

Cortés, F., Rubalcaba, M.R. (1987). *Métodos estadísticos aplicados a la investigación en Ciencias Sociales*. México: Colegio de México.

## I. Introducción

El propósito fundamental que guía este trabajo es presentar el análisis de asociación como una técnica estadística de apoyo para la investigación social. La búsqueda de este objetivo nos remite de lleno al tema de la relación entre teoría e información en la investigación social. Sin embargo, éste es un tema tan vasto y complejo que supera con creces no sólo los límites de este libro, que más bien corresponde al ámbito del análisis estadístico, sino los de nuestro propio conocimiento del tema. A pesar de ello es imprescindible, para situar el instrumental estadístico que se presenta en los capítulos subsecuentes, referirnos a algunos aspectos involucrados en el vínculo entre teoría e información estadística. En consecuencia, la exposición será parcial y no podrá gozar de todo el rigor con que debiera tratarse. Nos conformaremos con no desvirtuar la imagen de esta relación ni inducir a sobresimplificaciones que puedan conducir a errores de apreciación sobre los diferentes problemas que aparecen en la investigación en ciencias sociales.

Expuesto de manera breve podríamos decir que el énfasis se pone en el estudio de la relación entre teoría y análisis de la información estadística. Interesa situarnos en aquellos puntos del proceso de la investigación que se constituyen como intersección de las operaciones teóricas y sus correspondientes operaciones empíricas ya que implican la movilización de nociones y técnicas definidas dentro del campo de la denominada estadística de atributos. Queremos destacar los vínculos que ligan en el proceso de investigación las elaboraciones propias del pensamiento teórico con los desarrollos característicos de las técnicas de análisis de la información, de modo que estemos en condiciones de orientar el estudio estadístico en función de las preocupaciones conceptuales y que al mismo tiempo las construcciones teóricas se basen en el material empírico acumulado en el proceso de investigación y no sólo en la consistencia interna de sus propios desarrollos. La relación entre teoría y estadística suele darse en varios momentos a lo largo de una investigación; a veces predominan las consideraciones teóricas y otras las estadísticas.

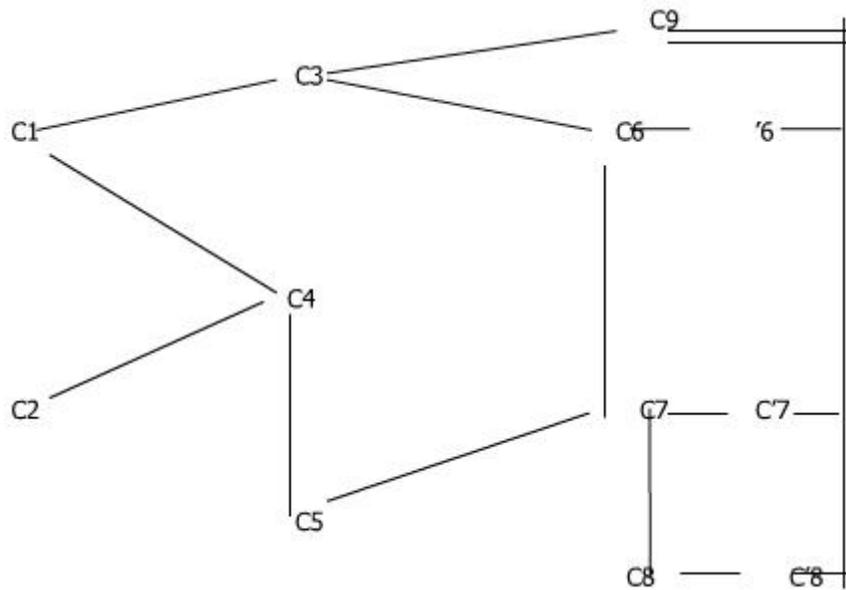
Examinemos con mayor detenimiento el tipo de relación que nos preocupa para sustentar nuestra aseveración. Si circunscribimos la discusión a las investigaciones que introducen lo empírico vía la estadística podremos encontrar una estructura que nos permita ligar las consideraciones teóricas con el manejo de la información estadística. Ahora bien, para un texto de estadística dedicado al estudio de las relaciones entre variables no interesan todos los puntos de contacto entre teoría e información sino aquellos que permitan ligar ciertas operaciones conceptuales con el material estadístico que se expone. En el proceso de investigación la interconexión de los análisis teórico y estadístico se puede dar en distintos momentos; algunas veces las preguntas que orientan el análisis estadístico surgirán desde inquietudes conceptuales y otras veces los resultados del análisis estadístico serán los que provoquen las preguntas teóricas. Pero sea cual fuere el origen o el predominio de la relación será necesario establecer el modo lógico a través del cual se la pueda capturar.

