

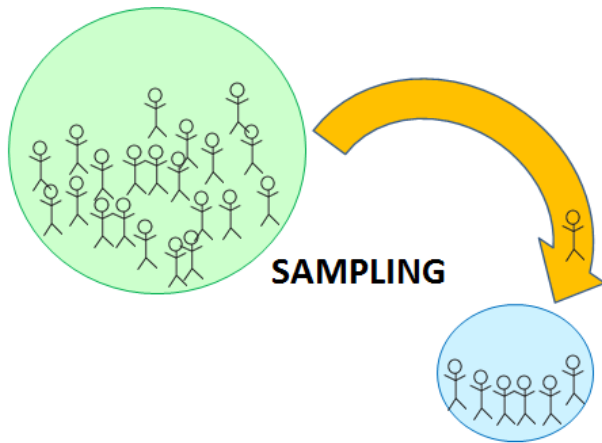
# Técnicas de Muestreo

---

IVAN OLMOS PINEDA

# Introducción

---



En términos prácticos, en la mayoría de los casos es inviable tratar de recolectar todos los elementos de una población para realizar algún estudio. Algunas de las causas son:

- Costo económico
- Tiempo
- Accesibilidad de los datos
- Temporalidad de los datos
- Variabilidad de los datos
- ...

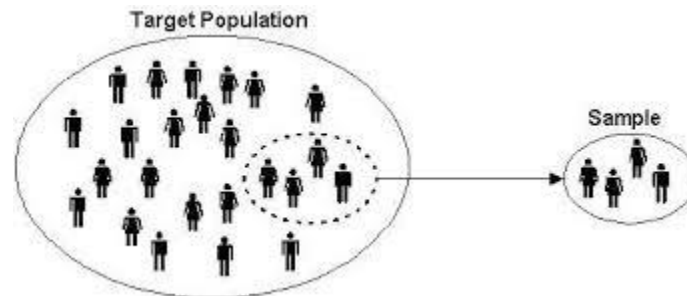
Por lo anterior, en la mayoría de las investigaciones se parte de recolectar un conjunto de datos (muestra) que permita hacer conjeturas (válidas) sobre lo que sucede en la población

# Población y muestra

---

La población es el conjunto total de elementos de los cuales se desea sacar alguna conclusión (tendencia, patrón de comportamiento, etc.)

La muestra es un subconjunto (en general limitado) de instancias o elementos de una población que serán recolectados para una investigación



# Tipos de muestreo

---

En general, a partir de una muestra se buscan obtener resultados que sean aplicables a la población (de donde proviene la muestra)

- No cualquier porción de una población tendrá las características necesarias para considerarla una muestra
- Las muestras deben de cumplir el principio de representatividad (cualquier patrón de comportamiento en los elementos de una población deben de estar presentes en la muestra)

El muestreo es una estrategia que permite seleccionar elementos de una población

Existen dos categorías de muestro:

- Muestreo probabilístico, basado en una selección probabilística de elementos \*
- Muestreo no probabilístico, guiado por creencias, conocimientos previos, sobre la población y aplicada en la selección de instancias en una muestra

