

Índice

Introducción	9
Capítulo 1. El análisis estadístico de los datos sociológicos	13
1.1. Introducción.....	13
1.2. Notas históricas sobre los métodos estadísticos.....	14
1.3. Funciones de la estadística: estadística descriptiva y estadística inferencial.....	18
1.4. Relaciones entre la estadística y la sociología: conceptos básicos	20
1.5. Teoría, estadística y medición de variables	24
1.5.1. Medidas nominales	26
1.5.2. Medidas ordinales.....	27
1.5.3. Medidas de intervalo.....	28
1.5.4. Medidas de cociente o proporción.....	29
1.5.5. Tipos de variables.....	30
Ejercicios	33
Capítulo 2. Estadística descriptiva univariable: la lógica del análisis comparativo	35
2.1. Importancia del análisis comparativo en sociología.....	35
2.2. Tipos de comparación	38
2.2.1. Comparaciones entre grupos.....	38
2.2.2. Comparación entre un grupo y un individuo.....	39

2.2.3.	Comparación entre el resultado de un estudio y un resultado estándar	40
2.3.	Operaciones básicas de comparación.....	41
2.3.1.	La organización de los datos	41
2.3.2.	Distribuciones	44
2.3.3.	Percentiles	51
2.3.4.	Razones y números índices.....	53
2.4.	Técnicas de representación gráfica.....	55
2.4.1.	Gráficos para variable discreta	55
2.4.2.	Gráficos para variables continuas.....	58
Ejercicios		61

Capítulo 3. Características de una distribución de frecuencias: tendencia central, dispersión y forma. La distribución normal 63

3.1.	Características de la distribución univariable.....	63
3.2.	La posición de una distribución: medidas de tendencia central	66
3.2.1.	Moda	66
3.2.2.	Mediana.....	67
3.2.3.	Media aritmética	68
3.2.4.	Tipos especiales de medias	69
3.2.5.	Relación y comparación entre los estadísticos de tendencia central	70
3.3.	Variación o dispersión de una distribución	72
3.3.1.	Recorrido	73
3.3.2.	Desviación media	74
3.3.3.	Desviación típica y varianza.....	74
3.4.	Forma de una distribución	78
3.4.1.	Características de la forma de una distribución: asimetría y apuntamiento	78
3.4.2.	Medidas de la forma de una distribución. Momentos	80
3.5.	La distribución normal.....	83
3.5.1.	La curva normal.....	84
Ejercicios		87

Capítulo 4. Estadística inferencial: probabilidades y tipos de muestreo 89

4.1.	Introducción a la estadística inferencial.....	89
4.2.	Probabilidad: nociones básicas y definición.....	91
4.3.	Aspectos generales del muestreo en la investigación sociológica	94
4.3.1.	Tipos de muestreo	96
4.3.2.	Muestreo aleatorio simple	97
4.3.3.	Estimadores y errores de muestreo	99
4.3.4.	Determinación del tamaño de la muestra	102
4.3.5.	Otros tipos de muestreo probabilístico	105
4.3.6.	Muestreo no probabilístico.....	111
Ejercicios		112

