

Determinación de audiencias: una aplicación multivariable

Carlos A. Guerrero Serón

Resulta suficientemente conocida para el lector la dificultad que entrañan los métodos usuales para la determinación de audiencias. En términos puramente instrumentales, la mayoría de ellos –si no todos– requieren disponer, en primer lugar, de una amplia base de datos –acaso lo más difícil– y, después, de un paquete estadístico muy especializado. La consecuencia de ello es clara: la mayoría de los usuarios tienen que conformarse con lo que los grandes centros de datos les ofrecen –y al precio que les pidan– o planificar intuitivamente.

1. Mi objetivo

En las líneas que siguen voy a tratar de ofrecer los resultados de una técnica de análisis que trata de sentar las premisas de un sistema de determinación de audiencias que, si bien no pretende llegar a la planificación de medios, al menos, puede constituir una buena herramienta que permita seleccionar mejor los medios más adecuados a un ‘target’ determinado.

En primer lugar, quiero señalar que mi «campo de estudio» está delimitado por los siguientes condicionantes:

Trato de buscar una metodología de análisis que,

a) permita explotar los datos contenidos en los estudios de audiencias disponibles,

b) el tratamiento de la información pueda realizarse con una simple calculadora y/o una hoja de cálculo.

En otras palabras, lo que pretendo es encontrar una técnica que, siendo estadísticamente válida, pueda ser utilizada por cualquier profesional con unos conocimientos mínimos de la materia.

Obviamente, como acabo de decir, no va a ser posible planificar medios con estas herramientas pero, en cualquier caso, espero ofrecer una aplicación relativamente sencilla que constituya una estimable ayuda para tantos jefes de medios de pequeñas agencias que no pueden acudir a las bases de datos centrales por falta de recursos económicos y tienen que conformarse con las «planificaciones» que les ofrecen algunos intermediarios (sobre todo si disponen de una simple hoja de cálculo en un PC).

2. Los métodos actuales

Los métodos actualmente disponibles para segmentar audiencias, en general, siguen unos pasos o etapas metodológicas en su construcción:

a) descansan sobre la premisa de la existencia de un «hábito» de exposición a los medios sin la cual sería imposible estimar la frecuencia de contacto. Es decir, «...el contacto entre los individuos y los soportes no se produce al azar sino que es consecuencia de una reglas, de unas costumbres, de unas ideologías, de unas preferencias, de unos niveles culturales, que se cristalizan en la idea de hábito de audiencia» (Martínez Ramos 1992: 13-6).

b) Pero, una vez solventadas las definiciones o postulados iniciales sobre lo que es una audiencia y lo que constituye un soporte (una publicación concreta, una franja horaria, etc.) nos encontramos con la necesidad de determinar cuales son los «atributos» de los individuos que mejor definen la audiencia de un soporte. Entramos ahora en un nuevo problema, cual es, transformar esos atributos en variables que nos permitan trabajar estadísticamente con ellos y obtener resultados válidos para nuestros planteamientos.

Existen dos tipos fundamentales de variables o atributos, las cuantitativas –sexo, edad, formación, etc– y las cualitativas, basadas en los modos de vida de las personas, sus intereses, sus aficiones, etc.

c) En función del tipo de variables sobre el que se vayan a realizar las segmentaciones se construyen los modelos estadísticos. De esta manera, nos encontramos con métodos de Belson, de Agostini, de la Ghi–cuadrado, etc. para variables sociodemográficas y métodos de Areas de Interés, Estilos de Vida y basados en corrientes socioculturales que utilizan variables de tipo cualitativo (*Ib*: 18-24).

3. La técnica que propongo

Por mi parte, sin entrar en las dificultades de definición y las condiciones técnicas o características básicas que deben reunir las variables para ser operativas (Vidosa González 1992 , Soler y Perdiguier 1992, López Pintor 1989, Abascal y Grande 1989), quiero centrarme en la metodología de análisis utilizada para extraer resultados de la información que contienen las variables más comunes en nuestros estudios, esto es, variables de tipo nominal o, a lo máximo, ordinales.

Por ahora, mi trabajo se ha centrado en el análisis de Tablas de Contingencia utilizando el estadístico «d» de Davis, siguiendo las enseñanzas de J. J. Sánchez Carrión en su libro *Análisis de Tablas de Contingencia* (1989)

Este «sistema de análisis» parte de la utilización de variables nominales basado en un solo estadístico, la Diferencia de Porcentajes, a partir del cual se construye el sistema de análisis que desemboca en un sistema de ecuaciones que es el que mejor predice el comportamiento de la audiencia respecto a un determinado medio o soporte.

Con este sistema, podremos estudiar, además:

1. La existencia de causalidad. Es decir, si la relación entre dos variables es de tipo causal o se trata de una relación puramente estadística. Por ejemplo, es un hecho que en hábitats rurales se lee menos prensa que en los hábitats urbanos, sin embargo, no por ello podemos decir que la *causa* del menor índice de lectura es el hábitat. Es posible que en los núcleos poblacionales pequeños se den ciertas circunstancias –poco mercado, carencia de infraestructuras, estratos ocupacionales dis-

tintos, comunicaciones,... – que favorezcan una menor exposición a los medios, sobre todo a la prensa, pero también es posible que este resultado sea más una apariencia de causalidad (relación puramente estadística) que una realidad.

2. Estudio de la secuencia causal. Una vez asumida la relación causal entre dos variables, podremos conocer cuál es la *secuencia* causal entre ambas. Por ejemplo, podemos encontrarnos con que las personas de mayor edad (Variable Edad) suelen ocupar puestos de mayor nivel profesional (Variable Ocupación) que las más jóvenes. Esta situación puede deberse tanto a que, a igualdad de experiencia, las empresas prefieren a personas con mayor formación (efecto directo, independiente de cuál sea la edad), que al hecho de que tener una determinada edad facilita la posibilidad de acceder a mayores niveles formativos –los más jóvenes tienen muchas más oportunidades de estudiar que sus padres, por ejemplo. Esto es lo que llamamos efecto indirecto de Formación sobre Ocupación a través de Edad.

