

Unidad 8

- Los Métodos en Sociología.

El análisis de las relaciones entre variables

Una vez construidas las variables, se trata de analizar sus relaciones. Diversos estudios muestran por ejemplo que el suicidio es más frecuente entre los divorciados, el absentismo profesional más frecuente entre las mujeres, el nivel de aspiración más débil entre los niños criados en una familia de estructura autoritaria, las oportunidades de entrar en 6to de secundaria más débil para los niños oriundos de familias que pertenecen a medios de pocos recursos. Semejantes correlaciones son a menudo el punto de partida del análisis sociológico. Así, toda la obra *El Suicidio* de Durkheim es un esfuerzo para explicar ciertas relaciones primitivas de este tipo: relación entre tasa de divorcio y tasa de suicidio, entre longitud del día y suicidio, etc.

Los estudios respecto a la movilidad social siempre parten de relaciones entre comportamientos de movilidad y categorías socio profesionales. Es el caso, por ejemplo de los trabajos de A. Girard sobre la entrada en 6to o de R. Bourdon en *La Desigualdad de oportunidades* (1).

En la fase pre-durkheimiana de la sociología empírica, la tendencia era tomar relaciones así como un hecho e interpretarlos directamente. Esta práctica se basaba en el postulado erróneo según el que una correlación estadística entre dos variables sería siempre un signo de causalidad entre esas variables.

Así, los positivistas italianos, como Lombroso o Ferri, interpretaban la relación que las estadísticas permitían establecer entre tasa de suicidio y temperatura (las tasas de suicidio aumentan en efecto con la temperatura) como una relación de causalidad: el calor, decían, provoca un estado de excitación física que aumenta la propensión al suicidio. La desigualdad de oportunidades en los logros escolares o el éxito social es asimismo interpretada por autores como Galton y Pearson como el producto de factores hereditarios (2).

1. Explicaciones sociológicas y sistema de relaciones causales. - Durkheim es el primero en haber entendido que una relación estadística no puede ser interpretada sin grandes precauciones como una relación causal.

Tomemos el ejemplo de su análisis entre suicidio y temperatura. Los días más calientes, se fija, también son los días más largos. Pero ¿no indicará eso que el suicidio es más frecuente cuando la vida social es más intensa? Efectivamente, las estadísticas muestran que las tasas de suicidio varían no sólo según las temporadas, sino también según el ritmo de la actividad que caracteriza los días de la semana como las horas del día. Además, cuando el tiempo social varía según las categorías o los subgrupos de una sociedad, las tasas de suicidio

siguen esas diferencias de ritmo. Así, la tasa de suicidio femenino se eleva durante el fin de semana, mientras que el suicidio masculino disminuye en ese momento, debido a que -en la época en la que escribe Durkheim al menos - el fin de semana corresponde al máximo de las actividades sociales de la mujer y al mínimo de las actividades sociales del hombre. Asimismo, el tiempo social urbano era, en aquella época, mucho más regularizado que el tiempo social rural, que dependía más estrechamente del tiempo astronómico. Ese hecho explica, según Durkheim, que la tasa de suicidio sea, en varios países observados, regularmente en su punto más alto en verano, mientras que en los centros urbanos de esos mismos países, parece ser así ora en primavera, ora en verano. Además, mientras que en el conjunto de los países las tasas de suicidio se elevan del 50% aproximadamente sobre la temporada en que el suicidio es menos frecuente (invierno) a la en que es más frecuente, no se acrecienta más que del 25% a lo mucho en las grandes ciudades de esos mismos países. Ese resultado muestra de nuevo la regularización del tiempo social en las ciudades e indica que las variaciones de la tasa de suicidio según la temperatura y la longitud del día deben explicarse por el efecto del ritmo de la vida social sobre el suicidio.