# Muestreo probabilístico

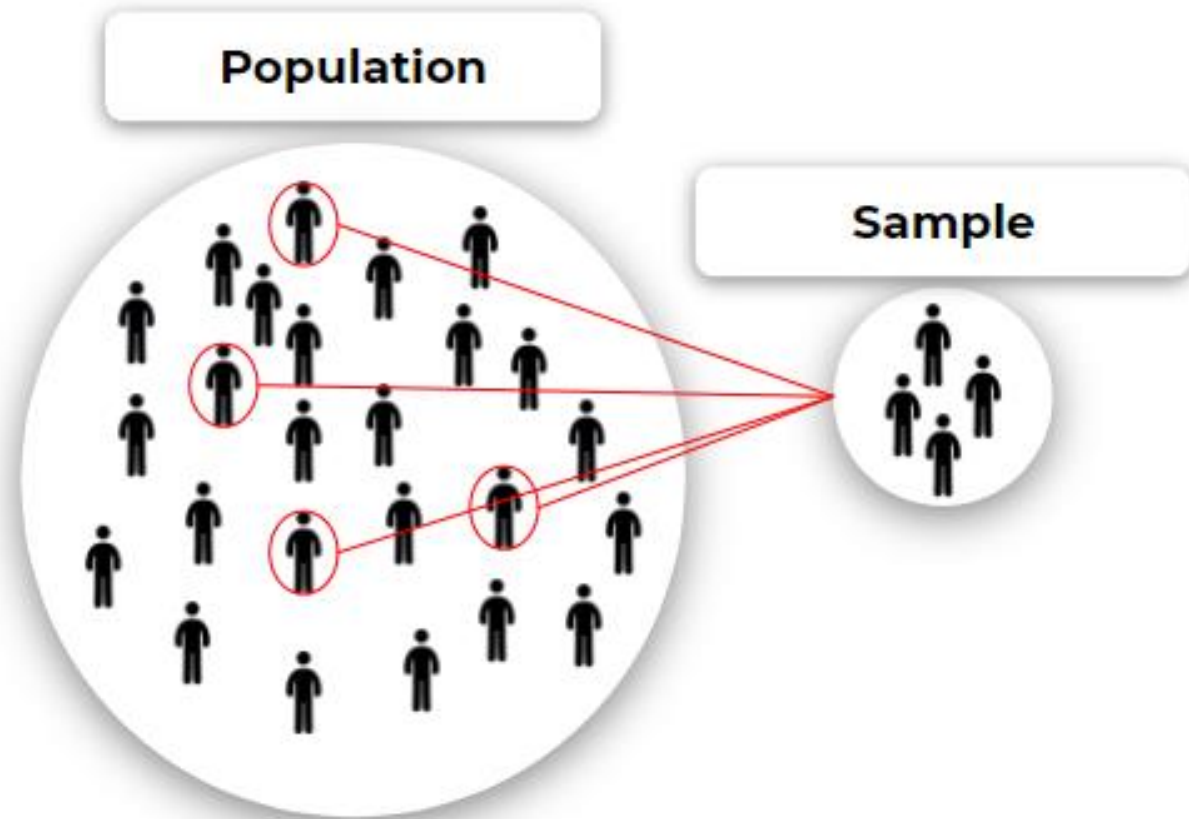
---

# Preguntas clave en un muestreo

Dado que el muestreo es el proceso de seleccionar elementos de una población, al momento de construir una muestra se deben de resolver dos preguntas claves:

¿Cómo se seleccionan elementos en la población?

¿Cuántos elementos se van a seleccionar?



# Seleccionando elementos de la población

---

Objetivo principal: seleccionar un conjunto de elementos de la población que mantenga las características de comportamiento de los datos en la población

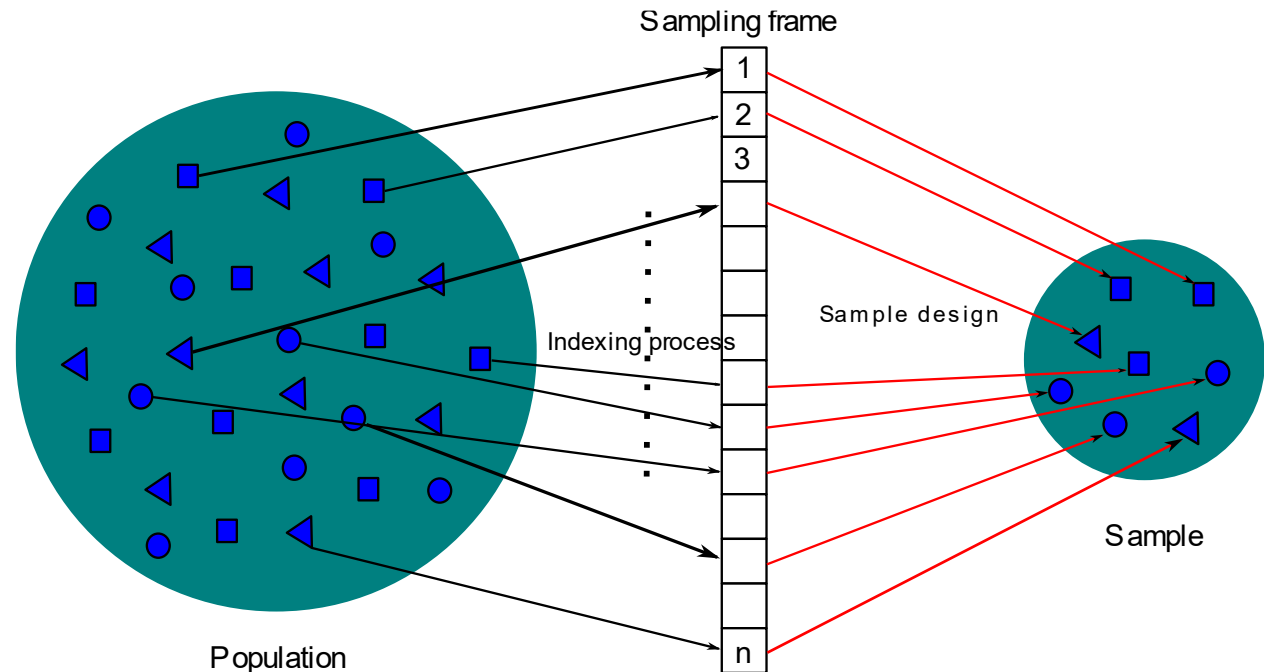
Para la construcción de la muestra, los investigadores se suelen apoyar en un **muestreo probabilístico**

- Se asume que cualquier elemento en la población tenga una probabilidad de ser seleccionado (idealmente igual)
- Existen diversos diseños de muestreo probabilístico (sample design), entre las que se encuentran:
  - Muestreo aleatorio simple\*
  - Muestreo sistemático\*
  - Muestreo estratificado\*
  - Muestreo por conglomerados (clusters)
  - Muestreo multi-etapa

\* Para aplicarlos, se debe asumir que se tiene una lista de los elementos que pueden ser seleccionados de la población (marco de muestreo ó sampling frame)

# Seleccionando elementos de la población

Los elementos en una muestra se obtienen a partir del diseño de un muestreo, es decir, como se seleccionan elementos de la población a partir de un marco de muestreo (listado de objetos de la población a los cuales se tiene acceso)





# Muestreo aleatorio simple

---

Técnica ampliamente utilizada en diferentes tipos de investigaciones

La técnica consiste en lo siguiente:

1. Asignar un número (único) a cada elemento en el sampling frame
2. A partir de un proceso estocástico (por ejemplo, un generador de números aleatorios) seleccionar elementos del sampling frame a partir de un número asignado

Ventajas:

- En muchos casos esta técnica construye muestras con una baja probabilidad de sesgos
- Fácil de implementar

■ Desventajas:

