ÍNDICE

Estudios de Economía Aplicada
Nº 15, 2000. Págs. 163-186

Análisis de la opinión pública española sobre la influencia ciudadana en las decisiones gubernamentales mediante modelos de asociación para variables categóricas

SÁNCHEZ RIVERO, M.
FAJARDO CALDERA, M.A.
Dpto. Economía Aplicada y Organización de Empresas
Universidad de Extremadura

Esta versión incluye todas las correcciones sugeridas por el evaluador, las cuales nos han parecido oportunas y por las que le quedamos muy agradecidos.

RESUMEN

La participación ciudadana en las labores de gobierno de los regímenes democráticos ha sido objeto de constante preocupación. Los autores de este trabajo analizan cómo afecta el nivel de estudios a la opinión pública española en relación a la influencia de los ciudadanos sobre las decisiones de gobierno. La metodología estadística más adecuada para conseguir este objetivo son los modelos de asociación Fila-Columna.

A partir de la información estadística contenida en una tabla de contingencia bidimensional con variables ordinales, se plantean diversas hipótesis relativas a los valores asignados a las categorías de filas y de columnas. Estas hipótesis pueden contrastarse mediante la tabla de "Análisis de Asociación", que emplea un formato muy similar al del Análisis de la Varianza con dos factores de variación.

Palabras clave: Democracia participativa, modelos de asociación, efectos de filas, efectos de columnas, tabla de "Análisis de Asociación".

ABSTRACT

Civic participation in government tasks of democratic systems has been object of constant worry. The authors analize how educational level has an effect on spanish public opinion about citizens influence on government decisions. The most suitable statistical methodology to get this objective are Row-Column association models.

From data presented in a two-way contingency table with ordered variables, several hypotheses are imposed on assigned scores of row and column c ategories. These hypotheses can be tested with Analysis of Association table that uses a format quite similar to two-way Analysis of Variance.

Keywords: Participant democracy, contingency tables, association models, row effects, column effects, Analysis of Association table.

Código UNESCO: 1209/12, 1209/03, 1209/09

Artículo recibido en noviembre de 1999. Revisado en febrero de 2000.

ÍNDICE

1. Introducción

La Constitución conforma al régimen político español como una Monarquía parlamentaria y, por consiguiente, como una democracia representativa en la que la participación ciudadana en el Gobierno del Estado se encauza principalmente a través de la elección de representantes populares en los órganos de gobierno, pero también mediante la presentación de 500.000 firmas, lo que convierte a los ciudadanos en sujetos de la iniciativa legislativa.

El debate entre democracia representativa y democracia participativa está de candente actualidad, como confirman las investigaciones realizadas por diferentes instituciones y organismos dedicados al análisis de la estructura sociológica de la sociedad española, como el *Centro de Investigaciones Sociológicas* (CIS) o el *Centro de Investigaciones sobre la Realidad Social* (CIRES) que abordan la cuestión en varios de sus estudios.

Además, la popularidad que en los últimos años están alcanzando las nuevas tecnologías de la información, especialmente la red Internet, han convertido a estos canales de información en un medio ideal para fomentar y facilitar la participación ciudadana.

Por consiguiente, el interés que suscita el mayor o menor carácter participativo de las democracias modernas está fuera de toda duda. Con la intención de reflexionar sobre la cuestión planteada, y ciñiéndose a un aspecto muy concreto de la misma, los autores de este trabajo pretenden profundizar en la interrelación existente entre el nivel de estudios del ciudadano y la mayor o menor confianza que le inspira la democracia española como régimen político participativo. Se ha considerado el nivel de estudios porque, en opinión de los autores, el nivel formativo del ciudadano influye decisivamente en su posicionamiento sobre la participación ciudadana en las tareas de gobierno.

Nos planteamos, por tanto, la siguiente cuestión: ¿condiciona el nivel de estudios de un individuo su opinión sobre el carácter participativo de la democracia española?

Tabla 1.1. Clasificación cruzada del grado de acuerdo con la frase "*En un sistema democrático como el nuestro los ciudadanos influyen realmente en las decisiones del Gobierno*" y el nivel de estudios del entrevistado

NIVEL DE ESTUDIOS	*Muy de* acuerdo	*Más bien* de acuerdo	*Más bien en* desacuerdo	*Muy en* desacuerdo
Menos de estudios primarios (no sabe leer)	10	101	74	18
Menos de estudios primarios (sabe leer)	92	572	503	106
Estudios primarios completos (cert. escolar)	135	883	889	283
Formación profesional (primer grado)	17	106	178	49
Formación profesional (segundo grado)	14	134	145	70
Bachiller elemental	34	302	275	80
Bachiller superior	35	276	340	109
Estudios de Grado Medio (escuela universitaria)	16	125	191	61
Universitarios o técnicos de grado superior	23	145	191	83

Fuente: elaboración propia a partir de datos del CIRES (1997)[1]. *Tamaño muestral:* 6.665 encuestas.

Para dar respuesta a esta pregunta se ha empleado la información contenida en el estudio titulado *La realidad social en España (septiembre 1995 - junio 1996)* del CIRES y, más concretamente, los datos agregados del ejercicio 1995/1996. En la pregunta 14 del citado estudio se solicita al entrevistado que manifieste su opinión respecto a la influencia que los ciudadanos ejercen sobre las decisiones gubernamentales. Al cruzar esta información con el nivel educativo del entrevistado se obtienen las frecuencias observadas que se recogen en la TABLA 1.1.

---

1. La tabla ha sido elaborada una vez eliminadas las respuestas indeterminadas (no sabe / no contesta) de los entrevistados, al considerarse que las mismas no arrojan ninguna información relativa a la asociación existente entre las dos variables categóricas ordinales que conforman la tabla.

Para extraer la mayor información posible de los datos que contiene la tabla anterior (que, como puede apreciarse, recoge la clasificación cruzada de dos variables categóricas de naturaleza ordinal) se pueden emplear diferentes métodos estadísticos. Pero los autores de este trabajo consideran que el método que proporciona un análisis más eficiente de la información estadística de la TABLA 1.1. es el que subyace en los llamados modelos de asociación RC, a cuyo tratamiento se dedicarán los siguientes apartados de este trabajo.

2. Modelos de Asociación

Dada una tabla de contingencia bidimensional de dimensión IxJ, sea m_{ij} la frecuencia esperada en la *i*-ésima fila y en la *j*-ésima columna de la misma. Si se considera una distribución multinomial[2], para cada una de las diferentes subtablas 2x2 formadas a partir de dos filas consecutivas (es decir, filas i e $i+1$) y de dos columnas consecutivas (esto es, columnas j y $j+1$) se define el cociente de ventajas local[3] a partir de las frecuencias esperadas de la siguiente forma:

$$\theta_{ij}=\frac{m_{ij}m_{i+1,j+1}}{m_{i,j+1}m_{i+1,j}} \quad \text{para } 1\leq i\leq I-1; 1\leq j\leq J-1 \qquad (2.1.)$$

Bajo el conocido modelo de independencia estadística entre las categorías de la variable fila y las categorías de la variable columna, se verifica que todos y cada uno de los (I-1)(J-1) cocientes de ventajas locales son iguales a 1 (Christensen, 1990, pag. 38), es decir:

$$q_{ij}=1 \quad \text{para } 1\leq i\leq I-1; 1\leq j\leq J-1 \qquad (2.2.)$$

El anterior modelo (llamado por Goodman (1979, pag. 539) modelo de **asociación nula**, o modelo O), se contrasta mediante los conocidos tests de Pearson y de

2. En cualquier caso, los modelos aquí tratados también son válidos para un muestreo producto-multinomial (cuando se consideran I distribuciones multinomiales independientes para cada una de las filas, o cuando se consideran J distribuciones multinomiales independientes para cada una de las columnas de la tabla) o para un muestreo de Poisson (cuando se consideran IxJ distribuciones de Poisson independientes para cada una de las casillas de la tabla), ya que el tipo de muestreo empleado no afecta ni a la validez de los tests estadísticos utilizados para contrastar los diferentes modelos ni a las propiedades asintóticas de estos estadísticos bajo la hipótesis nula.

