

### ESTADÍSTICA: DEFINICIÓN

El término Estadística tiene dos acepciones fundamentales. Por un lado, la Estadística como ciencia o método científico y por otro lado la estadística o estadísticas como conjunto o colecciones de datos. Este segundo concepto es muy usado hoy en día para referirnos a resultados ya elaborados en un estudio en el que se empleó la Estadística como método.

La Estadística constituye una poderosa herramienta para generar conocimiento y ha experimentado un vigoroso desarrollo desde sus orígenes hasta nuestros días. Actualmente, se aplica en todas las áreas del saber y, de manera determinante, en las Ciencias Sociales. Por ejemplo, en Administración de Empresas se utiliza la estadística para evaluar la aceptación de un producto antes de comercializarlo; en Economía, para medir la evolución de los precios mediante números índice o para estudiar los hábitos de los consumidores; en Sociología para realizar investigación social estudiando los perfiles y dinámica de colectivos sociales; en el ámbito de las Relaciones Laborales elaborando análisis de salarios, paros, accidentes, expedientes de regulación de empleo, etc.; en Criminología para el análisis de los rasgos de la delincuencia así como la prevención del crimen. En general, en las Ciencias Sociales la estadística se utiliza para medir las relaciones entre variables y hacer predicciones sobre ellas.

### ETAPAS DE UN ESTUDIO ESTADÍSTICO

El proceso de investigación supone un conjunto de etapas que podemos representar mediante el siguiente esquema:

- 1. Planteamiento del problema:** consiste en definir el objeto de la investigación y precisar el universo o población al que se refiere el estudio.
- 2. Planificación del trabajo de campo:** en esta etapa se toman decisiones con respecto a los procedimientos de entrevista, características del muestreo, diseño de herramientas, etc.

**3. Recopilación de información:** incluye la recogida de los datos y también la depuración de la información obtenida, es decir, tratar los problemas de la no-respuesta, los errores de campo, los errores de oficina, los datos desaparecidos y los datos anómalos.

**4. Análisis de los datos:** esta etapa presenta varias fases:

a) Análisis descriptivo: organización (mediante tablas y gráficos) y resumen (en pocos valores que proporcionen la máxima información posible) los datos disponibles para extraer la información relevante en nuestro estudio.

b) **Inferencia estadística:** en el caso de trabajar con muestras, la inferencia se encarga de generalizar los resultados de la muestra a la población y obtener conclusiones generales. La inferencia se basa en suponer que la población sigue un modelo o una distribución conocida y los datos que tenemos son realizaciones aleatorias de ese modelo. Para cuantificar la fiabilidad de estos resultados se recurre al cálculo de probabilidades.

c) Validación del modelo: finalmente es necesario diagnosticar la validez de los supuestos del modelo que nos han permitido interpretar los datos y llegar a conclusiones sobre la población.

**5. Interpretación:** obtener las conclusiones de los resultados obtenidos.

En el proceso descrito hemos definido las tres ramas fundamentales de la Estadística como ciencia y que son objeto de estudio en las asignaturas habituales de Estadística:

- **Estadística Descriptiva.**
  
- **Cálculo de probabilidades.**
  
- **Inferencia Estadística.**

## CONCEPTOS BÁSICOS

Generalmente el objetivo de un estudio estadístico es conocer información sobre alguna característica de cierto conjunto de elementos. Este conjunto de elementos normalmente tiene un tamaño excesivamente grande para abarcarlo en su totalidad, por lo que nos centraremos en la obtención de información y estudio sobre algunos de ellos, un subconjunto, que podremos hacer extensible al total. Para matizar este problema, se suelen emplear los siguientes términos:

- **Población:** Conjunto de personas, objetos, ideas o acontecimientos sometido a una observación estadística.
- **Individuo o elemento:** Cada uno de los elementos de la población.
- **Muestra:** Subconjunto de una población.
- **Carácter o variable:** Cada una de las propiedades, rasgos o cualidades que poseen los elementos de una población y que son objeto de estudio. Haremos la siguiente clasificación:
  - **Variable cualitativa:** Los valores que toman no se pueden cuantificar. Cada uno de estos valores se denomina categoría, clase o modalidad. Pueden ser Ordinales o Nominales dependiendo de si se puede establecer un orden entre las diferentes categorías o no. Ejemplos de este tipo de variables son el carácter rango militar y el sexo, respectivamente.
  - **Variable cuantitativa o medibles:** Los valores que toman se pueden cuantificar o medir. Pueden ser Discretas (los valores que pueden tomar son aislados) o Continuas (pueden tomar cualquier valor de la recta real o de un intervalo). Ejemplos de este tipo de variables son el carácter número de hermanos y el precio de unas acciones respectivamente.
- **Parámetro:** es un valor numérico calculado a partir de una fórmula matemática, **obtenida a partir de datos de la población**. Por ejemplo: el porcentaje de españoles a favor de autorizar la eutanasia en ciertos casos.
- **Estimador o Estadístico:** es un valor numérico **derivada de un conjunto de datos de una muestra**, con el objetivo de estimar o inferir características de una población o modelo estadístico. Se trata por tanto de una variable aleatoria que

depende de la muestra. Por ejemplo: el porcentaje de los españoles encuestados a favor de autorizar la eutanasia en ciertos casos.

### ¿QUE ES EL MUESTREO?

El muestreo es una herramienta de la investigación científica, cuya función básica es determinar que parte de una población debe examinarse, con la finalidad de hacer inferencias sobre dicha población.

La hora de elegir una muestra hay que prestar mucha atención a dos aspectos:

- Tamaño de la muestra, deberá ser el adecuado.
- La forma de seleccionarla, los elementos de la muestra deberán ser elegidos aleatoriamente, todos los elementos deberán tener la misma probabilidad de ser elegidos.

Si se cumplen estas dos condiciones, diremos que la muestra es **representativa**. En caso de que la muestra esté mal elegida y la selección no fuera aleatoria, se producen errores adicionales e incontrolables, se dirá que muestra es **sesgada**.

### Ejemplos:

1. Para estimar el número de erratas que contiene un libro de 600 páginas, se quiere seleccionar 20 de sus páginas y contar el número de erratas en ellas.  
En este caso la población la forman las 600 páginas y la muestra las 20 páginas que vamos a seleccionar. Para esta selección podemos seguir básicamente dos procedimientos:  
Elegir al azar, mediante un sorteo 20 de las 600 páginas.  
Dividir 600 entre 20 y de las 30 primeras elegir una página al azar. Las 19 restantes se eligen sumando de 30 en 30. Por ejemplo, si la primera elegida fue la 19, las páginas elegidas sería : 19, 49,79,...589.  
En cualquiera de los dos casos a muestra se puede considerar representativa.
2. En lo que conocemos como sondeos o encuestas electorales, los elementos de la muestra (votantes) se escogen convenientemente (tomando aquellos afines al

partido o idea política del medio de comunicación) o se toman muy pocos elementos de la muestra y los resultados no tienen mucho significado ya que la muestra no es representativa.

## **TÉCNICAS O TIPOS DE MUESTREO**

La forma de obtener una muestra depende, fundamentalmente, de la estructura de la población de la que pretendemos obtener información.

Podemos diferenciar entre: muestreos no aleatorios y muestreos aleatorios.

### **MUESTREO NO ALEATORIO**

En ocasiones es imposible elegir la muestra mediante un proceso aleatorio lo que hace tener que trabajar con una muestra no aleatoria y por lo tanto no representativa de la población y obtendremos errores. Por ejemplo, en un control de calidad de un contenedor de cajas de pescado congelado, la muestra 7 cajas, que se eligen nada más abrir el contenedor. En este caso la muestra es una parte de la población que es accesible.

### **MUESTREO ALEATORIO**

En el muestreo aleatorio los componentes de la muestra son elegidos al azar y todos tienen la misma probabilidad de ser elegidos para formar parte de la muestra y, por consiguiente, todas las posibles muestras de tamaño “n” tienen la misma probabilidad de ser elegidas. Por este procedimiento las estimaciones obtenidas son fiables. A continuación, veremos las técnicas más habituales para seleccionar una muestra de forma aleatoria.

