

CONCEPTOS BÁSICOS DE ESTADÍSTICA

Jorge M. Galbiati Riesco

La Estadística está constituida por un conjunto de métodos de análisis de datos que pueden agruparse en tres categorías: La Estadística Descriptiva, el Análisis Exploratorio de Datos y la Inferencia Estadística, que se describen a continuación:

Estadística Descriptiva. Es un conjunto de procedimientos que tienen por objeto resumir masas de datos por medio de números calculados a partir de ellos, o medidas de resumen, tablas y gráficos.

Análisis Exploratorio de Datos. Conjunto de procedimientos que tienen por finalidad la búsqueda de estructuras en una masa de datos, que conduzcan a la percepción de las características que posee la fuente de donde provienen.

Inferencia Estadística. Conjunto de procedimientos que permiten confirmar o concluir propiedades de una población fuente de información, de la cual sólo conocemos una parte representativa, o muestra.

Variabilidad. Nombre que se da a las diferencias en el comportamiento de todo fenómeno observable que se repite bajo iguales condiciones, debidas a cambios en factores no controlables, que influyen sobre él. Estas diferencias pueden ser casi imperceptibles, como en el caso de experimentos de laboratorio, donde hay un alto grado de control sobre los factores que influyen sobre el fenómeno; pueden ser pequeñas, como en el caso de en procesos industriales, y pueden ser grandes, como en el caso de fenómenos en que está involucrado el comportamiento humano, como los fenómenos psicológicos, sociológicos y económicos.

La variabilidad existente en los fenómenos se puede reducir, se puede explicar parcialmente, pero no se puede eliminar.

Los siguientes conceptos se relacionan con el acto de extraer datos de una fuente de información, lo que comúnmente se denomina medir:

Medición. Es la observación de un fenómeno o propiedad, y la asignación de un número o categoría, como forma de representar ese fenómeno. Suele usarse el término *observar* como sinónimo de medir.

Por ejemplo, pesar un objeto es observar una propiedad, su peso, y asignarle un número, el número de kilogramos que pesa. Determinar el estado civil de una persona también es efectuar una medición, que consiste en clasificar esa propiedad en una de varias categorías: soltera, casada, etc.

Medida. Es el número o categoría que se asigna al medir un fenómeno. Un sinónimo de medida es *observación*.

Escala de Medida. Una escala de medida es el conjunto de valores que puede tomar una determinada medida. Se pueden clasificar en los siguientes tipos:

Escalas de medida Categóricas : Nominales.

 Ordinales.

Escalas de medida Numéricas ; De intervalos.

 De razón.

Las escalas categóricas sólo asignan una categoría, o clasifican el fenómeno o propiedad que se mide. Las Ordinales se distinguen de las nominales, en el hecho que sus valores tienen un orden natural. Las escalas numéricas asignan números. Si la escala tiene un cero absoluto, la escala es de razón, si el cero es arbitrario, la escala es de intervalo. Por ejemplo, al determinar el color de ojos de una persona se está midiendo en una escala categórica nominal, su fecha de nacimiento se mide en escala categórica ordinal. Su temperatura corporal, en grados centígrados, se mide en una escala numérica de intervalos. Su peso, en escala numérica de razón. En esta escala tiene sentido decir que una persona tiene el doble de peso que la otra.

Instrumento de medida. Es un elemento sensible al fenómeno que se desea medir, y que se emplea para medir.

Por ejemplo, una regla graduada, un examen de laboratorio, una encuesta, el juicio de una persona, etc.

Validez de un instrumento de medida. Un instrumento de medida es válido respecto de una propiedad, si es relevante a ella, es decir, un cambio en la propiedad, determina un cambio en la medida, y un cambio en la medida obedece sólo a un cambio en la propiedad. Un instrumento válido proporciona una medida válida. Por ejemplo, para medir pesos se debe usar una balanza, para medir cortesía se debe usar un instrumento adecuado, como una encuesta, si se desea que sean válidas las medidas.

Exactitud de un instrumento de medida. Un instrumento de medida es exacto si tiene la capacidad de describir el fenómeno sin sobrevalorarlo o subvalorarlo sistemáticamente. Si un instrumento no es exacto, se dice que tiene sesgo. Un instrumento de medida exacto proporciona medidas exactas. Por ejemplo, un examen puede no rendir una medida exacta de los conocimientos de los alumnos si es extremadamente fácil o extremadamente difícil, en relación al conocimiento que se quiere medir.

Confiabilidad de un instrumento de medida. Un instrumento de medida es confiable si tiene la capacidad proporcionar medidas similares cuando se aplica reiteradas veces al mismo fenómeno. Como sinónimo se usa el término *precisión*. Un instrumento confiable proporciona medidas confiables.

Los conceptos siguientes se relacionan con las fuentes de información y los conjuntos de observaciones disponibles para el análisis:

Población. O *población objetivo*, es el conjunto de todos los valores de un fenómeno o propiedad que se quiere observar. También se usa el nombre de *variable* para designar a este conjunto. Por ejemplo, las edades de los escolares de enseñanza media del país, las preferencias de marca de jabón manifestadas por un conjunto de consumidores, los diámetros de los ejemplares de un objeto producido por una máquina, etc.