3. La expresión original en inglés es "*local odds-ratio*".

razón de verosimilitud, que siguen sendas distribuciones chi-cuadrado asintóticas con (I-1)(J-1) grados de libertad.

El modelo de asociación nula puede formularse también en términos de logaritmos de las frecuencias esperadas como sigue:

$$log\ \boldsymbol{m}_{\boldsymbol{ij}} = \tau_0 + \tau_i^A + \tau_j^B \tag{2.3.}$$

donde t_0 es la **media global** de todos los logaritmos de las frecuencias esperadas, mientras que t_i^A y t_j^B reciben el nombre de **efectos primarios** y recogen las diferencias en los valores marginales de las filas y de las columnas de la tabla, respectivamente.

Cuando una o las dos variables que forman la tabla poseen categorías ordenadas, será preciso reflejar dicha ordenación mediante la asignación de **valores fijos** conocidos a las categorías de las variables. Sean $\boldsymbol{x}_i$ (*i* = 1, 2, ..., I) (siendo $\boldsymbol{x}_1 \leq \boldsymbol{x}_2 \leq \ldots \leq \boldsymbol{x}_I$) los valores (*scores*) asignados a las categorías de la variable fila, y sean $\boldsymbol{y}_j$ (*j* = 1, 2, ...,J) (siendo $\boldsymbol{y}_1 \leq \boldsymbol{y}_2 \leq \ldots \leq \boldsymbol{y}_J$) los valores asignados a las categorías de la variable columna.

Para este último caso, y en el supuesto de que el modelo de asociación nula no se verifique, Goodman (1979, pag. 539) propone los siguientes modelos alternativos:

2.1. *Modelo de asociación uniforme*

Considérese, en primer lugar, un modelo en el que se verifica la siguiente condición:

$$\theta_{ij} = \theta \text{ para } 1 \leq \boldsymbol{i} \leq I-1; 1 \leq \boldsymbol{j} \leq J-1 \tag{2.4.}$$

siendo q un valor no especificado. Este modelo, que establece que la asociación local es la misma en todas las regiones de la tabla, se denomina **modelo de asociación uniforme** (*uniform association model*) o modelo U, el cual asume, con carácter general, que los valores elegidos para las filas y las columnas son equidistantes (es decir, se verifica que

$$\boldsymbol{x}_2 - \boldsymbol{x}_1 = \boldsymbol{x}_3 - \boldsymbol{x}_2 = \ldots = \boldsymbol{x}_I - \boldsymbol{x}_{I-1} = \boldsymbol{e}^*$$

y que

$$\boldsymbol{y}_2 - \boldsymbol{y}_1 = \boldsymbol{y}_3 - \boldsymbol{y}_2 = \ldots = \boldsymbol{y}_J - \boldsymbol{y}_{J-1} = \boldsymbol{e}^{**}$$) y,

particularmente, que esta equidistancia es unitaria (es decir, $\boldsymbol{x}_i = \boldsymbol{i}$; $\boldsymbol{y}_j = \boldsymbol{j}$). Dado que el modelo (2.4.) posee únicamente un parámetro más que el modelo O, los tests estadísticos para contrastar la validez de este modelo poseen (I-1)(J-1)-1 grados de libertad.

El modelo de asociación uniforme también puede reescribirse en términos de logaritmos de las frecuencias esperadas. Para llegar a esta expresión, considérese, por motivos de exposición, una generalización del anterior modelo de asociación uniforme, en la que se relaja la asunción de equidistancia unitaria entre los valores asignados a las categorías de las filas (x_i) y de las columnas (y_j) de la tabla. Este modelo más general se conoce con el nombre de **modelo de interacción lineal-lineal** (*linear-by-linear interaction model*, Clogg y Shihadeh, 1994, pag. 23), cuya expresión en términos de logaritmos de frecuencias esperadas es la siguiente:

$$\log m_{ij} = \tau_0 + \tau_i^A + \tau_j^B + \varphi x_i y_j \qquad (2.5.)$$

En la expresión anterior, el término $j x_i y_j$ puede ser considerado como una desviación de $\log m_{ij}$ del modelo de independencia completa ($t_0 + t_i^A + t_j^B$), de forma que $j x_i y_j$ es lineal en A para una categoría fija de B, y lineal en B para una categoría fija de A. Así, por ejemplo, para la columna *j*, la desviación del modelo de independencia es una función lineal de A (representada por los valores x_i) con pendiente $j y_j$. Esta circunstancia justifica el calificativo "lineal-lineal" de este modelo de asociación.

Dada la expresión (2.5.), el logaritmo de q_{ij} puede escribirse como sigue:

$$\phi_{ij} = \log \theta_{ij} = \varphi d_{x(i)} d_{y(j)} \qquad (2.6.)$$

donde $d_{x(i)} = x_{i+1} - x_i$ es la distancia entre las categorías consecutivas *i* e *i*+1 de la variable A y $d_{y(j)} = y_{j+1} - y_j$ la distancia entre las categorías consecutivas *j* y *j*+1 de la variable B.

La expresión (2.6.) permite apreciar que la intensidad de la asociación existente en cada región de la tabla está determinada por tres fuentes diferentes de asociación:

1ª) La asociación *intrínseca* (j).
2ª) La distancia entre categorías consecutivas de la variable fila A ($d_{x(i)}$).
3ª) La distancia entre categorías consecutivas de la variable columna B ($d_{y(j)}$).

A partir de (2.5.), la expresión del modelo de asociación uniforme en términos de los logaritmos de las frecuencias esperadas es inmediata, sin más que tener en cuenta la hipótesis de equidistancia unitaria que asume el citado modelo, ($d_{x(i)} = e^* = 1 \quad \forall i$; $d_{y(j)} = e^{**} = 1 \quad \forall j$):

ÍNDICE

$$log\ \boldsymbol{m}_{ij} = \tau_0 + \tau_i^A + \tau_j^B + \varphi \boldsymbol{i}\, \boldsymbol{j} \qquad (2.7.)$$

Para este modelo, la expresión (2.6.) queda reducida a:

$$log\ \theta_{ij} = \varphi \qquad (2.8.)$$

que establece que la asociación existente entre las variables ordinales A y B de la tabla está cuantificada por el parámetro j , de forma que si dicho parámetro es igual a 0, la asociación entre las variables será nula, es decir, se verificará el modelo de independencia completa ($\varphi = 0 \Rightarrow \theta_{ij} = 1 \forall \boldsymbol{i}, \boldsymbol{j}$).

