la tabla original es el siguiente:

$$\mu + \lambda^{L} + \lambda^{H} + \lambda^{A} + \lambda^{LH} + \lambda^{LA} + \lambda^{HA}$$

Es decir en los datos empíricos de la Tabla 2 existen los siguientes efectos y relaciones entre las variables:

—Efecto total
—Los tres efectos de los marginales
—El efecto debido a la asociación de las variables Actividad laboral y Abandono
—El efecto debido a la asociación de las variables Actividad laboral y Horario
—El efecto debido a la asociación de las variables Horario y Abandono se encuentra en un límite que permite su inclusión o no en el modelo dependiendo de las exigencias del investigador.
—Y por último, no existe el efecto debido a la interacción de las tres variables.

Estos resultados son idénticos a los obtenidos en el artículo sobre el Sistema de ecuaciones y grafos. También allí los impactos causales más fuertes eran:

—de la variable Actividad laboral sobre el Abandono
—de la variable Actividad laboral sobre el Horario

Por el contrario, el impacto causal del Horario sobre el Abandono era débil.

En el análisis del sistema de ecuaciones y grafos, la decisión de introducir el efecto indirecto de la Actividad laboral sobre el Abandono, por medio del Horario, dependía de cuán fino hilásemos. También en el análisis logarítmico lineal, la decisión de incluir la asociación de las variables Horario-Abandono y la interacción de las variables Actividad laboral-Horario-Abandono, depende de las exigencias del investigador y de los niveles de significación estadística que esté dispuesto a aceptar.

2) Rechazamos el modelo. Esto implica que no seguimos adelante reduciendo el número de efectos o parámetros y volvemos al modelo saturado, incluyendo la interacción de las tres variables.

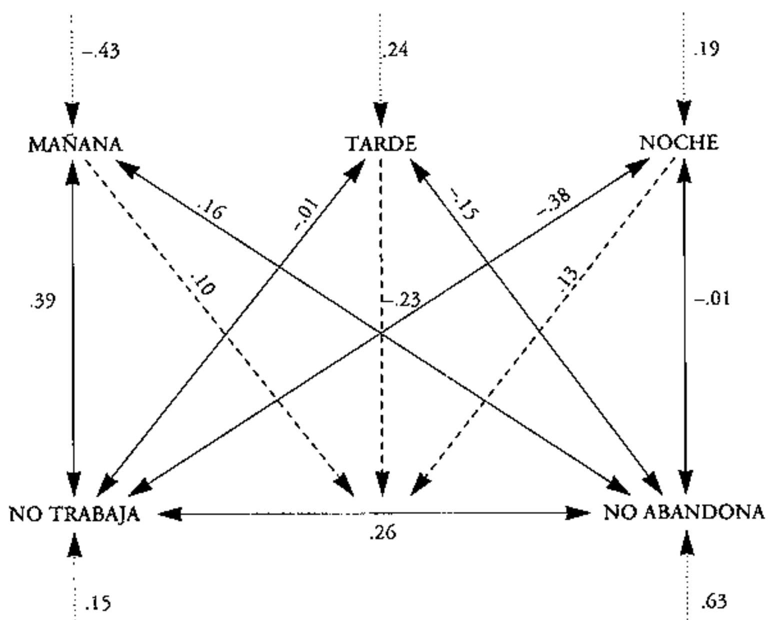
Esta decisión nos va a permitir incluir todos los efectos y, por tanto, calcular e interpretar un mayor número de resultados. De esta manera, cumplimos mejor los objetivos que nos habíamos propuesto en este artículo.

Esta decisión implica admitir como mejor modelo el saturado:

$$\mu = \lambda^{L} + \lambda^{H} + \lambda^{A} + \lambda^{LH} + \lambda^{LA} + \lambda^{HA} + \lambda^{LHA}$$

Procederemos como hemos indicado en el apartado anterior a cuantificar

Gràfico 1. Efectos del modelo saturado



los parámetros del modelo con el fin de determinar su impacto. Para facilitar la interpretación se presenta el gráfico número 1, donde aparece el valor de todos los parámetros:

—Flechas de puntos: Efectos de los marginales de cada variable (λ^{L}, λ^{H}, λ^{A}).
—Flechas de rayas continua: Efectos de asociación entre dos variables (λ^{LH}, λ^{LA}, λ^{HA}).
—Flechas de rayas discontinuas: Efectos de interacción entre tres variables (λ^{LHA}).

En función de los resultados que aparecen en el gráfico número 1, observamos que existe: 1) el efecto de la distribución desigual de los marginales; 2) asociaciones entre cada par de variables y 3) una interacción de tercer orden entre las variables «Horario», «Actividad laboral» y «Abandono». La interpretación es la siguiente:

Efectos de los marginales

El efecto de la distribución desigual de los entrevistados en cada variable que más influye proviene de la variable «Abandono» (.63), y la que menos influye en el valor que toman las casillas de la tabla de contingencia proviene de la variable «Actividad laboral» (.15).

