

ANÁLISIS MULTIVARIABLE DE TABLAS DE CONTINGENCIA: SISTEMAS DE ECUACIONES Y GRAFOS¹

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 de análisis multivariable denominada Sistema de ecuaciones y grafos, basada en niveles de medición nominal u ordinal. Se trata de un modelo causal en el que es necesario determinar previamente las relaciones que existen entre las variables y definir las variables dependientes, independientes, espúreas e intervinientes. Se parte de una exposición del análisis tradicional de tablas de contingencia (presentación, lectura e interpretación de las tablas y prueba del Gi-cuadrado) y a continuación se presentan las tablas multidimensionales y su análisis mediante el Sistema de ecuaciones y grafos.

Resum

A l'hora de validar qualsevol model hipotètic en el seu esquema més elemental l'investigador es troba amb el problema de la introducció de la tercera variable. El mètode i el càlcul que es presenta permeten aclarir i calcular, d'una manera senzilla, els efectes directes, indirectes i globals per a un model hipotitzat, així com trobar els gràfics pertinents per millorar la representació dels dits models.

Abstract

In validating any hypothetical model in its basic scheme, the investigator has the problem of introducing the third variable. The method and calculus presented clarify and compute in a simple way the direct, indirect and global effects for an hypothesized model, as well as giving us the relevant graph for better representation of these models.

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

INTRODUCCIÓN

Las técnicas estadísticas de análisis de datos se pueden clasificar según el nivel de medición de las variables en dos categorías:

- Aquellas que se basan en variables cuyo nivel de medición es nominal u ordinal.
- Aquellas que se basan en variables cuyo nivel de medición es de intervalo o razón.

El objeto de este artículo es explicar las técnicas que se basan en niveles de medición nominal u ordinal. Estas técnicas se conocen con el nombre de *análisis tabular* o análisis de tablas de contingencia, y como todo análisis estadístico se basan en unos supuestos e implican un modelo de comportamiento de la realidad.

La mayoría de las investigaciones por encuesta utilizan casi exclusivamente variables nominales o clasificatorias, por lo que el *análisis tabular* es especialmente útil en la investigación social empírica. Así, gran parte de los informes finales de investigaciones basan exclusivamente su análisis en las tablas de contingencia: calculan los porcentajes verticales u horizontales, según proceda; efectúan el test de G^2 -cuadrado, para determinar la significación de la relación; y por último, calculan alguno de los coeficientes de asociación entre variables más conocidos: V de Cramer, Coeficiente de Contingencia, Φ , Λ .

Este tipo de análisis tabular, a pesar de que está muy extendido en el quehacer de los sociólogos empíricos, tiene el inconveniente de que se analiza tan solo la relación que existe entre dos variables. Sin embargo, en el contexto de una investigación normal existen muchas más variables, que pueden influir en cada par de asociaciones bivariadas.

Aunque es posible, por medio del análisis tradicional de tablas de contingencia, examinar tablas de más de dos variables, en la práctica resulta bastante complicado, por el número tan elevado de tablas que hay que construir. Por esta razón, el análisis tradicional se limita normalmente a construir tablas de dos variables. Para poder ir más allá, se recurre a otros modelos multivariados.

En los últimos años se ha desarrollado un modelo denominado *Sistema de ecuaciones y grafos*, a partir de los trabajos de Lazarsfeld y Rosemberg (1955); Rosemberg (1968); Davis (1976, 1980); Heise (1975); Goodman (1972, 1973); y en España Sánchez Carrión (1984).

El Sistema de ecuaciones y grafos es un modelo causal en el que el investigador define, previamente, la relación que existe entre las variables. Se

representa mediante un grafo donde se indican los efectos directos, indirectos y espúreos que pudieran existir entre las variables. Implica, por tanto, que hay que definir las variables dependientes e independientes.

Una tabla multidimensional (de n variables) se puede transformar, según este modelo, en un sistema de ecuaciones, donde los coeficientes representan el impacto causal que tienen unas variables sobre otras. A estos coeficientes se les denomina diferencia de proporciones o «d».

También se ha desarrollado en los últimos años otro modelo multivariable denominado *Modelo logarítmico lineal*. El objetivo de este análisis estadístico es encontrar el mejor modelo que explique las frecuencias observadas en la tabla de contingencia y determine las asociaciones e interacciones que son significativas.