2.2. *Modelo de efectos de filas y modelo de efectos de columnas*

Considérese seguidamente un modelo en el que se verifica lo siguiente:

$$\theta_{ij} = \theta_{i.} \text{ para } 1 \leq \boldsymbol{i} \leq I - 1; 1 \leq \boldsymbol{j} \leq J - 1 \qquad (2.9.)$$

donde $q_{i.}$ no está especificado. Este modelo recibe el nombre de **modelo de asociación de efectos de filas** (*row-effect association model*) o modelo R (Goodman, 1979, pag. 539). De la expresión (2.9.) se deduce que este modelo posee (I-1) parámetros más que el modelo de asociación nula, por lo que los grados de libertad para contrastar la bondad de ajuste del modelo R son (I-1)(J-2).

El modelo de efectos de filas se expresa en términos de logaritmos de las frecuencias esperadas como sigue:

$$log\ \boldsymbol{m}_{ij} = \tau_0 + \tau_i^A + \tau_j^B + \varphi \mu_i \boldsymbol{y}_j \qquad (2.10.)$$

donde m_i son valores *desconocidos* asociados a las categorías de la variable fila A y denominados *efectos de filas*[4] que verifican las restricciones $\sum_{i=1}^{I} \mu_i = 0$ y $\sum_{i=1}^{I} \mu_i^2 = 1$.

4. Como puede apreciarse, el modelo R no es más que un caso particular del modelo de asociación lineal-lineal en el que se han sustituido los valores ordenados por los efectos de filas, lo que también demuestra que el modelo R es, en realidad, un modelo log-lineal (lineal en los parámetros), a pesar de que la expresión (2.10.) incluya un producto de parámetros.

También es posible considerar un **modelo de asociación de efectos de columnas** (*column-effect association model*) o modelo C (Goodman, 1979, pag. 539), en el cual se verifica la siguiente condición:

$$\theta_{ij} = \theta_{.j} \text{ para } 1 \le \boldsymbol{i} \le I-1; 1 \le \boldsymbol{j} \le J-1 \qquad (2.11.)$$

en la que $q_{.j}$ es un valor no especificado. Los grados de libertad para contrastar el anterior modelo son iguales a (I-2)(J-1).

La expresión del anterior modelo en términos de logaritmos de las frecuencias esperadas sería la siguiente:

$$log\ \boldsymbol{m}_{ij} = \tau_0 + \tau_i^A + \tau_j^B + \varphi \boldsymbol{x}_i \varepsilon_j \qquad (2.12.)$$

En este caso, los valores y_j asignados a las categorías de la variable columna son sustituidos por parámetros desconocidos, denominados *efectos de columnas* (e_j), que están sometidos a las restricciones $\sum_{j=1}^{J} \varepsilon_j = 0$ y $\sum_{j=1}^{J} \varepsilon_j^2 = 1$.

2.3. *Modelo de efectos de filas y de columnas*

Por último, considérese un modelo que incluye tanto efectos de filas como efectos de columnas sobre los q_{ij}. Pueden definirse dos versiones de este modelo.

La primera versión viene dada por la siguiente condición:

$$\theta_{ij} = \theta_{i.} \theta_{.j} \text{ para } 1 \le \boldsymbol{i} \le I-1; 1 \le \boldsymbol{j} \le J-1 \qquad (2.13.)$$

donde $q_{i.}$ y $q_{.j}$ no están especificados. Este modelo recibe el nombre de **modelo de asociación de efectos de filas y de columnas aditivo** (o modelo R+C, o *Modelo I* en Goodman, 1979, pag. 539).

La segunda versión establece que:

$$log\ \theta_{ij} = \phi_{i.} \phi_{.j} \text{ para } 1 \le \boldsymbol{i} \le I-1; 1 \le \boldsymbol{j} \le J-1 \qquad (2.14.)$$

siendo $f_{i.}$ y $f_{.j}$ cantidades no especificadas. Este modelo se denomina **modelo de asociación de efectos de filas y de columnas multiplicativo** (o modelo RC, o *Modelo II* en Goodman, 1979, pag. 539).

ÍNDICE

Ambas versiones del modelo de efectos de filas y de columnas poseen (I-2)(J-2) grados de libertad y pueden formularse también en términos de logaritmos de las frecuencias esperadas. Así, tendremos que:

Para el modelo I (R+C):

$$\log \boldsymbol{m}_{ij} = \tau_0 + \tau_i^A + \tau_j^B + \varphi\left(\mu_i + \varepsilon_j\right) \qquad (2.15.)$$

Para el modelo II (RC):

$$\log \boldsymbol{m}_{ij} = \tau_0 + \tau_i^A + \tau_j^B + \varphi\mu_i\varepsilon_j \qquad (2.16.)$$

Esta última expresión, cuyos parámetros verifican que $\sum_{i=1}^{I}\mu_i = \sum_{j=1}^{J}\varepsilon_j = 0$ y que $\sum_{i=1}^{I}\mu_i^2 = \sum_{j=1}^{J}\varepsilon_j^2 = 1$, permite escribir el logaritmo de q_{ij} como sigue[5]:

$$\log \theta_{ij} = \varphi\left(\mu_{i+1} - \mu_i\right)\left(\varepsilon_{j+1} - \varepsilon_j\right) \qquad (2.17.)$$

En cualquier caso, debe advertirse que, de la misma forma que el modelo R+C es un modelo log-lineal en sus parámetros, el modelo RC ya no es un modelo log-lineal, puesto que el logaritmo de las frecuencias esperadas es una función multiplicativa de los parámetros m_i y e_j, lo que justifica los calificativos de *log-multiplicativo* o *log-bilineal* del modelo RC. Esta circunstancia podría dar lugar a que la función de verosimilitud no sea cóncava y a que posea máximos locales, en lugar de globales[6].

---

5. De la expresión (2.17.) se deduce que el modelo RC estima los valores óptimos de los parámetros m_i y e_j que maximizan el ajuste de la hipótesis de asociación lineal-lineal a los datos observados. También puede apreciarse que el modelo R se puede obtener a partir del modelo RC cuando las distancias $d_{B(j)} = e_{j+1} - e_j$ no dependen de *j*, es decir, cuando los efectos de columnas e_j son equidistantes. Los mismos comentarios podrían realizarse sobre el modelo C cuando en el modelo RC todos los efectos de filas equidistan entre sí.