Ese tipo de análisis, que uno vuelve a encontrar en todo *El Suicidio* puede ser, desde un punto de vista lógico, considerado como una verdadera revolución. Muestra que una relación estadística entre dos variables sólo puede ser interpretada generalmente si uno la introduce en el modelo causal. Se puede traducir gráficamente el análisis durkheimiano de la relación entre suicidio y temperatura (fig. 1).

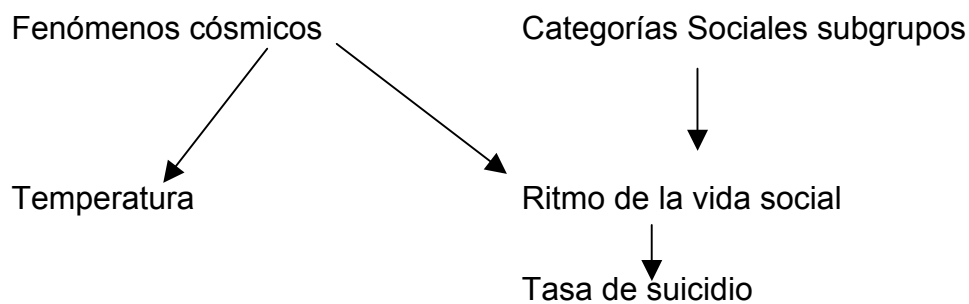


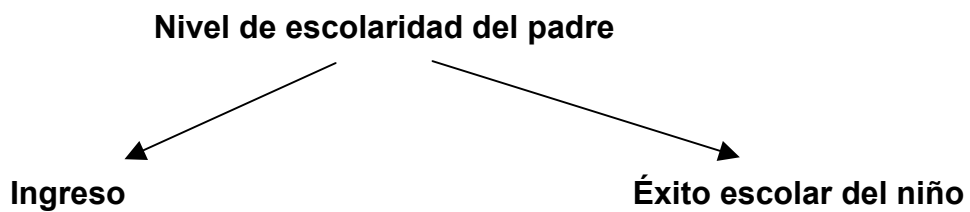
Figura 1. El análisis durkheimiano de la relación entre suicidio y longitud del día

Uno ve mediante este esquema que la relación entre temperatura y tasa de suicidio se explica por la doble influencia de los fenómenos cósmicos sobre la temperatura y sobre el ritmo de la vida social. Pero la explicación demuestra al mismo tiempo que la temperatura no ejerce ninguna influencia sobre la tasa de suicidio.

De los análisis durkheimianos, se desprende la idea de que la explicación de una relación estadística consiste en introducir variables adicionales de modo que se pueda desprender el modelo causal en el cual está inscrita.

A. Girard usa un procedimiento semejante para explicar la relación entre el ingreso de los padres y el desempeño escolar (1). Esta relación es positiva. Dicho de otra manera, entre más elevado sea el ingreso, más grande el éxito escolar. Pero ¿qué es el significado de ese *hecho*? ¿Es menester interpretarlo directamente y admitir por ejemplo que una mayor seguridad económica llevan a los padres a vislumbrar estudios más largos y a estimular más el celo escolar del niños? Es una interpretación posible. Ésta se encuentra sin embargo corregida por el *hecho de* que, cuando uno considera niños oriundos de familias que presentan un nivel cultural semejante, la relación entre ingreso y éxito escolar desaparece. Ese resultado indica que la relación entre ingreso y éxito se debe al hecho de que un nivel de escolaridad más elevado corresponde en promedio a un nivel de ingreso más elevado. Pero es en realidad el nivel cultural de la familia el que causa el éxito escolar del niño. En ese caso, la relación primitiva halla su explicación en un modelo que puede representar mediante el esquema de la figura 2.