El análisis de asociación es un instrumento estadístico concebido para analizar relaciones entre variables ligadas a las relaciones entre conceptos planteadas en el campo teórico (que no agotan todo el dominio de las preocupaciones teóricas pero son las que nos interesa destacar pues es la parte que se vincula con el análisis de asociación). En consecuencia nuestra inquietud se puede replantear en términos de encontrar las ligas entre las relaciones que involucran conceptos con las que relacionan variables.

En estadística se da el nombre genérico de variable a la resultante del proceso que conecta los conceptos teóricos con los indicadores o índices que se sitúan en el mundo de la información y que para los propósitos que nos animan debemos considerar como información susceptible de tratarse estadísticamente, la que no se limita sólo a las estadísticas sino que cubre también hechos históricos, acontecimientos, información periodística, etc. Por tanto, hay una conexión entre la variable estadística y el concepto que nos servirá para establecer el vínculo que nos interesa (aunque más adelante se tocará este tema, no lo expondremos en profundidad porque además de ser complejo nos desviaría de nuestro propósito central). En efecto, si a partir del discurso teórico habitualmente expresado en lenguaje natural podemos formalizar sus contenidos, es decir, si a partir de los desarrollos teóricos establecemos el conjunto de conexiones lógicas que les corresponden (que denominaremos proposiciones teóricas) y hacemos explícitos los pasos que ligan los conceptos con las variables estadísticas (es decir, con los índices e indicadores) entonces a cada relación entre dos o más conceptos corresponderá una relación análoga entre dos o más variables (que en este texto denominamos proposición empírica).

El problema del vínculo entre la teoría y el análisis estadístico (limitado al análisis de asociación) se manifiesta de dos maneras: primero, hay una dificultad lógica para conectar las aseveraciones referidas a un cuerpo teórico, pero expresadas en lenguaje natural con lo que hemos denominado proposiciones teóricas, y segundo, hay que establecer las conexiones entre estas últimas y las proposiciones empíricas que expresan relaciones entre variables. Una vez que se ha construido la proposición empírica hay que proceder a estudiar su articulación con el análisis estadístico de la información. Es exactamente este punto el que constituye el tema central de este libro. Hemos sostenido que en el proceso de investigación es usual que haya múltiples puntos de contacto entre la teoría y el manejo estadístico de la información, lo que entra en abierta contradicción con la idea ampliamente extendida de que la investigación empieza con el desarrollo del marco teórico y culmina con el análisis de la información. En efecto, nuestra imagen del proceso es muy diferente; para nosotros la investigación empírica (es decir, el tipo de investigación que incorpora a los indicadores e índices provenientes de la realidad de manera sistemática y no a título de mero ejemplo) avanza sobre la base de la discordancia entre teoría e información.

El desacuerdo entre las proposiciones teóricas y las empíricas correspondientes, así como las inconsistencias entre éstas y las observaciones, conduce a un replanteamiento de las relaciones que a veces implica la necesidad de mayores refinamientos teóricos otras un reexamen del proceso que ligó a los conceptos con sus variables y en otras oportunidades la construcción de nuevos datos estadísticos. También suele acontecer que se pongan en cuestión simultáneamente todos los procesos involucrados en la investigación. En nuestra concepción el vínculo que liga teoría y realidad es esencialmente dinámico. El orden de la exposición de los resultados de una investigación no constituye su historia real. Es una costumbre generalizada situarlos en el punto terminal de una cadena que se inicia en el marco teórico. Se empieza por la teoría y se culmina en la presentación de la evidencia; sin embargo, el proceso de investigación está lleno de avances y retrocesos, de caminos sin salida, de exploración de nuevas vías y del cierre de otras. La investigación raras veces avanza linealmente; los obstáculos que hay que vencer no siempre se traducen en adelantos sino que la mayoría de las veces implican rodeos que, en el mejor de los casos, exigen un mayor grado de problematización teórico y metodológico. Es por ello que si deseáramos representarlo gráficamente deberíamos recurrir a una espiral.