6. Para resolver los inconvenientes derivados de este hecho, algunos autores (Goodman, 1979, pag.550) han propuesto *linealizar* el modelo RC fijando un conjunto de parámetros (por ejemplo, $m_i^* = j\, m_i$) y estimando seguidamente el otro conjunto de parámetros (ε_j) a partir de un modelo log-lineal. En el siguiente paso, se fijan los parámetros ε_j en los valores obtenidos en el paso anterior y se estiman los parámetros m_i^* a partir de un modelo log-lineal. Estas últimas estimaciones volverían a fijarse para reestimar los parámetros ε_j , y así sucesivamente. Se trata, por tanto, de un algoritmo en el que cada

La estimación de los parámetros de todos estos modelos se lleva a cabo mediante procedimientos iterativos de máxima verosimilitud. La maximización del logaritmo de la función de verosimilitud plantea un sistema de ecuaciones que no tiene una solución directa, pero que puede ser resuelto mediante métodos iterativos de estimación, como el algoritmo de Newton-Raphson, que consiste en realizar desarrollos de series de Taylor mediante aproximaciones sucesivas de un polinomio de segundo grado a la función de verosimilitud. Para más información sobre la estimación máximo-verosímil de los parámetros de todos estos modelos de asociación, puede consultarse Agresti (1984, pp. 79-80, 86) y Goodman (1979, pp. 549-55).

3. Modelos log-lineales vs modelos de Asociación

Llegados a este punto, podría plantearse la discusión sobre la idoneidad de los modelos de asociación para el análisis de tablas de contingencia, especialmente si se consideran los conocidos modelos log-lineales nominales. Es cierto que la modelización logarítmico-lineal ha demostrado ser una herramienta de gran utilidad para el análisis estadístico de tablas de contingencia. También es verdad que la formulación log-lineal permite una generalización inmediata de su planteamiento y de la contrastación de sus hipótesis de independencia a tablas multidimensionales. En síntesis, la descomposición del logaritmo de las frecuencias esperadas de la tabla en una serie de términos (efectos primarios y efectos de interacción) hace posible verificar hipótesis relativas a la independencia (completa, parcial o condicionada) entre las variables que forman la tabla, sin más que contrastar la significatividad estadística de los términos de interacción correspondientes. Agresti (1990), Christensen (1990) y Andersen (1990), entre otros, abordan la metodología log-lineal con suficiente profundidad.

Sin embargo, la modelización log-lineal presenta algunas limitaciones importantes. En concreto, considera que todas las variables categóricas que forman la tabla de contingencia son nominales, lo que provoca que tanto la estimación de los parámetros del modelo como los estadísticos de bondad de ajuste (test estadístico de Pearson y test de la razón de verosimilitud) sean invariantes ante una ordenación de las categorías de las variables. En consecuencia, si una o más de las variables que conforman la tabla de contingencia son de naturaleza ordinal, la formulación

---

ciclo consta de dos pasos, de forma que cada paso consiste en el ajuste iterativo del modelo R y del modelo C, respectivamente. En cualquier caso, tanto los errores estándar de los estimadores como otras medidas de precisión podrían ser erróneas, ya que este procedimiento no tiene en cuenta la aleatoriedad de ambos conjuntos de valores estimados y, aunque generalmente así ocurre cuando el modelo se ajusta bien a los datos, no hay garantía de que este algoritmo converja a estimaciones máximo-verosímiles.

logarítmico-lineal ignora la información que contiene la citada ordenación. Además, cuando las variables ordinales están presentes en el análisis, la presencia de elevados valores de los residuos ajustados en modelos de independencia confirma únicamente un ajuste inadecuado a los datos observados, pero sin tener en cuenta el hecho de que esa falta de ajuste puede estar explicada por la ordinalidad de las categorías de las variables. Este cúmulo de circunstancias explica, en parte, las limitaciones de los modelos log-lineales *nominales* para el análisis de datos ordinales.

Frente a los inconvenientes de los modelos nominales, las ventajas derivadas de explotar, cuando proceda, la ordinalidad de las variables categóricas son evidentes, como apunta Agresti (1990, pag. 262):

1º) Los parámetros de los modelos ordinales de asociación identifican tendencias y su interpretación es más sencilla que la de los parámetros de los modelos nominales.

2º) La gama de modelos existente entre el modelo de independencia completa y el modelo saturado es más amplia cuando se trabaja con modelos ordinales que cuando se hace con modelos nominales, hasta el punto de que existen modelos ordinales no saturados en situaciones en las que los modelos nominales son saturados.

3º) Los modelos ordinales estructuran la asociación y las interacciones entre las variables en un menor número de parámetros, por lo que retienen más grados de libertad que los modelos nominales.

4º) Los tests estadísticos basados en modelos ordinales incrementan la posibilidad de detectar ciertos tipos de interacción y asociación que, en los modelos nominales, pueden quedar ocultos.

Además de las ventajas citadas por Agresti, pueden mencionarse también las siguientes:

a) La metodología empleada por los modelos de asociación para tablas bidimensionales (que es la que aborda este trabajo) puede generalizarse al caso multidimensional, dando lugar a los denominados modelos de *asociación condicionada* y de *asociación parcial*.

b) Los modelos de asociación también se pueden utilizar para detectar diferencias significativas entre grupos en el análisis de distribuciones univariantes. Para ello, bastará simplemente que las categorías de la variable fila se refieran a diferentes grupos o poblaciones y que se compare la distribución de otra variable (ubicada en columnas) entre los grupos.

c) Finalmente, la asociación detectada por estos modelos en tablas de contingencia bidimensionales puede representarse gráficamente mediante el *análisis de correspondencia*. Los valores asignados a las filas y a las columnas de la tabla se representan mediante puntos, de forma que las posiciones de los mismos indicarán la mayor o menor intensidad de la asociación.

ÍNDICE

Por las razones expuestas, consideramos que la metodología más adecuada para analizar la asociación entre las dos variables de la TABLA 1.1. (ambas ordinales) son los modelos de asociación que fueron tratados en el epígrafe anterior.

4. La tabla de "Análisis de la Asociación"

La propiedad de anidamiento[7] que verifican los modelos de asociación tratados en epígrafes anteriores hace posible examinar de forma secuencial el conjunto de modelos (O, U, R, RC) o el conjunto de modelos (O, U, C, RC).

Un primer ejemplo de anidamiento es el que se refiere a los modelos O y U. Designando por $L^2(O)$ y $L^2(U)$ a los tests de la razón de verosimilitud[8] del modelo O y del modelo U, respectivamente, y teniendo en cuenta la propiedad de particionabilidad de este estadístico (Agresti, 1990, pag. 211), la diferencia $L^2(O/U)=L^2(O)-L^2(U)$, que sigue una distribución chi-cuadrado con 1 grado de libertad, puede emplearse para contrastar la hipótesis de asociación uniforme.

También están anidados los modelos O y R, de forma que, asumiendo que el modelo de efectos de filas se verifica, la hipótesis de independencia (que se cumpliría si $m_1=m_2=\ldots=m_I$) puede ser contrastada comparando ambos modelos mediante el estadístico $L^2(O/R)=L^2(O)-L^2(R)$, que sigue una distribución chi-cuadrado con (I-1) grados de libertad. Sin embargo, y puesto que el modelo de interacción lineal-lineal es un caso especial del modelo de efectos de filas, podría resultar más intere-

---

7. Se dice que un modelo M_2 está *anidado* con otro modelo M_1 cuando el primero es un caso especial del segundo, es decir, cuando M_2 es un modelo más simple que M_1, de manera que cuando M_2 se verifica, necesariamente M_1 también se verifica (Agresti, 1990, pag. 211).