Muestra. Es la parte de la población que efectivamente se mide, con el objeto de obtener información acerca de toda la población. La selección de la muestra se hace por un procedimiento que asegure en alta grado que sea representativa de la población. Los métodos de selección de muestras se describen más adelante.

Censo. Es el proceso de observar la población completa. Es decir, tomar una muestra igual a la población.

Unidad Muestral. Es cada una de los miembros individuales de una población. Cada unidad muestral proporciona una medida.

Marco muestral. Es la parte de la población desde donde se selecciona la muestra. Idealmente el marco muestral coincide con la población. Sin embargo, por razones de costo, se suele no considerar una parte de la población, al seleccionar la muestra. Las conclusiones que se obtienen como resultado de un estudio estadístico aplicado a los datos obtenidos de una muestra, se refieren a toda la población, aún cuando la muestra haya sido obtenido de un marco muestral que no es toda la población. Por ejemplo, la población es el conjunto de todos los trabajadores de una empresa. Para efectuar un estudio, se extrae una muestra de los trabajadores de la oficina central, dejándose fuera del marco muestral, pero no de la población, a los de una sucursal que se encuentra geográficamente distante.

Parámetro. Nombre dado a una característica global de una población. En general, un parámetro no es conocido. Por ejemplo, la edad promedio de una población de habitantes de una región.

Estadístico. Es una característica de la muestra. Es un valor conocido, que varía de una muestra a otra. Se utiliza para obtener conclusiones acerca de la población. Por ejemplo, el promedio de edad de los habitantes seleccionados en una muestra es un estadístico. Se puede utilizar para estimar la edad promedio de la población de la que se obtuvo la muestra.

Errores muestrales. Son errores causados por el acto de obtener la información a partir de una muestra, en lugar de la población completa. Se deben a la variación existente entre una muestra y otra.

Errores no muestrales. Son errores no asociados al hecho de obtener información a partir de una muestra. Por ejemplo, datos faltantes, errores de recolección de datos, errores de respuesta, errores de transcripción, errores de proceso.

Para que un plan de muestreo produzca muestras representativas de la población, la selección debe ser al azar, en el sentido que se pueda conocer el modelo probabilístico que describe el conjunto de todas las distintas muestras que es posible obtener. A continuación se describen algunos procedimientos para obtener muestras, o *métodos de muestreo*:

Muestreo Aleatorio Simple. Es un procedimiento de toma de muestra, en el que todas las muestras posibles, de un tamaño fijo, tienen igual probabilidad de ser seleccionada. Para obtener una muestra por este procedimiento, se enumeran todas las unidades muestrales, y se sortean unidades hasta completar el número requerido.

Muestro Sistemático. Se practica cuando se dispone de una lista de todas las unidades muestrales, en un orden independiente de la variable que se desea medir. Supóngase que el tamaño de la población es **N**, y el tamaño seleccionado para la muestra es **n**. Sea **M** el entero más cercano al valor del cociente **N/n**. El procedimiento de muestreo sistemático consiste en seleccionar al azar, un número entre **1** y **M**, correspondiendo éste al orden de la primera unidad seleccionada. Luego se recorre la lista y se selecciona una de cada **M**, hasta completar las **n** unidades muestrales requeridas.

Muestro Estratificado. Consiste en dividir la población en grupos homogéneos, o estratos, de acuerdo a los valores de una variable relacionada con la característica bajo estudio, y aplicar el método de muestreo aleatorio simple en cada estrato. Por ejemplo, si se desea medir las preferencias de una población de consumidores, se separa la población en estratos de acuerdo al nivel socioeconómico, y se muestrea en cada estrato separadamente.

Muestreo por Conglomerados. Consiste en dividir la población en pequeños grupos, o conglomerados, obtener una muestra aleatoria simple de conglomerados, y practicar un censo dentro de cada conglomerado seleccionado en la muestra. Por ejemplo, para realizar una encuesta de hogares, en una ciudad, en lugar de obtener una muestra de hogares, se obtiene una muestra de manzanas. Se envía un encuestados a cada manzana seleccionada, con el objeto de encuestar todos los hogares de las manzana, lo que resulta más eficiente.

PREGUNTAS

1. Describa los tipos de escalas de medida que conoce.
2. Se diseñó un instrumento de medida y se piensa que tiene las propiedades de validez, confiabilidad y exactitud. Explique estas tres características.
3. Una encuesta para medir la calidad de atención recibida por el público que acude a una oficina debe tener estas tres propiedades: validez, confiabilidad y exactitud. Explique en qué consiste cada una de estas propiedades.
4. ¿Qué es medir?
5. Describa lo que una población objetivo, un marco muestral y una muestra
6. ¿Qué diferencia hay entre los errores muestrales y los errores no muestrales?
7. Describa lo que un parámetro y que es un estadístico.

¿Cuándo una muestra es un censo?

8. Se selecciona una muestra de hogares, de la siguiente forma:
Se sortean 20 manzanas, al azar, de un plano de la ciudad. Se envían encuestadores a cada manzana, con las siguientes instrucciones:

Parta encuestando la primera casa que vea. Luego encueste una de cada tres, hasta dar vuelta la manzana. ¿Qué métodos de muestreo se están empleando?

9. ¿Qué característica fundamental debe cumplir un plan de muestreo para que la muestra obtenida sea representativa de la población?