#### ➤ **Muestreo aleatorio simple:**

De entre todos los elementos de la población, el muestreo aleatorio simple selecciona a cada elemento de la muestra de forma aleatoria teniendo todos los elementos las mismas probabilidades de ser escogidos. Este muestreo es adecuado cuando la población es homogénea en la característica que se va a

estudiar. Si el tamaño de la población es  $N$  tendremos que seleccionar  $n$  de forma aleatoria. El procedimiento empleado es el siguiente:

**1º** Asignamos un número a cada individuo de la población.

**2º** A través de alguno medio mecánico (bolas dentro de una bolsa, tablas de número aleatorios, o generados por la calculadora u ordenador...) elegimos tantos elementos como sean necesarios para completar la muestra.

**Ejemplo:**

Una empresa tiene 200 trabajadores. Si se quiere saber el nivel de satisfacción de los trabajadores, ya que encuestar a todos es demasiado laborioso, se pretende encuestar únicamente a 30.

Por tanto: población (200 trabajadores), muestra (30 trabajadores)

Para escoger a estos trabajadores se va a emplear muestreo aleatorio simple y para ello se enumeran a los 200 trabajadores y se escogen aleatoriamente a 30.

➤ **Muestreo aleatorio sistemático:**

Este proceso exige como el anterior, numerar todos los elementos de la población, pero solamente se selecciona un elemento al azar y, a partir de él, se van seleccionando los componentes de la muestra igualmente espaciados.

**Ejemplo:**

En una maratón participan  $N=200$  atletas. La organización decide realizar un control antidopaje a una muestra de  $n=10$  atletas.

Ordenamos a los 200 atletas por su número de dorsal. Calculamos el coeficiente de elevación, en este caso  $k = \frac{200}{10} = 20$ . Por sorteo elegimos un número del 1 al 20, por ejemplo, si sale el 11, este será el origen. Los siguientes se eligen de 20 en 20 partiendo desde el 11, así la muestra elegida será: 11,31,51,71,91,111,131,151,171,191.

➤ **Muestreo aleatorio estratificado:**

Este muestreo, se usa cuando la población estudiada no es homogénea y queremos que la muestra contenga elementos de cada parte de la población, para así obtener resultados más fiables. En el muestreo aleatorio estratificado,

la población, de tamaño  $N$ , se divide en subpoblaciones homogéneas respecto a alguna característica, denominadas estratos.

Una vez decidido el tamaño  $n$  de la muestra total, el tamaño de la muestra en cada estrato poder ser:

Muestreo aleatorio estratificado constante, cuando todas las muestras de cada estrato tienen el mismo tamaño.

Muestreo aleatorio estratificado proporcional cuando el número de elementos seleccionados de cada estrato es proporcional al tamaño del estrato.

### **Ejemplo:**

En un instituto de 950 alumnos, deseamos hacer una encuesta a una muestra de ellos sobre sus futuros estudios. Si la muestra elegida es de  $n = 60$  alumnos, ¿cuántos alumnos de cada estrato formarán parte de la muestra?

Se decide formar tres estratos, por los diferentes niveles que encontramos en el instituto, y elegimos la muestra proporcionalmente:

Estrato 1: Alumnos de 1º y 2º ESO 300 alumnos

Estrato 2: Alumnos de 3º y 4º ESO 250 alumnos

Estrato 3 Alumnos de bachillerato 400 alumnos

Como la muestra tiene que ser de 60 alumnos, calculamos el número que se ha de tomar en cada estrato, de forma proporcional:

$$\text{Estrato 1: } \frac{300}{950} \cdot 60 \approx 19 \text{ alumnos}$$

$$\text{Estrato 2: } \frac{250}{950} \cdot 60 \approx 16 \text{ alumnos}$$

$$\text{Estrato 3: } \frac{400}{950} \cdot 60 \approx 25 \text{ alumnos}$$

## **TECNICAS PARA OBTENER UNA MUESTRA ALEATORIA EN UNA POBLACIÓN FINITA**

### **Elección mediante extracción**

En una caja se introducen tantas bolas o papeletas como individuos tiene la población ( $N$ ). Estos fueron previamente numerados del 1,2,...  $N$ . Se escoge al azar tantas

papeletas como individuos ha de tener la muestra ( $n$ ). Esta operación podrá realizarse de dos formas:

- Sin reemplazamiento: se eligen simultáneamente o bien una la una, las  $n$  papeletas.
- Con reemplazamiento: se eligen una a una las  $n$  papeletas, pero después de cada extracción, la papeleta elegida y anotada, se devuelve la caja.

