$\boxed{\text{Introducción a las técnicas de muestreo}}$

1. **Introducción.**
2. **Análisis de poblaciones por muestreo.**
3. **Sesgos del muestreo.**
4. **Muestreo aleatorio simple.**
5. **Muestreo sistemático.**
6. **Muestreo estratificado.**
7. **Muestreo por conglomerados.**

Introducción a las técnicas de muestreo

Estudio de una población

Una <u>población</u> es un conjunto de elementos o universo de objetos que se desea estudiar.

> Debe definirse sin ambigüedad de manera que siempre pueda clasificarse un elemento como perteneciente o no a la población.

<u>Ejemplo 1</u>. Como parte de un proyecto de mejora ambiental en el campus en el que se propone la eliminación de la circulación de vehículos, se quiere incluir un estudio sobre la opinión de las personas vinculadas a la universidad. Por ejemplo, se desea conocer ¿cuánta gente está a favor de prohibir la entrada de coches en el campus?

<u>¿Cómo definimos la población?</u>

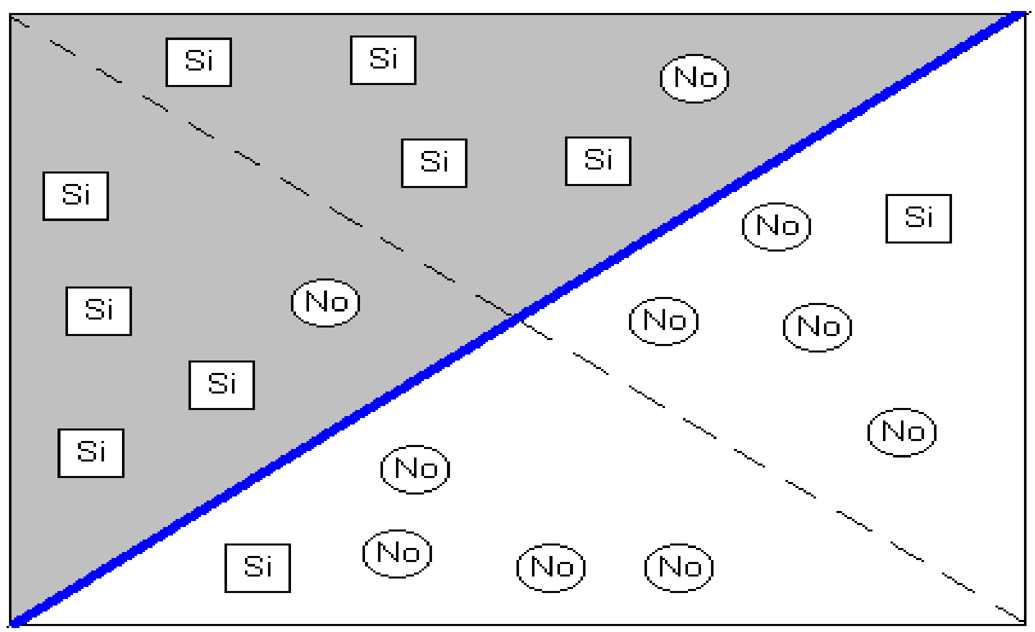
1. Todos los estudiantes matriculados durante este curso en los tres ciclos.

2. Todo el personal de administración y servicios (incluidos los dependientes de alguna contrata) con un contrato de seis meses como mínimo.

3. Todos los profesores a tiempo completo.

Estudio de una población

Ejemplo 1 (reducido). Supongamos que en el siguiente gráfico hemos representado las respuestas de todos los elementos de la población a la pregunta ¿estaría de acuerdo en prohibir la entrada de coches en el campus?. En el triángulo superior (en gris) representamos las respuestas de los alumnos de la facultad A y en el triángulo inferior las respuestas de los alumnos de la facultad B.



Con líneas discontinuas dividimos los alumnos de primer y de segundo/tercer ciclo de las respectivas facultades.

Suponemos que en cada elemento de la población se ha definido una *variable aleatoria*, **y queremos conocer su distribución entre los elementos de la población.**

Puede ocurrir:

1. **La distribución la conocemos por estudios anteriores o viene prefijada por la forma de recoger la información.**

2. **La forma de la distribución la conocemos pero depende de los valores que toman algunos parámetros ó características que sirven para determinarla, (por ejemplo, media y varianza en una distribución normal).**

3. **No tenemos ninguna información sobre la distribución.**

¿Podemos estudiar la variable en toda la población?