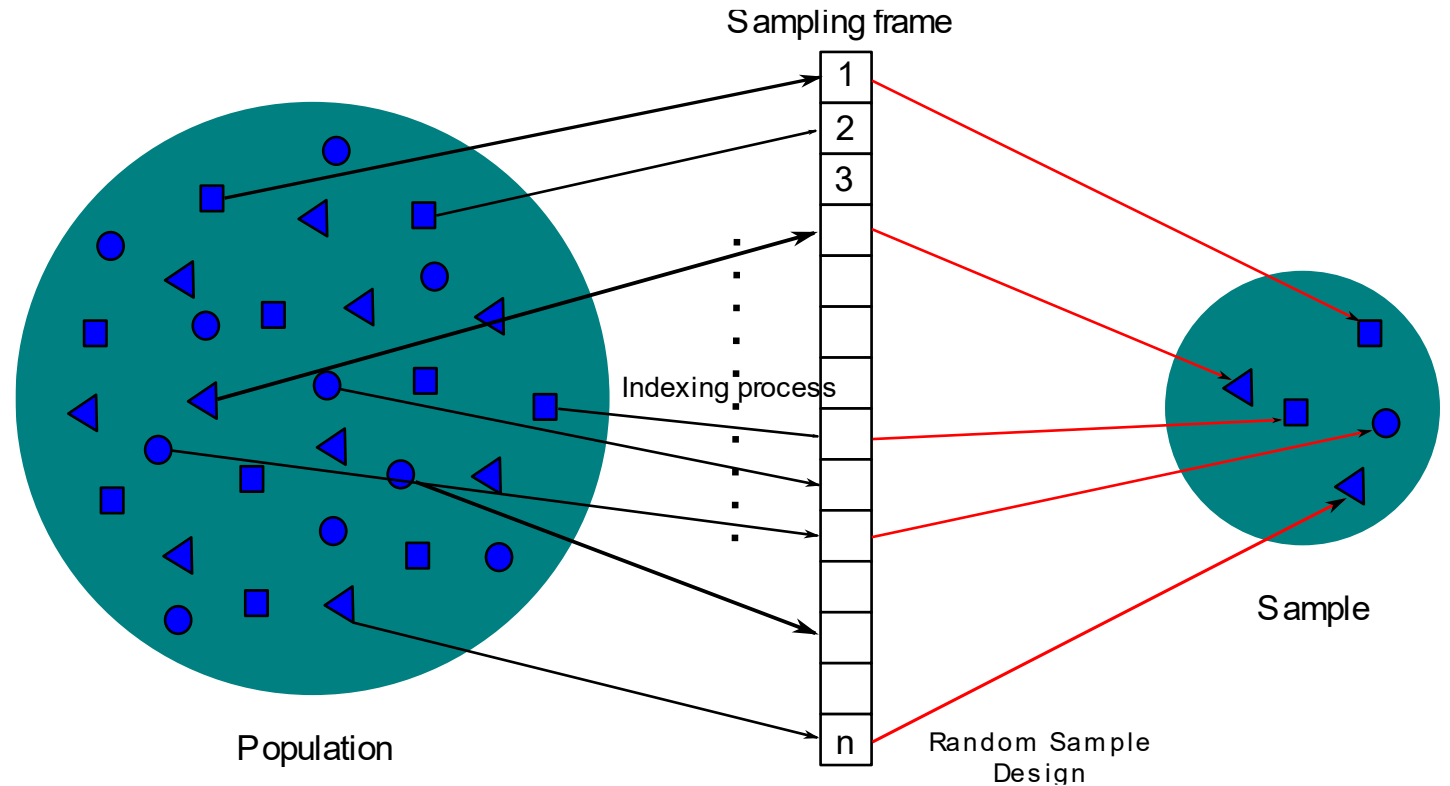
- Se debe de garantizar que no existen sesgos en el sampling frame (del tipo temporal, espacial, etc.)
- Los generadores de números aleatorios no son totalmente estocásticos

# Muestreo aleatorio simple

En este muestreo, se seleccionan a partir de un proceso estocástico elementos considerando dentro del marco de muestreo.

Se generan “n” números aleatorios, a partir de los cuales se determinan que objetos de la población serán considerados en la muestra

Método muy usado cuando no existen sesgos en la población.



# Muestreo sistemático

---

En este tipo de muestreo se procede como sigue:

1. Se asume que los elementos de la población ( $N$ ) son indexados en una lista (marco de muestreo), considerando que el indexado es aleatorio. Se define un tamaño de muestra a crear ( $n$ )
2. Se define un intervalo de distancia de muestreo  $\Delta = N / n$  (se divide los elementos en el sample frame entre el tamaño de muestra deseado)
3. Se elige un punto de inicio arbitrario " $s$ " entre 1 y  $\Delta$
4. Los elementos en la muestra son aquellos que se encuentran en posiciones múltiplos de  $\Delta$  a partir del punto de inicio predefinido, esto es, en las posiciones " $s + \Delta k$ " ( $0 \leq k < n$ )

## Ventajas

- Fácil de implementar y con aleatoriedad en la selección de los elementos en la muestra

