

La interpretación de las relaciones estadísticas como propiedad de investigación. El rol de las variables-test

Paul Lazarsfeld

Comunicación presentada en el congreso de la Sociedad Americana de Sociología de Cleveland (1946). Los principales puntos de esta comunicación han sido recogidos en artículos posteriores. Publicado en Lazarsfeld y Rosenberg (eds.): The language of social research, pp. 115-125

El punto de partida de la presente discusión es un procedimiento cuya aplicación es casi automática en la investigación empírica; cuando se ha constatado la presencia de una relación entre dos variables, generalmente se intenta analizar el rol de las variables suplementarias. Este procedimiento puede ser concretizado por un conjunto de datos que provienen, con una forma un poco estilizada, de una serie de estudios sobre las preferencias en materia de emisiones radiofónicas.

Si relacionamos la edad del informador y el tipo de emisión que escucha regularmente, se observa que las personas de mayor edad buscan preferentemente los programas religiosos y políticos. Por el contrario, no se observa ninguna diferencia entre los grupos de edad con respecto a las emisiones de música clásica.

Cuadro 1
Proporciones de oyentes pertenecientes a dos grupos de edad

	Jóvenes % de oyentes	Viejos % de oyentes
Programas religiosos	17%	26%
Tribunas políticas	34%	45%
Programas de música clásica	30%	29%

Todo el mundo sabe que la edad está ligada al nivel cultural; en efecto, a causa de la extensión creciente de la enseñanza, se observa generalmente en una comunidad un nivel cultural más elevado entre los jóvenes que entre los viejos. La relación entre edad e instrucción observada en nuestra población figura en el Cuadro 2 (para la variable instrucción, la clasificación está determinada por el diploma de fin de estudios secundarios; para la variable edad, dividimos en 40 años).

Cuadro 2
Relación entre edad y nivel de instrucción

	Jóvenes	Viejos	Total
Nivel de instrucción superior	600	400	1.000
Nivel de instrucción inferior	400	900	1.300
Total	1.000	1.300	2.300

Así, pues, se obtienen razones sobre tres variables: edad, instrucción y emisiones preferidas. Para simplificar, hemos reducido cada variable a una dicotomía. El nivel de instrucción introducido para elaborar y clarificar la relación original se llama *variable-test* (t). La edad corresponde a lo que convencionalmente llamamos *variable independiente* (x), y las preferencias en materia de emisiones radiofónicas a la *variable dependiente* (y). A veces, por razones de brevedad, utilizaremos los símbolos x, y, t, en el sentido indicado en la frase anterior. Aparte de este simbolismo, en esta exposición renunciamos al uso de las matemáticas.

Vemos fácilmente que, entre las tres variables anteriores, se pueden definir tres relaciones. La primera relaciona la edad y las preferencias (xy). La información correspondiente ha sido presentada en el primer cuadro para cada uno de los tres tipos de programas. En segundo lugar, tenemos una relación entre la edad y la variable-test, es decir, el nivel de instrucción (xt). Esta relación evidentemente es la misma para los tres tipos de emisión. La información correspondiente se presenta en el Cuadro 2: las cifras han sido redondeadas, pero reproducen datos reales. En último lugar tenemos la relación (ty), es decir la relación entre el nivel de instrucción y las preferencias. Esta relación varía evidentemente con los tres tipos de programas. Más adelante damos las estadísticas correspondientes a esta tercera relación.

En este punto, es importante comprender que el Cuadro 2 en su conjunto y una línea cualquiera del Cuadro 1 dan el mismo tipo de información. El contenido del Cuadro 2 puede resumirse en el resultado siguiente: un 60% de los jóvenes y un 31% de los viejos pertenecen al grupo caracterizado por un nivel de instrucción superior. Inversamente, es posible transformar cada línea del Cuadro 1 en un cuadro con cuatro casos, presentando, para los dos grupos de edad, el número de personas que escuchan o no un determinado tipo de emisión. Utilizaremos los dos tipos de presentación, según las necesidades del contexto.