Por ello, hay una secuencia que lleva de tener una determinada edad (más o menos disponibilidad de formación) a contar con mayores niveles de formación y de aquí a ocupar puestos profesionales de mayor cualificación.

3. Estudio de las relaciones *ocultas*. Introduciendo una tercera variable podremos descubrir relaciones ocultas entre otras dos. Por ejemplo, si a una determinada edad parece que existe una acusada relación con la exposición a un determinado soporte, este hecho puede deberse a la formación de los individuos, existiendo por lo tanto una relación oculta entre edad, formación y audiencia.

4. Estudio de la Interacción. Al añadir terceras variables podremos especificar bajo qué condiciones se produce una relación entre dos variables, viendo si esa relación es la misma o cambia según las circunstancias. En general, la Ocupación (variable Ocupación) es un factor que influye en la lectura de prensa (Variable Lectura) entre los individuos (relación entre Ocupación y Lectura). Esta influencia es mucho más acusada entre los individuos que desarrollan ocupaciones de «cuello blanco» que entre los inactivos y trabajadores manuales, lo que nos da una fuerte relación entre Ocupación y Lectura de Prensa. Si introducimos la variable Formación, vemos cómo entre los individuos de formación Baja se

da mucha menos lectura de prensa que entre los de formación Media y, en éstos, mucha menos que entre los de formación Alta. Por lo tanto, la relación entre Ocupación y Lectura es mucho más intensa conforme se incrementa el nivel de formación en cualquiera de los estratos o categorías, es decir, la relación entre Ocupación y Lectura está condicionada por la formación que tiene cada individuo.

5. Estudio de los efectos relativos y conjuntos. Podremos estudiar la influencia conjunta de dos o más variables independientes sobre otra dependiente. Bien para comparar sus efectos (¿Qué influye más en la lectura: la Formación, la Edad, o el Hábitat?), o bien para estudiar su influencia conjunta (Si el efecto de Ocupación sobre lectura es 0,249 ¿Cuál será el efecto de Ocupación más Formación?). Por ejemplo, antes nos preguntábamos por qué los individuos con niveles ocupacionales más altos leen más prensa y ahora la pregunta sería ¿Por qué hay unos individuos que tienen un nivel ocupacional más alto que otros?.

3.1. Base de Datos

Para ilustrar esta serie de «beneficios» que nos proporciona el análisis multivariable en base a la construcción de tablas de contingencia, me sirvo de los datos contenidos en el *II Estudio sobre Hábitos de Exposición a Medios en Andalucía* (Guerrero Serón 1992). Esta fuente de datos tiene para mí varias ventajas. En primer lugar, la conozco perfectamente al haber sido el director del estudio. En segundo lugar, es la única que se refiere a Andalucía, que es el campo de estudio que me interesa en estos momentos. Además, cuento con una muestra de 3.900 elementos que me van a permitir realizar cruces múltiples sin quedarme sin sujetos.

Como materia de análisis para este artículo me voy a centrar en la lectura de prensa en el día de *ayer*.

3.2. Construcción de tablas de contingencia

Para seleccionar las categorías de análisis de las variables he utilizado la técnica de Análisis de los Residuos a todas las variables incluidas en el Estudio de Medios. Esta operación me ha permitido, por un lado, eliminar variables poco significativas para explicar la lectura 'Ayer' de

prensa y, por otro, reducir las categorías de las variables incluidas en función de su significatividad respecto a la variable dependiente.

En cualquier caso, también he tenido en cuenta la ‘operatividad’ práctica de las variables incluidas en el Estudio, esto es, su valor funcional para los planteamientos más comunes de los profesionales de la publicidad a la hora de efectuar sus planes de medios⁷. Con este criterio, se han mantenido para la construcción del modelo las variables Sexo y Rol Familiar, ya que suelen ser utilizadas con frecuencia para determinar los ‘target’ publicitarios. Con el mismo criterio práctico, he eliminado otras variables que muestran una gran redundancia con alguna de las expuestas. Por ejemplo, las categorías de la variable Estado Civil (Casado, Soltero y Otros) son equivalentes a Perceptor de principal de Ingresos, Ama de Casa, Hijos y Otros miembros, por lo que su inclusión en el modelo la considero redundante.

Por otro lado, aún a sabiendas de que el sistema de ecuaciones debe ser el resultado de los análisis que se efectuarán a continuación, sobre todo después de calcular los Coeficientes de Predecibilidad y establecer el ‘camino’ de mayor ganancia, he elaborado este sistema de ecuaciones, tras los análisis de residuos y con los criterios operativos que se acaban de exponer, con el fin de, por un lado, proporcionar al lector unos elementos más didácticos con vistas a la comprensión de las elaboraciones que se van a abordar de inmediato y, por otro, poder ofrecer tablas de contingencia ya elaboradas en base a los criterios operativos.

Concretamente, las variables que voy a utilizar son:

–Sexo: Hombres y Mujeres.

–Edad: agrupadas en las siguientes categorías: Hasta 25 años (–25 años), de 25 a 44 años (25/44 años) y de 45 o más años (45 ó +).

–Hábitat: agrupados en dos categorías, Rural (hasta 50.000 habitantes) y Urbano (Más de 50.000 habitantes)

–Formación: Baja (hasta estudios primarios), Media (EGB, FP, etc.), Alta (Universitarios).

–Ocupación: Inactivos (Pensionistas, Rentistas, Estudiantes, sus labores,...), Trabajadores Manuales (cualificados o no), Técnicos y Profesionales por cuenta ajena, Empresarios y Profesionales por cuenta propia.

–Rol Familiar: Perceptor principal de ingresos, Encargado/a del

Hogar y Otras situaciones.