SI $\Rightarrow$ **hacemos un CENSO, que es un estudio exhaustivo en todos los elementos de la población.**

NO $\Rightarrow$ **cogemos una MUESTRA, que es un conjunto representativo de elementos de la población.**

Cuando la muestra está bien escogida podemos obtener una información similar a la del censo con mayor rapidez y menor coste.

$\boxed{\text{Análisis de poblaciones grandes por muestreo}}$

Problema. Cuando la población contiene muchos elementos no es posible ni deseable medir la variable de interés en todos ellos, ya que:

- Puede resultar inviable económicamente.

- Aún siendo económicamente posible, estudiar todos los elementos llevaría tanto tiempo que la característica en estudio habría cambiado durante el período de estudio, por ejemplo si se desea estimar la tasa de paro.

- El estudio puede implicar alteraciones o destrucción del elemento en estudio, como ocurre en investigaciones biomédicas o industriales.

Solución: $\boxed{\text{Muestreo}}$.

$N \equiv$ tamaño de la población.

$n \equiv$ número de elementos en la muestra, o tamaño de la muestra.

$\frac{n}{N} \equiv$ fracción de muestreo (proporción de la población representada en la muestra).

$\frac{N}{n} \equiv$ factor de elevación (unidades en la población por cada elemento en la muestra).

<u>Estimador</u>: valor que puede calcularse a partir de los datos de la muestra y que proporciona información sobre el valor del parámetro

<u>Ejemplo 1</u>: Para cada individuo de los N que hay en la población se define una variable $X_i = \begin{cases} 1 & \text{si responde Si} \\ 0 & \text{si responde No} \end{cases}$, y se supone que $X_i \sim Bernoulli(p)$. Es decir, el parámetro p es la proporción de personas a favor de prohibir la entrada de coches en el campus.

En el *ejemplo reducido* **sabemos que $p = \frac{\#\{X_i=1\}}{N} = \frac{10}{20} = 0.5$, pero en general p es desconocido y deseamos estimarlo mediante una <u>muestra</u>.**

Elegimos una *muestra aleatoria con reemplazamiento* **de n individuos, por tanto $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes $Bernoulli(p)$:**

> **Sea $T \equiv$ número de personas a favor entre las n encuestadas, entonces $T = \sum_{i=1}^{n} X_i \sim Bin(n,p)$ y sabemos que $\mathbf{E}[T] = np$.**

Podemos estimar p como la proporción de respuestas positivas en la muestra, $\dfrac{T}{n}$.

¿Cómo elegimos la muestra para que sea representativa de toda la población?

Sesgos del muestreo

El riesgo principal de un procedimiento de muestreo es obtener una muestra sesgada, que significa que no es representativa de la población:

- Sesgo de selección: algunos miembros de la población tienen una probabilidad más alta que otros de estar representados en la muestra.

  En el ejemplo: si sólo preguntamos a las personas que salen de la estación de tren a primera hora tendríamos una *muestra sesgada* pues no incluye a los que tienen coche, a los que vienen en autobús, ó los que tienen clases por la tarde . . .

  En el ejemplo reducido: si solo preguntamos a los estudiantes de la facultad A.

  Solución: Diseñar la muestra con un procedimiento objetivo que garantice la representación de todos los individuos de la población.

- Sesgo por no respuesta: una parte de la población no esta representada porque no proporciona respuesta.

  En el ejemplo: si enviamos un cuestionario, puede que no contesten los que no tienen coche propio por sentirse poco afectados tendríamos una *muestra sesgada*.

  Solución: no siempre se puede evitar, pero al menos se debe controlar (incluir preguntas tipo ¿tienes coche propio?).

Muestreo aleatorio simple - M.A.S.

Diremos que hemos extraído una muestra aleatoria simple cuando su proceso de extracción garantice:

- Cada elemento de la población tiene la misma probabilidad de ser elegido.

- Los elementos se seleccionan de uno en uno y con reposición (reemplazamiento), de manera que la población permanece idéntica en todas las extracciones.

  - Observacion: cuando la fracción de muestreo (n/N) es pequeña, será indiferente realizar el muestreo con o sin reposición.

◇ Se utiliza cuando la población es homogénea respecto a la característica a estudiar.