Se han construido varios coeficientes para medir el grado de ligamen entre dos variables dicotomizadas. La única información que nos interesa en este caso es la de saber si entre las dos variables existe una relación físicamente significativa. Por lo

tanto, siendo el mejor índice el más simple al nivel de generalidad en que nos situamos, utilizaremos el producto cruzado estandarizado. Para el Cuadro 2, por ejemplo, tenemos:

$$(xt) = \frac{600 \times 900 - 400 \times 400}{2300^2} = 0,08$$

(el valor de este producto es, digámoslo de paso, casi igual a una cuarta parte del valor del coeficiente de correlación r_{xy}). Hay, pues, una relación entre la edad y el nivel de instrucción, porque de otra forma el producto cruzado sería nulo.

El procedimiento que presentamos tiene, pues, como punto de partida una relación primaria (r_{xy}), a partir de la cual se introduce una variable-test, que hace posible la formación de dos nuevas relaciones (r_{xt} y r_{ty}). Pero los resultados más importantes obtenidos a partir de las variable-test son dos *relaciones condicionales*. En efecto, nos podemos plantear la siguiente pregunta: si se estudian por separado las personas con un nivel de instrucción homogéneo, inferior o superior, ¿qué sucede con la relación entre edad y preferencias radiofónicas? La respuesta nos la da el Cuadro 3.

Cuadro 3
Relación entre edad e interés por emisiones religiosas en función de la instrucción

	Nivel de instrucción superior			Nivel de instrucción inferior		
	Jóvenes	Viejos	Total	Jóvenes	Viejos	Total
Escuchan	55	45	100	115	285	400
No escuchan	545	355	900	285	615	900
Total	600	400	1000	400	900	1300

El cuadro anterior se refiere a las emisiones religiosas. Para facilitar la comparación con la primera línea del Cuadro 1, hemos transformado a continuación los datos brutos en porcentajes: en el Cuadro 4 vemos que, en cada uno de los dos grupos correspondientes a un nivel de instrucción dado, la relación entre edad e interés por las emisiones religiosas prácticamente ha desaparecido. (La primera línea del Cuadro 4 reproduce la información presentada en el Cuadro 1).

Cuadro 4
Proporción de oyentes que escuchan emisiones religiosas

Jóvenes 17%		Viejos 26%	
Nivel de instrucción superior		Nivel de instrucción inferior	
Jóvenes	Viejos	Jóvenes	Viejos
9%	11%	29%	32%

Es posible reproducir el mismo análisis sobre los otros dos tipos de emisiones. Más abajo presentamos los resultados omitiendo las etapas intermedias. Es importante señalar que la variable-test juega un papel muy diferente según los ejemplos. Presentamos en primer lugar los resultados relativos al interés por las tribunas políticas.

Cuadro 5
Proporción de oyentes que escuchan tribunas políticas

Jóvenes 34%		Viejos 45%	
Nivel de instrucción superior		Nivel de instrucción inferior	
Jóvenes	Viejos	Jóvenes	Viejos
40%	55%	25%	40%

Los datos del Cuadro 1 se reproducen en la primera línea del Cuadro 5. El Cuadro 5 muestra que, en el interior de cada grupo de instrucción, la diferencia debida a la edad es más fuerte que en la muestra total.

En lo que ahora concierne al interés por los programas de música clásica se puede ver en el primer cuadro que la edad no juega ningún papel en este caso. Es de señalar, sin embargo, si nos referimos a la nota al final del capítulo, que las personas pertenecientes al grupo de instrucción superior escuchan con preferencia este tipo de programa y sabemos que los jóvenes tienen, por término medio, un nivel de instrucción superior.

Si se analizan con detalle estas relaciones, se obtiene la estructura relativamente compleja presentada en el Cuadro 6.

Cuadro 6
Proporción de oyentes que escuchan los programas de música clásica

Jóvenes 30%		Viejos 29%	
Nivel de instrucción superior		Nivel de instrucción inferior	
Jóvenes	Viejos	Jóvenes	Viejos
32%	52%	28%	19%

El Cuadro 6 muestra que la edad juega un rol diferente en los dos grupos de instrucción. En un medio más cultivado, la madurez provoca mayor atención a las manifestaciones culturales y, en particular, a la música clásica. En un ambiente más tosco, el interés máximo por estas manifestaciones parece coincidir por el contrario con el momento en que prevalece la influencia escolar; con la edad, el interés cultural parece declinar.

