

Capítulo 1

EL ANALISIS ESTADISTICO DE LOS DATOS SOCIOLOGICOS

1.1. INTRODUCCIÓN

La estadística, en términos amplios y generales, puede considerarse como la ciencia de las regularidades que se observan en conjuntos de fenómenos naturales. Puede considerarse a la estadística también como la colección de métodos científicos que permiten el análisis e interpretación de la información numérica. El campo de la estadística tiene aplicaciones cada vez más amplias en una diversidad de disciplinas, desde la ingeniería a la zoología, desde la sociología a la economía de la empresa. Con todo, esta joven disciplina, en plena expansión, no es popular ni entre los estudiantes de ciencias humanas y sociales ni entre muchos profesionales de tales ciencias. La imagen de la estadística como una rama de las matemáticas, que sirve para manipular números y que es de difícil comprensión, por su carácter abstracto, va acompañada, sin embargo, de otra imagen, un tanto contradictoria con la primera, que hace de la estadística una especie de instrumento mágico que permite que los estudios o investigaciones con un regular aparato estadístico sean o aparezcan más «científicos». Quizá porque muchas personas tienen una actitud previa claramente temerosa hacia las matemáticas y todo aquello que signifique cuantificación, el estudio de la estadística como asignatura básica en los programas de ciencias sociales no suele ser bien recibido.

Este síndrome de actitudes e imágenes poco favorables hacia la estadística se produce en un contexto social en el que la información cuantitativa invade los aspectos más íntimos de la vida cotidiana. Vivimos inmersos en un mar de cifras, sobre los temas más diversos: paro, divorcio, natalidad, enfermedades, gasto público, salario mínimo, inflación, etcétera. De ahí que cada vez sea más necesario para el estudiante y estudioso de las ciencias humanas y sociales poseer unos conocimientos básicos y rigurosos sobre el contenido y alcance de la estadística, que le permitan comprender y evaluar apropiadamente esa realidad social que se presenta abrumadoramente cuantificada. Pero antes de pasar a

estudiar tales conceptos básicos se hace preciso que realicemos unas consideraciones iniciales de carácter histórico, que nos ayuden a situar la estadística con la perspectiva suficiente en el marco del desarrollo científico contemporáneo. Sólo así, quizá, podemos entender el alcance y las limitaciones de los métodos estadísticos para el trabajo e investigación en las ciencias humanas y sociales.

1.2. NOTAS HISTÓRICAS SOBRE LOS MÉTODOS ESTADÍSTICOS

No se puede situar con precisión el origen histórico de la estadística, y ello es así porque, como señala acertadamente Kendall (1978, pág. 1093), en la corriente de conocimientos que constituyen en la actualidad el amplio campo de la estadística teórica, confluyen diversas aportaciones que han formado, a lo largo de los dos últimos siglos, lo que es hoy la disciplina. Los conceptos de la probabilidad surgen de los juegos de azar; la necesidad de reclutar soldados y obtener dinero para financiar la guerra lleva a los gobiernos a obtener sistemáticamente datos estadísticos; los seguros marítimos se originan como reacción ante la acción imprevisible de la piratería en el Mediterráneo; los estudios estadísticos sobre la mortalidad son provocados por las plagas, que tantas vidas se cobran en Europa a lo largo del siglo XVII; los astrónomos contribuyeron con la teoría de errores; la teoría de la correlación se origina en los estudios biológicos, como será la agronomía la que propicie la teoría del diseño experimental; la teoría de las series temporales surge de los estudios de series de datos en economía y meteorología. Finalmente, la psicología y la sociología también contribuyen al desarrollo contemporáneo de la estadística, por medio de la teoría del análisis de componentes y de la teoría del chi-cuadrado, respectivamente.

Pero, a pesar del amplio desarrollo que tiene en la actualidad la estadística, no será hasta alrededor de 1850 cuando la palabra «estadística» comience a utilizarse en el sentido amplio que tiene ahora. La primera vez que aparece impresa la palabra «estadística» es en un libro editado en Londres en 1770. Aparentemente, proviene del latín *status* o estado. En sus orígenes, el concepto de «estadística» aparece ligado con la actividad gubernamental, y el término «estadístico», con el de estadista o político. Y es que, en efecto, el primer y principal uso de las estadísticas fue realizado por los gobernantes, que deseaban conocer la extensión de sus dominios, la población residente en ellos y la cantidad de impuestos que podían esperar obtener de dicha población. Ahora se continúa utilizando el término «población» en estadística, pero para referirse de una forma genérica al conjunto de unidades individuales o elementos —sean o no seres humanos— en un grupo, clase o categoría que se estén estudiando.

En la actualidad, la teoría de la estadística va unida en buena medida a la teoría de las probabilidades, pero durante mucho tiempo esto

no fue así. Los jugadores de dados, naipes y otros juegos de azar en la antigua Grecia y Roma ya habían observado las regularidades que se presentan al lanzar repetidas veces, por ejemplo, un dado. Más tarde se descubrieron regularidades en otros campos. La astronomía fue la primera ciencia en desarrollar la idea de las regularidades en la medición de los fenómenos. Galileo (1564-1642) realizó importantes contribuciones a las leyes de la probabilidad, como lo hicieron Pascal (1623-1662) y Fermat (1601-1665), y, ya en el siglo XVIII, Simpson descubrió la distribución continua, mientras que, a finales de siglo, Laplace y Gauss descubrieron la más importante de las distribuciones: la distribución normal.

Este retraso en la formulación matemática de las regularidades de los sucesos, que habían sido detectadas desde hacía tanto tiempo, se explica entre los modernos historiadores de la estadística (Kendall, 1978; Kruskal, 1978; Lazarsfeld, 1961) por las ideas religiosas y filosóficas imperantes hasta el período renacentista en el mundo occidental. Desde la óptica cristiana, como antes desde la óptica pagana, cada cosa ocurría por voluntad divina y, en tal sentido, no había azar; resultaba, pues, impío o irreverente suponer que los fenómenos obedecían a leyes de probabilidad. Por todo ello, hasta que Europa no superó la teología y filosofía medievales no fue posible desarrollar el cálculo de las probabilidades.

Otro paso importante lo dieron Quetelet y otros cuando introdujeron el concepto de que el mundo viviente, y no sólo el mundo inorgánico de los dados, naipes o estrellas, también ofrecía distribuciones de frecuencia. Desde finales del siglo XIX comienzan a descubrirse regularidades en disciplinas tan diversas como la genética, biología, meteorología, economía, psicología, sociología e incluso en las artes. Los conceptos de distribución de frecuencias y de distribución de probabilidades se utilizan ya intercambiamente, al confluir definitivamente la teoría clásica de la probabilidad y la estadística como teoría de la información numérica.

Por otro lado, conviene señalar el papel tan fundamental que han jugado la demografía y las compañías de seguros en el desarrollo de la estadística. En la segunda mitad del siglo XVIII comienzan a recogerse en algunos países los primeros censos de población. En España, el primer censo de población se refiere al año 1860. El esfuerzo que supone obtener una información censal, y el carácter periódico que los censos pronto adquieren, significaron un magnífico impulso para crear una importante infraestructura administrativa y burocrática de índole estadística. Además, la extensión de los seguros de vida y la necesidad de construir tablas de esperanza de vida convirtieron a aquellos en una actividad científica exacta.