Capítulo 5. El empleo de las pruebas de decisión estadística en la investigación social. Distribuciones muestrales	115
5.1. Introducción.....	115
5.2. El uso de las pruebas de decisión estadística en la investigación	118
5.2.1. Formulación de las hipótesis estadísticas	118
5.2.2. Elección de una prueba estadística	120
5.2.3. El nivel de significación y el tamaño de la muestra	120
5.2.4. La distribución muestral	125
5.2.5. La región de rechazo. Pruebas unilaterales y pruebas bilaterales	132
5.3. Elección de una prueba de decisión estadística apropiada	134
Ejercicios	136
 Capítulo 6. Pruebas de decisión estadística para el caso de una sola muestra	 139
6.1. Introducción	139
6.2. La distribución binomial. La prueba binomial.....	140
6.3. La prueba de Ji-cuadrado (χ^2) para una sola muestra	141
6.4. Distribuciones muestrales de las medias	144
6.4.1. La distribución <i>t</i> de Student	145
6.5. Estimación puntual y por intervalo de parámetros	148
6.5.1. Estimación de proporciones. Intervalos de confianza	150
6.5.2. Estimación de medias.....	154
Ejercicios	155
 Capítulo 7. Estadística descriptiva bivariable: características de una asociación bivariable.....	 157
7.1. Distribuciones bivariantes: un ejemplo	157
7.2. Presentación y análisis de una tabla bivariable	160
7.2.1. Cálculo de los porcentajes en una tabla	162
7.3. Características de una asociación de dos variables	165
7.4. Medidas de asociación entre dos variables.....	170
7.4.1. Medidas simétricas para variables nominales	171
7.4.2. Medidas asimétricas de asociación para variables nominales.....	174
7.4.3. Medidas de asociación para variables ordinales	178
Ejercicios	186
 Capítulo 8. Pruebas de decisión para el caso de dos o más muestras	 191
8.1. La prueba de la diferencia entre dos proporciones.....	192
8.2. La prueba de la diferencia entre dos medias	194
8.3. El análisis de varianza	199
8.3.1. El análisis de varianza con un solo factor.....	200
8.4. El análisis de varianza para variables no paramétricas	207
8.4.1. El análisis de varianza por rangos de Kruskal-Wallis.....	207
Ejercicios	209

Capítulo 9. Medidas de asociación para variables de intervalo: regresión y correlación	213
9.1. Planteamiento general	213
9.2. Ecuaciones de regresión lineal	214
9.2.1. Relación entre dos variables estadísticas: ecuación de una recta	216
9.2.2. La ecuación de regresión y el ajuste por mínimos cuadrados	218
9.3. Correlación. Coeficiente R de correlación de Pearson	226
9.3.1. Interpretación del coeficiente de correlación	231
9.3.2. Correlación y regresión con valores típicos, Z	233
9.4. La matriz de correlaciones.....	234
9.5. Consideraciones finales sobre la selección e interpretación de las medidas de asociación	235
Ejercicios	237
Capítulo 10. El análisis multivariable	241
10.1. El análisis multivariable en la investigación sociológica	242
10.2. Definición de análisis multivariable	243
10.3. Nociones algebraicas elementales en las técnicas multivariadas.....	247
10.4. Clasificación de las técnicas de análisis multivariable.....	249
10.5. Breve panorama descriptivo de las técnicas multivariadas.....	255
10.5.1. Técnicas basadas en la dependencia entre las variables.....	255
10.5.2. Técnicas basadas en la interdependencia de las variables...	259
10.5.3. Técnicas basadas en la semejanza entre objetos.....	264
Ejercicios	266
Apéndice I. Tablas estadísticas	269
Bibliografía	279
Índice conceptual	285

Introducción

«La estadística expresa cierto estado del alma colectiva.»

(DURKHEIM, 1972: 38)

La enseñanza de la estadística en la licenciatura o el grado de sociología ha venido tropezando desde sus inicios con no pocas dificultades tanto en las universidades españolas como en otros países que han logrado institucionalizar un programa de estudios sociológicos universitario. La falta de preparación suficiente en estadística matemática de parte de quienes aspiran a graduarse en una facultad de ciencias sociales y, en no pocas ocasiones, la difusión de una cierta actitud preventiva hacia los «tecnicismos» y la «cuantificación» en sociología han impedido que la estadística se estudie en igualdad de condiciones académicas con otras materias sociológicas de tipo teórico o sustantivo.

Con todo, consideramos que existe en la actualidad un amplio consenso en aceptar el papel crucial que juega el análisis estadístico en la investigación empírica social, máxime en unos momentos en que la extensión de los bancos de datos sociales, el uso generalizado del ordenador y la revolución de Internet han provocado una transformación profunda en la forma de entender y ejercer el oficio de la sociología.

De nuestra experiencia como profesores universitarios de sociología e investigadores sociales hemos sacado la conclusión de que era necesario contar en nuestra universidad con un texto introductorio de estadística para que, sin entretenerse excesivamente en la derivación de fórmulas matemáticas, ofreciese una visión rigurosa y amplia del uso actual del

análisis estadístico en la investigación social. Y a este propósito responde la reedición de la presente *Socioestadística*.