Figura 2. - Explicación de la relación entre ingreso de la familia y éxito escolar del niño



Ese modelo muestra que la relación estadística entre ingreso y éxito escolar no debe ser interpretada directamente, pero que es consecuencia del sistema de relaciones que caracteriza las tres variables consideradas. De la misma manera que el concepto «variable» se encuentra en todos los estudios sociológicos, tanto en los estudios monográficos y cualitativos como en los estudios cuantitativos, así el postulado durkheimiano según el cual el análisis sociológico consiste en determinar las estructuras causales que explican las relaciones entre variables, es común la mayoría de las investigaciones sociológicas.

Pero sólo en caso de que dispongamos de medidas correspondientes a esas variables se puede determinar con precisión el sistema de relaciones causales explicativo de un fenómeno social. Este solo hecho da cuenta del desarrollo considerable de las encuestas de tipo cuantitativo.

2. Los problemas lógicos del análisis de relaciones entre variables: el análisis multivariado. - El análisis de relaciones entre variables plantea problemas lógicos delicados como uno lo verá en el siguiente capítulo. Por el momento, examinaremos con Lazarsfeld las diferentes situaciones que pueden lógicamente presentarse cuando uno introduce una variable adicional en una relación primitivamente observada entre dos variables (1),

El ejemplo que sigue Lazarsfeld en muy simple> se trata de las relaciones entre edad e interés por tres tipos de programas radiofónicos. Presentamos esas relaciones en la Tabla III .

Las personas mayores buscan en mayor medida los dos primeros tipos de programas. En cambio, los porcentajes son más o menos idénticos en lo que se refiere al tercer tipo de programas

Tabla III. Relaciones entre la edad y el interés por 3 tipos de programas radiofónicos

	Jóvenes - % radioescuchas	Viejos - % radioescuchas
Programas religiosos	17	26
Tribunas políticas	34	45
Programas de música clásica	30	29
(Número de casos)	(1000)	(1300)

Uno puede intentar interpretar directamente esas relaciones. Una teoría posible a propósito de la primera de esas reacciones sería ver el signo de un apego menor de parte de los jóvenes hacia los valores religiosos. La ausencia de reacción observada en el tercer caso, por su lado, puede ser interpretada por el hecho de que el interés hacia la música es un mero asunto de gustos.

Vamos a ver que la introducción de una variable adicional, que Lazarsfeld llama variable de *prueba*, puede contribuir a esclarecer la interpretación. La variable de prueba es el nivel de educación. He aquí los resultados que se obtienen cuando se analiza las relaciones precedentes al repartir la población observada en dos grupos caracterizados respectivamente por un nivel de escolaridad "superior" e "inferior". Examinemos primero el caso de los programas religiosos (tabla IV).

En este caso, uno ve que el homogeneizar los grupos en relación al nivel de escolaridad, la relación observada primitivamente se atenúa de manera considerable al punto de volverse prácticamente nula. En otras palabras, la relación entre edad e interés por ese tipo de programa se debe al hecho de que, en una sociedad sometida a un proceso de democratización de la enseñanza, los jóvenes tienen en promedio un nivel de educación más elevado.

Tabla IV. Interés por los programas religiosos en función de la edad y del nivel de escolaridad

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

Como el interés por los programas religiosos depende del nivel de educación (así como podemos darnos cuenta al comparar las columnas 1 y 3 y las columnas 2 y 4 de la Tabla IV), resulta que se da una relación entre edad e interés por los programas religiosos. La relación desaparece cuando uno homogeneiza los grupos en relación al nivel de educación, la influencia de esta última variable es la única real. El análisis puede ser resumido por la estructura causal representada en la figura 3.

Figura 3. - Estructura causal correspondiente a la tabla IV

Edad → Nivel de escolaridad → interés por los programas religiosos
 (-) (-)

El signo (-) colocado abajo de las flechas indica que la relación es negativa, si la edad es más elevada, el nivel de escolaridad es en promedio más bajo; si el nivel de escolaridad es más elevado, el interés por los programas religiosos es en promedio más bajo.