El período que transcurre desde la última década del siglo XIX hasta el comienzo de la Segunda Guerra Mundial significa el definitivo desarrollo y consolidación de la teoría estadística. Galton, Pearson, Yule y,

más tarde, Fisher propician un gran despliegue de la estadística en el campo de la ciencia. La curiosidad científica y la fe en el progreso humano conducen a muchos científicos, matemáticos y estadísticos a buscar regularidades en la vida humana y social. La ciencia estadística era, para aquellos ilusionados y curiosos prohombres, un nuevo y excitante instrumento para la exploración del mundo viviente; de este modo, las ciencias de la conducta comenzaron a dar signos de que sus estructuras admitían el análisis matemático.

Es así como se desarrolla el estudio de las relaciones entre variables, y se introducen los conceptos de asociación, contingencia, correlación y regresión, primero en el campo de la biología y, después, en el resto de las ciencias. Con el desarrollo de la teoría de las muestras se hace posible realizar estimaciones precisas de los valores, lo que significó un importante impulso para el desarrollo de las ciencias sociales. Los problemas de la estimación, de la prueba de decisiones estadísticas y el contraste (prueba o test) de hipótesis, son áreas que experimentan un fuerte avance motivado por las necesidades científicas de las ciencias sociales.

Como señala Kendall (*op. cit.*, pág. 1098), fue tanto lo que ocurrió en el campo de la estadística entre 1920 y 1940, que se hacía preciso un período de reposo para asimilar todo lo descubierto. La moderna estadística, que es sobre todo inglesa en su etapa de máximo desarrollo, es admitida por todos los científicos después de la Segunda Guerra Mundial. Las Administraciones públicas y privadas y, sobre todo, los gobiernos pueden recurrir ya con seguridad al campo de la estadística para encontrar soluciones a los problemas de índole numérica que plantea una vida social y pública cada vez más complejas. En los momentos actuales, la estadística y los estadísticos se encuentran firmemente establecidos en la vida académica y en el mundo de las burocracias públicas y privadas, a la vez que se consolidan los campos de estudio iniciados por los Pearson, Yule, Fisher, etc., y se desarrollan nuevas áreas de estudio, tales como el análisis secuencial, el análisis multivariable, los métodos aparamétricos y otros, todos ellos aplicados también al tratamiento y análisis de datos sociológicos, como tendremos ocasión de ver a lo largo del presente libro.

1.3. FUNCIONES DE LA ESTADÍSTICA: ESTADÍSTICA DESCRIPTIVA Y ESTADÍSTICA INTERFERENCIAL

Si la estadística se define, tal como se está haciendo aquí, de una manera amplia, esto es, como una teoría de la información numérica, resulta difícil oponerse a su utilidad, y necesidad. Con todo, la disciplina ha despertado, y continúa despertando, bastantes críticas, basadas muchas veces en la propia concepción que el crítico tiene de la estadística, aunque dicha concepción no tenga que ver mucho con las fun-

ciones que tanto de una forma teórica como aplicada, realiza la estadística.

Quizá muchos críticos de la estadística comparten el punto de vista que el siglo pasado formulara el famoso político conservador, primer ministro de la Reina Victoria, Benjamín Disraeli, que dijo en cierta ocasión que habían tres clases de mentiras: «mentiras, mentiras despreciables y estadísticas» (en inglés, *lies, damned lies, and statistics*). Y es que, en efecto, se pueden utilizar, como de hecho se utilizan, datos numéricos —«estadísticas»— para apoyar razonamientos falsos. Pero esto es así no porque los números mientan, sino porque se hace un mal uso —intencionado o no— de la lógica que permite la descripción cuantitativa y sintética de una población, y la inferencia de propiedades de dicha población a partir de la observación sistemática de una muestra de la misma.

Como señala Kruskal (*op. cit.*, pág. 1073), la vida no es estable o simple, aunque hay aspectos de ella que sí lo son. La ciencia trata, en general, de elucidar estos últimos y la estadística ofrece métodos generales para encontrar pautas recurrentes y construir modelos que faciliten su estudio. Siguiendo a este mismo autor, podemos enumerar las siguientes funciones generales que realizan los métodos estadísticos:

- a) Resumen de los datos y extracción de información relevante de los mismos, esto es, de las mediciones observadas, sean éstas numéricas, clasificatorias, ordinales, o de otro tipo;
- b) búsqueda y evaluación de los modelos y pautas que ofrecen los datos, pero que se encuentran ocultos por la inherente variabilidad aleatoria de los mismos;
- c) contribuir al diseño eficiente de experimentos y encuestas;
- d) facilitar la comunicación entre los científicos, ya que siempre será más fácil comprender la referencia a un procedimiento estándar, sin necesidad de mayor detalle.

Aparte de estas cuatro funciones básicas, Kruskal señala la existencia de otras dos que, desgraciadamente, facilitan la crítica contra la estadística. Una sería la de suministrar el requisito de «autenticación científica» —así, por ejemplo, hay editores de revistas y directores de tesis que insisten en la utilización de ciertos procedimientos estadísticos, sean o no apropiados— y otra es la de tratar de ofuscar, mitificar o impresionar —así, por ejemplo, hay publicaciones de trabajos sociológicos que contienen tal masa, no digerida, de datos estadísticos que no sirven para otro objetivo que para mostrar lo mucho que ha trabajado su autor—. Para prevenirse contra tales usos de la estadística, Kruskal recomienda, por un lado, la utilización por los científicos estadísticos de un código ético que evite abusos de dicha índole, y por otro lado, que el científico empírico sólo utilice aquellos méto-

dos estadísticos de los que comprende bien su lógica subyacente, renunciando a emplear aquellos otros cuya lógica no domina.

Las anteriores funciones conducen a la división de la estadística en dos partes bien diferenciadas. Por un lado, la *Estadística Descriptiva*, que consiste en el conjunto de instrumentos y temas relacionados con la descripción de colecciones de observaciones estadísticas —se refieren tanto al total de la población como a una muestra de la misma—, y por otro lado, la *Estadística Inferencial o Inductiva*, que se ocupa de la lógica y procedimientos para la inferencia o inducción de propiedades de una población en base a los resultados obtenidos de una muestra conocida. Como señala Blalock (1960, pág. 5), la estadística inferencial demanda un tipo de razonamiento mucho más complejo que el que se emplea en la estadística descriptiva, pero cuando se comprende y emplea adecuadamente, se convierte en una herramienta importante para el desarrollo de una disciplina científica. La estadística inductiva se basa, como veremos con mayor detalle en los capítulos siguientes, en la teoría de las probabilidades que, como se sabe, es una rama de las matemáticas actuales. Es de esta forma como una disciplina deductiva, las probabilidades matemáticas, suministra la base racional para un razonamiento inductivo, la estadística inferencial. No existe otra base racional, desde nuestro punto de vista, para el razonamiento inductivo que aspire a ser científico.