En este artículo expondremos el *Sistema de ecuaciones y grafos*. Para ello, partiremos del análisis tradicional de tablas de contingencia de dos variables:

- Presentación, lectura e interpretación de las tablas.
- Prueba de significación de Gi-cuadrado.

A continuación analizaremos las tablas multidimensionales, utilizando el Sistema de ecuaciones y grafos. La exposición que hacemos de este modelo no es minuciosa y profunda, sino general y panorámica, ya que los objetivos son conocer cómo se utiliza esta técnica y presentar las ventajas de la misma.

ANÁLISIS DE LAS TABLAS DE CONTINGENCIA

PRESENTACIÓN DE LAS TABLAS, LECTURA E INTERPRETACIÓN

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 sobre el abandono de los estudios universitarios, en la Facultad de Ciencias Políticas y Sociología. Esta investigación fue presentada como tesina en octubre de 1982 (Latiesa 1982, 1983).

Si partimos de las variables Actividad laboral y Abandono de los estudios, cuyas distribuciones de frecuencias aparecen en la Tabla 1, y procedemos a la tabulación cruzada de las mismas, obtenemos la Tabla 2. La variable Actividad laboral aparece en las *filas* y la variable Abandono, en las *columnas*. Los *marginales* (totales) son las distribuciones de frecuencias de cada una de las variables y las *casillas* son el cruce de cada pareja de categorías.

TABLA 1

Distribución de frecuencias absolutas y relativas de las variables Actividad laboral y Abandono de los estudios

<i>Actividad laboral</i>	<i>N</i>	<i>%</i>	<i>Abandono estudios</i>	<i>N</i>	<i>%</i>
no trabaja	290	61	no	367	77
sí trabaja	184	39	sí	107	23
Total	474	100	Total	474	100

TABLA 2

Cruce de las variables Actividad laboral y Abandono de los estudios

<i>Actividad laboral</i>	<i>Abandono</i>		<i>Total</i>
	<i>no</i>	<i>sí</i>	
no	245	45	290
sí	122	62	184
Total	367	107	474

Para leer e interpretar la tabla, el procedimiento habitual que se utiliza es hallar los porcentajes. Estos pueden ser de fila, de columna o totales, según tomemos como base los marginales de la variable Actividad laboral, de la variable Abandono, o del total. Las tres tablas se presentan en las tablas 3, 4 y 5.

La interpretación de las tablas difiere según consideremos la Tabla 3, 4 ó 5. Luego, surge la pregunta: ¿qué tabla es la que debemos utilizar? La res-

TABLA 3

Cruce de las variables Actividad laboral y Abandono de los estudios (Porcentajes horizontales)

<i>Actividad laboral</i>	<i>Abandono</i>		<i>Total</i>
	<i>no</i>	<i>sí</i>	
no	84.5	15.5	100
sí	66.3	33.7	100
Total	77.4	22.6	100

TABLA 4

Cruce de las variables Actividad laboral y Abandono de los estudios
(Porcentajes verticales)

<i>Actividad laboral</i>	<i>Abandono</i>		<i>Total</i>
	<i>no</i>	<i>sí</i>	
no	66.8	42.1	61.2
sí	33.2	57.9	38.8
Total	100	100	100

TABLA 5

Cruce de las variables Actividad laboral y Abandono de los estudios
(Porcentaje total)

<i>Actividad laboral</i>	<i>Abandono</i>		<i>Total</i>
	<i>no</i>	<i>sí</i>	
no	51.7	9.5	61.2
sí	25.7	13.0	38.8
Total	77.4	22.6	100.0

puesta depende del investigador y la hipótesis o explicación que sustente. Siguiendo las recomendaciones de Zeisel, los porcentajes se deben calcular en la dirección de la variable independiente y se deben comparar en la dirección de la variable dependiente. Luego, la Tabla más apropiada es la 3, ya que nos indica si la Actividad laboral (considerada variable independiente) influye en el Abandono (considerada variable dependiente). La interpretación de esta tabla es la siguiente: los alumnos que no trabajan abandonan en menor proporción que los alumnos que trabajan. Intuitivamente, parece que la relación es significativa, pero se suele utilizar el X^2 para ver si esta relación es realmente significativa. En el apartado siguiente vamos a realizar esta prueba de contraste.