Vemos, en particular, que la relación no puede ser interpretada como traduciendo un apego menor de los jóvenes hacia los valores religiosos, ya que ese apego es el mismo sin importar la edad, mientras el nivel de educación sea el mismo.

Observemos ahora que los resultados referentes al interés por los programas políticos (Tabla V),

Tabla V

Interés por los programas políticos en función de la edad y de la escolaridad

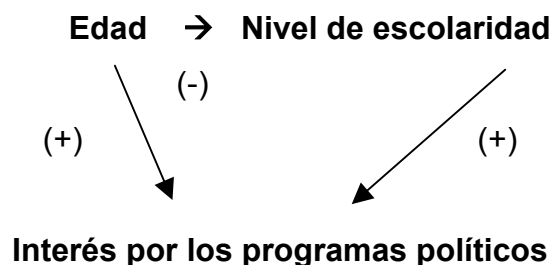
Jóvenes 34 %		Viejos 45 %	
Nivel de educación superior		Nivel de educación inferior	
Jóvenes	viejos	Jóvenes	viejos
40 %	55 %	25 %	40 %

La situación es completamente distinta en ese caso, En efecto, la relación entre edad e interés persiste cuando uno homogeneiza los grupos en relación a la escolaridad. Las dos variables, edad e instrucción, tienen un efecto sobre el interés hacia los programas políticos. La estructura causal correspondiente es representada en la figura 4.

La introducción de la variable de prueba permite, otra vez, afinar la interpretación. La escolaridad tiene como efecto el aumentar el interés hacia la política tal vez porque induce en el individuo el sentimiento de que es capaz de entender los acontecimientos políticos, que tiene derecho de juzgarlos, y que puede actuar sobre ellos. Pero tal vez, también el interés más débil de la gente

menos escolarizada se debe al hecho de que ocupan en general una posición social más baja, que albergan un sentimiento de exclusión y desarrollan una actitud de retraimiento. Para decidir entre esas dos interpretaciones, habría que introducir nuevas variables de prueba. En cuanto a la edad, hay un efecto independiente del de la escolaridad. Eso tal vez indique que el tipo de inserción social del adulto conlleva en éste el sentimiento de que los acontecimientos políticos le conciernen más.

Figura 4. - Estructura causal correspondiente a la Tabla V



En el caso de los programas de música clásica, la estructura de las relaciones es otra vez distinta, como lo muestra La Tabla VI.

Tabla VI

Interés por la música clásica en función de la edad y de la escolaridad

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

En este caso, la relación primitivamente nula entre la edad y el interés por los programas de música clásica se vuelve positiva en el grupo de escolaridad superior y negativa en el grupo de escolaridad inferior. Podemos imaginar que esta estructura traiciona el hecho de que en las personas de escolaridad inferior, el momento en el que el contacto con la *cultura* es más directo es el de la escolaridad; luego, resultan pertenecer a entornos socio profesionales en los que la cultura clásica no es percibida como un valor central. Correlativamente, el hecho de que la relación entre edad e interés por la música clásica sea positiva en el caso del grupo de escolaridad superior puede indicar que la música clásica es un valor cultural más importante para las personas de más edad, mientras compite más con otras formas de cultura (cine, música moderna, jazz) entre los más jóvenes.

Sea como sea, vemos que la introducción de la variable de prueba pone en evidencia fenómenos relativamente complejos, que era imposible detectar al nivel de la relación primitiva. La nulidad de esta relación es, como podemos verlo, el producto de efectos que se compensan. La estructura causal correspondiente puede ser descrita por los esquemas de la figura 5.