Algunos autores no distinguen entre las dos ramas mencionadas de la estadística y quizá desde un punto de vista estricto no exista tal división. Así, Kruskal (1978, pág. 1072) considera que la estadística es la disciplina que se ocupa de los problemas de inferencia a partir de un conjunto de datos empíricos. De este modo, la estadística suministra los instrumentos que permiten el salto desde las observaciones a la inferencia o decisión, siendo la descripción y resumen de las observaciones un paso previo para la operación fundamental de la estadística, la inferencia. Pero desde nuestro punto de vista, y aun aceptando esta definición de la disciplina, nada impide que a efectos de una presentación pedagógica de la utilización de las técnicas estadísticas en la investigación sociológica de tipo empírico, se diferencien claramente los procedimientos estadísticos descriptivos de lo que son, formalmente, inductivos o inferenciales.

Y eso es lo que hemos hecho en el presente libro. Primero, presentar los instrumentos de análisis estadístico-descriptivo, para pasar en posteriores capítulos a tratar los problemas que comporta la inducción de propiedades a partir de los resultados obtenidos sobre muestras representativas. Conviene, a este respecto, realizar una aclaración terminológica. Un segundo significado técnico de la palabra *estadística* hace referencia a cualquier descripción de una *muestra* de observaciones estadísticas. Mientras que si el conjunto de observaciones estadísticas que se describen es una *población* estadística, en tal caso una des-

cripción de tales observaciones se denomina un *parámetro*. Así, por ejemplo, el promedio de años de escolaridad que tiene la población española comprendida entre los quince y los veinte años de edad, según datos del Censo de Población de 1980, se puede tratar como un *parámetro* poblacional, mientras que el promedio de años de escolaridad que tiene la muestra de jóvenes estudiada en la Encuesta de la Juventud (Linz, 1978), puede ser tratado como un *estadístico*. Obsérvese que un parámetro y un estadístico pueden consistir en el mismo tipo de descripción —un promedio, en el ejemplo anterior—; lo que los diferencia es el objeto que describen ambos, esto es, una población o una muestra, respectivamente.

1.4. RELACIONES ENTRE LA ESTADÍSTICA Y LA SOCIOLOGÍA: CONCEPTOS BÁSICOS

El papel de la estadística en el proceso de investigación sociológica está claramente determinado. La estadística se utiliza para operar con números, que reflejan valores de mediciones que se supone satisfacen determinados supuestos. Esto es, las consideraciones estadísticas se introducen tan sólo en la fase analítica del proceso de investigación, después de haber obtenido los datos frecuentemente a partir de una muestra. En un sentido estricto, no es competencia del estadístico el diseño de un cuestionario, o la construcción de escalas de medición a partir de valores concretos de las variables. Estos son problemas metodológicos del diseño de la investigación, que deben de plantearse, y resolverse, a partir de una concepción clara y del establecimiento de un marco teórico pertinente.

Si el problema de investigación que nos ocupa no está teóricamente bien definido, de poco servirá la utilización de un gran aparato estadístico, ya que los resultados no van a mejorar por ello. La estadística hay que considerarla como un auxiliar en el proceso de investigación, un auxiliar ciertamente imprescindible y que cuando es utilizado correctamente, conduce a la utilización más detallada de la teoría y a la elaboración más precisa del modelo que se va a seguir en la investigación. Por supuesto, las cosas no pueden separarse con tanta precisión analítica en la realidad de la investigación social, que debe contemplarse como un proyecto unitario y global, aunque a efectos de iniciarse en el progresivo aprendizaje del trabajo científico, resulta conveniente que el estudiante aprenda a valorar el papel y funciones que desempeñen los diferentes instrumentos teóricos y metodológicos que se utilizan.

Debe, pues, quedar claro desde un principio para el estudiante de sociología, que la estadística es siempre una buena ayuda, pero nunca un sustituto, para un buen razonamiento teórico y un buen quehacer metodológico. En la historia de la sociología como ciencia, el desarrollo

del proceso de cuantificación ha jugado un papel ciertamente importante en la configuración de la sociología como una disciplina científica. Los fundadores de la sociología, desde Marx a Durkheim y Weber, pasando por Quetelet y Le Play, reconocieron la importancia de la obtención de información cuantitativa relevante sobre los fenómenos sociales, y de su tratamiento estadístico, para construir una ciencia sobre la sociedad*. Durkheim fue de los primeros en reconocer el fundamental papel de la estadística para «aislar apropiadamente los hechos sociales de los hechos particulares». En los promedios, en los porcentajes sobre natalidad, matrimonios, suicidios, etc., veía Durkheim, creo que acertadamente, el hecho social, separado «de todo lo que está mezclado en él». Durkheim expresó en bellas palabras la importancia metodológica de la estadística para el desarrollo de la ciencia de la sociedad, cuando afirmó que «la estadística expresa cierto estado del alma colectiva» (Durkheim, 1972, pág. 38). Por supuesto, Durkheim se estaba refiriendo, al hablar de estadística, a los datos estadísticos y a lo que actualmente denominaríamos indicadores sociales, como elementos de información necesarios para una precisa conceptualización en Sociología, y no tanto a los instrumentos de análisis estadístico que forman parte rutinaria hoy en día de la investigación empírica de nuestra disciplina.

En cierto modo, la situación que se planteaba a los primeros sociólogos de finales del siglo XIX y principios del XX, era en buena medida opuesta a la actual. Disponían de poca información estadística para poder contrastar sus teorías, pero algunos de ellos estaban repletos del genio creador que les condujo a establecer las bases teóricas de la Sociología. En la actualidad ocurre lo contrario. La información cuantitativa sobre los hechos sociales acumulada en los bancos de datos sociológicos es muy importante, se encuentra en pleno desarrollo el movimiento de los indicadores sociales**, pero los avances teóricos son muy pobres. Los sociólogos actuales, inmersos en un mar de información cuantitativa, encuentran dificultades muy grandes en ordenar conceptual y teóricamente la disciplina. Sin embargo, es mi opinión que los nuevos desarrollos de la estadística, como por ejemplo el análisis

* Para un tratamiento más detallado de la historia de la cuantificación en la sociología y de los primeros intentos de Marx, Weber y Durkheim para trabajar con datos cuantitativos, véase mi trabajo: Manuel GARCÍA FERRANDO, *Sobre el método*, Madrid, C.I.S., 1979, especialmente las págs. 27-71.

** Los indicadores sociales, como mediciones repetidas en el tiempo de diversos fenómenos sociales, son cada vez más utilizados no sólo por los científicos sociales, sino también por los políticos y los planificadores. Desde mediados de los años setenta, los gobiernos de diversos países han comenzado a publicar, de forma regular, informes basados en indicadores sociales. La oficina del censo de los Estados Unidos publica desde 1973 un informe titulado *Social Indicators*. En España, el Instituto Nacional de Estadística ha comenzado a publicar, por primera vez en 1975, una *Panorámica social*. Para un tratamiento actualizado de la cuestión se recomienda la lectura del libro de ROSI y GILMARTIN *Hand-book of Social Indicators*, New York, Garland STPM Press, 1980.

sis multivariable, una vez se generalice su uso entre los sociólogos, va a facilitar enormemente la necesaria tarea de síntesis que en la actualidad requiere la sociología. En capítulos posteriores de este libro, trataremos con mayor detalle dichos problemas. Ahora, continuemos con la discusión y presentación de los conceptos básicos de índole estadística aplicados a la sociología.