GI-CUADRADO

Siguiendo con el ejemplo del cruce de las variables Actividad laboral y Abandono de los estudios, las frecuencias observadas y esperadas (bajo la hipótesis nula) se presentan a continuación:

<i>Frecuencias observadas</i>				<i>Frecuencias esperadas</i>			
<i>Actividad laboral</i>	<i>Abandono</i>		<i>Total</i>	<i>Actividad laboral</i>	<i>Abandono</i>		<i>Total</i>
	<i>no</i>	<i>sí</i>			<i>no</i>	<i>sí</i>	
no	245	45	290	no	225	65	290
sí	122	62	184	sí	142	42	184
Total	367	107	474	Total	367	107	474

$$\chi^2 = \frac{\sum (F_{oij} - F_{eij})^2}{F_{eij}} = 20.1^*$$

A este valor de χ^2 le corresponde una significación de 0.000, con dos grados de libertad. Luego, existe una asociación significativa entre las variables Actividad laboral y Abandono de los estudios. Los alumnos que trabajan abandonan significativamente más que los que no trabajan.

Habitualmente, en el análisis de tablas de contingencia se procede a calcular otras asociaciones bivariadas, tomando como variable dependiente la misma variable, en este caso el Abandono de los estudios. De este modo procederíamos al cruce de las variables Horario, Sexo, Edad, etc., con el Abandono. Vamos a ilustrar otro ejemplo, con el cálculo de χ^2 , en el cruce de la variable Horario y Abandono. Las frecuencias observadas y esperadas se presentan a continuación.

<i>Frecuencias observadas</i>				<i>Frecuencias esperadas</i>			
<i>Horario</i>	<i>Abandono</i>		<i>Total</i>	<i>Horario</i>	<i>Abandono</i>		<i>Total</i>
	<i>no</i>	<i>sí</i>			<i>no</i>	<i>sí</i>	
Mañana	117	17	134	Mañana	104	30	134
Tarde	120	45	165	Tarde	128	37	165
Noche	130	45	175	Noche	135	40	175
Total	367	107	474	Total	367	107	474

$$\chi^2 = 11.2^*$$

De nuevo el X^2 es significativo, por lo que rechazamos la hipótesis nula de ausencia de relación y aceptamos la hipótesis de que existe asociación significativa entre el Horario y el Abandono de los estudios. Los alumnos de la mañana abandonan significativamente menos y los de la tarde significativamente más.

Así seguiríamos estableciendo las asociaciones entre la variable dependiente y multitud de variables independientes.

El problema de este tipo de análisis, es que determina las asociaciones brutas entre las variables, pero de este modo no podemos saber si la asociación entre dos variables es espúrea o está modulada por otras variables independientes. En el ejemplo, la relación entre Horario y Abandono podría deberse a que los alumnos trabajadores se matriculan por la tarde, y por eso, obtenemos que abandonan más los de este horario. En este caso, que incluimos una variable de control, podríamos construir dos nuevas tablas de contingencia y detectar una posible interacción, pero no si el efecto es espúreo. Por otra parte, cuando las variables que intervienen son más de tres, el intento de especificar otros tipos de relaciones se complica considerablemente, por lo que, normalmente, el análisis de tablas de contingencia se suele detener en el análisis bivariado de varias variables independientes con una dependiente.

Para ir más allá de estas limitaciones, necesitamos utilizar análisis multivariados de tablas de contingencia. A continuación, pasamos por tanto a exponer uno de los análisis que incluyen múltiples variables: el Sistema de ecuaciones y grafos.

SISTEMA DE ECUACIONES Y GRAFOS

Es un modelo causal en el que previamente hay que establecer la relación que existe entre las variables. Se definen las variables dependientes, independientes e intervinientes, y los efectos directos, indirectos y espúreos que pudieran existir. El inconveniente que presenta es que es difícil establecer modelos causales, especialmente cuando intervienen muchas variables.

Una vez que el investigador ha establecido el modelo causal, se procede a la transformación de la tabla de contingencia en un sistema de ecuaciones, en el que los coeficientes de la ecuación miden el impacto causal de unas variables sobre otras. Estos coeficientes se denominan diferencia de proporciones o «d».