Con el objeto de precisar algunas nociones centrales involucradas en la relación entre la teoría y el análisis estadístico de la información, acéptese que es posible detener la espiral de la investigación en un instante. En ese momento estarán presentes con mayor o menor énfasis los componentes del siguiente diagrama: C9 C3 C1 C6 '6 C4 C2 C7 C'7 C5 C8 C'8 A la izquierda de la doble línea vertical hemos representado el campo de la teoría y a la derecha el de la realidad. En el primero denotamos con una letra C los conceptos que forman parte de la teoría. Las líneas continuas que los unen, representan las relaciones constitutivas de la teoría, que para los propósitos de este trabajo nos basta con concebir como un conjunto de conceptos teóricos relacionados. Generalmente, entre el cuerpo teórico y la realidad se encuentran los conceptos observables u objetivados simbolizados por C'. Nos interesa remarcar algunas características de la gráfica para formarnos una idea del conjunto de pasos intermedios para conectar los conceptos teóricos con la información proveniente de la realidad:

- 1) En el dominio de la teoría hay una serie de conceptos que no pueden observarse o que no es posible poner en correspondencia con lo empírico; tal es el caso de C1, C2, C3, C4 y C5, en tanto que hay otros que sí son observables ya sea directamente (como C9) o a través de conceptos observables relacionados con ellos (como C6, C7, y C8,).
- 2) Estos últimos sólo son susceptibles de ponerse en correspondencia con la información a través de su objetivación en conceptos observables (C'6, C'7, C'8). El vínculo entre un concepto teórico y su correspondiente objetivado se representa con una línea punteada.
- 3) El concepto observable (que lo puede ser directamente o bien uno objetivado) se liga con la realidad a través de indicadores. A partir de este tipo de conceptos se recorta un segmento de la realidad y de la interacción surge la información. La doble línea representa, por una parte la forma como se observa la realidad de determinado concepto y por la otra los indicadores, que son el producto de la mirada conceptual sobre la realidad. La información estadística (susceptible de tratarse con el análisis de asociación) es el resultado del registro empírico de los indicadores.
- 4) Hay que notar que la noción de indicador se refiere a un elemento que surge de la realidad y que está vinculado a un concepto observable. Sus propiedades son que existe en

el mundo empírico y que es susceptible de registrarse. Adicionalmente hay que tomar en cuenta que todo indicador forma parte de una relación. La materia prima para la estadística son los indicadores, a los que se denomina variables. En primera instancia se podría pensar que ambas palabras son sinónimas, pero no sería del todo correcto. En efecto, la estadística, con su pretensión de generalidad, es decir, de servir como instrumento de cualquier ciencia sin apegarse a ninguna en particular, no considera el componente relacional del indicador. Esto quiere decir que para utilizar el conocimiento estadístico en el proceso de la investigación en general, y de la social en particular, es necesario llenar de contenido teórico a las variables. El buen uso de la estadística exige diferenciar los conceptos de las variables y establecer con claridad la naturaleza de su relación. Hasta ahora hemos supuesto implícitamente que a cada concepto le corresponde sólo un indicador, pero en la investigación social es lo habitual que a un concepto se le asocie más de un indicador. Si éste fuera el caso, para vincular a cada concepto con una medida única habría que proceder a construir unos índices que resumieran los indicadores. Si bien dentro del arsenal de conocimiento técnico a disposición del científico social se hace una clara distinción entre indicador e índice ésta no tiene lugar en la estadística, para la cual sólo existe el término variable. Éste engloba tanto a los índices como a los indicadores.