A continuación, se construyen las tablas que se corresponden con estas relaciones bivariadas:

TABLA PA.H.1.: Relaciones entre Variables Hábitat y Prensa (Leyó Ayer)
(% Horizontales)

<u>Hábitat</u>	<u>Leyó</u>	<u>No leyó</u>	<u>Total</u>	<u>Hábitat</u>	<u>Leyó</u>	<u>No leyó</u>	<u>Total</u>
Rural	317	1756	2127	Rural	0,174	0,826	1
Urbano	599	1179	1778	Urbano	0,337	0,676	1
Total	970	2935	3905	Total	0,248	0,752	1

TABLA PA.S.1.: Relaciones entre Variables Sexo y Prensa (Leyó Ayer)
(% Horizontales)

<u>Sexo</u>	<u>Leyó ayer</u>	<u>No leyó</u>	<u>Total</u>	<u>Sexo</u>	<u>Leyó ayer</u>	<u>No leyó</u>	<u>Total</u>
Hombres	607	1269	1876	Hombres	0,324	0,676	1
Mujeres	363	1665	2028	Mujeres	0,179	0,821	1
Total	970	2935	3905	TOTAL	0,248	0,752	1

TABLA PA.E.1.: Relaciones entre Variables Edad y Prensa (Leyó Ayer)
(% Horizontales)

<u>Edad</u>	<u>Leyó</u>	<u>No leyó</u>	<u>Total</u>	<u>Edad</u>	<u>Edad</u>	<u>No leyó</u>	<u>Total</u>
-25 Años	206	745	951	25 Años	0,217	0,783	1
25/44 Años	394	925	1319	25/44 Años	0,299	0,701	1
45 ó + Años	370	1264	1634	45 ó + Años	0,226	0,774	1
Total	970	2935	3905	TOTAL	0,248	0,752	1

TABLA PA.F.1.: Relaciones entre Variables Formación y Prensa (Leyó Ayer)
(% Horizontales)

<u>Formación</u>	<u>Leyó</u>	<u>No leyó</u>	<u>Total</u>	<u>Formación</u>	<u>Leyó</u>	<u>No leyó</u>	<u>Total</u>
Baja	301	1666	1967	Baja	0,153	0,847	1
Media	447	1067	1514	Media	0,295	0,705	1
Alta	222	201	423	Alta	0,525	0,475	1
Total	970	2935	3905	Total	0,248	0,752	1

TABLA PA.RF.1.: Relaciones entre Variables Rol Familiar y Prensa (Leyó Ay)
(% Horizontales)

<u>Rol fam.</u>	<u>Leyó</u>	<u>No leyó</u>	<u>Total</u>	<u>Rol fam.</u>	<u>Leyó</u>	<u>No leyó</u>	<u>Total</u>
Aporta In	450	920	1370	Aporta In	0,328	0,672	1
Enc.Hogar	184	1073	1257	Enc.Hogar	0,146	0,854	1
Otras Sit	336	942	1278	Otras Sit	0,263	0,737	1
Total	970	2935	3905	TOTAL	0,248	0,752	1

TABLA PA.O.1.: Relaciones entre Variables Ocupación y Prensa (Leyó Ayer) (Frecuencias)
(% Horizontales)

<u>Ocupación</u>	<u>Leyó</u>	<u>No leyó</u>	<u>Total</u>	<u>Ocupación</u>	<u>Leyó</u>	<u>No leyó</u>	<u>Total</u>
Inactivos	329	1571	1900	Inactivos	0,173	0,827	1
Trab.Manu	252	849	1101	Trab.Manu	0,229	0,771	1
T+Pf. C/A	252	319	571	T+Pf. C/A	0,441	0,559	1
E+Pf. C/P	137	193	330	E+Pf. C/P	0,415	0,585	1
Total	970	2932	3902	TOTAL	0,249	0,751	1

3.3. Cálculo de los estadísticos

Para nuestro caso, estamos considerando el sistema formado por las seis variables que mejor predicen la lectura.

En primer lugar estudiaremos la variable Lectura Ayer en función de las variables Sexo, Edad, Rol Familiar, Hábitat, Formación y Ocupación (influencia conjunta de una serie de variables independientes o predictoras sobre otra dependiente o predicha).

Después, una vez encontrado el ‘camino’ que nos produce más ‘beneficios’ predictivos, limitaremos el ejemplo a las variables incluidas en el sistema limitado por este ‘camino’.

3.3.1. Coeficiente de Predecibilidad

Tras contrastar que no todos los individuos leen prensa, queremos ver la influencia de las distintas variables en este comportamiento. Para ello tenemos que desarrollar,

1) Un estadístico que nos permita medir la influencia de las variables, y

2) Un procedimiento para seleccionar las variables que mejor explican la lectura de prensa.

El criterio que vamos a seguir para decidir cuándo una variable influye en otra es mirar la 'ganancia' que se obtiene en la predicción de un individuo en la variable dependiente antes y después de conocer su clasificación en la variable antecedente.

Tomemos los valores totales de cualquiera de las tablas anteriores. Según se desprende de ellas, la probabilidad de acertar si un individuo es lector de prensa es del 24,9% ó 0,249.

Si ahora construimos las tablas posibles con dos variables antecedentes respecto a la lectura de prensa «Ayer», podremos ver cuál es la 'ganancia' (capacidad de predicción) que se obtiene con el añadido de cada una de ellas. Para no desbordar las 'medidas' de un artículo, omito reflejar todas las tablas posibles, aunque no los resultados que se obtienen. Así, con la variable Sexo –es decir, si sabemos cuál es el sexo de un individuo– podemos alcanzar una probabilidad del 0,324 en la predicción si se trata de un hombre. Cuando, además, añadimos la variable Ocupación –si sabemos en qué trabaja– nuestra capacidad de predecir que lee prensa se incrementa hasta al 0,466 (46,6%) si se trata de un hombre cuya ocupación es de Técnico, Profesional o Empresario sin asalariados.

Pues bien, este número –el porcentaje de la categoría de la variable predicha para la combinación más favorable de categorías de las variables predictoras (esto es, la combinación que arroja un valor más alto)–, es el que tomamos como medida de la capacidad de predicción de un modelo.

Con el fin de utilizar una medida que tenga en cuenta los distintos marginales (totales de filas y columnas) de las variables predichas, podemos calcular una medida estándar de predicción, P_{\cdot} .

Si llamamos impredecibilidad a la diferencia entre 1 y el marginal (valor del conjunto) de la categoría de la variable que tratamos de predecir, es posible comparar nuestra ganancia en capacidad de predicción –la diferencia que obtenemos con la combinación de las categorías de las variables y el original– con respecto a este valor.