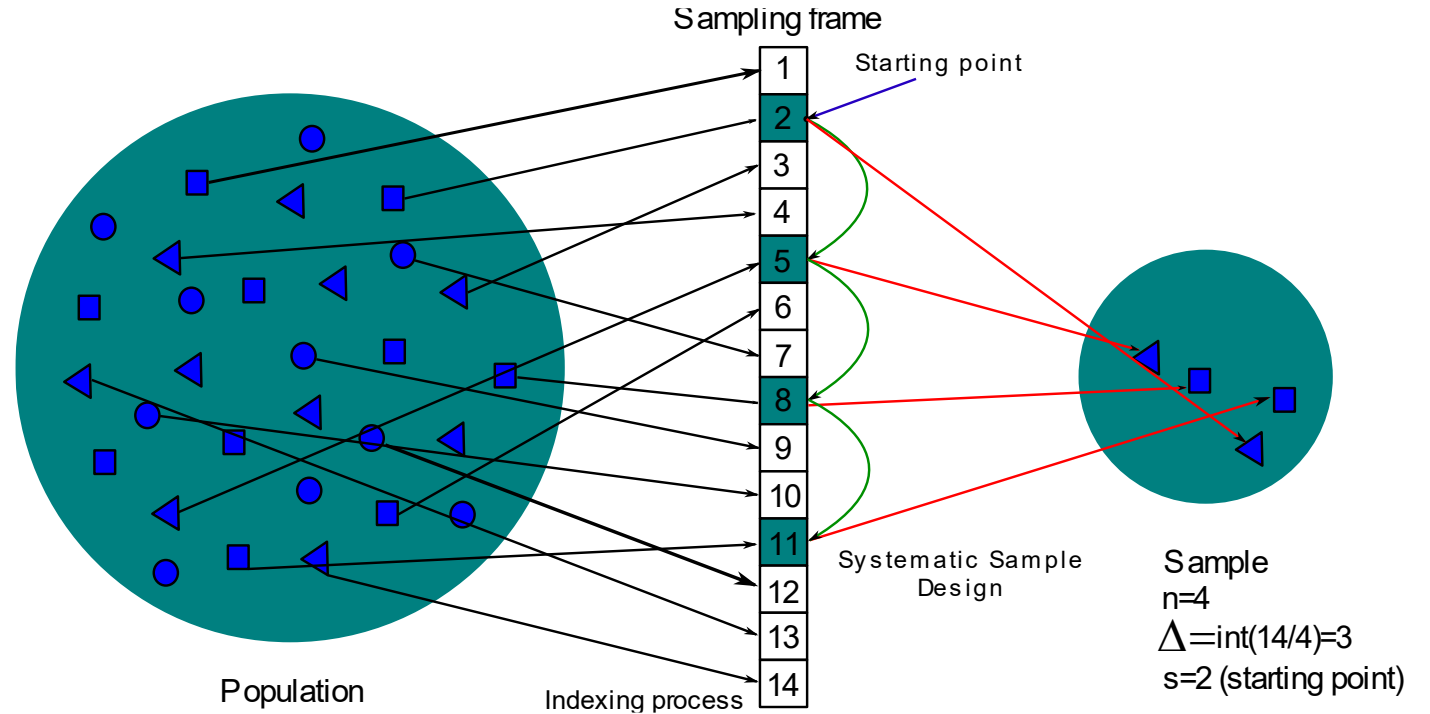
## Desventajas

- Se requiere verificar que no existan patrones de ocurrencia en los elementos del sample frame

# Muestreo sistemático

Se elige un punto pivote en el sampling frame y a partir de una distancia delta, se van eligiendo los elementos que formarán a la muestra

Es importante verificar que no existan patrones de ocurrencia en el sampling frame



# Muestreo estratificado

---

Se basa en crear grupos para lograr la representatividad:

1. Se asume que los elementos de la población son indexados (marco de muestreo), considerando que el indexado es aleatorio
2. Se divide el marco de muestreo en grupos disjuntos en base a algún criterio de asociación (estratificado)
3. Se aplica un muestreo aleatorio simple en cada grupo, seleccionando la cantidad de elementos requeridos

## Ventajas

- Permite el balanceo de muestras, se evita que subgrupos no queden fuera de la muestra por su baja representatividad