Subrayemos que el interés de los ejemplos precedentes reside no en su contenido sino en su forma. La introducción de la variable edad revela efectos diferentes en cada uno de los tres ejemplos del Cuadro 1: en el caso de los programas religiosos, reduce la diferencia inicial; en el de las tribunas políticas, aumenta por el contrario esta diferencia; en el caso de los programas de música clásica, la introducción de esta variable permite poner en evidencia dos tendencias contrarias que no eran perceptibles en los resultados iniciales.

Esquema general de análisis

La esencia de los tres ejemplos precedentes puede resumirse en una fórmula general. Para comprender su importancia, debemos orientar nuestra atención hacia las que hemos llamado asociaciones condicionales, cuadros condicionales o diferencias condicionales. Van siempre por pares, como se puede ver en el Cuadro 3 y en las segundas líneas de los Cuadros 4 y 5. La relación original (xy) se analiza en dos relaciones condicionales correspondientes a cada uno de los niveles de instrucción. Un simbolismo cómodo para estas dos relaciones es $(xy;t)$ y $(xy;t')$. Su significado es análogo al de la correlación parcial en estadística; sin embargo, es de señalar que esta última es de hecho una media ponderada de nuestras relaciones condicionales. Ahora bien, lo que aquí importa es precisamente el hecho de que puedan ser separadas y observar, en algunos casos, relaciones de naturaleza distinta para cada valor de la variable-test.

La estructura correspondiente, cuyos elementos están constituidos por dos variables

originales y una variable-test, puede ser formulada de la siguiente manera:

$$(1) \quad (xy) = (xy;y) \oplus (xy;t') \oplus (xt) (ty)$$

Esta fórmula muestra que la relación original entre x e y puede ser descrita como la suma de dos relaciones condicionales y de un término suplementario. Este último es el producto de dos términos habitualmente designados como *relaciones marginales* entre el factor test y cada una de las dos variables originales. Señalemos, con respecto a la aritmética de esta fórmula, que los signos $+$ han sido rodeados por un círculo para indicar que (xy) es una suma aritmética no simple, sino ponderada de las tres elementos $(xy;t)$, $(xy;t')$ y (ty) , como se puede ver fácilmente desarrollando la ecuación. Sin embargo, como esta ponderación no interviene en la discusión que sigue, es inútil complicar la fórmula escribiéndola en una fórmula explícita.

Apliquemos ahora esta fórmula a un cierto número de casos concretos. Se ha constatado que el número de niños nacidos en una comunidad determinada está en relación con el número de cigüeñas. Naturalmente, es fácil explicar esta curiosidad; basta, en efecto, con introducir como variable-test la distinción entre comunidades urbanas y comunidades rurales. Se constata evidentemente que el ligamen entre el número de cigüeñas y el de nacimientos desaparece en cada grupo de comunidades: el ligamen es nulo. La relación primitivamente observada corresponde a la convergencia de dos hechos: en las comunidades rurales las cigüeñas son más numerosas y el porcentaje de natalidad más elevado.

Los ejemplos de las primeras páginas pueden, igualmente, ser analizados con la ayuda de la ecuación anterior. En el caso de las emisiones religiosas, el producto de las relaciones (xt) y (ty) es positivo y las dos relaciones condicionales son débiles. En el caso de las tribunas políticas, el producto es negativo; por lo tanto, las relaciones condicionales deben ser superiores a la relación primitiva. En el caso de las emisiones de música clásica, el rasgo importante es la oposición de signo entre las dos relaciones condicionales. La primera línea de los Cuadros 4 al 6 corresponde a la relación (xy) ; la segunda, a las relaciones condicionales. La relación (xt) puede extraerse del Cuadro 4. El Cuadro 3 y la nota final del capítulo permiten obtener (ty) para cada uno de los tres tipos de emisiones.