5) Una idea general que se desprende de la gráfica es que las variables solo tienen sentido si están relacionadas con un cuerpo teórico. Los indicadores o los índices son referentes empíricos de los conceptos observables. Las variables que reflejan propiedades observables de las unidades para las cuales se registra la información sólo tienen pleno sentido cuando se llenan de contenido conceptual. Hay que recordar que el propósito central de este libro es situar el análisis de asociación en los puntos de intersección entre las operaciones teóricas y empíricas involucradas en el proceso de la investigación social. Con el único objetivo de definir y perfilar nociones tales como variable, indicador, índice e información recurrimos al supuesto de que es posible detener el proceso de investigación y distinguir analíticamente sus componentes. Con este artificio se pretende introducir la idea de una precedencia temporal en el proceso de investigación que se originaría en el campo de la teoría. Sin embargo, esta concepción empieza a desvanecerse cuando reincorporamos la dinámica por medio de la idea de una espiral que en sus distintos niveles tendrá momentos más cercanos a lo empírico y otros más próximos a lo teórico. Además nuestra separación entre lo teórico, lo técnico y la información no es en absoluto nítida cuando se investiga.

En cada acto de investigación se encuentran siempre presentes, aunque con pesos diferentes, las consideraciones teóricas, las metodológicas y las técnicas. A lo largo de todo el texto nos preocuparemos por destacar cómo las proposiciones empíricas orientan el análisis estadístico de la información. Esto quiere decir que sólo subsidiariamente nos dedicaremos a mostrar el modo específico como se relacionan las proposiciones empíricas con las teóricas. También significa que analizaremos en abstracto el punto en que se establece el vínculo entre el mundo de la teoría y el de la información. Por abstracto queremos significar no referirlo a ningún momento preciso dentro del proceso de investigación. Enfatizamos el análisis del modo en que la proposición empírica orienta el estudio estadístico de las relaciones entre variables independientemente del lugar de la espiral en que se encuentre el proceso. Hechas estas aclaraciones es necesario agregar una más. Para dar contenido sustantivo a los desarrollos estadísticos hemos tomado como ejemplos partes de una serie de investigaciones ya publicadas. El material está organizado de acuerdo con la lógica de la exposición de sus autores, siéndonos imposible recuperar la de la investigación. En consecuencia, los ejemplos parten del análisis de las proposiciones teóricas para establecer (sin mucha discusión) las proposiciones empíricas y de ahí en adelante se continúa examinando las consecuencias lógicas de éstas últimas sobre las distribuciones de frecuencias. El orden con que se presentan los ejemplos no se debe a que supongamos que en el proceso de investigación lo empírico esté determinado por lo teórico, sino que se justifica en gran medida porque es la óptica que nos pareció razonable para incorporar el material de investigación que sirve de base a los desarrollos propiamente

estadísticos y porque, además, es una manera simple de introducir al estudioso de la estadística en las complejidades del uso de esta técnica en la investigación en ciencias sociales.

Tomando en cuenta las anteriores consideraciones y con el propósito de dar contenido a las ideas expuestas veamos un par de ejemplos. Nun plantea que en el volumen I de *El capital*, Marx investiga las relaciones entre capital y trabajo en general con el propósito de arrojar claridad sobre las leyes de funcionamiento de la forma más pura del modo de producción capitalista y así determinar por qué y cómo la acumulación capitalista produce constantemente una población obrera excedente. En el capítulo 23, Marx teoriza sobre el funcionamiento del mercado de trabajo en una economía de libre competencia y "discierne una primera tendencia al rápido crecimiento del proletariado industrial; dada su forma, una segunda tendencia a liquidar las diferenciaciones de la fuerza de trabajo que habían sido propias del periodo de la manufactura y a promover, por tanto, una homogeneidad obrera cada vez mayor en términos de calificación y de ingreso" (p. 57). Nun continúa su desarrollo señalando que con estas previsiones Marx describió acertadamente las características que asumió el mercado de trabajo en Inglaterra y los Estados Unidos en el último cuarto del siglo XIX, pero que en esa época se inicia la tendencia hacia la concentración y la centralización del capital, que también Marx pronosticó.