En nuestro ejemplo de la Tabla PA.S.1, tenemos:

$$-\text{Impredecibilidad} = 1,000 - 0,248 = 0,752.$$

Por lo tanto, el coeficiente de Predecibilidad se calcula,

$$-\text{Coeficiente de predecibilidad (P_)} =$$

$$[(0,324 - 0,248) / 0,752] * 100 = 10,1\%$$

Este valor se puede interpretar como el tanto por ciento de impredecibilidad explicado por las variables del modelo. Es una medida de la capacidad de predicción de una variable o combinación de variables, ponderando en función del número de elementos de cada categoría (varía entre 0 y 100).

Este coeficiente también nos permite predecir, por supuesto, la otra categoría de la variable –la No lectura– y, en este caso, con mayor capacidad de certeza. Igualmente, se puede utilizar para comparar el efecto de añadir nuevas variables que predigan la lectura de un determinado soporte, etc.

Con relación al tema de la selección de las variables, a menos que haya un criterio particular para seleccionar unas u otras variables (en cuyo caso seleccionaremos las que nos parezcan más convenientes, cosa que ya hemos hecho en cierta medida), a cada paso seleccionaremos aquellas variables que nos ofrezcan un mayor Coeficiente de Predecibilidad, P_.

A continuación, realizamos los cálculos para algunas de las combinaciones posibles (las que hemos seleccionado previamente en base al Análisis de Residuos):

$$\text{Coeficiente de Impredecibilidad} = (1 - 0,248) = 0,752$$

Coeficientes de Predecibilidad:

Ocupación=	[(,441–,249)/,752]=	0,255 (T+Pf Ct/A)
Formación=	[(,525–,249)/,752]=	0,368 (Alta)
Rol Fam.=	[(,328–,249)/,0752]=	0,106 (Aporta Ing)
Edad=	[(,299–,249)/,0752]=	0,068 (24–44)
Sexo=	[(,324–,249)/,0752]=	0,101 (Hombre)
Hábitat=	[(,337–,249)/,0752]=	0,118 (Urbano)
Edad/Rol=	[(0,409–0,248)/0,752]=	0,214 (25–44/Ot.Sit)
Edad/Ocp=	[(0,469–0,248)/0,752]=	0,294 (+45/T.+Pf Ct/A)
Sexo/Ocp=	[(0,466–0,248)/0,752]=	0,290 (Hombre/E+Pf Ct/P)
Form/Ocp=	[(0,710–0,0,248)/0,752]=	0,614 (Alta/E+Pf Ct/P)
Hábitat/Form/Ocup=	[(0,844–0,249)/0,70 =	0,793 (Urbano/Alta/Ct/P)

[en este caso hemos sumado las dos últimas categorías de ocupación al tener la última muy pocos elementos]

Edad/Form/Ocup=

$$[(0,955-0,248)/0,752]= 0,610 (+45/Alta/T+Pf+Emp)$$

De acuerdo con estos datos, la variable que mejor predice la lectura de prensa es la Formación. Con dos variables, la mejor combinación es Formación/Ocupación. Y con tres variables tenemos la combinación Hábitat/Formación/Ocupación.

3.3.2. Cálculo de las d ponderadas

A continuación tenemos que

- 1) Especificar el modelo determinado por las variables,
- 2) Estimar sus parámetros, y
- 3) sacar las conclusiones pertinentes (deducir las propiedades del sistema).

Para especificar el modelo hemos de incluir todas las variables importantes, indicando la dirección de sus influencias.

Para los objetivos de este artículo vamos a considerar las variables Formación, Ocupación y Leyó Ayer, procediendo a recodificar Ocupación en dos categorías: (Técnicos, Profesionales y Empresarios por Cta. Propia) y (Inactivos y Trabajadores por cuenta ajena). De esta manera, vamos a comparar los individuos con estudios superiores (categoría Formación Alta) frente a los que tienen estudios inferiores a los universitarios (Formación Media-Baja) y los que tienen una profesión de tipo técnico o profesional tanto por cuenta ajena como por cuenta propia, frente a los inactivos y trabajadores manuales, estén o no cualificados.

Es conveniente volver a repetir en este momento, que esta operación puede realizarse agrupando las categorías de otra manera que más convenga a los objetivos del análisis que estemos realizando.

En este momento, tendríamos que construir el sistema de ecuaciones o grafos y establecer las variables antecedentes, consecuentes, intermedias, etc.

En nuestro caso, siguiendo las directrices aplicables construimos el sistema que se expone a continuación:

$$O(\text{Alta})=F*f_0+K_0$$

$$A=F*f_1+O*o_1+k_1$$

donde

O(Alta)= Ocupación de nivel alto.

F= Formación

f_0 = efecto directo (coeficiente) de Formación sobre Ocupación.

K_0 = constante (efectos no controlados) sobre Ocupación.

A= Leyó Ayer

f_1 = efecto directo (coeficiente) de Formación sobre Leyó Ayer.

o_1 = efecto directo (coeficiente) de Ocupación sobre Leyó Ayer.

K_1 = constante (efectos no controlados) sobre Leyó Ayer.

El sentido de la causalidad consideramos que sigue el mismo orden en el que se exponen las variables, esto es, Formación se configura como variable antecedente puesto que es previa a la Ocupación. Esta última, Ocupación, sigue a Formación y viene determinada por ella, mientras Leyó Ayer consideramos que es la variable dependiente puesto que no interviene en las anteriores y es la que tratamos de explicar en su comportamiento.

Para estimar los parámetros del modelo habrá que calcular los efectos, las constantes y los valores de las variables.

A la hora de calcular el efecto entre dos variables vamos a seguir la regla de controlar por todas las variables antecedentes e intervinientes. Ello quiere decir que:

–Para ver el efecto de Formación sobre Ocupación, calcularemos la diferencia de proporciones, puesto que, al haber aislado las variables del ‘camino’ (con fines didácticos) no tenemos en cuenta otras variables antecedentes ni intervinientes.