En un cuestionario diseñado para estudiar actitudes y opiniones generales de la población, se suelen incluir preguntas sobre la edad, el nivel de educación, religiosidad, preferencias ideológicas y otras cuestiones que hacen referencia a las actitudes y opiniones objeto de estudio. Cada una de tales preguntas se formula para medir una propiedad o característica de cada una de las personas entrevistadas. A propiedades tales como la edad, la preferencia ideológica o la opinión sobre un tema determinado, se les denomina *variables*, que reflejan aspectos en los que difieren entre sí los individuos y que son de interés para el investigador. Algunos autores se refieren a las puntuaciones o valores que toman las variables como *observaciones estadísticas*. Siempre que se pueda, las observaciones estadísticas se expresan en números, aunque éstos no son esenciales para el empleo de la estadística.

El tipo de «objeto» al que hace referencia la variable que se mide se denomina *unidad de análisis*. En los estudios por encuesta suelen ser los individuos entrevistados las unidades de análisis, aunque esto no tenga que ser así necesariamente. Es decir, que las unidades de análisis pueden ser también grupos de diferente tamaño, o cualquier tipo de instituciones o agregados sociales (familias, partidos políticos, asociaciones voluntarias, comunidades, etc.).

Anteriormente hemos señalado que debido a los orígenes de la estadística, el término «población» ha permanecido en la actualidad, pero con diferente significado que el que tuvo originalmente. Una *población estadística*, o *universo*, lo forman el conjunto de todos los valores de las variables que desee medir el investigador en todas las unidades de análisis. En una encuesta sobre temas de opinión pública, por ejemplo, la población estadística es *finita*, dado que, en un principio, se pueden contar las observaciones estadísticas en una población estadística de tal tipo. Pero esto no es siempre así, ya que a veces las observaciones estadísticas que son de interés para el investigador no pueden limitarse en el tiempo o en el espacio. Entonces, se tiene una población estadística *infinita*, llamada a veces un *universo conceptual*. Si se estudian, por ejemplo, los sistemas de estratificación de las sociedades, el sentimiento de anomia o las actitudes post-materialistas en las sociedades industriales avanzadas, las observaciones estadísticas no están limitadas en forma tan concreta. En tal caso, pues, el investigador trata con poblaciones estadísticas infinitas o universos conceptuales.

Ahora bien, sea finito o infinito el universo estudiado, los sociólogos raras veces se ocupan de todas las observaciones estadísticas, sino que seleccionan un subconjunto de valores llamado *muestra estadística*. Precisamente, a la obtención de muestras que sean suficientemente representativas de la población que se pretende estudiar, dedican los sociólogos notables esfuerzos, ya que si se trabaja con muestras en lugar del conjunto de la población, los ahorros de coste de todo tipo que se producen son muy importantes. El campo de la estadística inferencial o inductiva se ocupa de la forma en que se pueden obtener muestras fiables, y los resultados obtenidos en ellas pueden hacerse extensibles a la población en general.

Antes de seguir adelante en este capítulo dedicado a realizar una introducción a la disciplina, puede resultar conveniente que insistamos una vez más, para así dejarlo bien claro desde un principio, en la naturaleza dual de la estadística que, por un lado, se presenta como una rama de las matemáticas y que, por tanto, trabaja con conceptos abstractos. Pero por otro lado, y ahí es donde le interesa al sociólogo, la estadística se ocupa de datos reales de los fenómenos naturales y sociales, datos que siempre ofrecen algún tipo de resistencia a ser incorporados a las fórmulas matemáticas. Tal como ha señalado agudamente Mosimann (1957), el análisis estadístico real del mundo social, no se refiere al universo euclideo, frío y abstracto, sino que se refiere al mundo de las relaciones humanas, y en tal sentido, el análisis real del mundo social tiene más de arte que de ciencia. Y es que mientras que la estadística matemática está ausente de error (hace referencia a entidades matemáticas puras), la estadística real se basa en medidas, sujetas a error, ya que al medir cosas reales nunca se puede alcanzar la exactitud perfecta. Así, uno de los problemas más importantes que se le plantean al sociólogo al tratar de medir ciertos fenómenos sociales, es el de no influir excesivamente en la alteración de los mismos con su presencia y sus instrumentos de medida.

A la consecución de medidas válidas y fiables de las variables que intervienen en la investigación, ha de dedicar el sociólogo los mayores esfuerzos, ya que por sofisticados que sean los métodos estadísticos que emplee en los cálculos, los resultados no van a mejorar si los datos de partida no son de buena calidad. Loether y McTavish (1974, página 357) sugieren que cuando el investigador dispone de unos datos, la primera pregunta que debe formularse debe ser: ¿merecen estos datos un análisis? Sólo cuando se haya obtenido una respuesta afirmativa, se podrá pensar en el empleo de un instrumento estadístico u otro. De ahí que los problemas de medición de las variables sean centrales en el quehacer del trabajo sociológico.

1.5. TEORÍA, ESTADÍSTICA Y MEDICIÓN DE VARIABLES

Ya hemos dicho anteriormente que la estadística juega un papel concreto, y limitado, en el proceso de investigación en sociología, y que no existe ningún sustituto estadístico para una correcta conceptualización teórica. Antes de llegar a la fase del análisis estadístico en una investigación determinada, el sociólogo debe resolver una serie de problemas teóricos cuya solución permitirá hacer un buen uso de los instrumentos estadísticos. Como ejemplo, supongamos que se desea probar la proposición «A mayor juventud y mayor *status* socioeconómico de la población, corresponden índices más elevados de conciencia regional». En ella encontramos dos conceptos que merecen una definición más clara, «*status* socioeconómico» y «conciencia regional». Supongamos que definimos el primero como la posición relativa de las personas en una jerarquía de *status*, y que la conciencia regional se define como un fenómeno subjetivo que implica la existencia de una identidad regional, la percepción del hecho diferencial regional y de los intereses de la región. Con unas definiciones tales, resulta difícil utilizarlas para medir el *status* socioeconómico y la conciencia regional de una persona concreta o de un grupo social determinado.

Y es que las definiciones las hemos dado en términos teóricos y no en términos operacionales, es decir, especificando los procedimientos de medición que se han de seguir para dar valores o puntuaciones a la variable que se deduce del concepto definido operacionalmente. Un ejemplo de una definición operacional de *status* socioeconómico puede ser el índice de Warner para las características de *status*, y como definición operativa de conciencia regional puede utilizarse el conjunto de *items*, 38 en total, que utilizan los autores del estudio *La Conciencia Regional en España* (J. Jiménez Blanco *et al.*, 1977) para definir las cuatro dimensiones —administrativa, lingüística, económica y política— de dicho concepto.