Con el fin de familiarizarse con la fórmula general que acabamos de enunciar, el lector queda invitado a aplicarla a un tipo de razonamiento particularmente frecuente en Durkheim: "lejos de predisponer a ello [al suicidio], la idiotez parece ser más bien un preservativo; porque los idiotas son mucho más numerosos en el campo que en la

ciudad, mientras que con los suicidios ocurre al revés". Otro ejemplo es el fragmento en el que intenta probar que las personas menos instruidas tienen una propensión más débil al suicidio: "Hemos visto que, en todos los países del mundo, la mujer se suicida muchos menos que el hombre. Ahora bien, también está mucho menos instruida". Como puede verse, Durkheim no tiene nunca en cuenta las relaciones condicionales. Sería interesante preguntarse que efectos resultan de ello para la lógica de su argumentación.

En dos situaciones, la fórmula reviste una forma muy particular. El primer caso formaliza la noción de experiencia con un grupo de control: de dos grupos aparejados, uno está expuesto a un estímulo dado. Llamemos "x" al estímulo. La idea del aparejamiento puede ser formulada con la ayuda de la ecuación 1. En efecto, los dos grupos aparejados deben revelarse idénticos con respecto a toda variable-test: en este caso, debemos obtener $(xt) = 0$; por lo tanto, el tercer término del miembro de la derecha de la ecuación fundamental será siempre nulo. Evidentemente, los grupos sólo pueden aparejarse sobre variables-test que sean pertinentes en el contexto de un análisis dado, es decir variables susceptibles de estar en relación con el criterio "y". En una situación en que la variable-test es no pertinente, (ty) es nulo, y de nuevo el tercer término debe desaparecer.

En el caso de una experiencia con grupo de control, la introducción de una variable-test significativa corresponde, pues, a la siguiente particularización de la fórmula fundamental:

$$(2) \quad (xy) = (xy;t) \oplus (xy;t') \oplus (0) \cdot (ty)$$

El segundo caso particular interesante es aquél en el cual, siendo nulas las dos relaciones condicionales, la relación primitiva es igual al producto de las relaciones entre la variable-test y cada una de las variables originales. Se ha observado, por ejemplo, que los accidentes de automóviles son más frecuentes en los hombres que entre las mujeres. Pero, si se introduce la variable-test "kilometraje anual", se constata que el ligamen entre sexo y accidentes desaparece. En este caso, la relación primitiva es equivalente al producto de dos nuevas relaciones que, a su vez, podrían someterse a un análisis parecido al que se ha efectuado sobre la relación inicial. Estas dos *relaciones marginales*, aparecen cuando se cruza la variable-test con cada una de las dos variables primitivas; no recurren al cruzamiento simultáneo de tres variables: no requieren, en otras palabras, el conocimiento de las relaciones condicionales.

$$(3) \quad (xy) = (0) \oplus (0) \oplus (xt) \cdot (ty)$$

Es fácil referirse a la diferencia entre los dos casos particulares correspondientes a las formas 2 y 3. En la forma 2, las relaciones condicionales subsisten y la relación (xt) desaparece. Es cómodo, por lo tanto, llamar al primer caso, *elaboración por las relaciones parciales* (P), y al segundo, *elaboración por las relaciones marginales* (M). (El término relaciones marginales se justifica si se examina más de cerca el Cuadro 3: las relaciones (xt) e (yt) comparan la edad y las preferencias radiofónicas de dos grupos de instrucción, es decir, las sumas marginales correspondientes a los cuadros condicionales).

Si hay reparos en utilizar un lenguaje estadístico, se pueden expresar las anteriores reflexiones en otro lenguaje. En el caso P, la relación primitiva subsiste después de haber introducido una variable-test. En el caso M, desaparece y es reemplazada por dos nuevas relaciones marginales en las que interviene el factor-test.

Así, pues, hemos reducido a su armazón lógico un procedimiento de investigación familiar. Consiste en analizar una relación entre dos variables mediante la introducción de una tercera variable. Hasta ahora lo hemos aplicado a ejemplos derivados de estudios empíricos, pero sería fácil mostrar que ciertas teorías tienen una estructura lógica idéntica. Dollard sostiene, por ejemplo, que los blancos de clase acomodadas son más favorables a los negros que los blancos de clases inferiores, porque los primeros están menos expuestos a la competencia de los negros. Durkheim sugiere que el suicidio es menos frecuente entre los católicos que entre los protestantes, porque la comunidad católica está más cohesionada. Estos dos ejemplos representan teorías evidentemente distintas por su contenido, pero el esquema lógico de las mismas es idéntico: en los dos casos, dos variables son analizadas a la luz de una tercera.

