

CAPÍTULO 2-1 DESCRIPCIÓN E INDUCCIÓN ESTADÍSTICAS EN CIENCIAS SOCIALES⁸²

La estadística es esencialmente una colección de métodos matemáticos para procesar datos con el fin de resumir o bien generalizar la información que nos proporcionan. La estadística descriptiva es la parte de la estadística que sirve para resumir la información. El proceso de generalizar la información se llama inducción estadística.

2.1.1 ESTADÍSTICA DESCRIPTIVA

Al enfrentarse a un conjunto de datos aunque pequeño, el cerebro humano se ve incapaz de captar de manera inmediata toda la información detallada que contiene.⁸³ La solución a este problema consiste en dejar de lado los detalles para concentrar su atención en los grandes rasgos. La estadística descriptiva permite procesar metódicamente los datos con el objetivo de condensar la información que contienen. Con cal-

⁸² Blalock (1979); Wonnacott y Wonnacott (1992).

⁸³ Según Georges Ifrah (1994, tomo 1, p. 33-34, “Les limites de la perception directe des nombres”), el cerebro no puede aprehender de manera concreta, es decir, sin contar (lo que constituye una abstracción) más de cuatro objetos al mismo tiempo. Por ejemplo, sin contar, no podemos, con la pura mirada, diferenciar entre cinco o seis objetos.

cular porcentajes, promedios, desviaciones estándar o coeficientes de correlación, es posible tener una visión global de los datos. Sin embargo, debemos estar conscientes de que al resumir los datos de esta manera, perdemos parte de la información que contenían, y esto puede ser causa de errores al menos que seamos prudentes al momento de interpretar.

2.1.2 LA INDUCCIÓN ESTADÍSTICA

Es casi imposible, y particularmente en ciencias sociales, conseguir los datos pertinentes de la totalidad de un fenómeno en estudio. Es más probable que las observaciones que están a nuestra disposición se apoyen en una parte del fenómeno. Entendemos, entonces, la necesidad de generalizar a partir de una información incompleta. Sin embargo, para que se considere una generalización ciertamente científica, es necesario fundamentarla con principios epistemológicos.⁸⁴

Los métodos de inducción estadística son, de hecho, una expresión matemática de principios epistemológicos gracias a los cuales se pueden inferir proposiciones de alcance más general a partir de la información que se obtiene de un conjunto de datos particulares. La inducción estadística es, por lo tanto, una manera científicamente válida de pasar de lo particular⁸⁵ a lo general.

Concretamente, el proceso inductivo estadístico busca sacar conclusiones con relación a las diversas características de una población a partir de hechos observados en una muestra obtenida de ésta. La estadística emplea la palabra *parámetros* para designar las características de la población y la palabra *estadísticas* para designar las características de la muestra.

⁸⁴ La epistemología es una parte de la filosofía que consiste en el estudio crítico de las ciencias con el fin de determinar, de cada una, su origen lógico, su valor y su alcance.

⁸⁵ Hay una broma en inglés que ilustra el carácter particular de los datos: “‘Data’ is the plural of ‘anecdote’ ” (‘Datos’ es el plural de ‘anécdota’).

Es importante siempre recordar que se consideran, en una situación normal, los parámetros como unos valores fijos con relación a una población y que, por lo general, son desconocidos (puesto que no se conoce, en su totalidad, la población misma). Por lo contrario, dado que se puede obtener más de una muestra de una población específica, las estadísticas son valores que pueden variar de una muestra a otra; por otro lado, se conocen o se pueden calcular los valores de las estadísticas de una muestra dada. Sin embargo, no sabemos hasta que grado una muestra es representativa de la población en general, ni en qué medida una estadística calculada a partir de esta muestra se parezca al parámetro correspondiente a la población desconocida.

Ejemplos de inducción estadística:

1. Basándose en las respuestas obtenidas por sondeo efectuado a una muestra de la población de una ciudad, estimar la proporción de los ciudadanos que son favorables a algún proyecto urbanístico.
2. Partiendo de la hipótesis (del modelo), aceptada a priori, que la relación macroeconómica entre el ingreso de los hogares y las inversiones en construcción de vivienda privada se puede describir con la ecuación

$$I = a + bR$$

(donde I es el monto de las inversiones y R el ingreso agregado), estimar el valor de los parámetros a y b a partir de los datos publicados por INEGI para el Estado de Puebla de 1974 a 1994.⁸⁶

Es de hacer notar que las medidas que se usan en estadística descriptiva (promedio, desviación estándar...) se usan también en el contexto de la inducción estadística. No obstante, en estadística descriptiva no se distingue entre parámetros y estadísticas porque la estadística descriptiva no hace distinción entre población y muestra.

⁸⁶ Este ejemplo de modelo es, de toda evidencia, demasiado simplista.

2-1.3 LAS PROBABILIDADES Y LA INDUCCIÓN ESTADÍSTICA: LA RELACIÓN ALEATORIA ENTRE UNA MUESTRA Y LA POBLACIÓN

Con la inducción estadística dejamos el dominio de certidumbre. En efecto, la inducción estadística parte de una muestra entre tantas otras posibles que se pueden obtener de la población en estudio: si obtenemos de una población dada una muestra de un tamaño específico con un procedimiento determinado, en caso de repetir el proceso, la segunda muestra tiene mucha probabilidad de ser diferente a la primera. Para una población dada existe, por consiguiente, muchas muestras posibles. El conjunto de muestras posibles forma también una población en términos estadísticos: los “individuos” de esta población son las muestras.

Por ejemplo, el total de los abonados del teléfono en la ciudad de Montreal forma una población. Podríamos obtener de esta población una muestra de 1000 abonados seleccionados al azar en un directorio telefónico. Luego podríamos repetir el proceso y obtener una segunda muestra, luego una tercera, etc. (de hecho, podríamos repetir este proceso al infinito siempre y cuando se pueda seleccionar nuevamente los abonados que pertenecen a una muestra⁸⁷). El conjunto de todas las muestras posibles de 1000 abonados seleccionados al azar en el directorio es la población de las muestras.

Entre las muestras posibles, algunas son representativas de la población estudiada mientras que otras lo son menos. Puesto que no conocemos la población sino es por la muestra, nunca podemos saber con certeza hasta que grado la muestra específica que se obtuvo es representativa de la población en

⁸⁷ No se debe confundir este proceso con la muestra con reemplazo (sustitución) cuando se sortean los individuos que constituyen una muestra de manera secuencial y cuando un individuo que se sorteó es nuevamente elegible para el sorteo que sigue.

general. La relación entre la muestra y la población es, por lo tanto, esencialmente aleatoria (o sea, influenciada por el azar).

En resumen, está claro que no es posible saber con certeza en qué medida una estadística calculada a partir de datos de una muestra se parece al parámetro desconocido correspondiente en la población. Sin embargo, la teoría de las probabilidades nos facilita herramientas para evaluar la probabilidad de que la diferencia (el error de estimación) entre la estadística y el parámetro se sitúe en el interior de un cierto margen. Es por consiguiente con la teoría de las probabilidades que fundamentamos las reglas de la inducción estadística.