8. El test más empleado para contrastar la bondad de ajuste de los modelos de asociación es el llamado **test de la razón de verosimilitud**, que compara las frecuencias observadas, n_{ij}, con las correspondientes frecuencias esperadas estimadas bajo el modelo en cuestión, $\hat{m}_{ij}^{(M)}$. Su expresión matemática es la siguiente:

$$L^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\, log\left(\frac{n_{ij}}{\hat{m}_{ij}^{(M)}}\right)$$

Sustituyendo en la expresión anterior $\hat{m}_{ij}^{(M)}$ por $\hat{m}_{ij}^{(O)}$, $\hat{m}_{ij}^{(U)}$, $\hat{m}_{ij}^{(R)}$, $\hat{m}_{ij}^{(C)}$ y $\hat{m}_{ij}^{(RC)}$ se obtendrían, respectivamente, las expresiones de los tests $L^2(O)$, $L^2(U)$, $L^2(R)$, $L^2(C)$ y $L^2(RC)$. Como puede observarse, cuanto más próximo a cero esté el valor de L^2, mejor será el ajuste del modelo a las frecuencias observadas. Por otro lado, para tamaños muestrales suficientemente grandes, L^2 sigue una distribución chi-cuadrado asintótica, cuyos grados de libertad se obtienen por diferencia entre el número de casillas de la tabla de contingencia y el número total de parámetros (incluyendo la media global) a estimar en el modelo.

sante comprobar la validez del conjunto de valores elegidos para la variable fila en un modelo U. Este test puede realizarse, haciendo uso nuevamente de la propiedad de particionabilidad del estadístico L^2, comparando dichos valores con los correspondientes efectos de filas del modelo R mediante la diferencia de verosimilitud $L^2(U/R)=L^2(U)-L^2(R)$, que sigue, en este caso, una distribución chi-cuadrado con (I-2) grados de libertad.

Comentarios similares a los anteriores serían válidos para los modelos anidados O y C, por un lado, y U y C, por otro lado, considerando los tests estadísticos $L^2(\mathrm{O}/\mathrm{C})=L^2(\mathrm{O})-L^2(\mathrm{C})$ y $L^2(U/C)=L^2(U)-L^2(C)$, respectivamente, que siguen sendas distribuciones chi-cuadrado con (J-1) y (J-2) grados de libertad.

Por su parte, la comparación entre los modelos R y RC contrasta la hipótesis de parámetros e_j equidistantes[9]. Esta comparación se realiza calculando la diferencia de verosimilitud $L^2(\mathrm{R}/\mathrm{RC})=L^2(\mathrm{R})-L^2(\mathrm{RC})$, que seguirá una distribución chi-cuadrado con (J-2) grados de libertad. De forma similar, la comparación entre los modelos C y RC proporciona información sobre el cumplimiento o incumplimiento de la hipótesis de equidistancia de los valores asignados a las filas de la tabla. Bastará simplemente obtener la diferencia de verosimilitud $L^2(C/RC)=L^2(C)-L^2(RC)$, que sigue también una distribución chi-cuadrado, pero, en este caso, con (I-2) grados de libertad. Finalmente, es posible comparar los modelos U y RC para contrastar la hipótesis de equidistancia conjunta de los valores de filas y de los valores de columnas. Para este tercer caso, el estadístico $L^2(U/RC)=L^2(U)-L^2(RC)$ sigue una distribución chi-cuadrado con (I + J-4) grados de libertad[10].

A la vista de estas relaciones entre modelos de asociación, Goodman (1979, pp. 537-552) propuso la llamada tabla de *análisis de asociación* (o tabla ANOAS). Esta tabla particiona el estadístico chi-cuadrado de forma similar a como se particiona la suma de cuadrados en un análisis de la varianza bifactorial. El estadístico que se particiona en la tabla ANOAS es el test de la razón de verosimilitud por la ya mencionada propiedad de particionabilidad, que, sin embargo, no verifica el test estadístico de Pearson.

Para construir la tabla ANOAS deberá partirse del estadístico de bondad de ajuste del modelo de asociación nula, $L^2(O)$, que es el que cuantificará la variabilidad total

---

9. O, lo que es lo mismo, la validez de la asignación de valores enteros consecutivos a las categorías de la variable ordinal ubicada en las columnas de la tabla.

10. Debe tenerse en cuenta que cuando el modelo de asociación nula (modelo O) se verifica, los parámetros m_i y e_j del modelo RC no se definen, por lo que no tiene ningún sentido proponer un test condicionado de independencia a partir del estadístico $L^2(\mathrm{O}/\mathrm{RC})$. Además, Haberman (1981, pp. 1178-1186) demostró que este estadístico ya no sigue una distribución chi-cuadrado.

a explicar. Seguidamente, deberá examinarse cómo la asociación *no explicada* puede ser analizada mediante los efectos (de filas, de columas o ambos) considerados por otros modelos.

Para una tabla de contingencia en la que ambas variables poseen categorías ordinales, la tabla ANOAS es la que se muestra en la TABLA 4.1. Al objeto de examinar la contribución de los diferentes efectos, se puede proceder de diferentes formas. Así, se pueden analizar los efectos de filas o se pueden analizar los efectos de columnas, siendo, en cualquier caso, recomendable realizar ambos análisis para detectar las contribuciones relativas de ambos tipos de efectos. De esta forma, si se consideran primero los efectos de filas, deberemos analizar la secuencia de modelos O, U, R, RC, cuyos componentes vienen dados por las líneas 1, 2a, 3a y 4 de la TABLA 4.1. Si, por el contrario, se consideran en primer lugar los efectos de columnas, la secuencia de modelos será O, U, C, RC, cuyos componentes se reflejan en las líneas 1, 2b, 3b y 4 de la citada tabla. En ambos casos, los residuos vendrán representados por el modelo RC (o R+C).

Tabla 4.1. Tabla ANOAS para una tabla de contingencia bidimensional en la que ambas variables son ordinales

Efectos de asociación	*Modelos empleados*	*Grados de libertad*	*Valor chi-cuadrado*
1. Efecto global	O-U	1	$L^2(O/U)$
2a. Efectos de filas	U-R	(I-2)	$L^2(U/R)$
2b. Efectos de columnas	U-C	(J-2)	$L^2(U/C)$
3a. Efectos de columnas dados efectos de filas	R-RC	(J-2)	$L^2(R/RC)$
3b. Efectos de filas dados efectos de columnas	C-RC	(I-2)	$L^2(C/RC)$
4. Residual	RC	(I-2)(J-2)	$L^2(RC)$
5. Total	O	(I-1)(J-1)	$L^2(O)$

Fuente: Clogg y Shihadeh (1994, pag. 54).

En el supuesto de que sólo estén ordenadas las categorías de la variable columna, el modelo U no intenvendrá en el análisis porque este modelo asume la ordenación de las categorías de ambas variables, circunstancia que aquí no ocurre debido a que la variable fila no es ordinal. Por el mismo motivo (asunción de filas ordenadas), los modelos C y R + C tampoco serían relevantes en el análisis de asociación. Sin embargo, en este caso el modelo RC sí es relevante, ya que los efectos de filas se aislarán comparando los modelos O y R, mientras que los efectos de columnas, dados los efectos de filas, se pueden aislar comparando los modelos R y RC. De esta forma, la secuencia de modelos sería, en este caso, O, R, RC. Por el mismo motivo, si la variable fila es ordinal y la variable columna no, la secuencia de modelos sería O, C, RC.