Los coeficientes «d» se obtienen interpretando los porcentajes como probabilidades y calculando las diferencias, es decir, los coeficientes son dife-

rencias entre proporciones, y en una representación gráfica, esta medida se corresponde con la pendiente de la recta en una ecuación lineal.

Para ilustrar este modelo vamos a emplear dos ejemplos. *En el primero*, nos centraremos exclusivamente en las operaciones y cálculos que hay que efectuar, para transformar una tabla de contingencia en un sistema de ecuaciones. Es suficiente, por tanto, emplear una tabla bivariada. *En el segundo*, además de efectuar los cálculos, trataremos de ver la utilidad interpretativa del modelo en un ejemplo de tres variables.

PRIMER EJEMPLO

Siguiendo con el ejemplo del cruce de las variables Actividad laboral y Abandono de los estudios, reproducimos a continuación la Tabla 3 donde se expresan los porcentajes horizontales de la tabulación cruzada de las dos variables:

<i>Actividad laboral</i>	<i>Abandono</i>		<i>Total</i>
	<i>no</i>	<i>sí</i>	
no	84.5	15.5	100
sí	66.3	33.7	100

Tenemos que convertir esta tabla en una ecuación lineal, que se expresa así:

$$Y = K + d X$$

Y es la variables dependiente (Abandono de los estudios)

K es la constante

d es el coeficiente

X es la variable independiente (Actividad laboral)

Para convertir la tabla de contingencia en una ecuación lineal, tenemos que calcular el valor de dos variables, una constante y un coeficiente.

Si la ecuación es la siguiente:

$$Y \text{ no abandona} = K + d X \text{ no trabaja}$$

Veamos como se calculan cada uno de estos valores:

COEFICIENTE

El coeficiente «d» mide el impacto causal de la variable independiente

Análisis multivariable de tablas de contingencia: sistemas de ecuaciones y grafos

(Actividad laboral) sobre la variable dependiente (Abandono de los estudios).

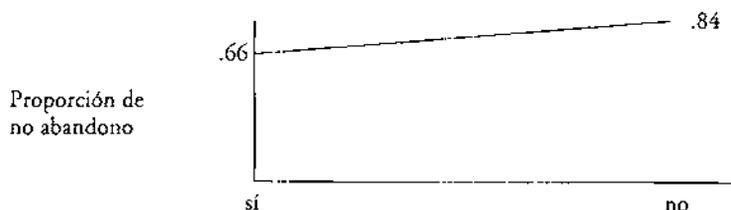
El coeficiente «d» es la diferencia de proporciones de «no abandono» según trabajen o no:

<i>Actividad laboral</i>	<i>Proporción de no abandono</i>
no	.84
sí	.66

$$d = (.84 - .66) = .18$$

La interpretación es la siguiente: un individuo que no trabaja es .18 más probable que no abandone, que un individuo que trabaja.

En un grafo cartesiano el valor de «d» se corresponde con la pendiente de una ecuación lineal:



$$\text{pendiente} = \frac{\Delta Y}{\Delta X} = \frac{.18}{1} = .18 = \text{coeficiente «d»}$$

CONSTANTE

La constante K es igual al valor de Y cuando la variable X es igual a 0, es decir, la proporción de casos de la variable dependiente cuando la independiente es igual a 0. Porque según la ecuación:

$Y = K + .18 X$; X vale 0 cuando la categoría es «sí abandona», luego Y es igual a la constante (K = .66).

VARIABLES

En una variable que toma los valores 0 y 1, la proporción se corresponde con la media. Así, la media de los valores «sí abandona» (0) y «no abandona» (1) es igual a la proporción del marginal de la categoría 1 ($367/474$) = .77; y la media de los valores «sí trabaja» (0) y «no trabaja» (1) es igual a ($290/474$) = .61

Luego:

$$X \text{ no trabaja} = 290/474 = .61$$

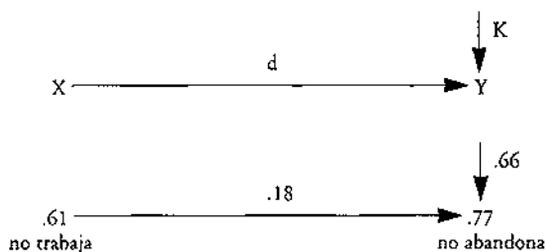
$$Y \text{ no abandona} = 367/474 = .77$$