Desde este momento estamos en condiciones de enumerar el conjunto de deducciones que es posible extraer de tres variables: están expresadas en nuestra fórmula fundamental. Por lo tanto, esta ecuación debe permitir determinar las diferentes formas que puede revestir una proposición teórica cualquiera construida a partir de dos variables primitivas y de una variable adicional. En otras palabras, debe ser posible clasificar en un pequeño número de tipos, a partir de las condiciones de la fórmula fundamental, las proposiciones teóricas que conciernen a un conjunto de tres variables. Vamos a ver que esta clasificación es realizable, pero antes querríamos introducir una consideración suplementaria.

(Para simplificar, hemos restringido el análisis al caso de tres variables. La

introducción de variables suplementarias no entraña ninguna modificación en la lógica del razonamiento, como se puede ver fácilmente. Por otra parte, nosotros sólo tenemos en cuenta las variables dicotómicas. La introducción de variables que pudieran tomar un número de valores superior a dos complicaría el argumento sin cambiar la forma del mismo).

Rol de la anterioridad temporal de las variables

El rol temporal de las variables introduce un aspecto nuevo en esta discusión. Muchas veces, los atributos se pueden colocar según un orden cronológico: la duración de los esponsales y la armonía conyugal, por ejemplo. En este caso, el orden es evidente. También ocurre que esta ordenación temporal, sin ser aparente, puede restituirse de modo indirecto: se ha observado una relación inversa entre la importancia del ingreso y la frecuencia de la pertenencia a organizaciones. Lo más verosímil es suponer una prioridad temporal al estatuto económico. Esta situación aparece también en los estudios que ponen de manifiesto una relación entre rasgos de personalidad relativamente estables y determinados criterios de éxito escolar o profesional.

Ciertas variables pueden ser utilizadas de varias maneras, en función del problema examinado, tener posiciones diferentes en la secuencia temporal. Un ejemplo típico es la edad. Las personas de 60 años se distinguen de los jóvenes de 20 años no solamente por las posibilidades físicas, sino también por el hecho de que han nacido y han sido educadas en otra época. En la proposición "las personas de edad tienen un nivel de instrucción inferior", lo importante es evidentemente la época en que estas personas han alcanzado la edad de la escolarización: en este caso, la edad es *anterior* a la educación. Si, por el contrario, se relaciona la edad de fallecimiento y el clima, la edad evidentemente debe considerarse como *posterior*. La misma distinción puede introducirse cuando se comparan las célibes con las personas casadas: el matrimonio puede referirse según el contexto, o bien al hecho de haber contraído matrimonio, o bien al hecho de vivir juntos en determinadas condiciones específicas.

Finalmente, existen variables cuyo orden temporal es indeterminado. Cuando se constata, por ejemplo, que los demócratas aparecen como más dirigistas que los republicanos, es imposible ordenar en el tiempo la simpatía política y la actitud ligada a ella. La misma indeterminación aparece cuando se observa que ciertas personas utilizan un producto dado y escuchan una cadena de emisoras de radio en la que no se hace publicidad de ese producto. Muy a menudo, se tienen que utilizar variables cuyo orden temporal es dudoso; ello puede ser debido a una negligencia en el plano

de observación o a tener que utilizar datos recogidos para otros trabajos. De hecho, siempre es teóricamente posible determinar el orden cronológico de las variables; el progreso de la investigación empírica consiste precisamente, entre otras cosas, en determinar las anterioridades entre variables. Evidentemente, sería interesante discutir los procedimientos de análisis que pueden utilizarse en el caso en que el orden temporal es dudoso, pero ello nos alejaría de nuestro tema. De ahora en adelante, admitimos que el orden de las variables es conocido.