Finalmente, si ninguna de las dos variables de la tabla de contingencia es ordinal, el análisis de asociación involucra únicamente a los modelos O y RC, de forma que el estadístico $L^2(O/RC)$ se puede emplear para cuantificar la influencia que los parámetros del modelo RC ejercen sobre la asociación existente entre las variables[11]. La razón por la que habría que considerar únicamente a los modelos O y RC es que estos son los dos únicos modelos en los que la magnitud de los valores chi-cuadrado no depende de la ordenación empleada y, por tanto, son los únicos válidos para analizar la asociación existente en una tabla nominal-nominal.

5. Aplicación de los modelos de asociación al análisis de la opinión sobre la influencia ciudadana en las decisiones gubernamentales

Volviendo a la TABLA 1.1., y una vez presentada la metodología más adecuada para el estudio de la misma, nuestro análisis empírico se iniciará estimando los diferentes modelos de asociación presentados anteriormente, al objeto de confirmar la posible presencia de efectos de filas y/o de efectos de columnas responsables de la asociación entre las dos variables de la tabla. Los resultados de estas estimaciones se muestran en la TABLA 5.1.

Una primera inspección de dicha tabla confirma la presencia de asociación entre las dos variables que la forman, ya que el modelo de asociación nula arroja un ajuste a los datos bastante insatisfactorio, lo que nos da pie a rechazar la hipótesis de independencia. Esta presencia de asociación es, si cabe, más evidente cuando se compara el modelo O con el modelo de asociación uniforme. La importante reducción en el valor del test de la razón de verosimilitud ($L^2(O/U)$ = 57,3065 con1g. l.) pone claramente de manifiesto que el parámetro del modelo U es estadísticamente significativo.

11. En cualquier caso, ya se ha comentado con anterioridad que este estadístico no sigue necesariamente una distribución chi-cuadrado, circunstancia que ha llevado a algunos autores (Haberman, 1981, pp. 1178-1186) a proponer alternativas basadas en la distribución de Wishart.

Tabla 5.1. Modelos de asociación estimados para los datos de la Tabla 1.1.

MODELO	g.l.	L^2	c^2
Asociación nula (modelo O)	24	121,6306	121,2959
Asociación uniforme (modelo U)	23	64,3241	64,7581
Efectos de filas (modelo R)	16	27,8649	27,777
Efectos de columnas (modelo C)	21	63,2218	63,6064
Efectos aditivos de filas y de columnas (modelo R+C)	14	26,4928	26,2708
Efectos multiplicativos de filas y de columnas (modelo RC)	14	22,7684	22,5212

Fuente: Elaboración propia a partir de los cálculos realizados por el programa LEM versión 1.0.

Para cuantificar la contribución relativa del parámetro j a la variabilidad total se ha construido la tabla ANOAS (TABLA 5.2.a.), en la que se aprecia que el citado parámetro representa algo más del 47 % del valor del test de la razón de verosimilitud del modelo base, que es el modelo de independencia completa. La TABLA 5.2.a. también desvela la presencia de efectos de filas y de efectos de columnas. Un análisis más profundo de estos efectos evidencia una presencia importante de efectos de filas, como confirma la sustancial reducción del valor L^2 cuando el modelo U es reemplazado por el modelo R ($L^2(U/R)$ = 36,4592 con 7 g.l.) o la similar disminución que experimenta este estadístico cuando el modelo C es sustituido por el modelo RC ($L^2(\mathrm{C}/\mathrm{RC})$ = 40,4534 con 7 g.l.). Por el contrario, los efectos de columnas no contribuyen sustancialmente a explicar la asociación existente entre las variables, ya que al sustituir el modelo U por el modelo C, la reducción de L^2 es muy escasa ($L^2(U/C)$ = 1,1023 con 2 g.l.), sucediendo algo similar cuando el modelo R es sustituido por el modelo RC ($L^2(R/RC)$ = 5,0965 con 2 g.l.). Para confirmar esta idea, obsérvese la TABLA 5.2.b., que demuestra que los efectos de filas representan algo más del 87 % de la variabilidad explicada por los efectos de filas y de columnas considerados conjuntamente, mientras que los efectos de columnas representan tan sólo un 12% de la misma.

En la TABLA 5.3. se recogen las estimaciones de los parámetros del modelo RC (que es, como se observa en la TABLA 5.1., el que proporciona un ajuste más satisfactorio a los datos), las cuales han sido obtenidas con los programas CDAS y LEM. A la vista de los datos de esta tabla, además de constatarse la existencia de asociación entre las dos variables que forman la tabla ($\hat{\jmath}$ = 1,0665), se deduce que la asignación óptima de valores a las 9 categorías de la variable fila es la siguiente:

-0,5429	-0,4817	-0,1171	0,2947	0,2836	-0,2287
0,1213	0,3220	0,3487			

Tabla 5.2.

a) Tabla ANOAS para los datos de la Tabla 1.1.

Efectos de asociación	g.l.	L^2	*Porcentaje*
Efecto general	1	57,3065	47,11
Efectos de filas y de columnas	9	41,5557	34,17
Efecto residual	14	22,7684	18,72
TOTAL	*24*	*121,6306*	*100,00*

b) Partición de los efectos de filas y de columnas para los datos de la TABLA 1.1.

Efectos de asociación	*Modelos empleados*	g.l.	L^2	*Porcentaje*
Efectos de filas	U-R	7	36,4592	87,74
Efectos de columnas, dados los efectos de filas	R-RC	2	5,0965	12,26
Efectos de filas y de columnas	*U-RC*	*9*	*41,5557*	*100,00*

Fuente: Elaboración propia a partir de los cálculos realizados por el programa LEM versión 1.0.

Un análisis de estos valores pone de manifiesto, entre otras cosas, lo siguiente:

1º) La categoría 3 (estudios primarios completos) difiere de forma sustancial de las categorías 1 (menos de estudios primarios, no sabe leer) y 2 (menos de estudios primarios, sabe leer), ya que la distancia estimada entre aquella y estas dos últimas categorías es de 0,4258 y 0,3646, respectivamente. Sin embargo, la distancia entre las categorías 1 y 2 es tan sólo de 0,0612, lo cual sugiere que las mismas son *internamente homogéneas*[12]. En consecuencia, la asociación

12. Esta expresión significa que si se construyese una subtabla con estas dos categorías de la variable fila y con las cuatro categorías de la variable columna, se verificaría el modelo de independencia entre las dos variables de la tabla.

ÍNDICE

entre el nivel de estudios y la variable columna será diferente para los individuos que poseen estudios primarios completos que para aquellos que no han completado sus estudios primarios. Además, para estos últimos individuos, el hecho de saber leer no introduce diferencias significativas en la asociación, por lo que ambas categorías de la variable “nivel de estudios” pueden ser combinadas o colapsadas.

Tabla 5.3. Estimaciones de los parámetros del modelo RC para los datos de la Tabla 1.1.