De esta forma, es decir, operacionalizando el concepto, se puede pasar a la fase de la medición y, por tanto, a la introducción de los métodos estadísticos en la investigación. Queda claro, pues, que el razonamiento teórico, fase previa y necesaria en toda investigación, no conduce por sí mismo a la medición de las variables y a la realización de pruebas estadísticas. La operacionalización de los conceptos teóricos aparece como una fase intermedia que une la formulación teórica de un problema y la medición de las variables relevantes para el tratamiento empírico de dichas variables.

El empleo de las herramientas estadísticas requiere que las variables sociológicas sean cuantificadas siguiendo el nivel de medición que las propiedades de las variables exigen. Siguiendo la formulación ya clásica de Stevens, se puede definir la *medición* como el procedimiento de asignación de numerales a objetos o acontecimientos de acuerdo con ciertas reglas (Stevens, 1951, 1). Siguiendo esta conceptualización de la

medición, se puede definir una *variable* como cualquier característica o propiedad de un objeto o acontecimiento que contenga dos o más categorías posibles en las que un «objeto o acontecimiento» puede ser clasificado potencialmente. El *status* social o la conciencia regional son dos variables de los individuos y de los grupos sociales. Si para medir el *status* social utilizamos el índice de Warner y hacemos que dicho índice varíe entre un mínimo de 10 y un máximo de 100, una medición de dicho índice para una persona determinada, de valor 70, representará una observación estadística y lo utilizaremos como el valor de la variable *status* para dicha persona.

Obsérvese que venimos utilizando los términos *indicador* e *índice* para referirnos a aspectos cuantitativos de los conceptos. En efecto, los indicadores son elementos, cuantitativos, que sirven para medir un concepto. Los años de escolaridad, el nivel de ingresos y el tipo de ocupación son los indicadores con los que Warner operativizó el concepto de *status* social. La combinación ponderada de los valores que toman cada uno de los tres indicadores forman un índice que tomará valores numéricos concretos, con lo que operará estadísticamente el investigador.

Cuando el investigador busca el procedimiento de medición más adecuado ha de fijarse, sobre todo, en dos aspectos del instrumento de medición. Dicho instrumento debe ser válido y fiable. La *validez* hace referencia a que el procedimiento utilizado mida lo que realmente se pretende medir —así, el índice de Warner es válido como medida del *status* social porque mide posiciones en una jerarquía de *status* y no otro concepto, como pudiera ser el prestigio social—, y la fiabilidad hace referencia a la propiedad del instrumento que permite, al ser utilizado repetidas veces bajo idénticas circunstancias, producir los mismos resultados. Los autores suelen distinguir entre *validez interna* y *validez externa*. La primera hace referencia a la cuestión siguiente: ¿se obtendrían resultados diferentes si se hubieran utilizado procedimientos diferentes?; mientras que la segunda hace referencia a otra cuestión: ¿cuán generalizable es el procedimiento utilizado? Se trata de problemas muy difíciles de resolver, sobre todo en sociología, ya que habitualmente los conceptos sociológicos más interesantes suelen ser lo bastante complejos como para impedir su operacionalización directa y simple.

El concepto de medición, tal como se utiliza en sociología y se define en el presente libro, es más amplio que el utilizado en las ciencias físicas, en las que el acto de medir se hace a las propiedades métricas del sistema, con lo que se excluyen de la medición las propiedades que se encuentran a nivel nominal e incluso ordinal, que son, por otro lado, el tipo de propiedades que con mayor frecuencia emplea el sociólogo. Se pueden distinguir *cuatro niveles de medición*, cada uno de los cuales requiere la utilización de un determinado conjunto de instrumentos estadísticos. En concreto, las pruebas o tests estadísticos que va a poder

utilizar el sociólogo para contrastar sus hipótesis dependerán de la escala o nivel de medida de los diferentes aspectos estudiados.

Cualquier procedimiento adecuado de medición debe ser exhaustivo, esto es, debe tener suficientes categorías en las que puedan clasificarse cada uno de los casos considerados. Así, por ejemplo, si disponemos de las categorías 0, 1, 2, 3 y 4 para clasificar las familias según el número de hijos, no podremos incluir en dicha clasificación a las familias que tengan cinco hijos o más. En tal caso se puede ampliar la última categoría a «4 o más», con lo que se podrá incluir en la escala cualquier tipo de familia. La categoría «otros» se suele utilizar con frecuencia para crear un sistema clasificatorio que sea lógicamente completo o inclusivo. En el diseño de cuestionarios, el listado de las posibles respuestas a cada pregunta suele incluir también las categorías «no sabe» y «no contesta», con el mismo fin. Su análisis plantea especiales problemas al sociólogo a la hora de interpretar los resultados —a este respecto, ver el excelente artículo de José Luis Martín Martínez «Ensayo de tipificación de los sin opinión (1981)».

Otra propiedad de un buen sistema clasificatorio hace referencia a que las categorías deben ser *mutuamente exclusivas*, esto es, que debe ser posible clasificar cada caso individual tan sólo en *una* categoría del sistema clasificatorio. Si, por ejemplo, utilizamos las categorías «25 años o menos» y «25 años o más» para clasificar a la población según la edad, no sabremos en qué categoría colocar a los individuos que tengan exactamente veinticinco años. Para que las categorías fueran mutuamente exclusivas habría que ensayar una clasificación tal como «menos de 25 años» y «25 años o más».

También resulta conveniente que el procedimiento de medición sea lo más *preciso* posible, es decir, que haga el mayor número posible de distinciones. Así, el esquema de medición de preferencia por los partidos políticos: «partidos de derecha», «partidos de izquierda» y «otros partidos», es menos preciso que este otro esquema: «partidos de derecha», «partidos de centro», «partidos de izquierda», «partidos nacionalistas» y «otros partidos». Las variables bien definidas se clasifican generalmente en uno de los siguientes cuatro tipos de medidas.

1.5.1. Medidas nominales

Se realiza una medida nominal cuando la propiedad estudiada en los objetos o acontecimientos sólo puede agruparse en categorías lógicamente exhaustivas y mutuamente exclusivas, de tal modo que pueden establecerse claramente equivalencias o diferencias. A cada una de las categorías se le asignan atributos diferentes, que pueden ser tanto nombres como números. Ahora bien, si asignamos números, por ejemplo, 1, 2, 3 y 4, sólo se pueden interpretar como que son diferentes entre sí, esto es, que $1 \neq 2 \neq 3 \neq 4$, sin que se pueda afirmar que uno es superior a otro y, por tanto, sin que se puedan ordenar.

Con las medidas nominales se consigue realizar la operación más simple y básica de toda ciencia, que es la clasificación. Por medio de la clasificación, los elementos se dividen en categorías, y se decide cuáles son más parecidas y cuáles son más diferentes. El objetivo es siempre el de obtener categorías que permitan una clasificación clara de los elementos y que sean homogéneas en relación a otras variables. Así, por ejemplo, si clasificamos a la población española según el nivel de religiosidad que manifiesta, esto es, católicos practicantes, católicos no practicantes, indiferentes y otras creencias, y lo relacionamos con las preferencias ideológicas de la misma población, esto es, izquierda, centro y derecha, encontraremos que los individuos indiferentes en materia religiosa tienden a ser más de izquierdas que los individuos más católicos. Si los individuos los hubiéramos clasificado, por ejemplo, según su región de origen, no hubiéramos encontrado una relación con la preferencia política tan clara como la anterior relación. Esto es, las diferencias entre los individuos dentro de cada región serían mayores —por lo que a preferencias ideológicas se refiere— que las diferencias que se pudieran encontrar entre los individuos clasificados según el tipo de religiosidad. Por eso, la clasificación de la población según su religiosidad es analíticamente preferible a la clasificación según la región de origen, a efectos de estudiar las preferencias ideológicas de la población.