Efectos de las asociaciones

—Actividad laboral y Abandono: Los alumnos que no trabajan es más probable que no abandonen al finalizar el primer año universitario (.26); o a la inversa , los trabajadores abandonan más.
—Horario y Abandono: Los alumnos de la mañana es más probable que no abandonen los estudios (.16), por el contrario los de la tarde es más probable que dejen sus estudios al finalizar el primer año (-.15).
—Actividad laboral y Horario: Los alumnos que no desempeñen actividad laboral es más probable que se matriculen por la mañana (.39), que por la tarde (-.01) o por la noche (-.38).

Efecto de la interacción

Existe interacción entre las variables, lo que implica que las relaciones que hemos descrito varían según las categorías que toman las tres variables. Expresado de forma más inteligible, podemos decir que la relación que existe entre la Actividad laboral y el Abandono es diferente en el horario de mañana y noche al de tarde, según muestran los valores de los parámetros (.10, .13, –.23).

CONCLUSIÓN

A lo largo de este artículo hemos tratado de poner en evidencia la necesidad de incorporar este análisis multivariable en el quehacer de los sociólogos empíricos. Podemos resumir las ventajas que aporta el modelo lineal logarítmico en las siguientes:

—Permite una mayor claridad en la exposición e interpretación de las tablas de contingencia multidimensionales.
—Permite cuantificar la interacción entre las variables.
—En las Ciencias Sociales, y en concreto en el contexto de una investigación por encuesta, este tipo de análisis es muy idóneo ya que:
  —La mayoría de las variables son nominales y ordinales.
  —Es necesario tratar *n* variables simultáneamente.
—No es necesario definir un modelo causal, ni variables dependientes e independientes, con la dificultad que estas definiciones previas implican.

BIBLIOGRAFÍA

Alvira, F. «Introducción al análisis de los datos», en García Ferrando, M.; Ibáñez, J.; Alvira, F. *Análisis de la realidad social. Métodos y Técnicas de Investigación.* Madrid, Alianza Universidad, 1986.

Alvira, F.; Marinez, E. «El efecto de los entrevistadores sobre las respuestas de los entrevistados en encuestas de opinión», en *Revista española de investigaciones sociológicas,* 29, Madrid, CIS, 1985.

Bishop, J.M., et al. *Bisaete Multivariate Analysis: Theory and Practic.* Cambridge University Press, 1975.

Blalock, H.M. *Estadística social.* Madrid, FCE, 1980.

García Ferrando, M. *Introducción a la Estadística en Sociología.* Madrid, Alianza, 1986.

Goodman, L.A. «The analysis of multidimensional contingency tables when some variables are posterior to others: a modifies path analysis approach», en *Biometrika,* 60, 1973.

Goodman, L.A. «A general model for the analysis of surveys», en *American Journal of Sociology,* 77, 1972.

Goodman, L.A. «Causal analysis of data from panel studies and other kinds of surveys», en *American Journal of Sociology*, 78, 1973.

Haberman, S.J. «Log Linear for contingency tables», en *Applied Statistics*, 21, 1972.

Knoke, D.; Burke, P.J. *Log-Linear Models*. Beverly Hills, Sage, 1980.

Latiesa, M. «Regularidad académica en la Facultad de CC.PP y Sociología», en *Educación y Sociedad*, 2, Akal, 1983.

Latiesa, M. «Abandono de los estudios en la Facultad de Ciencias Políticas y Sociología, sección de Sociología», en *Sociología, Revista de investigaciones sociológicas de la Asociación Castellana de Sociología*, 2, Especial Sociología de la Educación, ACS, 1983.

Latiesa, M. *Rendimiento académico, retraso escolar y abandono de los estudios en la F. de C.P. y Sociología*. Memoria para la obtención del grado de Licenciatura, F.C.P. y Sociología, 1982.

Nigel Gilbert, G. *Modelling Society: An Introduction to Log-Linear Analysis for Social Researchers*. Londres, George Allen Unwin, 1981.

Reynolds, H.T. *Analysis of nominal data*. Londres. Beverly Hills, Sage, 1977.

Ruiz-Maya, L. et al. *Metodología estadística para el análisis de datos cualitativos*. CIS, 1990.

Rosemberg, M. *The logit of Survey analysis*. Nueva York, Basic Books, 1968.

Sánchez Carrión, J.J. (ed.). *Introducción a las técnicas de análisis multivariante aplicadas a las Ciencias Sociales*. Madrid, CIS, 1984.

Simson, E.H. «The interpretation of interaction in contingency tables», en *Journal Royal Statistics Sociologica*, 1951.

Upton, G. «Contingency table analysis: log-linear models», en *Quality and Quanty*, 1, 1980.

Upton, G. *The analysis of cross-tabulated data*. Chischerter, Wiley, 1978.