La ecuación es la siguiente:

$$Y \text{ no abandona} = K + d X \text{ no trabaja}$$

$$Y \text{ no abandona} = .66 + ((.18) (.61)) = .77$$

Y el grafo es el siguiente:

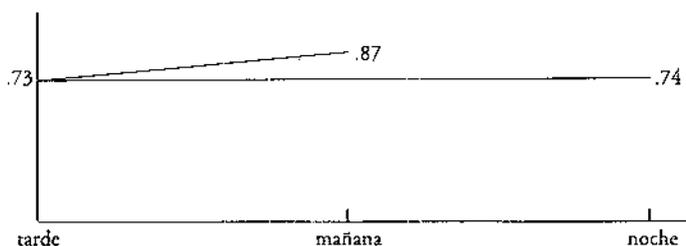


Si la variable independiente toma más de dos valores es necesario calcular más coeficientes, constantes y variables. En el caso de que tomase tres valores, es necesario escoger como base de comparación uno de ellos para calcular los coeficientes («d»). Veámoslo con el ejemplo del cruce de las variables Horario y Abandono de los estudiantes.

<i>Horario</i>	<i>Abandono</i>		<i>Total</i>
	<i>no</i>	<i>sí</i>	
Mañana	.87	.13	100
Tarde	.73	.27	100
Noche	.74	.26	100
	d		
Mañana	.14		
Noche	.01		
Base: Tarde	.73		

La interpretación es la siguiente:

Los alumnos de la mañana abandonan menos que los alumnos de la tarde (.14) y los alumnos de la noche abandonan prácticamente igual que los alumnos de la tarde (.01).



Al igual que hemos calculado dos coeficientes, hay que calcular dos ecuaciones, con sus correspondientes constantes y variables. Para ello, procederíamos de la misma forma que en el ejemplo anterior.

SEGUNDO EJEMPLO

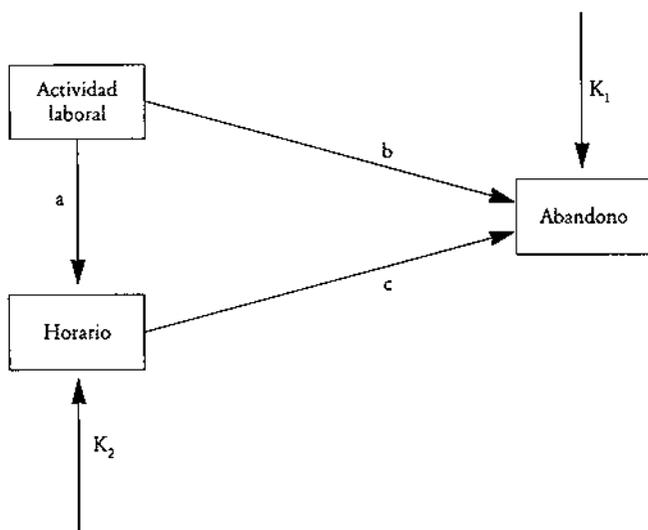
Con el fin de ver la utilidad y la ayuda interpretativa de este modelo, en un análisis multivariado, vamos a exponer un ejemplo con tres variables. En la Tabla 6 se presenta el cruce de las variables Horario, Actividad laboral y Abandono de los estudios.

Cuando se tienen más de dos variables el cálculo de las «d» se hace más complejo, puesto que las asociaciones brutas hay que descomponerlas en efectos directos, indirectos y espúreos. También es necesario especificar el modelo causal para el cálculo de los coeficientes y las constantes.

TABLA 6
Cruce de Actividad laboral con Horario y con Abandono

<i>Actividad laboral</i>	<i>Horario</i>	<i>Abandono</i>		<i>Total</i>
		<i>no</i>	<i>sí</i>	
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

GRÁFICO 1
Modelo Causal



El modelo causal se presenta en el gráfico 1 y es el siguiente:

- El Horario tiene una influencia directa sobre el Abandono.
- La Actividad laboral influye directamente en el Abandono de los estudios y en el Horario; también tiene un efecto indirecto sobre el Abandono por medio del Horario.