La clasificación o medida nominal es el nivel más bajo de medición. Algunos autores se refieren a este nivel de medición con el término *escala nominal*. En sociología, muchas variables aparecen medidas a nivel nominal, tales como el estado civil, el sexo, tipo de comunidad, tipo de religiosidad, tipo de relación de roles, etc.

Desde un punto de vista formal, las escalas nominales poseen la propiedad de la *relación de equivalencia*. Este tipo de relación engloba a la propiedad *reflexiva*, esto es, que $A=A$ para todo valor de A ; la propiedad *simétrica*, es decir, que si $A=B$, también ocurrirá que $B=A$, y la propiedad *transitiva*, que quiere decir que si $A=B$ y $B=C$, también se produce que $A=C$. Dicho de otra forma, que si A se encuentra en la misma categoría que B , que B se encuentra en la misma categoría que C , y que si A y B se encuentran en la misma categoría y B y C también en la misma categoría, entonces se tiene que A y C deben encontrarse en la misma categoría.

Hemos dicho anteriormente que con las escalas nominales no pueden realizarse las operaciones aritméticas que se realizan usualmente con los números —suma, resta, multiplicación y división—. A veces, sin embargo, se hace preciso, por necesidades del modelo estadístico que utilizamos, cuantificar todas las variables. En tal caso, se pueden crear variables ficticias o por medio de algún procedimiento arbitrario, tal como hacer corresponder una o más de las categorías con el número 1 y el resto de las categorías con el número 0. Más adelante, al estudiar

el análisis multivariable, tendremos ocasión de ver con más detalle este procedimiento.

1.5.2. Medidas ordinales

Se tiene una medida ordinal cuando, además de incluir las propiedades de la medida nominal, se incluye la propiedad de que las categorías pueden ser ordenadas en el sentido de menor que o mayor que. Con frecuencia, en sociología, las categorías de los fenómenos estudiados o los rasgos de los individuos se ordenan en términos de «mayor que», «más preferido», «más difícil», «más elevado», etc., o, inversamente, en términos de «menor que», «menos preferido», «menos difícil», «menos elevado», etc. Tales relaciones se pueden expresar mediante el signo $>$, que significa «mayor que», o mediante el signo opuesto $<$, que significa «menor que». Cuando la relación $>$ se mantiene para todas las parejas de categorías que se pueden formar en un ordenamiento de todas las categorías de una variable, tenemos una *escala ordinal*.

Este tipo de medición es, por supuesto, de más alto nivel que el que se obtiene cuando se utiliza una medida nominal, ya que no sólo se puede distribuir a los fenómenos o individuos en categorías diferentes, sino que, además, se pueden ordenar tales categorías. Desde un punto de vista lógico, la diferencia fundamental entre una medida o escala ordinal y una medida o escala nominal es que la primera lleva incorporada no sólo la relación de equivalencia (simbólicamente $=$), sino también la relación «mayor que» ($>$). Esta última relación goza de la propiedad *irreflexiva*, esto es, que no es cierto que para todo A sea $A > A$; de la propiedad *asimétrica*, esto es, que si $A > B$, entonces $B < A$, y de la propiedad *transitiva*, esto es, que si $A > B$ y $B > C$, entonces $A > C$.

En sociología, son muchas las variables de interés que aparecen medidas a nivel ordinal. Por ejemplo, el *status* socioeconómico, tal como ha sido definido por Warner, constituye una escala ordinal. Así, el prestigio social de un grupo de *status alto* es mayor que el prestigio de un grupo de *status medio*, y éste, a su vez, manifiesta un prestigio social mayor que el que tiene un grupo de *status bajo*. En general, cuando se mide una variable sociológica que refleja una determinada propiedad o atributo, en términos de alto, medio y bajo, o cuando en el estudio de actitudes y opiniones las respuestas posibles vienen dadas en términos de muy de acuerdo, bastante de acuerdo, poco de acuerdo o nada de acuerdo, se está tratando de establecer una medida o escala ordinal.

Nótese que el nivel de medición ordinal no ofrece ningún tipo de información sobre la magnitud de las diferencias entre las categorías. Si éstas vienen dadas en términos de alto, medio y bajo, y decidimos asignar a tales categorías los numerales 3, 2, 1, ello no quiere decir que exista una diferencia de la unidad entre cada par de categorías, sino simplemente que $3 > 2 > 1$. Por esta razón no se pueden, en general, rea-

lizar las operaciones aritméticas de la suma, resta, multiplicación y división con los valores de las variables medidas al nivel ordinal. En este sentido, resulta indiferente que asignemos unos números u otros a las categorías de una escala ordinal, con tal de que se mantenga la relación «mayor que» y «menor que».

1.5.3: *Medidas de intervalo*

En un sentido restringido, el concepto de medición habría que aplicarlo sólo a las medidas de intervalo, las cuales incluyen no sólo las propiedades lógicas de las escalas ordinales y nominales, sino, además, sus categorías se definen en términos de una *unidad de medición* estándar, tal como años de edad, años de escolaridad, pesetas de ingresos, etc. En otras palabras, se realiza una medida de intervalo cuando pueden asignarse al objeto o acontecimientos estudiados números que, además de poseer las características de la medida ordinal, permiten la interpretación de la diferencia entre dos medidas. En tal caso, la medición se ha alcanzado en el sentido de una *escala de intervalo*, esto es, que la asignación de números a las diferentes categorías de los objetos o acontecimientos es tan precisa que podemos conocer cuán amplios son los intervalos (distancias) entre todos los objetos de la escala. Lo que caracteriza a una escala de intervalo es la existencia de una unidad de medición común y constante, que permite asignar un número real a todos los pares de objetos del conjunto ordenado. En una escala de intervalo, el punto cero y la unidad de medición son arbitrarios, y el cociente de cualquiera de los intervalos es independiente de los mismos.

Las escalas de temperatura más conocidas, la centígrada y la Fahrenheit, son ejemplos de medidas y escalas de intervalo. Ambas escalas difieren tanto en sus respectivos puntos cero como en sus unidades de medición. Con todo, ambas escalas contienen el mismo tipo de información. Así, el punto de congelación del agua se produce a los 0 grados centígrados y a los 32 grados Fahrenheit, mientras que la ebullición tiene lugar a los 100 grados centígrados y a los 212 grados Fahrenheit. Sin embargo, el cociente o razón de las diferencias entre las temperaturas leídas en una escala es igual al cociente o razón entre las diferencias equivalentes en la otra escala. Por ejemplo, en la escala centígrada, el cociente de las

diferencias entre 30 y 10 y 10 y 0 es $\frac{30-10}{10-0}=2$. La misma lectura rea-

lizada en la escala Fahrenheit daría el siguiente cociente: $\frac{86-50}{50-32}=2$.

lo que pone de manifiesto que el cociente es idéntico en ambos casos.