Se trata de un libro que pretende recoger los conocimientos mínimos e imprescindibles que se debe tener para poder entender e iniciarse en el trabajo empírico de la sociología. No se trata de un recetario, ni tampoco un formulario que presenta, una tras otra, las expresiones estadísticas vulgarizadas de mayor uso. Por el contrario, el libro está compuesto por una serie conexas de 10 capítulos en los que se va introduciendo paulatina y lógicamente los fundamentos de la estadística descriptiva e inferencial tal como se utiliza ampliamente en la investigación sociológica.

De ahí, pues, el nombre de *Socioestadística* que le hemos dado al libro. Con esta expresión tratamos, sencilla y directamente, de expresar que los conceptos y expresiones estadísticos que aquí se estudian son los de mayor uso en sociología, presentados a través de ejemplos extraídos en su mayor parte de la investigación sociológica realizada en nuestro contexto académico y profesional.

Por supuesto, con este libro no se pretende simplificar algo que es, evidentemente, complejo por sí mismo, como es el análisis estadístico. Más bien, hemos tratado de presentar, de la forma más clara posible y a nivel introductorio, unos conceptos, unos métodos y unas técnicas de trabajo que son absolutamente necesarios para el trabajo científico en sociología. Y todo ello utilizando una presentación verbal de los conceptos, rehuendo, en la medida de lo posible, las derivaciones matemáticas, mediante la utilización de un lenguaje científico claro, inteligible y actualizado de nuestro idioma. En el texto, cuando utilizamos un término en español para expresar un vocablo estadístico que suele expresarse en inglés, ponemos a continuación, entre paréntesis, la expresión original inglesa para evitar confusiones. La razón estriba en que creemos que debemos esforzarnos en utilizar correctamente nuestra lengua en cualquier disciplina, máxime cuando contiene términos en esta que recogen el significado de las expresiones originales inglesas. De ahí el esfuerzo que se ha realizado por expresar en español el lenguaje que en su origen es mayoritariamente inglés.

Como hemos dicho anteriormente, el libro consta de 10 capítulos que presentan un panorama del análisis estadístico en su creciente complejidad. Comenzando con el análisis estadístico univariable y la lógica de la comparación se llega al estudio en el último capítulo de una panorámica del análisis multivariable, dedicando los capítulos intermedios al estudio asociativo y correlacional entre dos variables. Además, tanto en el estudio del análisis estadístico para una, dos o más variables, en primer lugar, se

presentan los conceptos y técnicas desde la perspectiva de la estadística descriptiva, para pasar a continuación a estudiarlos desde la perspectiva de la estadística inferencial.

De este modo, hemos tratado de lograr una continuidad en el estudio de la estadística, poniendo en relación las técnicas más complejas con las más sencillas; de ahí que resulte conveniente, cuando se utilice el libro en un curso introductorio, seguir el orden numérico de los capítulos. Pero de manera especial esperamos que la orientación pedagógica que nos ha guiado en la preparación de este libro contribuya positivamente a lo acordado en la Conferencia de Responsables Académicos del Área de Sociología de las universidades españolas celebrada en la Facultad de Ciencias Sociales de la Universidad de Valencia (21/22 de noviembre de 2013). Una Conferencia en la que se aprobaron diversas propuestas orientadas hacia la promoción profesional y mejora de la calidad normativa de los estudiantes matriculados en Sociología. Unas propuestas encabezadas por estos dos elementos:

- Reforzar la formación estadística de nuestras graduadas y graduados.
- Proporcionar más herramientas de sociología aplicada en la formación de los graduados y graduadas.

También confiamos que esta nueva edición actualizada y adaptada a los nuevos estudios de grado, que vuelve a aparecer en Alianza Editorial, tenga la misma acogida que las anteriores y que la utilización del libro en las aulas de clase y en los gabinetes de estudio responda a los objetivos para los que ha sido preparado. Para terminar, de nuevo, expresamos nuestra gratitud al CIS por haber dado todo tipo de facilidades a Alianza Editorial para que continuara su labor editora.