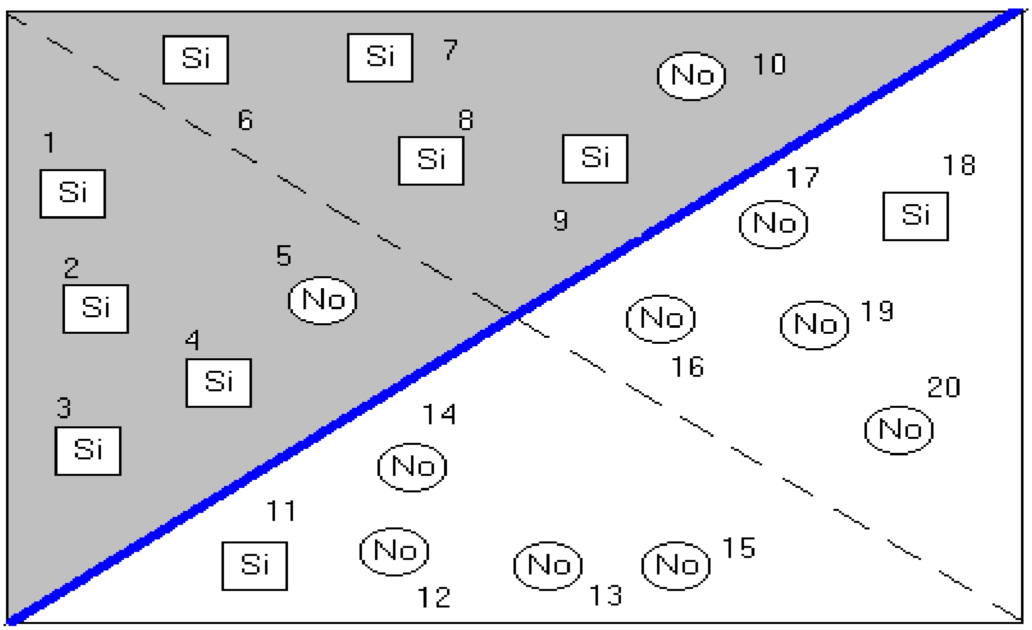
◇ Para diseñar la muestra se utilizan los números aleatorios.

Ejemplo 1 (reducido).

1. **Enumeramos los elementos de la población de 1 a N.**

2. **Seleccionamos n números aleatorios entre 1 y N, en el ejemplo: 1, 5, 12, 8.**

3. **Encuestamos a los individuos seleccionados, en el ejemplo: Si, No, No, Si.**



Pregunta. ¿Cuál es la probabilidad de seleccionar sólo elementos de la facultad A o sólo elementos de la facultad B en una muestra de tamaño $n = 4$?

$$\frac{10^4}{20^4} + \frac{10^4}{20^4} = 0.0625 + 0.0625 = 0.125$$

Muestreo sistemático - M.S.

Se utiliza cuando la población está ordenada en una lista.

1. **Elegimos al azar un individuo n_1 entre los primeros f_e, donde $f_e = \lceil N/n \rceil$.**

2. **Elegimos a continuación los elementos $n_1+f_e, n_1+2f_e, \ldots$, $n_1+(n-1)f_e$ hasta completar la muestra.**

En el ejemplo reducido: $f_e = 20/4 = 5$, tomamos un número aleatorio entre 1 y 5, supongamos que 2, entonces los elementos en la muestra son: 2, 7, 12, 17.

- **El M.S. es equivalente al M.A.S. si la lista está ordenada al azar.**

- **Si existe algún ciclo en la lista puede aparecer un sesgo de selección.**

Pregunta. ¿En el ejemplo reducido el muestreo sistemático es equivalente al muestreo aleatorio simple?

Muestreo aleatorio estratificado - M.A.E.

Cuando la población se divide en clases o estratos formados por elementos heterogéneos (entre estratos) respecto a la variable, la muestra se toma siguiendo:

1. Se asigna un número de elementos en la muestra a cada estrato:
   (a) Proporcionalmente al tamaño relativo del estrato en la población.
   (b) Proporcionalmente a la variabilidad que presenta la característica que estudiamos dentro de cada estrato.
2. Los elementos que forman la muestra se escogen por muestreo aleatorio simple dentro del estrato.