Por lo que a las ciencias sociales se refiere, con frecuencia se trata de alcanzar medidas de intervalo, pero sólo raras veces se consiguen. Y ello a pesar de las ventajas que acompañan al uso de medidas y es-

calas de intervalo, ya que es posible utilizar modelos estadísticos que son más poderosos y eficientes que los modelos que utilizan medidas nominales y ordinales. Con todo, la discusión sobre la viabilidad del empleo de escalas de intervalo en las ciencias sociales y humanas todavía continúa, ya que se argumenta que incluso cuando se dispone de una escala claramente de intervalo, como puede ser la distribución de los ingresos en la población, al tener en cuenta las consecuencias sociales y psicológicas de las diferencias de ingresos en los diversos niveles, se pierde la propiedad de la igualdad de las diferencias entre los intervalos. Esto es, que no es lo mismo una diferencia de 20.000 pesetas cuando nos estamos moviendo en los niveles bajos de ingresos que cuando la diferencia se produce entre los niveles altos. No tiene las mismas consecuencias sociales y psicológicas el pasar de un nivel de ingresos mensuales de 50.000 a 70.000 pesetas, que pasar de un nivel de ingresos de 500.000 a 520.000 pesetas. En ambos casos, las diferencias son de 20.000 pesetas, pero no resulta difícil aceptar que las consecuencias sociales y psicológicas son bien distintas en ambos casos. En otras palabras, no existe un paralelismo entre los ingresos medidos en pesetas y «el ingreso psicológico», como lo denomina Blalock (*op. cit.*, pág. 15). Sin embargo, hay que señalar que, desde el punto de vista de los cálculos estadísticos que se pueden realizar con la escala de ingresos, resultan irrelevantes las diferencias de percepción psicológica. Este es un hecho social que habrá que tener en cuenta al *interpretar los resultados*, no al realizar los *cálculos estadísticos*.

Con las escalas de actitudes, el problema es también muy importante. Aunque han sido muchos los intentos, desde la iniciativa de Thurstone (1928) por medir las actitudes, de alcanzar niveles seguros de medición, permanecen todavía las dudas de si realmente los valores que se obtienen al aplicar escalas de medición de actitudes alcanzan el nivel de intervalo o, simplemente, se han quedado en el nivel ordinal, o incluso, y simplemente, en el nivel nominal. Aunque no es tema que compete exclusivamente al estadístico, y menos a un libro introductorio de estadística como el presente, conviene advertir a los sociólogos de los errores y sesgos que se pueden introducir al emplear métodos estadísticos que requieren medidas y escalas de intervalo, con datos sociológicos que no sobrepasan los niveles nominales y ordinales de medición. No es que exista una imposibilidad de emplear modelos basados en supuestos de medición de intervalo, con datos ordinales y nominales. Lo que ocurre es que los resultados no pueden interpretarse de una manera rigurosa, sino de una forma limitada y aproximada.

Desde el punto de vista lógico, se puede demostrar que las operaciones y relaciones que produce la estructura de una escala de intervalo son tales que las diferencias en la escala son isomórficas con la estructura de los números aritméticos. Por ello, con los números asociados a la posición de los objetos en una escala de intervalo se pueden

realizar las operaciones aritméticas de la suma, resta, multiplicación y división. Al construir una escala de intervalo no sólo se ha de poder especificar la relación de equivalencia, como en las escalas nominales, y la relación «mayor que», como en las escalas ordinarias, sino que, además, se ha de especificar el cociente de cualquier par de intervalos.

1.5.4. *Medidas de cociente o proporción*

Cuando una medición tiene todas las características de una medida de intervalo y, además, se le puede asignar un punto de origen verdadero de valor 0, se tiene entonces una medida de cociente o proporción. En tal caso, se ha conseguido establecer una escala con cero absoluto o escala de cociente o proporción (en inglés, *ratio scale*).

La masa o el peso se miden mediante una escala de cociente o proporción, ya que la escala del peso, en gramos, contiene un punto cero verdadero, siendo el cociente de dos pesos independientes de la unidad de medida (sean kilogramos o libras, por ejemplo). La escala de tiempos, en segundos, también es una escala de cociente. Así, pues, el tiempo de reacción a un cierto estímulo será una medida de cociente.

Las medidas de cociente son isomórficas con la estructura aritmética. Por tanto, todas las operaciones aritméticas se pueden realizar con los números asignados en una escala de cociente. Las medidas de cociente no quedan afectadas por el hecho de multiplicarlas por una constante. En las medidas de cociente sólo es arbitraria la unidad de medida; no así el punto cero, que, como hemos dicho anteriormente, es absoluto o verdadero. La escala Kelvin de temperaturas es una escala de cociente, porque, a diferencia de la centígrada o la Fahrenheit, posee un cero absoluto.

En el trabajo sociológico suele ser difícil definir unidades de medición en la escala de cociente o proporción; de ahí que se encuentren muy pocos ejemplos en la literatura sociológica de empleo de dicha escala. En general, no se suele distinguir, cuando se trabaja con variables sociológicas, entre los niveles de medición de intervalo y los de cociente. Ejemplos de variables que se presentan medidas, indistintamente, en ambas escalas son: «tamaño de la familia», «tamaño del grupo», «tamaño del hábitat», «número de años de escolaridad», «número de niveles en una jerarquía organizacional» y, quizá también, «*status social*».

1.5.5. *Tipos de variables*

Según el nivel de medición que les sea aplicado, se pueden clasificar las variables, tal como hemos visto anteriormente, en nominales, ordinales y de intervalo. Para cada tipo de variable existen unos procedimientos estadísticos apropiados para hacer el mejor uso de la informa-

ción que contienen los valores de las variables. Si se utilizara un procedimiento estadístico apropiado para niveles bajos de medición con puntuaciones definidas a un nivel de medición más alto, no se cometería un error técnico, sino simplemente se produciría una pérdida de información, dado que las propiedades de los niveles de medición son acumulativas. Así, se puede utilizar un procedimiento estadístico apropiado para escalas ordinales con variables de intervalo. Pero el contrario no es cierto, ya que sí se produce un error al emplear procedimientos estadísticos diseñados para niveles altos de medición, con variables de niveles más bajos. En este caso, estaríamos actuando como si las puntuaciones de las variables contuvieran más información de la que realmente tienen.

El tema del empleo de los diferentes procedimientos estadísticos según el nivel de medición de las variables ha sido objeto de amplia discusión en las ciencias sociales*. La ventaja de los procedimientos estadísticos apropiados para niveles de medición elevados es que permiten una descripción más concisa de los datos. Algunos autores señalan que la utilización de un procedimiento estadístico apropiado para medidas de intervalo con variables ordinales no produce grandes errores en los resultados estadísticos.