## Desventajas

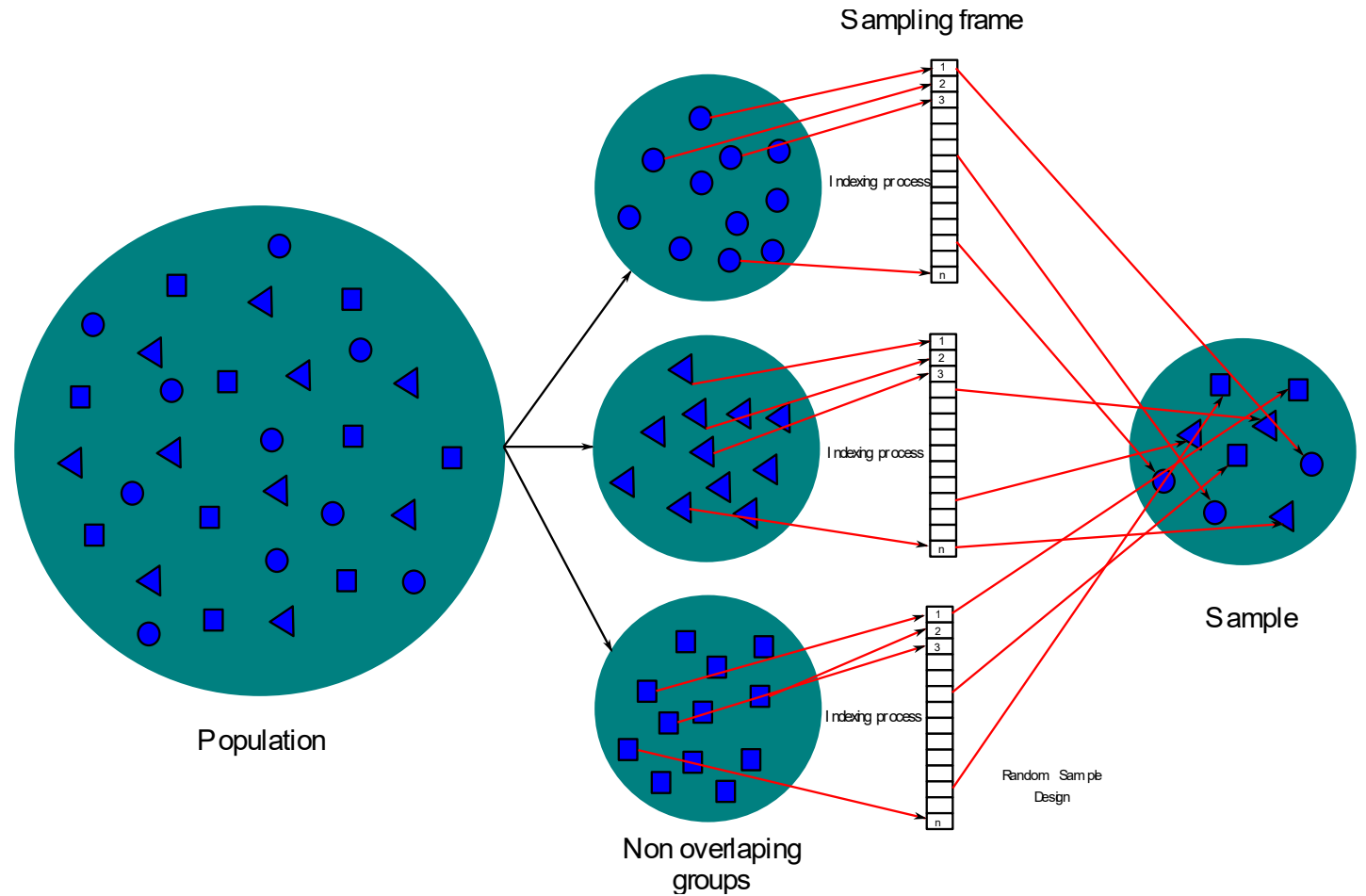
- Se requiere conocimiento previo sobre la población para crear los grupos

# Muestreo estratificado

Se agrupa el marco de muestreo en grupos disjuntos, usando algún criterio para la asociación.

Permite construir una muestra balanceada, gracias a que se definen cuantas muestras se requieren de cada grupo.

Se requiere a priori conocer criterios para la agrupación.



# Muestreo por grupos (clusters)

---

El muestreo por grupos es una de las técnicas ampliamente usada en diversas tareas, como son:

1. Censos de población, estudios socioeconómicos, estudios de salud, educación
2. Investigaciones en biomedicina, química
3. Big-data

En ocasiones, no es posible construir marcos de muestreo (por ejemplo, en encuestas sobre una población de personas), lo que imposibilita aplicar otras técnicas de muestreo

En lugar de ello, en la población se identifican grupos (conglomerados de datos). A partir de estos, se construye un marco de muestreo a nivel de grupos (cluster sampling)

# Muestreo por grupos

---

## Ventajas / Desventajas

- Permite trabajar con poblaciones muy grandes, o bien ubicadas en diversas zonas geográficas, o cuando la población esta naturalmente agrupada por alguna razón
- Si los grupos tienen tamaños muy diferentes, instancias de grupos grandes tendrán menos probabilidad de ser seleccionados. Para corregir este problema, se debe aplicar una probabilidad proporcional al tamaño

Existen diversas variantes al muestreo por grupos, siendo las más populares:

1. Muestreo simple de grupos de una etapa (simple one stage cluster sampling)
2. Muestreo de grupos de dos etapas (two-stage cluster sampling)



# Muestreo simple de grupos de una etapa

---

Se identifican las siguientes etapas en el muestreo:

