

Población y Muestras

Ya manejamos estos dos conceptos en el curso pasado, cuando estudiamos Estadística. En este curso vamos a profundizar en su estudio.

Supongamos que deseamos conocer una característica en un conjunto de personas, seres u objetos. Por ejemplo:

- En una fábrica de automóviles tenemos que inspeccionar las emisiones de CO/CO₂ de un determinado modelo.
- Necesitamos conocer cuántos españoles, en edad de trabajar, están ocupadas o están en paro.
- En una fábrica de bombillas queremos saber cuál es la duración media del modelo de 60 vatios.

La característica o propiedad que estudiamos se denomina variable aleatoria y el conjunto de individuos (que pueden ser personas, plantas, objetos, ...) en el que se estudia población.

Pero en muchas ocasiones no se puede estudiar esa característica en todos y cada uno de los individuos de la población. ¿Por qué? Puede ser debido a muchas causas. Veamos algunas:

- La población es muy numerosa y por ello se necesitaría un tiempo y un dinero del que no siempre se dispone.
- No es posible controlar la totalidad de los individuos, porque se trata de personas o animales en constante movimiento (imaginemos aves migratorias, animales de un ecosistema o los clientes de un centro comercial en un día concreto).
- El proceso de medición es destructivo (si encendemos todas las bombillas de 60w de una fábrica hasta que se fundan, para analizar su duración...)

En estos casos, ¿que hacer? Pues no queda más remedio que elegir sólo una parte de la población, que se denomina muestra, y realizar el estudio sólo en los individuos de esa muestra.

Lógicamente, este procedimiento tiene un riesgo: las conclusiones que sacamos se refieren a los individuos de la población que están en la muestra. ¿Nos servirán para todos los de la población? Porque esta es la idea: que el estudio de la muestra sirva para inferir características de toda la población.

Concluyamos, pues:

En el estudio de una característica:

- **Población** o **Universo** es el conjunto de todos los individuos objeto del estudio.
- **Muestra** es cualquier subconjunto extraído de la población.

El estudio de la muestra sirve para **inferir** características de toda la población.

Pensemos en un estudio que recoge datos de toda la población de un Estado, como es el Censo.

Este estudio precisa unos recursos enormes: elaboración de las preguntas, millones de formularios para completarlos, su distribución, cientos de miles de personas entregándolos en mano, comprobando de se rellenan correctamente los datos y entregándolos en su lugar, introducción de todos los millones de datos recogidos en sistemas informáticos, procesamiento de la información (separando poblaciones, edades, sexos, ...), análisis de resultados, etc, etc.

☞ **Ejemplo:** queremos conocer cuántos habitantes de nuestra ciudad, que tiene 13 000 habitantes fuman. Elegimos para averiguarlo una muestra de 100 individuos y obtenemos que el 31% de ellos fuman

Tenemos así que la población tiene un tamaño $N = 13\ 000$, y que la muestra tiene un tamaño $n = 100$. Como hemos obtenido que un 31% de los individuos de la muestra son fumadores, inferimos que también lo será idéntico porcentaje en la población.

Concluimos que hay $13\ 000 \times 0,31 = 4\ 030$ fumadores en nuestra ciudad.

2. Tipos de muestreo

En este epígrafe analizaremos los conceptos y el vocabulario, de forma resumida, en lo que respecta al procedimiento de elección de los individuos que constituirán una muestra.

Supongamos que queremos conocer una determinada característica en una población, pero que nos vemos obligados a restringirnos a su estudio en una muestra. Al elegir una muestra tendremos en cuenta su **tamaño** y la selección de sus individuos.

Se dice que una muestra es **representativa** cuando está bien elegida, pudiendo obtenerse conclusiones razonables para toda la población a partir de ella. En cualquier caso, se producirán errores imprevistos e incontrolables, que se denominan **sesgos**.

Muestreo es el proceso de formación de la muestra.

Se dice que es **aleatorio** cuando todos los individuos se eligen al azar, de modo que todos tienen, a priori, la misma probabilidad de ser elegidos.

De forma simple, podemos distinguir los siguientes tipos de muestreos aleatorios:

- Aleatorio **simple**: se enumeran los individuos y se eligen al azar (por sorteo) tantos como sea el tamaño de la muestra.
- Aleatorio **sistemático**: se enumeran los individuos y se elige al azar sólo uno de ellos, que se denomina **origen**. A partir de éste se toman los siguientes mediante saltos numéricos idénticos. Estos saltos están determinados por un número denominado **coeficiente de elevación**.

Vamos a detallar un poco el procedimiento: partamos de que el tamaño de la población es N y el de la muestra es n .

El coeficiente de elevación será $c = N/n$

El origen se determinará eligiendo al azar un individuo x de la población de entre los c primeros.

De esta forma quedan:

$$x, x+c, x+2c, \dots, x+(n-1)c$$

Para obtener muestras aleatorias se hace uso de los llamados **números aleatorios**.