Podíamos haber elegido otros muchos modelos como por ejemplo los siguientes: 1) influencia de las variables independientes sobre la dependiente y ausencia de relación entre las independientes; 2) suponer que existe un efecto espúreo. No obstante, y como ya se ha indicado, la elección del modelo más idóneo es un problema previo, que concierne a los supuestos teóricos de los que parte la investigación y el que aquí hemos elegido nos parece suficientemente razonable.

Una vez establecido el modelo, colocamos las variables en orden causal y calculamos:

Análisis multivariable de tablas de contingencia: sistemas de ecuaciones y grafos

- 1) Las proporciones marginales de la variable Actividad laboral.
- 2) Las proporciones de Horario, para cada categoría de Actividad laboral.
- 3) Las proporciones de Abandono para cada combinación de Actividad laboral y Horario.

En la Tabla 7 aparecen estos cálculos.

TABLA 7
Proporciones recurrentes para el modelo descrito

X ₁ = Actividad laboral					
no	sí				
.61	.39	(474)			
X ₂ = Horario		Mañana	Tarde	Noche	
Actividad laboral	no	.38	.33	.29	(290)
sí	.13	.38	.49	(184)	
		.28	.35	.37	(474)
X ₃ = Abandono		Horario	Abandono		
Actividad laboral	no	Mañana	.91	.09	(110)
no	no	Tarde	.73	.26	(95)
no	no	Noche	.88	.12	(85)
sí	Mañana	Mañana	.71	.29	(24)
sí	Tarde	Tarde	.71	.29	(70)
sí	Noche	Noche	.61	.39	(90)
				.77	.23 (474)

A partir de esta tabla vamos a calcular las ecuaciones y el grafo:

- 1) En primer lugar, la influencia de la Actividad laboral sobre el Horario. Como la variable Horario tiene tres categorías, tenemos que tomar como base una de ellas. Obtenemos así dos coeficientes, como muestra la Tabla 8.

TABLA 8

Proporción de alumnos de mañana y noche según la Actividad laboral
(Base = Tarde)

Actividad laboral	Proporción (Mañana)
no	.38
sí	.13

$d = .25$

Actividad laboral	Proporción (Noche)
no	.29
sí	.49

$d = -.20$

También habrá dos constantes:

K (Mañana) = .13

K (Noche) = .49

Las medias son las siguientes:

Actividad laboral = .61

Horario de mañana = .28

Horario de noche = .37

2) A continuación calculamos el efecto de la Actividad laboral sobre el Abandono; para ello, hay que controlar el efecto entre Actividad laboral y Horario, haciendo que la proporción de alumnos para cada categoría del Horario, sea la misma entre los que trabajan y los que no. En la Tabla 9 se efectúa este cálculo, utilizando los marginales del Horario.

TABLA 9

Cruce de Actividad laboral con Horario,
en el supuesto de no asociación entre las variables

Horario	Actividad laboral		Total
	no	sí	
Mañana	.282 (82)	.282 (52)	134
Tarde	.348 (101)	.348 (64)	165
Noche	.368 (107)	.368 (68)	175
Total	1.000 (290)	1.000 (184)	(474)

Considerando estos valores absolutos construimos el cruce de las tres variables en la Tabla 10.

TABLA 10

Cruce de Actividad laboral con Horario y con Abandono, en el supuesto de no asociación entre Actividad laboral y Horario

<i>Actividad laboral</i>	<i>Horario</i>	<i>Abandono</i>		<i>Total</i>
		<i>no</i>	<i>sí</i>	
No	Mañana	.91 (75)	.09 (7)	82
No	Tarde	.73 (74)	.26 (26)	101
No	Noche	.88 (94)	.12 (13)	107
Sí	Mañana	.71 (37)	.29 (15)	52
Sí	Tarde	.71 (45)	.29 (19)	64
Sí	Noche	.61 (41)	.39 (27)	68

Para calcular el efecto directo de Actividad laboral sobre Abandono, construimos el cruce de Abandono con Actividad laboral en la Tabla 11, colapsando la Tabla 6.