–Para ver el efecto de Formación sobre Leyó Ayer, es necesario calcular la diferencia de proporciones controlando por la variable interviniente Ocupación.

–Para calcular el efecto de Ocupación sobre Leyó Ayer, calcularemos la diferencia de proporciones controlando por la variable antecedente Formación.

Como resultado del cálculo de los efectos puede ocurrir una situación especial que rompe los supuestos de aditividad y linealidad subyacentes a los sistemas que estamos considerando. En el caso de que haya una inte-

racción importante entre las variables, en opinión de Sánchez Carrión (1989) es que se ha descubierto una información de interés, sin que sea necesario continuar con el análisis.

Cuando la interacción no sea estadísticamente significativa o, siéndolo, tenga un tamaño reducido, se hará necesario calcular una diferencia única, resumen de las diferencias parciales correspondientes a las categorías de la(s) variable(s) de control.

Una vez calculados los parámetros, será necesario deducir las propiedades del sistema. A tal fin es útil:

1) Analizar los efectos de unas variables sobre otras (cálculo de los caminos), y

2) Descomponer la asociación entre las variables en sus componentes principales (análisis de los efectos).

Comencemos a realizar los cálculos:

Para calcular el efecto de Formación sobre Ocupación, como no hay variables antecedentes ni intervinientes, obtenemos de la tabla-base los valores, lo que nos da los siguientes resultados:

TABLA F.O.2.:

Efecto directo de Formación (Alta) s/Ocupación(T+Pf+Emp)

% OCUPACION

FORMACION	(T+Pf+E)	Total
Alta	0,643 (272)	(423)
Media-Baja	0,181 (629)	(3481)
Total	0,231 (901)	(3904)

$$f_0 = 0.643 - 0.181 = 0.462$$

Tomando Formación alta y Ocupación alta como categorías de comparación, vemos que el efecto de Formación sobre Ocupación es igual a 0,462 o, lo que es lo mismo, los individuos con formación universitaria tienen el 46,2% más de probabilidades de obtener un empleo de nivel alto que aquellos otros que no han llegado a la Universidad. O, también, que hay un 46,2% más de individuos con formación universitaria ocupando puestos de nivel alto que de individuos con formación no universitaria.

A la hora de calcular el efecto de Formación sobre Leyó Ayer hay que controlar por la variable interviniente Ocupación. Para ello, primeramente, calculamos las diferencias de proporciones para cada una de las categorías de la variable Ocupación –a estas diferencias las llamamos *parciales*– para mostrar, a continuación, cómo se procede a la hora de resumir los parciales en un solo número, al que llamamos *diferencia de proporciones ponderada*.

Elaboramos los datos de las tablas–base y podemos construir la Tabla PA.F.O.2., de la cual podemos deducir la existencia de una cierta interacción que, más adelante, veremos si es significativa. De acuerdo con los resultados obtenidos, el efecto de Formación sobre Lectura de prensa Ayer es algo mayor entre los individuos Inactivos y Trabajadores Manuales que entre los que ocupan mayores niveles ocupacionales (0,245 de los primeros frente al 0,214 de los segundos), esta situación nos sugiere una influencia de la formación en la lectura de prensa bastante significativa, a falta de comprobación estadística, lo cual nos pone sobreaviso respecto a la existencia de relaciones de causalidad e, incluso, relaciones ocultas que más adelante trataremos de determinar.

Por el momento, si comprobamos que la interacción no es estadísticamente significativa o que su valor no es elevado, es conveniente obtener una sola diferencia de proporciones que sustituya a las diferencias parciales.

TABLA PA.F.O.2.: Relación entre Formación y P.Ayer controlando por Ocupación.

FORMACION	OCUPACION (T+Pf+Emp)		
	Leyó	No leyó	Total
Alta	0,581 (158)	0,419 (114)	1'– (272)
Media–Baja	0,367 (231)	0,633 (398)	1'– (629)
Total	0,432 (389)	0,568 (512)	1'– (901)

$$d(0.581-0.367)=0,214$$

FORMACION	OCUPACION (Inac+Man)		
	Leyó	No leyó	Total
Alta	0,424 (64)	0,576 (87)	1'- (151)
Media-Baja	0,179 (507)	0,821 (2332)	1'- (2839)
Total	0,191 (571)	0,809 (2419)	1'- (2990)

$$d (0.424-0.179)= 0,245$$

Para calcular la diferencia de proporciones ponderada, hemos de tener en cuenta el tamaño de la muestra y la variabilidad de cada diferencia parcial. Si calculásemos una media simple, estaríamos dando igual importancia a los 901 sujetos que componen la muestra de arriba frente a los 2990 de la de abajo. Igualmente, hemos de tener en cuenta cuál es la variabilidad de cada distribución para dar mayor importancia a los datos más homogéneos frente a los más dispersos.

Para calcular la diferencia ponderada (d') aplicaremos, por tanto, las reglas siguientes:

-Cuanto mayor sea el tamaño de la muestra de la categoría de la variable de control, mayor será el peso de sus diferencias de proporciones correspondientes.

-Cuanto mayor sea la variabilidad (la varianza muestral) de la diferencia obtenida para una categoría, menor será su peso.

De acuerdo con las reglas anteriores, calcularemos la d' ponderando inversamente la variabilidad de las d parciales:

$$d' = Sd_k * w_k$$

siendo K el número de categorías de las variables de control y W el peso atribuido a cada parcial.

En la Tabla PA.F.O.2.1. se muestra la disposición de los cálculos para la d' de Formación/Leyó Ayer controlando por Ocupación:

TABLA PA.F.O.2.1.: Cálculo de la d' ponderada de Formación-P.Ayer controlando por Ocupación

d(k)	Varianza		W(k)	W(k)*d(k)
	MtralV(k)	1/V(k)		
0,214	0,00126	791	0,569	0,122
0,245	0,00167	599	0,431	0,106
Total		1390	1	0,227

La d' estimada corresponde al coeficiente f_j , e indica el efecto medio ponderado de Formación sobre Leyó Ayer, controlando por Ocupación. Esto se puede interpretar diciendo que, si todos los individuos tuvieran la misma Ocupación, entre los individuos de formación 'alta' habría un 22,7% más de lectores de prensa que entre los de formación media-baja.