Figura 5. Estructuras causales correspondiente a la tabla VI



Esos esquemas sugieren que, en este caso, la influencia que tiene la edad sobre el interés por la música clásica depende del nivel de escolaridad y que, correlativamente, el efecto del nivel de escolaridad sobre el interés por la música clásica depende de la edad. Como lo vemos, no podemos aquí conformarnos con enunciar los efectos de las variables unas sobre otras. También hay que considerar efectos de índole más compleja, que las estadísticas llaman *efectos de interacción*. En el presente caso, hay interacción entre edad y nivel de escolaridad, ya que el efecto de cada una de esas variables sobre el interés por la música depende de otras variables. En otras palabras, no tiene sentido hablar en ese caso del efecto que tiene la edad sobre el interés por la música clásica, ya que este efecto, difiere completamente (cambia de sentido) según que uno está considerando las personas de escolaridad superior o las personas de escolaridad inferior. Asimismo, no se puede hablar de que tenga la escolaridad un efecto, ya que éste varía según la edad. Es precisamente esta imposibilidad de separar los efectos de dos variables lo que señala la expresión "efectos de interacción".

Es fácil demostrar que los tres tipos de estructura ilustrados por los ejemplos precedentes agotan situaciones que se pueden lógicamente encontrar cuando se analiza una relación primitiva entre dos variables a la luz de una tercera variable. Vemos además que el análisis puede, en ese caso, llevarse a cabo de manera intuitiva. En otras palabras, es posible, cuando se limita el análisis a tres variables, asociar de manera intuitiva una estructura causal interpretativa a los datos numéricos.

Pero la lógica misma del análisis (llamado análisis multivariado) que acabamos de exponer nos invita a generalizarla. Hemos visto que la interpretación se hallaba considerablemente enriquecida cuando introducíamos una tercera variable en el análisis de una relación primitiva.

Pero al mismo tiempo, las estructuras de tres variables sugieren nuevas variables intermedias. En un plano más general, es natural extender la lógica del análisis multivariado a cuatro, cinco y hasta un número de variables aun más elevado.

Tomemos un problema concreto que ya hemos evocado. Imaginemos que en un estudio sobre la elección profesional hayamos suministrado un cuestionario a una población de alumnos a nivel secundaria. Algunos de esos alumnos ya tienen una idea precisa de su porvenir profesional, otros no. Si el sociólogo se propone explicar ese fenómeno, muchas hipótesis van a parecerle plausibles. Tal vez la incitación a elegir es más grande en las familias de pocos recursos. Tal vez la elección tiene más chance de ser diferida en las familias que mantienen al niño en un estado de protección más firme. Tal vez el hecho de pertenecer a ciertas secciones más orientadas hacia la cultura general tiene como consecuencia dejar la elección para más tarde. Tal vez la elección tiene más chance de ser definida si el éxito escolar es más marcado en una asignatura que en otra, etc.

Podríamos así alargar la lista de las hipótesis. El ejemplo muestra que el sociólogo que analiza un fenómeno cualquiera es normalmente llevado a enumerar una serie de factores determinantes. Además, esos factores determinantes son generalmente ligados entre sí. Es muy conocido, por ejemplo, que la composición social de las clases a nivel liceo varían con las secciones: es más improbable aprender latín cuando uno pertenece a un medio humilde. Resulta de ello que el sistema de los factores determinantes hipotéticos forma generalmente una red compleja de relaciones enmarañadas.

El problema lógico que se plantea a propósito de un caso como éste es por ende establecer la estructura causal subyacente el conjunto de las variables hipotéticas. Pero un análisis intuitivo se vuelve difícil a partir del momento en que el número de variables pasa de cuatro o cinco.

Esas observaciones tienen como propósito introducir las generalizaciones matemáticas del análisis multivariado del que hablaremos un poco en el siguiente capítulo.

Notemos también que, en ciertos casos, se puede utilizar una metodología más poderosa que la del análisis multivariado: consiste en partir de un modelo hipotético-deductivo para reproducir, y allí interpretar, datos estadísticos. En la primera etapa, se formaliza el comportamiento de los individuos, y luego, por deducción, determinan las consecuencias estadísticas de esos comportamientos. Véase, para un ejemplo de ese tipo de metodología, B. Boudon, *La desigualdad de las oportunidades* (1).