1. El análisis estadístico de los datos sociológicos

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 alumnado de las facultades universitarias de ciencias humanas y sociales poseer unos conocimientos básicos y rigurosos sobre el contenido y alcance de la estadística, que 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: 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 tres 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 han contribuido al desarrollo contemporáneo de la estadística, por medio de la teoría del análisis de componentes y de la teoría del Ji-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, en tanto que Gauss (1809) y Laplace (1812) trabajaron con 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 historiadores clásicos 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 (1796-1874) y otros autores cuando introdujeron la idea 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, me-

teorología, 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. Un proceso de superación que también acabó influyendo en la incorporación de mujeres con aptitudes matemáticas al tratamiento científico de la estadística, un ámbito que durante siglos había estado dominado por hombres. Hubo que esperar a 1873 para que Florence Nightingale (1820-1910), pionera entre otras cosas en el uso de representaciones visuales de la información y en gráficos estadísticos, fuera elegida la primera mujer miembro de la Royal Statistical Society y, posteriormente, miembro honorario de la American Statistical Association.

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 (1822-1911), Pearson (1857-1936), Yule (1871-1951) y, más tarde, Fisher (1890-1962) 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 de uno y otro género a buscar regularidades en la vida humana y social. La ciencia estadística era, para ilusionadas y curiosas mentes, un nuevo y excitante instrumento para la exploración del mundo viviente; de este modo, las ciencias del comportamiento humano 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 impor-

tante 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 (1978: 1098), fue tanto lo que ocurrió en el campo de la estadística entre 1920 y 1940, que se hacía preciso un periodo de reposo para asimilar todo lo descubierto. La moderna estadística, que es sobre todo inglesa en su etapa de máximo desarrollo, es admitida en todos los ámbitos 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 compleja. En los momentos actuales, la estadística y sus profesionales 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 no-paramétricos y otros, todos ellos aplicados también al tratamiento y análisis de datos sociológicos, como tendremos ocasión de ver a lo largo del presente libro.

La historia de la estadística en el siglo xx está repleta de figuras que han desarrollado estadísticos, pruebas y técnicas a las que han dado nombre. Entre ellas, se encuentran Weibull (1887-1979), Wilcoxon (1892-1965), Mahalanobis (1893-1972), Hotelling (1895-1973), Cox (1900-1978), McNemar (1900-1986), Smirnov (1900-1966), Wald (1902-1950), Kolmogorov (1903-1987), Wilks (1906-1964), Cochran (1909-1980), Haenszel (1910-1998), Bartlett (1910-2002), Kish (1910-2000), Wallis (1912-1998), Tukey (1915-2000), Guttman (1916-1987), Cronbach (1916-2001), Box (1919-2013), W. H. Kruskal (1919-2005), Mantel (1919-2002), Theil (1924-2000), Akaike (1927-2009), J. B. Kruskal (1928-2010), Goodman (1928-), Gabriel (1929-2003), Shepard (1929-), Shapiro (1930-) y Jenkins (1933-1982).

Más recientemente, sobresalen McFadden (1937-), que realizó relevantes aportaciones en las regresiones logísticas, y Heckman (1944-), con importantes contribuciones en la evaluación de las políticas económicas. Ambos destacan también porque compartieron el Premio Nobel en Economía en el año 2000. Finalmente, nacidos tras la Segunda Guerra Mundial, Hausman (1946-) propuso en el campo de la econometría las pruebas de especificación a fin de detectar la omisión de importantes

variables y Arellano (1957-) y Bond (1963-) han aportado importantes avances en el análisis de los datos longitudinales de panel.

1.3. Funciones de la estadística: estadística descriptiva y estadística inferencial

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 funciones 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, Benjamin Disraeli, que dijo en cierta ocasión que había 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 (1978: 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 en el mundo científico, 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 métodos 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 a partir de los resultados obtenidos de una muestra conocida. Como señala Blalock (1960: 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. 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: 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, 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 los 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 descripción 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 los datos de un Censo de Población, se puede tratar como un parámetro poblacional, mientras que el promedio de años de escolaridad que tienen las muestras de jóvenes estudiadas en una encuesta de la juventud 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 cues-

tionario, 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 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 alumnado aprenda a valorar el papel y funciones que desempeñan los diferentes instrumentos teóricos y metodológicos que se utilizan.