Después de referirse a una serie de estudios, Nun nos dice que: El resultado de todo esto es una verdadera reversión de aquellas tendencias iniciales al crecimiento y a la homogeneización de un proletariado fundamentalmente industrial: a la vez que cae su peso relativo en el conjunto de la población económicamente activa, esta última aparece cada vez más fragmentada y diferenciada, de manera análoga a lo que sucediera en el periodo de la manufactura (p. 60). Dice que es a raíz de estos fenómenos que han surgido una serie de estudios que, en oposición a la ortodoxia neoclásica, procuran distinguir la existencia en Estados Unidos de un mercado de trabajo primario y uno secundario. Nun dice que para varios autores este dualismo aparece ligado al corte entre los monopolistas y el resto de los empleadores y que hay suficiente evidencia de que tienen diferentes comportamientos en el mercado de trabajo, por lo menos en lo que se refiere a salarios y beneficios sociales (p. 60). Una vez que ha marcado los cambios de tendencias en el mercado de trabajo concomitantes al pasaje del capitalismo de su estadio competitivo al monopolístico y que ha documentado la segmentación del mercado, Nun describe cómo el capitalismo penetró en el agro de los Estados Unidos; agrega que si bien es cierto que la mayoría de los trabajadores del sector secundario pueden definirse como un excedente respecto del mercado primario, estas ocupaciones no son supervivencias de la organización social anterior sino un emergente del nuevo estadio del desarrollo capitalista (p. 61). A continuación entra de lleno al tema central de su trabajo. Lo reproduciremos textualmente para disponer de la base que apoye nuestra lectura particular: Según las ramas, el uso capitalista de la maquinaria continúa sometiéndolo al empleo industrial a mercados vaivenes. Pero, ¿cuál es la configuración de la modalidad flotante de la superpoblación relativa a una industria fuertemente concentrada y dispensadora de ocupaciones predominantemente "primarias"?

La disminución proporcional de la fuerza de trabajo fabril en el conjunto de los asalariados no impide que haya aumentado su volumen absoluto: así, éste creció alrededor de 13% entre 1960 y 1972. ¿Vuelven entonces a incorporarse a la industria los obreros desplazados por las fluctuaciones cíclicas o por los progresos de la mecanización? En otras palabras, la característica saliente de la modalidad flotante sigue siendo la analizada por Marx, con todas sus implicaciones potenciales en términos de solidaridades obreras, de acumulación de experiencias y de reivindicaciones a nivel de la fábrica, y, presumiblemente ahora, de cristalización de una aristocracia del trabajo en el sector monopolista? Nuevamente, es difícil dar una respuesta comprensiva y, en todo caso, la literatura que conozco no la provee. Braverman (1974; 386-387) considera que el mecanismo de expulsión/absorción opera hoy en áreas geográficas cada vez más extendidas, de manera que la cesantía obliga

a menudo al obrero a "romper los lazos con la localidad y con la comunidad".

Pero también sugiere que buena parte de la mano de obra desalojada por "la rápida mecanización de la industria" no reingresa a plantas fabriles sino que suministra brazos baratos a los llamados "servicios", o sea, que pasa al mercado de trabajo secundario. Es muy probable que esta última generalización resulte tanto más válida cuanto menor sea el nivel de calificación del despedido; y es igualmente conocida la incidencia que tienen sobre las posibilidades de reemplazo variables individuales como la edad, la educación formal, el sexo y el origen étnico (pp. 61-62). Después de exponer dos estudios sobre el tema de la superpoblación flotante, realizados en un país central (la industria automovilística inglesa) y en uno periférico (el petróleo en Trinidad y Tobago), y mostrar las diferencias, estudia el caso de la superpoblación flotante en la industria automovilística argentina. Al centrar el interés en las tendencias que muestra el mercado de trabajo en la fase del capitalismo monopólico deberíamos esperar un cierto grado de heterogeneidad en la fuerza de trabajo, como consecuencia de la reversión de las tendencias en la etapa competitiva del capitalismo, así como de los procesos de centralización, concentración y segmentación del mercado. Para observar la homogeneidad/heterogeneidad de la fuerza de trabajo Nun toma como indicador el grado de calificación que tenían al momento del despido los obreros de algunas plantas automovilísticas de Buenos Aires, usa el tipo de ocupación al momento del estudio como indicador de la segmentación del mercado de trabajo. La calificación laboral origina una variable que clasifica a los obreros despedidos según tengan o no calificación (variable dicotómica), mientras que la ocupación distingue entre ocupaciones en la industria y en los servicios.