D) Para calcular el efecto de Ocupación sobre Leyó Ayer hay que controlar por la variable antecedente Formación. Para ello, actuamos como en el punto anterior, esto es, primeramente, calculamos las diferencias de proporciones para cada una de las categorías de la variable Formación –a estas diferencias las llamamos *parciales*– para calcular a continuación la d' diferencia de proporciones ponderada. Los cálculos son análogos a los realizados anteriormente:

TABLA PA.F.O.3.:

Relación entre Ocupación y P.Ayer controlando por Formación

OCUPACION	FORMACION (ALTA)		Total
	Leyó	No leyó	
(T+Pf+Emp)	0,581 (158)	0,419 (114)	1'– (272)
(Inac+Man)	0,424 (64)	0,576 (87)	1'– (151)
Total	0,525 (222)	0,475 (201)	1'– (423)

$$d(0.581-0.424)=0,157$$

OCUPACION	FORMACION (Media-Baja)		Total
	Leyó	No leyó	
(T+Pf+Emp)	0,367 (231)	0,633 (398)	1'– (629)
(Inac+Man)	0,179 (507)	0,821 (2332)	1'– (2839)
Total	0,213 (738)	0,787 (2730)	1'– (3468)

$$d(0.367-0.179)=0,189$$

En este caso, el efecto de Ocupación sobre lectura es más alto entre los individuos con formación media-baja que entre los universitarios.

Si todos los individuos tuvieran el mismo nivel de formación, el efecto de Ocupación sobre lectura sería más alto entre las personas con media-baja formación que entre las que tienen estudios universitarios. Este dato, al margen de las elaboraciones que estamos desarrollando, nos indica la existencia de un factor de status social (¿símbolo de status?) en la lectura de periódicos a tener en cuenta en el momento de realizar el pertinente análisis de resultados.

Otro hecho a tener en cuenta es que si hallamos las diferencias de las d parciales de los dos últimas tablas [(0,157-0,189=-0,032) y (0,214-0,245=-0,031)] obtenemos un mismo valor -las diferencias se deben a los redondeos-, es decir, que la interacción que obtenemos al ver la relación de Ocupación sobre lectura, controlando por Formación, es la misma que se obtiene al calcular la relación de Formación sobre lectura, controlando por Ocupación. Ello significa que la interacción entre tres variables es siempre la misma, no importa en qué forma se calculen las relaciones y se efectúen los controles (Guerrero Serón, 1992).

A partir de las d parciales, repitiendo el procedimiento seguido en la Tabla PA.F.O.2.1., obtenemos la d' correspondiente al valor de $-o_1-$:

TABLA PA.F.O.3.1.:

Cálculo de la d' ponderada de Ocupación-P.Ayer controlando por Formación

$d(k)$	$V(k)$	$1/V(k)$	$W(k)$	$W(k)*d(k)$
0,157	0,00251	398	0,144	0,023
0,189	0,00042	2375	0,856	0,162
Total		2773	1'-	0,184

La d' estimada corresponde al coeficiente $-o_1-$, e indica el efecto medio ponderado de Ocupación sobre Leyó Ayer, controlando por Formación. Si todos los individuos tuvieran la misma Formación, entre los individuos de Ocupación 'alta' habría un 18,4% más de lectores de prensa que entre los Inactivos y Trabajadores Manuales. Parece ser, por lo tanto, que el nivel ocupacional es en sí mismo un factor importante

para la lectura de prensa, que se potencia con la formación. En efecto, anteriormente vimos que la influencia de la 'categoría' laboral es más importante para la lectura de prensa entre los de formación media-baja; ahora, este resultado se complementa cuando vemos que, a igualdad de formación, se lee más conforme se asciende en la categoría laboral.

Con los resultados obtenidos estamos en condiciones de construir el grafo –o ecuación– definitivo de la relación entre las 3 variables (en nuestro caso lo teníamos hecho anteriormente por las razones aducidas). Aplazamos su interpretación hasta que hayamos visto la significatividad de los estadísticos.

3.4. Significatividad de los estadísticos y estimación de los parámetros poblacionales

De igual manera que cuando analizábamos la relación entre dos variables, ahora con 3 variables también hay que determinar cuales son los parámetros poblacionales para ver si son significativamente diferentes de cero. Además, cuando se utilizan 3 o más variables hay que añadir un nuevo contraste que nos permita ver si la interacción es estadísticamente significativa.

Comencemos por este contraste:

En la Tabla PA.F.O.3.2. se disponen los cálculos con el fin de ver:

1) si la diferencia de proporciones de Formación-Leyó Ayer para cada categoría de Ocupación es significativa –en cuyo caso diríamos que 'hay interacción' entre las 3 variables.

2) Al mismo tiempo. se calcula un intervalo de confianza para la d' del efecto de Formación sobre Leyó Ayer, a partir del cual, se puede ver si la d' es estadísticamente significativa.

Las siete primeras filas son prácticamente una reproducción de las Tablas PA.F.O.2 y PA.F.O.2.1., La fila (8) nos da la respuesta a la pregunta sobre la interacción. En ella se resta cada diferencia parcial de la diferencia ponderada, el resultado se eleva al cuadrado y se divide por la varianza de la d parcial. Según Goodman (1963) el estadístico que se obtiene en esta operación tiene una distribución G_i -cuadrado con $K-1$ grados de libertad –siendo K el número de categorías de la(s) variable(s) de control (Sánchez Carrión, 1989).

Veamos los resultados:

TABLA PA.F.O.2.2.: Efecto de Formación s/P.Ayer controlando por Ocupación. Estimación «d'», significatividad e interacción de las tres variables (fila 8)

Formación	Ocupación	TOTAL	
	(T+Pf+E)	(I+Man.)	
(1) Alta	0,581	0,424	0,109
	272	151	423
(2) Media-Baja	0,367	0,179	0,891
	629	2839	3468
(3) d_h	0,214	0,245	
(4) V_k	0,001264	0,001668	
(5) $1/V_k$	791	599	1390
(6) peso W_k	0,569	0,431	1,000
(7) $d=S[W_k*d_k]$	0,122	0,106	0,227
(8) $(d_k-d)/V_k$	6,707	11,666	18,373
(9) V_k*W_k	0,000409	0,000310	0,000719

El valor que resulta de (8) es superior a 3,841 que se obtiene en las Tablas Gi-cuadrado para un grado de libertad y un nivel de significación del 5%, por lo que la interacción que se observa 'si' es significativa.