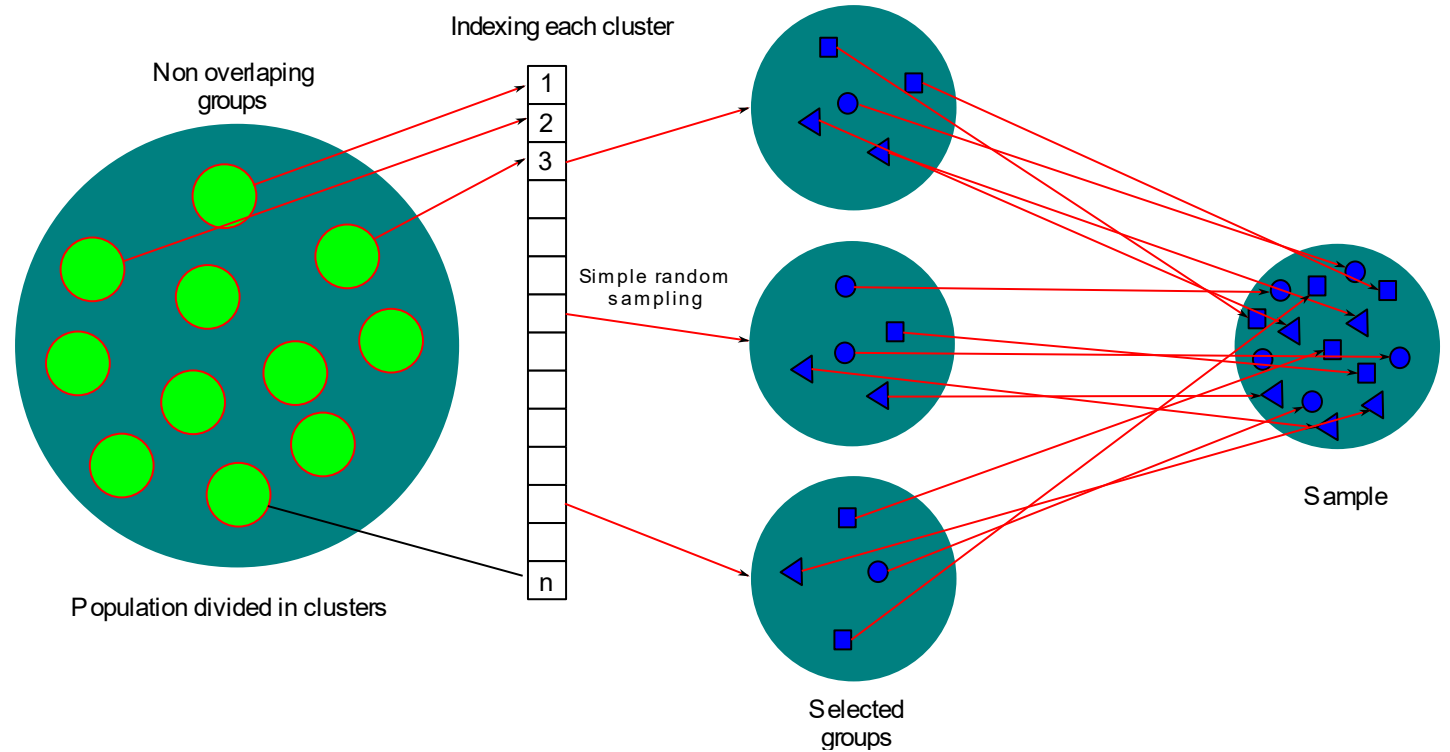
1. Se modela la población para identificar los posibles grupos a generarse
2. Se divide la población en grupos. La calidad de los grupos determinará la calidad de la muestra.
  1. Cada grupo de la población debe ser lo más diverso posible (se debe incluir cualquier potencial característica de la población)
  2. Cada grupo debe de tener una distribución de características similares de la población
  3. Los grupos deben de cubrir la totalidad de la población
  4. No debe de existir intersección entre grupos
3. Se seleccionan los grupos a ser considerados con un muestreo aleatorio simple
4. Se construye la muestra a partir de las instancias pertenecientes a cada grupo seleccionado

# Muestreo simple de grupos de una etapa

Primero se identifican grupos en la población, para luego ser seleccionados los grupos a partir de un muestreo simple.

Se recuperan las instancias pertenecientes a los grupos seleccionados para integrarse a la muestra

Se debe de cuidar que los clusters sean representativos de los posibles comportamientos que presente la población (a partir de conocimiento previo, observación etc.)



# Muestreo simple de grupos de dos etapas

---

También conocido como “Two-stage cluster sampling”

Esta estrategia se aplica en situaciones como:

- Una alta dimensión de los grupos
- Cuando los grupos son muy homogéneos, lo que provocará una alta redundancia

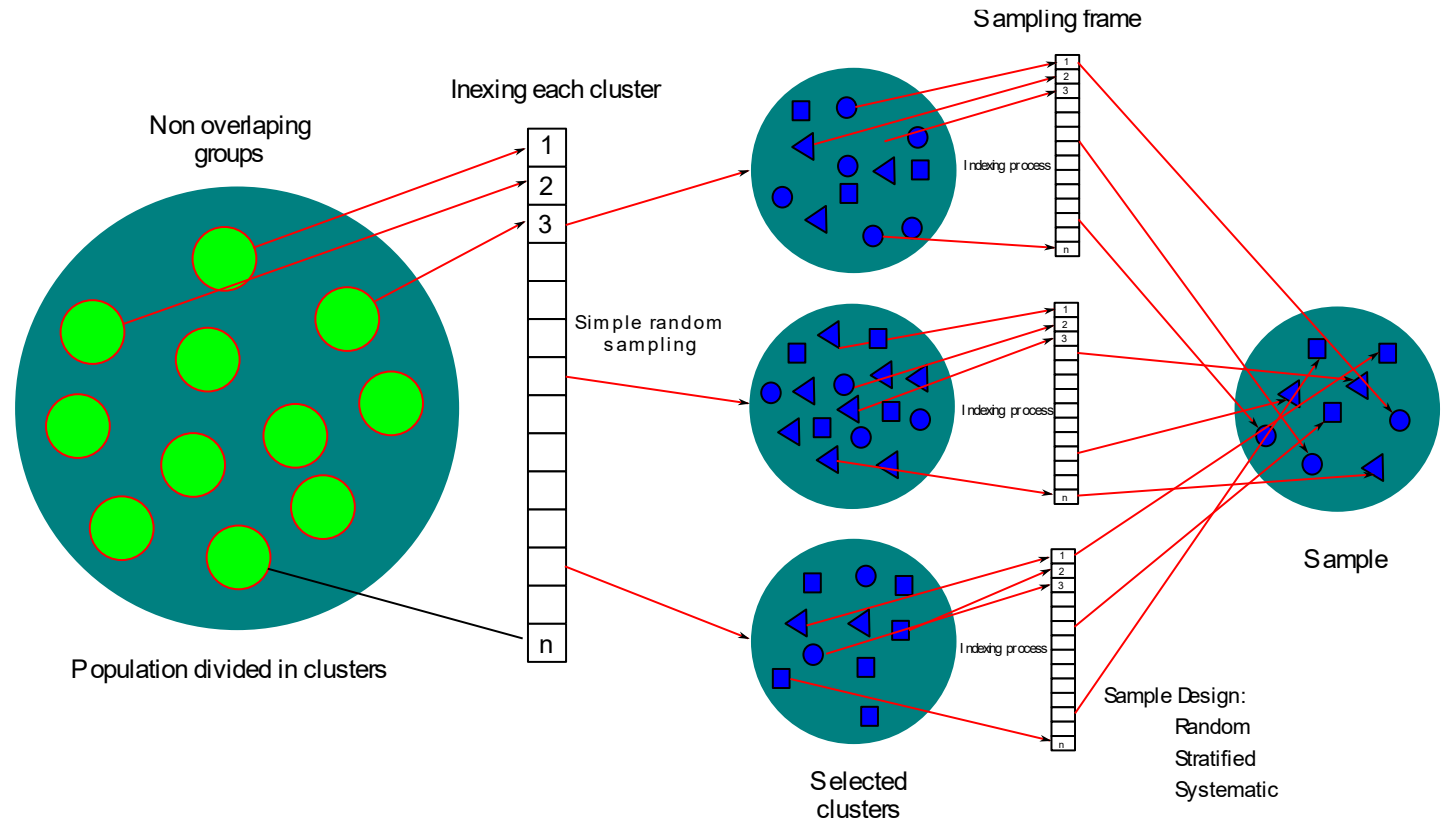
Para corregir estos problemas se aplica un muestreo de grupos de dos etapas:

1. En la primera etapa se aplica un muestro a nivel de grupos
2. En la segunda etapa se crea un marco de muestreo para seleccionar las instancias de los clústeres seleccionados previamente

# Muestreo simple de grupos de dos etapas

En la primera etapa se aplica un muestreo a nivel de clusters, para seleccionar aquellos que se tomarán como representantes

En la segunda etapa se aplica un marco de muestreo con alguna de las estrategias previas, con la finalidad de seleccionar las instancias que integrarán a la muestra.



# Muestreo multi-etapa

---

Es la combinación de uno o más métodos descritos anteriormente

Procedimiento general:

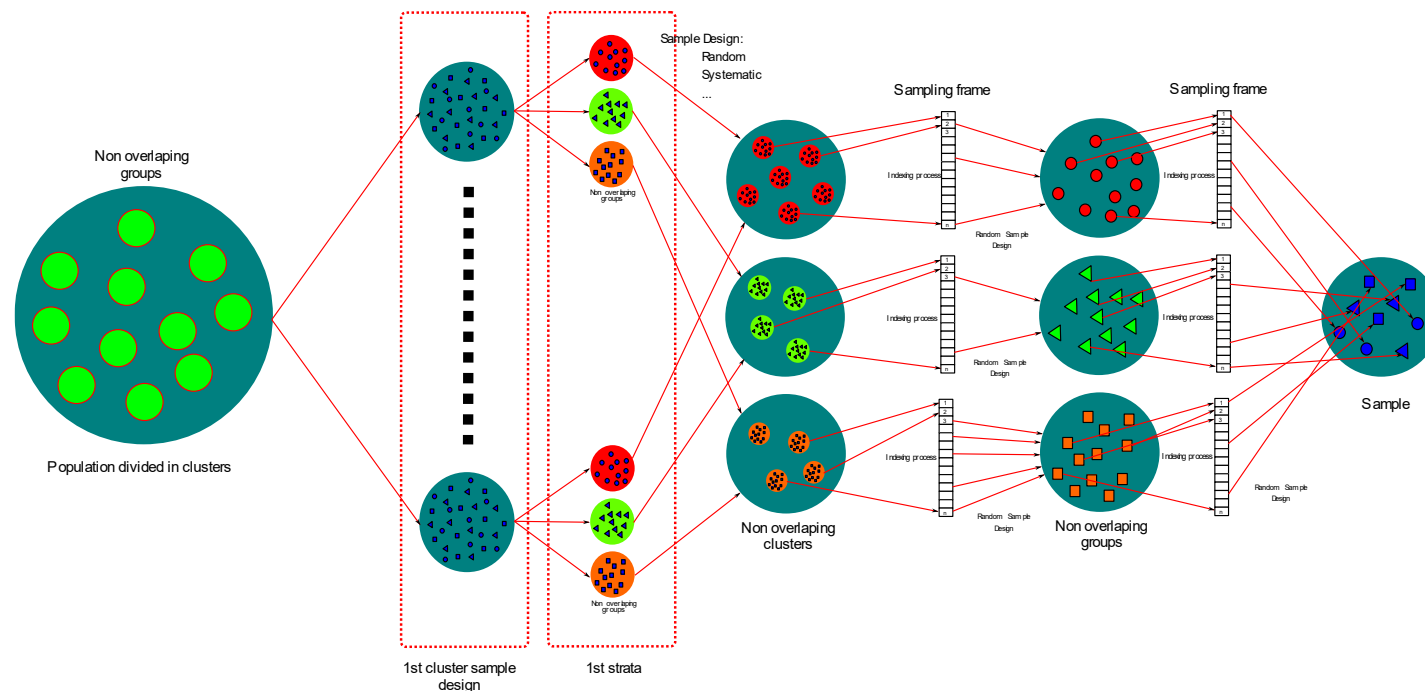
1. La población es dividida en grupos
2. Cada uno de los grupos es dividido en subgrupos (muestreo estratificado), usando criterios predefinidos de agrupación
3. Los subgrupos estratificados son agrupados en nuevos grupos
4. Se seleccionan aleatoriamente uno o más grupos estratificados (obtenidos del paso 3)
5. El proceso puede continuar hasta que no existan criterios de división aplicables y de interés para el estudio

# Muestreo multi-etapa

Se aplica un muestreo por clustering, para luego dividir los grupos bajo algún criterio (estratificación), para luego seleccionar los clusters estratificados (con algún criterio, por ejemplo aleatorio)

Estos clusters se seleccionan algunos para formar finalmente la muestra.