Estadísticos:

Número de iteraciones: 7

Valor L^2: 22,7684 (0,0641)

Valor c^2: 22,5212 (0,0685)

Indice de similaridad: 0,0212

Grados de libertad: 14

Logaritmo de la función de verosimilitud: -20.535,74242

Número de parámetros estimados: 22

Criterio de información de Akaike (L^2): -5,2316

Criterio de información de Akaike (log-verosimilitud): 41.113,4848

Parámetros log-lineales:

t_0 = 4,6563

t_i^A:

$i=1$	-1,0582
$i=2$	0,7840
$i=3$	1,3502
$i=4$	-0,4861
$i=5$	-0,4491
$i=6$	0,1920
$i=7$	0,2943
$i=8$	-0,3716
$i=9$	-0,2555

t_j^B:

$j=1$	-1,2389
$j=2$	0,7234
$j=3$	0,8331
$j=4$	-0,3176

Efectos de asociación:

Asociación intrínseca (j): **1,0665**

Efectos de filas (m_i):

$i=1$	**-0,5429**
$i=2$	**-0,4817**
$i=3$	**-0,1171**
$i=4$	**0,2947**
$i=5$	**0,2836**
$i=6$	**-0,2287**
$i=7$	**0,1213**
$i=8$	**0,3220**
$i=9$	**0,3487**

Efectos de columnas (e_j):

$j=1$	**-0,5032**
$j=2$	**-0,4108**
$j=3$	**0,1738**
$j=4$	**0,7402**

Tabla 5.3. Estimaciones de los parámetros del modelo RC para los datos de la Tabla 1.1. (cont.)

	Filas/Columnas	Valor observado (esperado)	Filas/Columnas	Valor observado (esperado)
Cocientes de ventajas	1,2 / 1,2	0.6156 (1.0060)	5,6 / 1,2	0.9280 (0.9507)
	1,2 / 2,3	1.2002 (1.0388)	5,6 / 2,3	0.8415 (0.7266)
	1,2 / 3,4	0.8664 (1.0376)	5,6 / 3,4	0.6026 (0.7339)
	2,3 / 1,2	1.0520 (1.0366)	6,7 / 1,2	0.8878 (1.0351)
	2,3 / 2,3	1.1449 (1.2552)	6,7 / 2,3	1.3528 (1.2438)
	2,3 / 3,4	1.5106 (1.2464)	6,7 / 3,4	1.1020 (1.2353)
	3,4 / 1,2	0.9533 (1.0415)	7,8 / 1,2	0.9907 (1.0200)
	3,4 / 2,3	1.6679 (1.2927)	7,8 / 2,3	1.2404 (1.1333)
	3,4 / 3,4	0.8648 (1.2824)	7,8 / 3,4	0.9962 (1.1289)
	4,5 / 1,2	1.5350 (0.9989)	8,9 / 1,2	0.8070 (1.0026)
	4,5 / 2,3	0.6444 (0.9931)	8,9 / 2,3	0.8621 (1.0168)
	4,5 / 3,4	1.7537 (0.9933)	8,9 / 3,4	1.3607 (1.0162)

Fuente: Elaboración propia a partir de los cálculos realizados por los programas LEM versión 1.0. y CDAS versión 3.50.

2º) Los valores estimados para las categorías 4 (Formación Profesional, primer grado) y 5 (Formación Profesional, segundo grado) son virtualmente idénticos (la distancia entre ellos es 0,0111), por lo que estas dos categorías también podrían ser colapsadas.

3º) Las categorías 8 (estudios de grado medio) y 9 (estudios universitarios) también son susceptibles de combinación, al ser la distancia entre los valores asignados a las mismas igual a 0,0267.

A pesar del escaso peso de los efectos de columnas en el análisis de la asociación entre las variables de la TABLA 1.1., señalaremos que los valores estimados para las cuatro categorías de la variable columna son, respectivamente, -0,5032, -0,4108, 0,1738 y 0,7402, lo que, en principio, descarta la posibilidad de combinar o colapsar algunas de estas categorías.

Para determinar si las hipótesis de colapsabilidad de determinadas categorías de la variable fila son ciertas, bastará calcular la diferencia entre los tests L^2 del modelo RC para la tabla original (nivel de estudios con 9 categorías), o modelo M_1, y el modelo RC para la tabla colapsada (nivel de estudios con 6 categorías), o modelo M_2. En este caso, esta diferencia es igual a $L^2(M_1/M_2)$= 22,7684 - 6,4047 = 16,3637 con 6 grados de libertad, lo que confirma que el modelo colapsado, M_2, mejora significativamente el ajuste del modelo RC original y, en consecuencia, la propuesta de colapsabilidad comentada anteriormente puede ser admitida. Las estimaciones de los efectos de filas para el modelo RC colapsado son, respectivamente,

-0,6621, -0,1427, 0,4232, -0,2964, 0,1897 y 0,4883, las cuales mantienen entre sí distancias lo suficientemente significativas como para no apuntar otra propuesta de colapsación de categorías. Por su parte, los efectos de columnas estimados para el modelo M_2 son -0,5097, -0,4041, 0,1745 y 0,7392, respectivamente.

Por otra parte, al calcular el producto de parámetros estimados $\hat{j}\hat{m}_i\hat{e}_j$ se estará en condiciones de determinar en qué regiones de la TABLA 1.1. se registra una mayor asociación (tanto positiva como negativa) entre las variables que la forman. Se observa que la mayor asociación positiva se produce entre las categorías "menos de estudios primarios (no sabe leer)" y "menos de estudios primarios (sabe leer)" de la variable nivel de estudios y las categorías "muy de acuerdo" y "más bien de acuerdo" de la variable columna (0,2913, 0,2378, 0,2585 y 0,2110, respectivamente); y también entre las categorías "estudios de grado medio (escuela universitaria)" y "universitarios o técnicos de grado superior" del nivel de estudios y la categoría "muy en desacuerdo" de la variable columna (0,2542 y 0,2753, respectivamente). Por su parte, la mayor asociación negativa se detecta entre las categorías "menos de estudios primarios (no sabe leer)" y "menos de estudios primarios (sabe leer)" de la variable fila y la categoría "muy en desacuerdo" de la variable columna (-0,4285 y -0,3803, respectivamente); al igual que entre las categorías "estudios de grado medio (escuela universitaria)" y "universitarios o técnicos de grado superior" de la variable fila y la categoría "muy de acuerdo" de la variable columna (-0,1729 y -0,1871, respectivamente). Esta circunstancia confirma, por tanto, que cuanto menor es el nivel de estudios del entrevistado mayor es el grado de acuerdo con la frase analizada, y que cuanto más elevada es su formación, mayor es también el porcentaje de individuos que discrepan abiertamente con la frase.

Para finalizar el análisis, se estimará la probabilidad de que el entrevistado se sitúe en una de las cuatro categorías de la variable columna ("muy de acuerdo", "más bien de acuerdo", "más bien en desacuerdo", "muy en desacuerdo") dado su nivel educativo. El cálculo de esta probabilidad estimada puede realizarse mediante la siguiente expresión:

$$\hat{p}_{j/i} = \frac{\hat{m}_{ij}}{\sum_{j=1}^{J} \hat{m}_{ij}}$$

donde $\hat{m}_{ij}$ es la frecuencia esperada estimada de la casilla (*i*,*j*) de la tabla bajo el modelo de asociación RC.