Para estimar la D poblacional de Formación sobre Leyó Ayer, o para contrastar si es significativamente diferente de cero, tenemos que calcular la desviación típica muestral de la d obtenida (la d' ponderada). Con el fin de tener en cuenta el tamaño de la muestra sobre la que se calcula cada d parcial y su variabilidad, se calcula una varianza —o su raíz cuadrada la desviación típica— que pondere cada d parcial inversamente proporcional a su varianza. La fila (9) muestra los cálculos y el resultado, 0,000719.

De acuerdo con el resultado obtenido en la fila (9), podemos calcular un intervalo de confianza del 95% para la d' de Formación sobre Leyó Ayer, controlando por Ocupación. Los cálculos se muestran abajo:

$$\begin{aligned} & \text{Intervalo de Confianza para la } d' \text{ ponderada de} \\ & \text{Formación s/P.Ayer, control.por Ocupación} \\ & D = d \pm (z * Sd) = 0,227 \pm (1,96 * 0,000719) \\ & D = d + (z * Sd) = 0,227 + 0,001408 = 0,228408 \\ & D = d - (z * Sd) = 0,227 - 0,001408 = 0,225592 \end{aligned}$$

El intervalo estará comprendido entre 0,280 y 0,175; puestos en términos porcentuales, entre el 28% y el 17,5%. Puesto que el valor 0 no está comprendido en el intervalo, concluimos que el efecto de Formación sobre Leyó Ayer, controlado por Ocupación, es estadísticamente significativo.

Para estimar el intervalo de confianza de Ocupación sobre Leyó Ayer, controlando por Formación, realizaríamos las mismas operaciones, como se muestra en la Tabla PA.F.O.3.2.

TABLA PA.F.O.3.2.:Efecto de Ocupación s/P.Ayer controlando por Formación. Estimación «d'», significatividad e interacción de las tres variables (fila 8)

Ocupación	Formación		TOTAL
	Alta	Media-Baja	
(1) (T+Pf+E)	0,581	0,367	0,232
	272	629	901
(2) (Inac+Man)	0,424	0,179	0,768
	151	2839	2990
(3) d_k	0,157	0,189	
(4) V_k	0,002512	0,000421	
(5) $1/V_k$	398	2375	2773
(6) peso W_k	0,144	0,856	1,000
(7) $d=S[W_k*d_k]$	0,023	0,162	0,184
(8) $(d_k-d)/V_k$	7,200	1,742	8,942
(9) V_k*W_k	0,000051	0,000308	0,000360

Intervalo de Confianza para la d'ponderada de Ocupación s/P.Ayer, controlando por Formación

$$D=d\pm(z*Sd)=0.184\pm(1.96*![(9)])$$

$$D=d+(z*Sd)=0.184+ 0,037222 = 0,221$$

$$D=d-(z*Sd)=0.184-0,037222=0,147$$

En el caso de la Variable Ocupación sobre Leyó Ayer, controlando por Formación, obtenemos un intervalo de confianza del 95% entre 0,221 y 0,147, por lo que concluimos que el efecto, al igual que en el caso anterior, es significativo.

Nos queda calcular el intervalo de confianza para el efecto directo de Formación sobre Ocupación. En la Tabla F.O.2.1. se exponen los resultados:

TABLA F.O.2.1.: Cálculo de la Varianza Muestral del efecto de Formación s/Ocupación

FORMACION	OCUP (T+Pf+E)	Varianza
Alta	0,643 423	0,000542
Media-Baja	0,181 3481	0,000042
diferencia	0,462	0,000585

$$D = d \pm (z * Sd) = 0,462 \pm (1,96 * \sqrt{0,000585}) =$$

$$D = 0,462 + 0,047413 = 0,510$$

$$D = 0,462 - 0,047413 = 0,415$$

El valor de f_0 estará comprendido entre 0,510 y 0,415. Ahora se puede representar la ecuación completa considerando los intervalos de confianza recién calculados, pasando ya a la interpretación de los resultados obtenidos.

3.5. Interpretación de los resultados

Después de obtener los parámetros de nuestro modelo, estamos en condiciones de dar respuesta a los objetivos del trabajo.

a) En primer lugar, ya hemos visto cómo las tres variables (Hábitat, Formación y Ocupación) explicaban el 79,2% de la impredecibilidad de la lectura Ayer. La mayor probabilidad de leer prensa Ayer se daba entre los individuos residentes en hábitats mayores de 50.000 habitantes, de formación universitaria y con ocupación en el sector terciario como técnicos o profesionales por cuenta ajena o por cuenta propia.

En el supuesto de tener que seleccionar las variables por orden de importancia –son importantes las variables que permiten predecir el comportamiento de los individuos–, elegiríamos primero la variable Formación (categoría Alta), después, seguida de Ocupación (categoría Técnicos, Profesionales y Empresarios) y, en último lugar, Hábitat (categoría Urbano).

b) También hemos comprobado que existe una interacción significativa entre las tres variables, aunque no pueda decirse que esta interacción sea muy fuerte.

c) Para ver por qué los individuos con Formación Alta leen más prensa que los de formación Media-Baja hemos calculado el efecto de Formación sobre Leyó Ayer, controlando por Ocupación. De esta manera, el efecto bruto o total queda descompuesto en dos partes:

TABLA PA.F.O.4.
Descomposición del efecto bruto o total entre Formación y Leyó Ayer

	Efectos Formación-Leyó Ayer
directo	0,227
indirecto	$(0,462 \cdot 0,184) = 0,085$
TOTAL	$0,227 + 0,085 = 0,312$

Existe un efecto directo de Formación sobre Leyó Ayer, cuyo valor es 0,227 puntos. Este efecto se interpreta como la influencia de Formación sobre Leyó Ayer, independientemente de cuál sea la Ocupación del entrevistado. Aún cuando dos individuos tuvieran la misma Ocupación, el de estudios universitarios sería 0,227 puntos más probable que leyera prensa que el de formación no universitaria. Se podría decir que es la parte ‘académica’ de la influencia –aquello que hay en la formación, diferente al hecho de que permita ocupar mejores puestos de trabajo, que hace que los individuos lean más la prensa. Porcentualmente, podríamos decir que, si hacemos 100 la influencia total, el 72,7% de esta influencia es directa $(0,227/0,312 \cdot 100)$.

