

TEMA 11 – LAS MUESTRAS ESTADÍSTICAS

11.1 – EL PAPEL DE LAS MUESTRAS

POBLACIÓN o Universo es el conjunto de todos los individuos objeto del estudio.

MUESTRA es un subconjunto extraído de la población. Su estudio sirve para inferir características de toda la población.

¿POR QUÉ SE RECURRE A LAS MUESTRAS?

- La población es excesivamente numerosa
- La población es muy difícil o imposible de controlar
- El proceso de medición es destructivo
- Se desea conocer rápidamente ciertos datos de la población y se tardaría demasiado en consultar a todos.

11.2 - ¿CÓMO DEBEN SER LAS MUESTRAS?

Hay dos aspectos de las muestras a los que debemos prestar mucha atención:

- **Su tamaño:** es claro que si la muestra es demasiado pequeña no podremos obtener de ella ninguna conclusión que merezca la pena.
- **Cómo se realiza la selección de los individuos que la forman:** Al sustituir el estudio de la población por el de la muestra, se cometen errores. Pero con ellos contamos de antemano y pueden controlarse. Sin embargo, si la muestra está mal elegida (no es representativa) se producen errores adicionales imprevisibles e incontrolables (**sesgos**). La elección de la muestra se llama **muestreo**.

TIPOS DE MUESTREOS

- **Muestreo aleatorio:** cuando todos los individuos de la muestra se eligen al azar, de modo que todos los individuos de la población tienen, a priori, la misma probabilidad de ser elegidos.
 - **Muestreo aleatorio simple:** Es el tipo de muestreo más sencillo y en él se basan todos los demás. Para obtener una muestra, se numeran los elementos de la población y se seleccionan al azar los n elementos que deben contener la muestra.
 - **Muestreo aleatorio sistemático:** Se numeran los individuos y, a partir de uno de ellos elegido al azar, se toman los siguientes mediante “saltos” numéricos iguales. El salto se llama coeficiente de elevación, h , y se obtiene mediante el cociente entero entre el número de individuos de la población, N y el número de individuos de la muestra: $n \rightarrow h = N/n$. El primer elemento, llamado origen, se elige al azar entre los números $1, 2, 3, \dots, h$. Esta forma de muestreo sólo es válida si el criterio por el que se han numerado los individuos de la población no tiene nada que ver con la característica que se quiere estudiar con la muestra.
 - **Muestreo aleatorio estratificado :** Si la población puede dividirse en estratos (por ejemplo: por edades, menores de 18 años, de 18 a 50, más de 50), a veces conviene elegir la muestra fijando de antemano el número de individuos de cada estrato. Cuando estos números son proporcionales a los tamaños de los estratos se dice que el muestreo es estratificado con reparto proporcional. $n/N = n_1/N_1 = \dots = n_k/N_k$

TÉCNICAS PARA OBTENER UNA MUESTRA ALEATORIA DE UNA POBLACIÓN FINITA

Para obtener una muestra aleatoria se sortean los individuos de la población para decidir al azar cuáles de ellos forman parte de la muestra. El sorteo puede realizarse de diversas formas:

- **Elección mediante extracción:** En una caja se introducen tantas bolas o papeletas numeradas como individuos hay en la población (N). Estos han sido previamente numerados ($1,2,3,\dots,N$). Se escogen al azar tantas papeletas como individuos ha de tener la muestra (n). Esta operación puede realizarse de dos formas distintas:
 - Sin reemplazamiento: Se eligen simultáneamente, o bien una a una, las n papeletas.
 - Con reemplazamiento: Se eligen una a una n papeleta pero, después de cada extracción, la papeleta elegida (y anotada) se devuelve a la caja.

Con ambos métodos se consigue una muestra aleatoria, pues todos los elementos de la población tienen, a priori, la misma probabilidad de ser elegidos. Sin embargo, si la elección se realiza con reemplazamiento, podríamos obtener algún individuo repetido que habría que desechar y realizar otra extracción. Por eso, cuando se procede por extracción, se debe realizar sin reemplazamiento, pues, además, el proceso es más cómodo.

- **Obtención de números aleatorios:** La calculadora tiene una tecla “RAN” que se llama **generadora de números aleatorios** con la cual se obtiene al azar un número decimal comprendido entre 0,000 y 0,999. Si multiplicamos el número obtenido por N (número de elementos de la población), obtenemos un número decimal cuya parte entera está comprendida entre 0 y $N-1$. Por tanto, si tomamos la parte entera del número obtenido mediante la secuencia $N \times \text{“ran”} + 1 =$, obtenemos un número al azar entre 1 y N . Si repetimos n veces el proceso obtendremos una muestra de n elementos similar a la que se obtendría mediante una extracción con reemplazamiento. Tendríamos que completar la operación suprimiendo los elementos repetidos (que acaso haya) y obteniendo nuevos elementos que los reemplacen. Para poblaciones numerosa este es, evidentemente el método más cómodo, pues no hay que andar preparando papeletas en grandes cantidades. Si la población tuviera más de 1000 elementos, habría que obtener los números aleatorios con ordenador, pues necesitaremos que tengan más de tres cifras decimales para poder separar todos los elementos de la población.

En resumen:

- Si se obtiene la muestra por **insaculación** (extracción de papeletas) debe 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 y ser sustituidos por otros.