Las primeras se abren en ocupaciones dependientes en plantas industriales, grandes, que proveen los puestos estables y bien remunerados, y en plantas pequeñas en las que se pagan salarios bajos y los empleos son poco estables. El primer tipo de ocupación es el característico del mercado de trabajo primario y el segundo el del secundario. Si interesase analizar por separado los procesos de segmentación del mercado de trabajo y el de homogeneidad/heterogeneidad de la fuerza de trabajo bastaría con analizar la información de cada indicador también por separado. Sin embargo, en el último párrafo de la larga cita de Nun se plantea la hipótesis de que mientras mayor sea la calificación del obrero despedido, más probable es que reingrese a las plantas fabriles. Esto quiere decir que al nivel de los conceptos se plantea una clara relación entre la heterogeneidad de la fuerza de trabajo y su acceso a los mercados; es esta relación la que permite comprender el estatus del cuadro 2 del estudio que nos sirve de ilustración.

Como por el momento sólo estamos interesados en destacar la relación entre los conceptos y sus indicadores y en señalar cómo los vínculos teóricos especifican las correspondientes asociaciones entre las variables estadísticas nos limitaremos únicamente a presentar la estructura de una tabla que se deriva de los cuadros 2 y 4 del trabajo de Nun (pp. 74 y 77):

TABLA I Ocupación actual, por calificación en empresa de origen

TABLA I

Ocupación actual, por calificación en empresa de origen

	Calificación	
	No calificados	Calificados
Ocupaciones dependientes en:		
1. Industria		
I.1. Plantas grandes		XXXXXXXXXX
1.2. Plantas pequeñas	XXXXXXXXXX	
2. En servicios	XXXXXXXXXX	

Es a partir de estas consideraciones que Marín se preocupa por el análisis empírico de las proposiciones ideológicas dominantes. Basa su estudio en información periodística que le permite dar cuenta de 8 509 hechos armados acontecidos en Argentina de mayo de 1973 a marzo de 1976. A pesar de que somete a contrastación varias proposiciones ideológicas nosotros sólo incluiremos a título de ilustración la que subyace al cuadro 3 del mencionado estudio. La primera proposición que analiza (proposición teórica) establece una relación entre fuerza social y tipo de violencia. El primer concepto lo objetiva en el de acción social y el segundo en acción contra el cuerpo. De manera que la proposición empírica consistente con la proposición ' teórica sostendría que existiría una relación entre la acción social y la acción contra el cuerpo.

El concepto observable acción social lo dicotomiza en acción subversiva y antsubversiva y la acción contra el cuerpo la divide en dos categorías: las que producen bajas y las que no las producen. La proposición ideológica concreta se plantea al nivel de los indicadores: la gran mayoría de los hechos que producen las fuerzas subversivas son hechos con muertos y heridos; en cambio, las fuerzas legales concentran la mayoría de sus hechos armados deteniendo y no matando. La proposición empírica conduce a la construcción de la siguiente tabla:

TABLA 2

Total de hechos armados producidos en el periodo de mayo de 1973 a marzo de 1976, distinguiéndolos según produzcan bajas o no

	Hechos armados que producen bajas	Hechos armados que no producen bajas
Total de hechos armados producidos por la acción "subversiva"	XXXXXXXXXX	
Total de hechos armados producidos por la acción "antsubversiva"		XXXXXXXXXX

En la tabla hemos marcado con la letra X la distribución esperada de las observaciones de acuerdo con las proposiciones planteadas. En este libro aplicaremos constantemente la lógica que hemos seguido en los ejemplos. Estableceremos las proposiciones teóricas y las relaciones con sus correspondientes proposiciones empíricas y a partir de estas últimas se derivará la distribución esperada de las observaciones que contrastada con la distribución observada, nos entregará criterios para juzgar el grado de concordancia entre la teoría y los datos.