Las estimaciones obtenidas se muestran en la TABLA 5.4., cuya inspección no viene sino a confirmar lo comentado hasta ahora. Destaquemos, en cualquier caso, los siguientes comentarios, que constituyen el punto final del análisis de asociación realizado:

ÍNDICE

a) Las mayores probabilidades de estar muy de acuerdo con la frase corresponden a los entrevistados que poseen “menos de estudios primarios” (tanto los que no saben leer - 6,98 % - como los que saben leer - 6,79 % -) y aquellos que poseen “Bachiller elemental” (6,05 %). Frente a ellos, los individuos que poseen Formación profesional de primer o segundo grado y los que poseen estudios de grado medio o superior poseen probabilidades de estar muy de acuerdo con la afirmación objeto de análisis que no superan, en el mejor de los casos, el 4,6 %.

Tabla 5.4. Probabilidades estimadas ($\hat{p}_{j/i}$) del grado de acuerdo con la frase dado el nivel de estudios del entrevistado

NIVEL *DE ESTUDIOS*	*Muy de acuerdo* (*j* = 1)	*Más bien de acuerdo* (*j* = 2)	*Más bien en desacuerdo* (*j* = 3)	*Muy en desacuerdo* (*j* = 4)
Menos de estudios primarios (no sabe leer) (*i* = 1)	0,0698	0,4706	0,3743	0,0853
Menos de estudios primarios (sabe leer) (*i* = 2)	0,0679	0,4610	0,3810	0,0901
Estudios primarios completos (certific. escolar) (*i*=3)	0,0572	0,4024	0,4174	0,1230
Formación profesional (primer grado) (*i*=4)	0,0457	0,3351	0,4493	0,1699
Formación profesional (segundo grado) (*i*=5)	0,0460	0,3369	0,4486	0,1685
Bachiller elemental (*i*=6)	0,0605	0,4205	0,4069	0,1121
Bachiller superior (*i*=7)	0,0505	0,3633	0,4373	0,1489
Estudios de Grado Medio (escuela universitaria) (*i*=8)	0,0450	0,3307	0,4510	0,1733
Universitarios o técnicos de grado superior (*i*=9)	0,0443	0,3263	0,4526	0,1768

Fuente: Elaboración propia a partir de los cálculos realizados por el programa LEM versión 1.0.

ÍNDICE

b) El análisis de las probabilidades asociadas a los individuos que están “más bien de acuerdo” con la frase arroja comentarios muy similares a los anteriores. Las probabilidades calculadas oscilan entre el 33 % de los entrevistados con mayor nivel educativo o con formación profesional y el 47 % de los que no poseen estudios primarios.

c) Con carácter general, se puede afirmar que el grado de desacuerdo con la frase es mayor cuanto más elevado es el nivel educativo del individuo, circunstancia que es apreciable en aquellos sujetos que están “muy en desacuerdo” con la frase, ya que frente a probabilidades próximas al 9% de los ciudadanos que no poseen estudios primarios, encontramos probabilidades que superan el 17% en aquellos sujetos que poseen estudios universitarios de grado medio o superior.

d) Obsérvese, por último, la gran similitud entre las probabilidades asociadas a los individuos que poseen “menos de estudios primarios (no sabe leer)” y aquellos que poseen “menos de estudios primarios (sabe leer)”. Esta circunstancia también se aprecia para aquellos sujetos que han cursado “Formación profesional (primer grado)” y los que han finalizado “Formación profesional (segundo grado)”; y para los entrevistados con “estudios de grado medio” y los “universitarios o técnicos de grado superior”. Esta identidad casi exacta entre las probabilidades es una consecuencia directa de la condición de colapsabilidad de las categorías adyacentes consideradas.

6. Conclusiones

A la vista del análisis empírico realizado, se pueden extraer las siguientes conclusiones finales:

1º) El modelo de asociación que mejor analiza la dependencia entre la opinión sobre el carácter participativo de la democracia española y el nivel de estudios del entrevistado es el modelo de efectos de filas y de columnas multiplicativo.

2º) La asociación intrínseca (j) entre las variables analizadas representa el 47,11% de la variabilidad total, mientras que los efectos de filas y de columnas explican el 34,17% de dicha variabilidad.

3º) Un análisis más detallado de los efectos conjuntos de filas y de columnas evidencia que los efectos de filas representan casi el 88 % de dichos efectos conjuntos, mientras que los efectos de columnas representan sólo un 12%.

4º) Las categorías “menos de estudios primarios (no sabe leer)” y “menos de estudios primarios (sabe leer)” de la variable nivel de estudios son *internamente homogéneas*, lo que significa que las mismas pueden ser combinadas o *colapsadas*. El mismo comentario es válido para las categorías “Formación profesional (primer grado)” y “Formación profesional (segundo grado)”, por un lado, y “estudios de grado medio” y “universitarios o técnicos de grado superior”, por otro lado.

ÍNDICE

5º) El análisis de asociación realizado mediante el modelo RC confirma que, con carácter general, cuanto más elevado es el nivel educativo del entrevistado, mayor es el escepticismo sobre el carácter participativo de la democracia española. Por contra, los sujetos con menor nivel formativo son los que más confían en la influencia de la ciudadanía sobre las decisiones gubernamentales.

Bibliografía

ANDERSEN, E.B. (1990): *The statistical analysis of categorical data*. Ed. Springer-Verlag. Berlín.

CENTRO DE INVESTIGACIONES SOBRE LA REALIDAD SOCIAL (1997): *La realidad social en España (septiembre 1995 - junio 1996)*. Datos agregados del ejercicio 1995/1996. Bilbao.

CHRISTENSEN, R. (1990): *Log-linear models*. Ed. Springer-Verlag. New York.

CLOGG, C.C. (1982): “Some models for the analysis of association in multiway cross-classifications having ordered categories”. *Journal of the American Statistical Association*, volumen 77, pp. 803-815.

CLOGG, C.C. y SHIHADEH, E.S. (1994): *Statistical models for ordinal variables*. Sage Publications. London.

ELIASON, S.R. (1990): *Categorical Data Analysis System, Version 3.50. User´s manual*. Departamento de Sociología. Universidad de Iowa.

GOODMAN, L.A. (1978): *Analyzing qualitative/categorical data*. Abt Books. Massachusetts.

GOODMAN, L.A. (1979): “Simple models for the analysis of association in cross-classifications having ordered categories”. *Journal of the American Statistical Association*, volumen 74, pp. 537-552.

GOODMAN, L.A. (1983): “The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds”. *Biometrics*, nº 39, pp. 149-160.

GOODMAN, L.A. (1984): *The analysis of cross-classified data having ordered categories*. Harvard University Press. Massachusetts.

HABERMAN, S.J. (1981): “Test of independence in two-way contingency tables based on canonical correlation and on linear-by-linear interaction”. *Annuals of Statistics*, nº 9, pp. 1178-1186.

ISHII-KUNTZ, M. (1994): *Ordinal Log-Linear Models*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-097. Thousand Oaks, CA: Sage Pubns.

VERMUNT, J. (1997): *Log-linear and event history analysis with missing data, Version 1.0. User´s manual*. Tilburg University. Holanda.