Debe, pues, quedar claro desde un principio para el alumnado 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 muy 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, creemos 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: 38).

* 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 operar con datos cuantitativos, véase García Ferrando (2014: 23-71).

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 actualmente de la investigación empírica de nuestra disciplina.

En cierto modo, la situación que se planteaba en la sociología a finales del siglo XIX y principios del XX era en buena medida opuesta a la actual. Se disponía de poca información estadística para poder contrastar las teorías, pero algunos autores 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 alcances 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 nuestra opinión que los nuevos desarrollos de la estadística, como por ejemplo el análisis multivariable, una vez generalizado su uso en el trabajo sociológico, van 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 suele 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

* Los indicadores sociales, como mediciones repetidas en el tiempo de diversos fenómenos sociales, son cada vez más utilizados tanto por los científicos sociales como por los políticos y los planificadores. Desde mediados de los años setenta del siglo XX, los gobiernos de diversos países comenzaron a publicar, de forma regular, informes basados en indicadores sociales como así hizo la Oficina del Censo de los Estados Unidos en 1973 con el informe titulado *Social Indicators*. En España, el Instituto Nacional de Estadística publicó, por primera vez en 1975, *Panorámica Social*. Para un tratamiento actualizado de la cuestión se recomienda la lectura del libro de C. Land et al. (eds.), *Handbook of Social Indicators and Quality of Life Research*, Dordrecht, Springer, 2012.

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 difiere la población y que son de interés para la investigación. 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 estos 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 las personas entrevistadas 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 *censo* lo forman el conjunto de todos los valores de las variables que desea 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 postmaterialistas en las sociedades industriales avanzadas, las observaciones estadísticas no están limitadas en forma tan concreta. En tal caso, pues, la investigación sociológica trata con poblaciones estadísticas infinitas o universos conceptuales.

Ahora bien, sea finito o infinito el universo estudiado, en la sociología raras veces se tratan todas las observaciones estadísticas; en cambio, se selecciona un subconjunto de unidades llamado *muestra estadística*. Precisamente, a la obtención de muestras que sean suficientemente representativas de la población que se pretende estudiar dedica la sociología 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 es la que 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.

A la consecución de medidas válidas y fiables de las variables que intervienen en la investigación se han de dedicar los mayores esfuerzos, ya que por sofisticados que sean los métodos estadísticos que se empleen en los cálculos, los resultados no van a mejorar si los datos de partida no son de buena calidad. Loether y McTavish (1974: 357) sugieren que cuando se 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, se 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 estatus socioeconómico de la población, corresponden índices más elevados de conciencia autonómica». En ella encontramos dos conceptos que merecen una definición más clara, «estatus socioeconómico» y «conciencia autonómica». Supongamos que definimos el primero como la posición relativa de las personas en una jerarquía de estatus, y que la conciencia autonómica se define como un fenómeno subjetivo que implica la existencia de una identidad autonómica, así como la percepción del hecho diferencial autonómico y de los intereses de la comunidad autónoma. Con unas definiciones tales, resulta difícil utilizarlas para medir el estatus socioeconómico y la conciencia autonómica 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. Sólo cuando se

haya operacionalizado 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 (1951: 1), se puede definir la *medición* como el procedimiento de asignación de numerales a objetos o acontecimientos de acuerdo con ciertas reglas. 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 estatus social, por ejemplo, es una variable de las personas y de los grupos sociales, respectivamente. Si para medir el estatus social utilizamos un índice y hacemos que dicho índice varíe entre un mínimo de 10 y un máximo de 100, una medición de dicho índice para un individuo determinado, de valor 70, representará una observación estadística y lo utilizaremos como el valor de la variable estatus 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, entre otros, los indicadores con los que suele operativizar el concepto de estatus social. La combinación ponderada de los valores que toma cada uno de los tres indicadores forma un índice que tomará valores numéricos concretos, con lo que se operará estadísticamente en la investigación.

Cuando se busca el procedimiento de medición más adecuado han de tenerse en cuenta, sobre todo, 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 anteriormente referido es válido como medida del estatus social porque mide posiciones en una jerarquía de estatus y no otro concepto, como pudiera ser el prestigio social—, y la *fiabilidad* hace referen-