En este muestreo, es posible aplicar varias etapas de clusters estratificados y variar el método de selección de los mismos.



# Consideraciones finales

---

Un caso especial se presenta cuando la dimensión de los clústeres seleccionados no es similar

- Si se aplica un muestreo que obtenga un porcentaje similar de cada clusters, entonces algunos de los clusters pueden tener pocos representantes en la muestra final
- Por otro lado, equilibrar el número de instancias final que aportará cada clúster puede provocar que se de un mayor peso (desproporcionado) a los clusters con pocas instancias

En estos casos se recomienda

1. Seleccionar una estrategia que equilibre proporcionalmente la aportación de instancias que cada cluster aportará a la muestra
2. En caso de no tener clusters homogéneos, es necesario introducir una mezcla de estrategias, como por ejemplo, muestreo estratificado seguido de un muestro aleatorio simple con un equilibrio proporcional a la dimensión

# Muestreo no probabilístico

---



# Generalidad

---

No es un muestreo al azar. Este tipo de muestreo depende en gran medida de la experiencia del investigador (para evitar introducir sesgos no deseados en la muestra)

1. A diferencia de un muestreo probabilístico, en un muestreo no probabilístico no todos los miembros de la población tienen la oportunidad de participar en el estudio
2. Este tipo de muestreo sólo se debe de aplicar cuando no es posible (por alguna razón sustentada) aplicar un muestreo probabilístico
3. Este tipo de muestreo es útil en investigaciones exploratorias (como la encuesta piloto)

Este tipo de muestreo es ampliamente usado en investigaciones cualitativas

# Categorías

---

Entre las técnicas más relevantes en esta categoría se encuentran:

- Muestreo por conveniencia
- Muestreo consecutivo
- Muestreo por cuotas
- Muestreo intencional o por juicio
- Muestreo “bola de nieve”

# Muestreo por conveniencia

---

Las muestras de la población son seleccionadas sólo porque están convenientemente disponibles para el investigador

- El investigador asume que las conclusiones sólo son aplicables al conjunto de datos recolectado
- Es la técnica no probabilística más común, debido a factores como el tiempo (rapidez) y costo (disponibilidad de la muestra)

Por ejemplo, si el investigador desea realizar un análisis exploratorio y tiene a su alcance un grupo de estudiantes, entonces construye su muestra con el grupo disponible

# Muestreo consecutivo

---

Basado en el muestreo por conveniencia

1. El investigador elige una muestra inicial en un periodo de tiempo específico y desarrolla su investigación con dicha muestra
2. En un tiempo futuro (la temporalidad la define el investigador), se selecciona otra muestra por conveniencia y se aplica la misma investigación (si es necesario esto se puede repetir “n” veces)

Ventajas: Los resultados pueden ser contrastados (por periodos de tiempo), por lo que se pueden obtener resultados más robustos que con un muestreo con conveniencia simple

Desventajas: Se requiere una mayor inversión de tiempo para llevar acabo el estudio

# Muestreo por cuotas

---

El muestreo por cuotas consiste en aplicar una estratificación a los datos accesibles (conjunto de datos que se pueden recolectar de la población)

1. Se definen los estratos (clases) de datos a trabajar
2. Se define un número de elementos por estrato, cuidando que el número no sobrepase lo disponible en los datos accesibles
3. Se eligen aleatoriamente los elementos que conformarán a cada estrato en la muestra

Aplicable a estudios dónde se desea obtener conclusiones sobre estratos que se conoce existen a priori en la población (p.e., evaluar el diseño ergonómico de un teclado de computadora de acuerdo a la perspectiva de hombres y mujeres en un grupo escolar)

# Muestreo intencional o por juicio

---

En este tipo de muestreo, los elementos de la muestra son seleccionados usando únicamente el juicio del investigador. Esto es, el investigador elige únicamente muestras que considera de interés.

1. No es recomendable que lo apliquen profesionales en formación, ya que en general no tendrán la madurez suficiente para la selección de la muestra
2. Tiene la desventaja de existir una alta ambigüedad en la investigación, pero si la experiencia del investigador es de muy alto nivel en su área, los resultados pueden llegar a ser de gran interés

Por ejemplo, este tipo de muestreo es usado en estudios piloto

# Muestreo “bola de nieve”

---

Aplicable a situaciones en las cuales no es fácil recolectar muestras:

1. Se identifican inicialmente un conjunto de muestras de interés
2. Localizadas las muestras iniciales, se aplica una búsqueda local sobre las muestras iniciales, para tratar de identificar nuevas instancias de interés. La búsqueda local varia dependiendo del entorno de la investigación.

Por ejemplo, si se busca encontrar a un grupo de personas con ciertas características, primero se localizan muestras iniciales válidas, para luego apoyarse en los sujetos localizados para identificar si conocen otros sujetos similares a sus características