Los principales tipos de elaboración

Desde este momento estamos en condiciones de presentar el punto esencial de nuestra argumentación: a partir de dos variables primitivas y de una variable-test, se pueden efectuar cuatro operaciones fundamentales. El hecho de que estas operaciones se realicen sobre datos empíricos o que correspondan a análisis teóricos no tiene ninguna importancia. Si se analiza la relación entre dos variables a la luz de una tercera, el análisis corresponderá siempre a una u otra de estas cuatro operaciones o a una combinación de las mismas. Una vez aclarado este esquema lógico, poco importa que se llame a este análisis interpretación, explicación, teoría o cualquier otro nombre.

Para introducir estas cuatro operaciones, admitamos en primer lugar que tenemos una relación entre las dos variables x e y , y que x es temporalmente anterior a y . Hagamos intervenir ahora una variable-test t . Son posibles dos relaciones temporales; o bien la variable-test es posterior a x y anterior a y , o bien es anterior a x (o contemporánea de x). O bien tenemos $(xt) = 0$, o bien $(xt) \neq 0$. Si $(xt) \neq 0$, admitiremos para simplificar que las relaciones condicionales entre x e y son nulas. Vemos, pues, que aparecen cuatro configuraciones fundamentales, descritas en el siguiente esquema por símbolos MA, PA, etc. Pasamos ahora a la descripción y a la ilustración de estas cuatro estructuras.

Cuadro 7

	$(xt) = 0$	$(xt) \neq 0$
Posición de «t»	$(xy;t) \neq 0$	$(xy;t) = 0$
Anterior	PA	MA
Intermedia	PI	MI

En el caso correspondiente al tipo PA, llamamos generalmente "condición" a la

variable "t". Es fácil imaginar ejemplos de este tipo, aunque sean relativamente raros en la práctica. El efecto de propaganda de un film, por ejemplo, es mayor entre las personas con un nivel de instrucción poco elevado. Las consecuencias nefastas de la depresión han sido más notables en las familias autoritarias que en las otras.

Con respecto a los resultados o razonamientos de este tipo, podemos hacer tres observaciones generales: a) corresponde al orden ya conocido estímulo-disposición-respuesta en el que x es el estímulo, y en el que t es anterior a x y corresponde a la disposición; b) ese tipo de razonamiento tiene una función de especificación: uno de los dos ligámenes condicionales debe ser necesariamente mayor que el ligamen primitivo. En otras palabras, el análisis tiene como función la de especificar las circunstancias en las cuales el ligamen original es más notable; c) generalmente, a partir de este punto, nos preguntaremos por qué esta relación es más intensa para uno de los valores de la variable-test. Esta pregunta nos podría llevar a otros tipos de análisis. Durkheim utiliza una argumentación del tipo PA cuando se pregunta por qué las personas casadas tienen propensión más débil al suicidio que las solteras. Introduce como variable-test una tendencia al suicidio "que la familia neutralizaría o impediría que saliera a la luz". Esto es exactamente un argumento del tipo PA. Pero no puede considerarse que tenga un valor explicativo, porque la variable-test hipotética introducida aparece como tautológica. El problema consiste más bien en preguntarse por qué la familia reprime el desarrollo de la tendencia al suicidio. Más adelante veremos que esto nos conduce a un argumento de tipo MI.

Es fácil producir ejemplos de tipo PI. Estudiando la relación entre éxito profesional y el ambiente, liberal o autoritario, de la educación, se ha podido observar que los jóvenes educados en una atmósfera liberal tienen menos éxito en una situación profesional de estructura autoritaria.

Aquí, la relación entre tipo de educación y éxito profesional está elaborada por un factor-test intermedio: la atmósfera profesional. Siguiendo el ejemplo de Paul Horst, llamamos "contingencia" a este factor. Este autor muestra que en muchos estudios de predicción, el valor predicho depende de circunstancias ulteriores no ligadas al que hace la predicción. Otro ejemplo de este tipo lo da Merton, en su estudio sobre el ligamen entre estatuto profesional y la participación en las actividades comunitarias en un conjunto residencial. La participación de los empleados está en razón inversa, y la de los obreros en razón directa, a su satisfacción profesional.

Recordemos que los tipos PI y PA tienen entre ellos una relación simple. Si t es una condición (tipo PA), x es una contingencia. Si t es una contingencia (tipo PI), x es una

condición.

Tenemos, pues, de hecho un tipo único de elaboración para el cual el vocablo "especificación" es sin duda el más apropiado. La consideración de la anterioridad temporal entre x y t introduce dos subtipos de especificación.