TABLA 11

Cruce de Actividad laboral con Abandono, controlando el Horario

<i>Actividad laboral</i>	<i>Abandono</i>		<i>Total</i>
	<i>no</i>	<i>sí</i>	
no	243	46	290
sí	123	61	184

Y procedemos a calcular la «d» o efecto directo en la Tabla 12.

TABLA 12

Proporción de «No abandona» según la Actividad laboral

<i>Actividad laboral</i>	<i>Proporción (No abandona)</i>
no	.838
sí	.668
d = .170	

3) Para calcular la influencia directa de Horario sobre Abandono construimos la Tabla 13, colapsando la Tabla 6.

TABLA 13

Cruce de Horario con Abandono, controlando la Actividad laboral

Horario	<i>Abandono</i>		Total
	no	sí	
Mañana	112	22	134
Tarde	119	45	164
Noche	135	51	186

Y procedemos a calcular «d», o efecto directo, en la Tabla 14.

TABLA 14

Proporción de «No abandona» según Horario

Horario	Proporción (No abandona)
Mañana	.876
Tarde	.726
d = .110	

Horario	Proporción (No abandona)
Noche	.727
Tarde	.726
d = .001	

4) Cálculo de las constantes:

La constante de Abandono es igual a la proporción de los que no abandonan, cuando el Horario y la Actividad laboral son iguales a 0. Es, por tanto, la proporción de los alumnos que no abandonan del Horario de tarde y que trabajan.

$$K \text{ Abandono} = .71$$

5) Y por último calculamos la media (proporción de Abandono):

$$367/474 = .774$$

6) Con los resultados obtenidos ya podemos expresar el sistema de ecuaciones siguiente:

$$\text{Horario noche} = .49 + ((- .20) (.61)) = .37$$

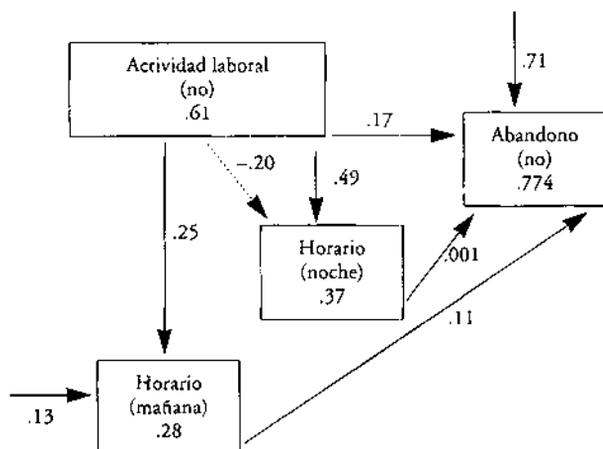
$$\text{Horario mañana} = .13 + ((.25) (.61)) = .28$$

$$\text{Abandono (no)} = .71 + ((.17) (.61)) + ((.11) (.28)) + ((.001) (.37)) = .78$$

Y el grafo correspondiente se presenta en el Gráfico 2.

GRÁFICO 2

Modelo Causal (Variables, constantes y coeficientes)



La interpretación de los resultados es la siguiente:

1. Existe relación directa entre el Horario y el Abandono, pero el impacto causal de esta relación es el más bajo de todos los que aparecen en el Gráfico 2.

Los alumnos de la mañana abandonan menos que los de la tarde (.11).

Los alumnos de la noche tienen las mismas probabilidades de abandonar que los de la tarde (.001).

2. Existe relación directa entre la Actividad laboral y el Horario:

Los alumnos de mañana es más probable que no trabajen que los alumnos de tarde (.25).

Por el contrario, es menos probable que no trabajen los del nocturno que

los de la tarde (-.20). En otras palabras, existen más trabajadores por la noche que por la tarde, por eso el coeficiente es negativo.

3. Existe relación entre la Actividad laboral y el Abandono.

Efecto Directo. Es .17 más probable que un alumno que no trabaja no abandone sus estudios, o a la inversa, es más probable que abandonen los alumnos que trabajan.