En una población de tamaño N que está enumerada del 1 al N , un individuo elegido al azar es:

$$E(N \cdot \text{aleatorio} + 1)$$

donde

$$0 \leq \text{aleatorio} < 1$$

es un número generado aleatoriamente.

- Aleatorio **estratificado**: la población se considera dividida en capas denominadas **estratos**.

Al estudiar la muestra, elegimos aleatoriamente de cada uno de los estratos un número de individuos **proporcionalmente** al peso del estrato en la población.

- ☞ **Ejemplo**: Tenemos una asociación formada por 500 socios y hay que elegir al azar, para las elecciones del Equipo de Gobierno, a cinco de ellos para que compongan la mesa electoral.

Veamos un procedimiento simple para obtenerlos:

Primero enumeramos los socios del 1 al 500.

Y luego generamos con una calculadora cinco números aleatorios distintos:

$$0,75, 0,263, 0,052, 0,21, 0,835$$

Terminamos determinando los números de los socios elegidos:

$$N_1 = E(500 \times 0,75 + 1) = E(376) = 376$$

$$N_2 = E(500 \times 0,263 + 1) = E(132,5) = 132$$

$$N_3 = E(500 \times 0,052 + 1) = E(27) = 27$$

$$N_4 = E(500 \times 0,21 + 1) = E(106) = 106$$

$$N_5 = E(500 \times 0,835 + 1) = E(418,5) = 418$$

Resumiendo, la mesa electoral la formarán los socios cuyos números son los siguientes:

$$27, 106, 132, 376, 418$$

- ☞ **Ejemplo**: En una población hay censados 15 400 habitantes. Vamos a elegir una muestra de 200 individuos para cierto estudio.

Vamos a explicar cómo realizar un muestreo aleatorio sistemático.

En primer lugar suponemos que hemos enumerado todos los individuos de la población, del 1 al 15 400.

En segundo lugar, calculamos el coeficiente de elevación:

$$c = \frac{N}{n} = \frac{15400}{200} = 77$$

En tercer lugar, hemos de hallar el origen: será un individuo elegido al azar de entre los 77 primeros. Para ello hacemos uso de los números aleatorios:

$$x_0 = E(77 \times \text{aleatorio} + 1) = E(77 \times 0,242 + 1) = 19$$

Concluimos ya, eligiendo los 200 individuos:

$$19; 19 + 77; 19 + 2 \cdot 77; \dots; 19 + 199 \cdot 77 \rightarrow 19; 96; 173; \dots; 15342$$

Observa que hemos generado un número aleatorio entre 0 y 1 para elegir el origen, tal y como se explicó anteriormente.

Una vez que tenemos el origen y el coeficiente de elevación, formamos la sucesión numérica que indica los números de los elegidos para la muestra.

Fíjate que el último se obtiene añadiendo al origen el coeficiente de elevación multiplicado por $n - 1$.

☞ **Ejemplo:** En un I.E.S. que cuenta con 1.000 alumnos se realiza un estudio sobre el consumo de drogas (tabaco, alcohol, estupefacientes, ...) por parte de los estudiantes. Para ello se eligen a 50 de ellos.

Deseamos diferenciar los estudiantes de E.S.O. de los de Bachillerato y separar la respuesta por sexos. En Secretaría nos dan esta información:

	<i>ESO</i>	<i>Bachillerato</i>
<i>Chicas</i>	250	225
<i>Chicos</i>	350	175

Seguiremos un muestreo aleatorio estratificado. La población se ha dividido en estratos: sexo y nivel de estudios. Debemos hacer lo mismo con la muestra.

Primero enumeramos la población del 1 al 1000.

Segundo calculemos la composición de la muestra. Debemos garantizar la representatividad respetando la proporción de cada estrato en la composición de la población:

$$\begin{aligned}
 \text{Chicas ESO} &= \frac{250}{1000} \cdot 50 = 12,5 & \text{Chicas Bach} &= \frac{225}{1000} \cdot 50 = 11,25 \\
 \text{Chicos ESO} &= \frac{350}{1000} \cdot 50 = 17,5 & \text{Chicos Bach} &= \frac{175}{1000} \cdot 50 = 8,75
 \end{aligned}$$

Hay un empate a la hora de elegir un chica o un chico de la ESO: lo elegimos a cara o cruz y punto.

La composición de la muestra queda así:

	<i>ESO</i>	<i>Bachillerato</i>
<i>Chicas</i>	13	11
<i>Chicos</i>	17	9

Ahora sólo queda elegir de cada estrato de la población, por muestreo aleatorio simple, el número de individuos que indica la tabla anterior.

Ahora detalla tú el proceso de elección de las chicas de bachillerato de forma sistemática y de los chicos de bachillerato de forma simple.

☞ **Ejemplo:** En la población anterior, donde hay censados 15.400 habitantes, se desea ahora estudiar separadamente los sexos. Sabemos que es:

<i>Hombres</i>	<i>Mujeres</i>
7 124	8 276

¿Qué tipo de muestreo debe realizarse? ¿Cómo debemos proceder ahora para elegir la muestra de 200 individuos para cierto estudio.