Junto a este efecto directo hay otro indirecto, vía Ocupación: los individuos con estudios universitarios acceden a puestos de trabajo de mayor nivel profesional y los individuos con mayor nivel profesional leen más prensa que los Inactivos y Trabajadores manuales. Este efecto se calcula multiplicando los efectos que hay en el camino entre Formación y Leyó Ayer y su valor se eleva a 0,085 puntos, lo que, en términos porcentuales, supone que un 27,2% de la influencia de la Formación sobre la lectura de prensa es indirecta.

Hablando en términos de sistemas, puesto que el signo de los efectos entre Formación y lectura es el mismo, diremos que estamos en presencia de un sistema consistente o reforzador, cuyo efecto bruto (0,312) es mayor que su efecto neto (0,227).

d) Mirando la relación entre Formación y Leyó Ayer, controlando por Ocupación se puede comprobar que no hay efectos ocultos –situación lógica desde el momento que el sistema es consistente– siendo su efecto total mayor que el efecto neto. (Es posible que existan efectos ocultos en otras variables del sistema completo, que no hemos incluido en este artículo por razones de espacio)

e) Para ver la naturaleza de la relación entre Ocupación y Leyó Ayer tenemos que controlar por Formación. Recordemos que,

–una relación es causal cuando su valor se mantiene al controlar por la tercera variable;

–es espuria (falsa, estadística) cuando desaparece.

La tabla PA.O.F.4. muestra la descomposición del efecto total entre Ocupación y Leyó Ayer cuando introducimos la Formación.

TABLA PA.O.F.4.
Descomposición del efecto bruto o total entre Ocupación y Leyó Ayer.

<u>Efectos Ocupación–Leyó Ayer</u>	
Causal	
directo	0,184
indirecto	
Espurio	(0,462*0,227)= 0,105
TOTAL	0,184+0,105 = 0,289

Según estos datos, porcentualmente, el 36,3% de la influencia original de la Ocupación sobre la lectura es espuria, atribuible al efecto conjunto de Formación sobre estas dos variables. Ello significa que si se incrementara el colectivo de Técnicos, Profesionales y Empresarios en un determinado porcentaje, la cantidad de lectores de prensa se incrementaría en 0,184 puntos en vez de los 0,289 puntos como parecía indicar la relación original.

Para construir un sistema de ecuaciones, sin embargo, nos queda

miten ver la información subyacente a la tabla de 3 dimensiones considerada en este ejemplo.

4. Una aplicación real

A partir de este sistema de ecuaciones, puede estimarse, por ejemplo, la lectura de prensa 'Ayer' para las diversas provincias andaluzas, o en una zona concreta, etc. ... con solo saber que proporción de universitarios y de técnicos, profesionales y empresarios hay en cada zona en estudio.

Veamos una aplicación real: (Si el lector dispone del «II Estudio sobre Hábitos de Exposición a Medios en Andalucía: Provincia de Sevilla» podrá comprobar los resultados).

En la provincia de Sevilla, el porcentaje de personas con formación universitaria es de, aproximadamente, el 12% y el de y el de Empresarios, Profesionales y Técnicos, alrededor del 25%. Si sustituimos estos valores en la ecuación anterior, tenemos:

$$\text{Ocupación (T+Pf+E)} = (0,12*0,462)+0,367=0,4213$$

$$\text{Leyó Ayer} = (0,227*0,12)+(0,184*0,421)+0,179=\backslash$$

$$\backslash = (0,027)+(0,077)+0,179= 0,2832, \text{ esto es, el } 28,32\%.$$

Comprobamos que en el «II Estudio sobre Hábitos de Exposición a Medios» el porcentaje de lectores 'Ayer' es del 29,1% con lo que hemos obtenido una desviación de la proyección sobre la cifra real del 0,78%. Si tenemos en cuenta el mayor error que soportan los datos al tratarse de sola una provincia más el error que llevan de por sí los parámetros de Ocupación y Formación (Alta) calculados para la provincia, hemos de concluir que el sistema nos ha proporcionado una buena predicción. Animo al lector a que contraste la bondad del sistema si puede disponer de datos sobre las variables antecedentes de otra provincia o zona de Andalucía.

REFERENCIAS BIBLIOGRAFICAS

- ABASCAL, E., y GRANDE, I. (1989): *Métodos multivariantes para la investigación comercial*. Barcelona, Ariel.
- GUERRERO SERON, C. (1992): *II Estudio sobre hábitos de audiencia en Andalucía* (Difusión restringida a profesionales de la publicidad y anunciantes).
- LOPEZ PINTOR, R. (1989): «El análisis de los datos de una encuesta», en M. GARCIA FERRANDO *et al*: *El análisis de la realidad social*. Madrid, Alianza.
- MARTINEZ RAMOS, E. (1992): *El uso de los medios de comunicación en marketing y publicidad*. Madrid, Akal.
- PIÑUEL RAIGADA, J. L. (1987): *El consumo cultural*. Madrid, Fundamentos.
- SANCHEZ CARRION, J. J. (1988): *Análisis de tablas de contingencia*. Madrid, CIS.
- SOLER, P., y PERDIGUER, A. (1992): *Prácticas de investigación de mercado*. Bilbao, Deusto.
- VIDOSA GONZALEZ, J. (1990): «Análisis factorial», en E. ORTEGA MARTINEZ: *Manual de investigación comercial*. Madrid, Pirámide.