Otra diferencia entre las variables, además del nivel de medición, hace referencia a si las variables vienen definidas según una escala de medición *continua* o *discreta*. Una *variable continua* es aquella para la que los individuos pueden tener, en principio, infinitos valores fraccionados, esto es, valores en cualquier punto de una escala ininterrumpida. Por el contrario, una *variable discreta* viene definida de tal modo que sólo se puede alcanzar un determinado conjunto de valores. En otras palabras, la escala de medición está interrumpida por espacios en la escala numérica que, en un principio, no contienen casos medidos de ningún tipo.

Gráficamente, se pueden visualizar las diferencias entre una variable continua —por ejemplo, la edad— y una variable discreta —por ejemplo, el tamaño de la familia— como sigue:

Variable continua:								
edad en años	15	16	17	18	19	20	21	22
Variable discreta:								
tamaño de la familia	1	2	3	4	5	6	7	8

* Para una detallada exposición sobre el uso de modelos estadísticos para los diferentes niveles de medición, véase Cletus J. BURKE, «Measurement Scales and Statistical Models», cap. 7, en Bernhardt LIEBERMAN, *Contemporary Problems in Statistics: A Book of Readings for the Behavioral Sciences*, New York, Oxford University Press, 1971.

Para una variable continua, y dados dos valores, siempre se puede encontrar un tercer valor que esté incluido entre los dos primeros. En sociología, variables tales como edad, alienación, segregación y clase social se definen usualmente como variables continuas.

Para una variable discreta, y dentro de un determinado espacio, existe tan sólo un número concreto de posibles categorías. Así, una variable como el tamaño de la familia no puede tomar más que valores discretos, como 1, 2, 3, etc., pero no tiene sentido hablar de una familia de tamaño 2,5. La mayor parte de las variables nominales son discretas.

Finalmente, conviene destacar una tercera forma en la que pueden diferenciarse las variables sociológicas entre sí, y es en el uso que se hace de ellas en la investigación. La clasificación más simple diferencia entre *variables independientes* y *variables dependientes*.

Frecuentemente, el investigador trata de explicar las variaciones que se producen en un determinado fenómeno en función de determinados factores o elementos causales. Así, puede preguntarse por los factores que determinan el tipo de preferencia política, o por los factores que inciden en el éxito o fracaso de los estudios universitarios o de los matrimonios. Pues bien, la variable que atrae primordialmente la atención del investigador, y cuya variación trata de explicar, se llama variable dependiente, porque se supone que los valores que toma la variable dependen de los valores que presentan otras variables. Estas variables que se supone influyen en los valores que toma la variable dependiente son las variables independientes o variables explicatorias, ya que permiten conocer por qué varía la variable dependiente de la forma que lo hace en una determinada población.

Al elaborar la relación entre dos variables en sociología se suele introducir una tercera variable, que se denomina *variable interviniente*, porque se supone que tiene un efecto determinado sobre la variable dependiente que puede ser controlado o modificado por la variable independiente. Si pensamos en términos de causa/efecto, la relación entre los tres tipos de variables puede expresarse por medio de unas flechas que señalan el camino desde la variable independiente a la dependiente, pasando por la variable interviniente, como sigue:

variable independiente \longrightarrow variable interviniente \longrightarrow variable dependiente

El papel que representa una variable en una investigación y en un marco teórico determinado puede alterarse cuando cambiamos de investigación o de marco teórico. Así, por ejemplo, el estado civil o la situación matrimonial son variables independientes en el estudio de Durkheim sobre el suicidio, ya que las tasas de suicidio varían, como mostró Durkheim en su conocida investigación, según los valores que toman aquéllas. Por el contrario, en un estudio sobre las causas del divorcio,

la situación matrimonial sería la variable dependiente que habría que explicar a partir de otras variables independientes. En general, las investigaciones sociológicas requieren la consideración simultánea de más de dos o tres variables, en cuyo caso hay que utilizar procedimientos estadísticos complejos, que tendremos ocasión de estudiar en próximos capítulos del presente libro.

Esperamos que haya quedado suficientemente claro en esta breve introducción la importancia e íntima relación entre el marco teórico, el nivel de medición alcanzado por las variables y los procedimientos estadísticos que es posible utilizar con los valores que presentan las variables. El papel crucial que juega el proceso de medición en la sociología es cada vez más evidente, en la medida que la rápida evolución de los procedimientos estadísticos y las enormes facilidades de cálculo que permite el uso generalizado de los ordenadores van derribando obstáculos para alcanzar una eficaz y pronta investigación cuantitativa de los fenómenos sociales. En estos momentos, uno de los mayores desafíos que se le presentan a los investigadores sociales, es el de desarrollar y encontrar procedimientos apropiados de medición de las variables sociológicas, que permitan su inclusión en los cada vez más complejos modelos estadísticos que se diseñan, con el fin de encontrar explicaciones realmente causales de los fenómenos sociales, objetivo último de toda explicación científica.

BIBLIOGRAFIA

- BLALOCK, Hubert M.: *Social Statistics*, New York, McGraw-Hill, 1960.
- BURKE, Cletus J.: «Measurement Scales and Statistical Models», capítulo 7, en Bernhardt LIEBERMAN: *Contemporary Problems in Statistics: A Book of Readings for the Behavioral Sciences*, New York, Oxford University Press, 1971.
- DURKHEIM, Emile: *Las reglas del método sociológico*, Buenos Aires, La Pléyade, 1972 (e. o., 1895).
- GARCÍA FERRANDO, M.: *Sobre el método. Problemas de investigación empírica en Sociología*, Madrid, C.I.S., 1979.
- JIMÉNEZ BLANCO, José, et al.: *La conciencia regional en España*, Madrid, C.I.S., 1977.
- KENDALL, Maurice G.: «The History of Statistical Method», en W. H. KRUSKAL y J. M. TANUR (eds.): *International Encyclopedia of Statistics*, New York, Free Press, 1978, págs. 1093-1101.
- KRUSKAL, William H.: «The Field of Statistics», en W. H. KRUSKAL y J. M. TANUR (eds.): *International Encyclopedia of Statistics*, New York, Free Press, 1978, páginas 1071-1091.
- ISIS, 52, 1962, págs. 277-233.
- LINZ, Juan José: *Informe de la Encuesta sobre la Juventud 1977*, Madrid, Instituto de la Juventud, 1978.

- LOETHER, Herman J., y Donald G. MCTAVISH: *Descriptive Statistics for Sociologists*, Boston, Allyn and Bacon Inc., 1974.
- MARTÍN MARTÍNEZ, José Luis: «Ensayo de tipificación de los sin opinión», *Revista Española de Investigaciones Sociológicas*, 16, 1981, págs. 9-37.
- MOSIMANN, Thomas F.: «Mathematical Statistics and Real Statistics», *IASI, Estadística*, junio 1957, págs. 390-394.
- STEVENS, S. S.: «Mathematics, Measurement and Psychophysics», en S. STEVENS (ed.): *Handbook of Experimental Psychology*, New York, Wiley, 1951, págs. 1-30.
- THURSTONE, L. L.: «Attitudes can be measured», *American Sociological Review*, vol. 33, 1928, págs. 529-554.