En el ejemplo: sospechamos que la opinión puede ser diferente entre distintas facultades, seleccionamos m.a.s. dentro de cada estrato con tamaño proporcional a la dimensión relativa del estrato $\frac{n_j}{N_j} = \frac{n}{N}$, donde n_j es el tamaño de muestra del estrato j-ésimo que tiene tamaño N_j.

Tenemos que $N_1 = N_2 = 10$, entonces $n_1 = n_2 = 2$.

♢ Necesitamos disponer de información sobre qué estratos existen (no siempre es sencillo).

♢ En general, es más preciso que el M.A.S.

Muestreo aleatorio estratificado - M.A.E.

Pregunta. ¿Cuál es la probabilidad de seleccionar sólo elementos de la facultad o sólo elementos de la facultad B en una muestra de tamaño $n = 4$ utilizando M.A.E.?

Pregunta. ¿Cuál es la probabilidad de que T = 1, si $n = 2$ en un M.A.S.? y en un M.A.E.?

M.A.S. **En el ejemplo, tenemos que bajo un muestreo aleatorio simple $T \sim Bin(n = 2, p = 0.5)$, por tanto:**

$$\Pr\{T = 1\} = \binom{2}{1} 0.5^1 0.5^{2-1} = 0.5.$$

M.A.E. **En el ejemplo, tenemos que bajo un muestreo aleatorio estratificado, $T = T_1 + T_2$ con $T_1 \sim Bin(n = 1, p = 0.8)$ y $T_2 \sim Bin(n = 1, p = 0.2)$, por tanto:**

$$\begin{aligned}
\Pr\{T = 1\} &= \Pr\{T_1 = 1 \cap T_2 = 0\} + \Pr\{T_1 = 0 \cap T_2 = 1\} \\
&= \tbinom{1}{1} 0.8^1 0.2^{1-1} \times \tbinom{1}{1} 0.2^0 0.8^{1-0} \\
&\qquad + \tbinom{1}{1} 0.8^0 0.2^{1-0} \times \tbinom{1}{1} 0.2^1 0.8^{1-1} \\
&= 0.8^2 + 0.2^2 = 0.64 + 0.04 = 0.68.
\end{aligned}$$

Muestreo por conglomerados - M.C.

Cuando la población se clasifica en unidades amplias de agrupación física, en el espacio o en el tiempo, que llamamos conglomerados, e idealmente los elementos dentro de cada conglomerado son heterogéneos como en la población total respecto a la variable que se estudia (los conglomerados son unidades representativas).

El muestreo se hace por etapas:

1. Seleccionamos aleatoriamente algunos conglomerados.

2. Si dentro de cada conglomerado existen nuevos conglomerados, seleccionamos aleatoriamente algunos de ellos.

3. Dentro de cada conglomerado analizamos todos los elementos o una muestra aleatoria simple.

En el ejemplo: las opiniones dentro de cada ciclo educativo son representativas de lo que ocurre en toda la población. El muestreo lo hacemos seleccionando aleatoriamente uno o varios ciclos (conglomerados).

En el ejemplo reducido: seleccionamos al azar un ciclo educativo, por ejemplo, el primer ciclo, y a continuación 4 estudiantes al azar (1, 4, 14, 15): Si, Si, No, No.

Muestreo polietápico

Para estudiar poblaciones complejas se combinan las ideas de estratificación y conglomerados en procedimientos de muestreo polietápicos.

1. **Se agrupan los conglomerados en estratos y seleccionamos una muestra estratificada de los conglomerados (la probabilidad de elegir un conglomerado dentro de cada estrato es proporcional a su tamaño).**

2. **Repetimos el proceso si dentro de cada conglomerado seleccionado hay otra estructura de conglomerados o estratos.**

3. **Se toman uno o varios elementos de la unidad final por M.A.S.**

En el ejemplo: tenemos que los estudiantes de primer ciclo y los estudiantes de segundo/tercer ciclo forman dos conglomerados, y en ellos podemos formar dos estratos definidos por la pertenencia a las facultades A o B.

	Ciclo 1	Ciclo 2 y 3
Facultad A	1, 2, 3, 4, 5	6, 7, 8, 9, 10
Facultad B	11, 12, 13, 14, 15	16, 17, 18, 19, 20

En el ejemplo reducido: tomamos un conglomerado al azar en cada estrato, por ejemplo el primer ciclo en A y el segundo/tercer ciclo en B, y en los conglomerados seleccionados tomamos una muestra aleatoria simple de 2 elementos.