### **Obtención de números aleatorios**

La calculadora tiene una tecla RAN#, que se le llama generadora de números aleatorios, con la cual se obtiene al azar un número decimal comprendido entre 0,000 y 0,999. Se multiplicamos uno de esos números por el tamaño de la población, obtenemos un número decimal con parte entera comprendida entre 0 y  $N-1$ . Por lo tanto, si tomamos la parte entera del número obtenido mediante la secuencia  $N \times \text{RAN\#} + 1$ , y nos quedamos con la parte entera, obtenemos un número elegido al azar entre 1 y  $N$ .

### **Obtención de una muestra mediante números aleatorios**

Si repetimos  $n$  veces el proceso anterior, obtendremos una muestra de  $n$  elementos similar a que se los obtendría mediante una extracción con reemplazamiento. Completaríamos la operación suprimiendo los elementos repetidos y obteniendo nuevos elementos. Para poblaciones numerosas esto es, evidentemente, el método más cómodo. Si la población es muy grande se obtendrán los números aleatorios con el ordenador, pues necesitaremos que tengan más de tres cifras decimales para poder “separar” todos los elementos de la población.

### **En resumen:**

- Si la muestra la obtenemos por insaculación (extracción de papeletas) deberá realizarse sin reemplazamiento.
- Si se recurre a los números aleatorios, se consigue una muestra como si fuera con reemplazamiento. Puede haber elementos repetidos que deben suprimirse por otros, también elegidos aleatoriamente.

**EJERCICIOS PARA PRACTICAR LO APRENDIDO**

1. Con la estadística inferencial se pretende:

- A) Realizar un análisis descriptivo de los datos de la investigación.
- B) Inferir los parámetros de la población a partir de los estadísticos muestrales.
- C) Describir los estadísticos de una muestra.

2. La representatividad de la muestra es la característica que permite:

- A) Generalizar los estudios de la muestra a la población.
- B) Conocer las distribuciones de probabilidad de las variables.
- C) Calcular los estadísticos.

3. Identificar la población, muestra y variable estadística evaluada en cada caso:

- a) En una escuela se quiere saber cuál es el deporte más practicado por los alumnos. Se realiza la encuesta a 5 alumnos de cada curso.
- b) Se desea conocer cuál es la estatura de los alumnos de una escuela. Se miden 10 alumnos por curso.
- c) Un fabricante de tornillos desea hacer un control de calidad. Para ello toma 1 de cada 100 tornillos producidos y analiza si es defectuoso, su longitud y su diámetro.

- 4.** En cada uno de los siguientes casos, di sí el colectivo es población o muestra:
- a) En una floristería, añaden al arroyo de todas las macetas unas gotas de cierto producto para probar su eficacia.
  - b) En un invernadero, seleccionara 200 plantas que serán regadas con unas gotas de un producto para analizar su eficiencia.
- 5.** En cierta cadena de centros comerciales trabajan 150 personas en el departamento de personal, 450 en el departamento de ventas, 200 en el departamento de contabilidad y 100 en el departamento de atención al cliente. Con el objeto de realizar una encuesta laboral, se quiere seleccionar una muestra de 180 trabajadores.
- a) ¿Qué tipo de muestreo debería utilizar para la selección de la muestra se queremos que incluya trabajadores de los cuatro departamentos mencionados?
  - b) ¿Qué número de trabajadores tendría que seleccionar en cada departamento atendiendo a un criterio de proporcionalidad?
- 6.** De una población de 543 individuos, queremos extraer una muestra de tamaño 40 mediante números aleatorios. Obtén los cinco primeros elementos de dicha muestra.