El tipo MA se utiliza sobre todo en la interpretación de correlaciones equívocas: la importancia de los desperfectos varía con el número de extintores en un accidente. Para elaborar esta relación sorprendente se introducirá naturalmente como factor-test la gravedad del incendio. Las relaciones condicionales quedan anuladas y el resultado primitivo aparece como producto de dos relaciones marginales.

Cuando encontramos una *relación sorprendente*, generalmente nos detenemos en ella, pero el mismo modo de elaboración puede utilizarse en situaciones distintas: los suicidios son más numerosos en verano que en invierno -señalemos que para ser completamente correctos, habría que enunciar la proposición de manera ligeramente distinta y decir: se contabilizan más suicidios en verano que en invierno. Durkheim sugiere, como factor t destinado a elaborar la relación anterior, el crecimiento de las actividades sociales durante el verano. Señalemos de paso cómo una formulación concreta contribuye a clarificar el orden temporal; la intensidad de la vida social precede a la vez a la observación del suicidio mismo.

El corazón del problema es ahora la relación (ty), es decir, que a una intensificación de la vida social corresponde un aumento de la proporción de suicidios. Tendríamos una prueba de que la interpretación de Durkheim es correcta si se pusiera de manifiesto una desaparición de las relaciones condicionales. Habría que demostrar que cuando la intensidad de la vida social se mantiene constante, la proporción de suicidios ya no aparece ligada a la estación del año.

Al no disponer de datos sobre este punto, Durkheim recurre a otros tipos de situaciones que se puedan suponer ligadas a una intensidad de la vida social. Observa de esta forma que los suicidios son más numerosos de día que de noche y explica de nuevo esta relación por el mismo factor. Pero esto nos lleva a la cuestión general de la probabilidad de las inferencias, lo cual desborda el marco de este artículo.

Es conveniente, creemos, precisar un aspecto de terminología. Utilizamos el término "elaboración" para designar el procedimiento de investigación expresado en nuestra ecuación fundamental. También hemos visto que algunos tipos de argumentación son formalmente idénticos a esta ecuación; la única diferencia consiste en que la argumentación no recurre a datos cifrados sino que usa afirmaciones que implican

esencialmente que determinadas relaciones, condicionales o marginales, son inferiores o superiores a otras, o que se acercan a cero, o que son distintas de cero.

El término "elaboración" es útil en la medida en que es suficientemente neutro como para evitar, tanto como sea posible, toda asociación que pueda ocultar el significado preciso que le damos aquí. Pero la terminología es más delicada cuando se trata de distinguir los cuatro tipos de elaboración. Estos últimos están rigurosamente definidos como procedimientos de investigación o tipos de argumentación. Hay que evitar, *nombrándolos*, que se olvide este punto esencial. Parece razonable, por ejemplo, llamar *explicación* al tipo MA, porque este concepto parece corresponder exactamente a la operación MA. Pero el mismo término ha sido utilizado con acepciones diferentes. Por lo tanto, es importante recordar que, en el contexto actual, el término "explicación" es idéntico a la expresión "elaboración de tipo MA". El hecho de saber si esta acepción es preferible a otra es una cuestión puramente secundaria. Si este término conduce a confusiones, podría abandonarse sin alterar por ello la esencia del argumento.

La dificultad terminológica aparece inmediatamente cuando nos referimos al tipo MI, al cual muchos autores han asociado y sin duda asociarán el término "explicación". Nosotros mismos hablaremos de "interpretación" para designar este tipo de operación. La diferencia entre la *explicación* y la *interpretación* en este contexto está ligada a la prioridad cronológica. En una interpretación, t es una variable intermedia situada entre x e y en la secuencia temporal.

A lo largo de este artículo hemos dado en varias ocasiones dos ejemplos del tipo MI: hay menos suicidios en el campo que en la ciudad; la mayor intimidad de la vida rural se introduce como variable intermedia. Si se dispusiera de una buena medida de la cohesión, se tendría que observar que el tipo de hábitat está ligado al grado de cohesión y que este último está ligado, a su vez, a la tasa de suicidio. Pero, evidentemente, ciertas comunidades rurales presentan un grado de cohesión menor que ciertas comunidades urbanas. Si esta variable se mantuviera constante, se debería observar que la relación condicional entre tipo de hábitat y tasa de suicidio es aproximadamente igual a cero.