Efecto Indirecto. La Actividad laboral influye por medio de la variable Horario sobre el abandono:

- Los que *no trabajan* se matriculan más por la mañana que por la tarde; y los de la mañana abandonan menos que los de la tarde.
- Los que *no trabajan* se matriculan menos por la noche que por la tarde, pero abandonan igual que los de la tarde.
- Los que *trabajan* se matriculan menos por la mañana que por la tarde; y los de la mañana abandonan menos que los de la tarde.
- Los que *trabajan* se matriculan más por la noche que por la tarde, pero abandonan igual que los de la tarde.

El efecto indirecto de la Actividad laboral sobre el Abandono es igual al producto de los efectos Actividad laboral sobre Horario y Horario sobre Abandono:

$$((- .20) (.001)) + ((.25) (.11)) = .027$$

Efecto Total. Es igual al efecto directo más el efecto indirecto.

$$.17 + .027 = .197$$

O lo que es lo mismo, el efecto «d» de la asociación bruta de las dos variables:

<i>Actividad laboral</i>	<i>Proporción de «No Abandono»</i>
no	.855
sí	.663
d = .192	

Las discrepancias entre .197 y .192 se deben a que existe una pequeña interacción entre las variables.

CONCLUSIONES

El análisis tradicional de tablas de contingencia permite examinar las relaciones entre cada par de variables, determinando si existe relación mediante la prueba del Gi-cuadrado y en caso afirmativo, determinando la intensidad de la relación mediante los coeficientes de asociación.

La necesidad en la investigación social de introducir otras variables de control ha obligado a desarrollar nuevas técnicas que permitan el tratamiento simultáneo de varias variables. El modelo causal que hemos presentado permite, mediante la transformación de los coeficientes «d» en coeficientes de regresión estandarizados, similares a los utilizados con las variables de intervalo o razón, incluir varias variables y descomponer los efectos existentes entre ellas. Pero, como ocurre en todos los análisis causales, el problema consiste en determinar previamente el modelo teórico que establezca las relaciones entre las variables, ya que no siempre es posible hacer explícitas las hipótesis que orientan estas relaciones. Por otra parte, y dependiendo de los objetivos de la investigación, no siempre es necesario recurrir a este modelo. Por ejemplo, si nuestro objetivo fuese determinar las asociaciones e interacciones que existen entre las variables, no necesitaríamos definir previamente variables dependientes e independientes, ya que todas ellas las trataríamos como independientes; en este caso, el modelo multivariable más apropiado sería el logarítmico lineal.

BIBLIOGRAFÍA

- Blalock, H.M (ed.). *Measurement in the social Sciences*. McMillan, 1974.
- Blalock, H.M. *Estadística social*. FCE. Madrid, 1980.
- Davis, J.A. «Analizing contingency tables with linear flow graphs: D. Systems», en Heise, *op. cit.*, 1975.
- Davis, J.A. «Contingency tables analysis: proportions and flowgraphs», en James, Alt (ed.) *Quality and Quantity*, 1, 1980.
- Fienberg, S.E. *The analysis of crossclassified data*. Cambridge, M. Press, 1977.
- García Ferrando, M. *Introducción a la Estadística en Sociología*. Madrid, Alianza, 1986.
- Goodman, L.A. «A modified multiple regression approach to the analysis of dichotomous variables», en *American Sociology Review*, 37, 1972.
- Goodman, L.A. «Causal analysis of data from panel studies and other kinds of surveys», en *American Journal of Sociology*, 78, 1973.

- Goodman, L.A. «The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach», en *Biometrika*, 60, 1973.
- Heise, D.R. (ed.). *Sociological Methodology*. San Francisco, Jossey Bass, 1975.
- Heise, D.R. *Causal Analysis*. Nueva York, John Wiley & Sons, 1969.
- 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.
- 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. «Regularidad académica en la Facultad de CC.PP y Sociología», en *Educación y Sociedad*, 2, Akal, 1983.
- Lazarsfeld, P.F.; Rosenberg, M. *The language of Social Research: a Reader in the Methodology of Social Research*. Nueva York, Free Press, 1955.
- Reynolds, H.T. *Analysis of nominal data*. Beverly Hills, Sage, 1977.
- Rosenberg, 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.
- Simpson, E.H. «The interpretation of interaction in contingency tables», en *Journal Royal Statistics Sociologica*, 1951.
- Upton, G. *The analysis of cross-tabulated data*. Chischester, Wiley, 1978.