Quizá sea útil introducir un nuevo ejemplo con el fin de ilustrar la diferencia entre los tipos MA y MI: se ha observado que las mujeres casadas que trabajan en una fábrica presentan una tasa de absentismo más elevada que las solteras. A continuación describimos algunos tipos posibles de elaboración:

a) Factor-test: la mayor importancia de las responsabilidades familiares. Esta última variable es intermedia. Si, habiéndola introducido, se observa una desaparición de dos

relaciones condicionales, entre la situación familiar y el ausentismo, tenemos una elaboración de tipo MI.

b) Factor-test: la menor resistencia física, medida de modo sumario por la edad. Esta variable es temporalmente prioritaria. Si se observa que, cuando la edad se mantiene constante, la relación entre situación familiar y ausentismo desaparece, habremos explicado la relación primitiva por un análisis de tipo MA, que de esta forma aparecerá como una relación equívoca.

Este último caso nos sugiere además un punto importante. Después de haber explicado la relación primitiva, habríamos podido dirigir nuestra atención hacia la relación (ty), es decir, el hecho de que las personas de más edad tienen una tasa de abstinencia más elevada. Ello podría conducir a nuevos tipos de elaboración: ¿es cierto que las mujeres de mayor edad tienen menos resistencia, o es que tienen más dificultades en adaptarse a la disciplina profesional porque han sido educadas en una época en la que el trabajo femenino era menos corriente? En otras palabras, después de haber terminado una elaboración de determinado tipo, normalmente tendremos ocasión de realizar nuevos análisis. Pero es importante señalar que los procedimientos de análisis se reducen siempre a un número de tipos limitado.

En este punto debemos hacer una observación que se deduce de la aritmética de nuestra fórmula fundamental. En los tipos MA y MI, las dos relaciones derivadas son siempre mayores que la relación primitiva, como se puede ver fácilmente. Por lo tanto, toda elaboración pone de manifiesto al menos una correlación más elevada que la correlación primitiva. Este hecho tiene implicaciones lógicas y psicológicas importantes que no podemos discutir aquí.

Podemos aclarar un último punto, al menos en cierta medida, mediante este análisis: éste permite, en efecto, sugerir una definición clara y distinta del concepto de relación causal entre los atributos. Si se observa una relación entre x e y , y esta relación subsiste para cualquier factor-test *antecedente*, debe ser considerada como causal. Por otra parte, poco importa que las operaciones implicadas se efectúen realmente. Basta con que aparezcan como plausibles a partir de consideraciones generales, que consisten siempre, de cualquier forma, en utilizar una u otra de las cuatro operaciones fundamentales, salvo en el caso en que tendrán que resolver problemas de anterioridad temporal.

Esta definición reviste particular importancia en ciertas discusiones: el crimen es más frecuente en regiones muy pobladas. Es posible discutir el carácter causal de esta relación a partir de dos tipos de criterios distintos: se puede sugerir la hipótesis de la

intervención de una variable intermedia, como el aumento de las tensiones en un lugar superpoblado; de hecho, esta interpretación, si bien es correcta, no niega el carácter causal de la relación primitiva. Pero de esta relación también puede darse la interpretación de que los módicos precios de los alquileres en las zonas superpobladas explicaría la presencia de una gran proporción de elementos indeseables; en este caso, las características de los habitantes preceden a las de la zona. La relación primitiva aparece esta vez como equívoca y no podría ser considerada como causal.

Nota

Para completar nuestra información es necesario saber cuál es la relación entre nivel de instrucción y las preferencias por los distintos tipos de emisiones radiofónicas, o sea (ty). A continuación presentamos los datos correspondientes:

	Escuchan tribunas políticas			Escuchan programas de música clásica		
	si	no		si	no	
niv. de instr. sup	460	540	1000	niv. de instr. inf	280	1020 1300
niv. de instr. inf.	460	840	1300	niv. de instr. sup.	404	596 1000
total	920	1380	2300	total	648	1616 2300