

Estadística y Muestreo

Probabilidad y Estadística
Departamento de Matemática y Estadística
Ingeniería Industrial
FCA-UNNE

2020

Unidad Temática I:
Diseños para la obtención de
información empírica.

Estadística y Muestreo.

Muestreo. Objetivos y ventajas.

Métodos.

Tamaño de muestra.

Error de muestreo.

Aplicaciones.

OBTENCIÓN DE DATOS

Los datos pueden ser obtenidos a través de

Estudios observacionales: son aquellos donde el investigador observa a las unidades de análisis y registra las variables de interés.

Por ejemplo, cuando se analiza la oferta semanal de un producto a nivel regional.

Definiciones

- Población: Es un conjunto de elementos acotados en un tiempo y un espacio, definidos por alguna característica común observable.
- Tamaño poblacional: número de elementos de la población y lo denotaremos con N .
- Muestra: Es un subconjunto representativo de elementos de la población.
- Tamaño muestral: número de elementos de la muestra y lo denotaremos con n .

Mas definiciones

- Unidad muestral: es el lugar donde se realiza la observación.
- Variable: es la característica de interés que se registra en cada unidad muestral.

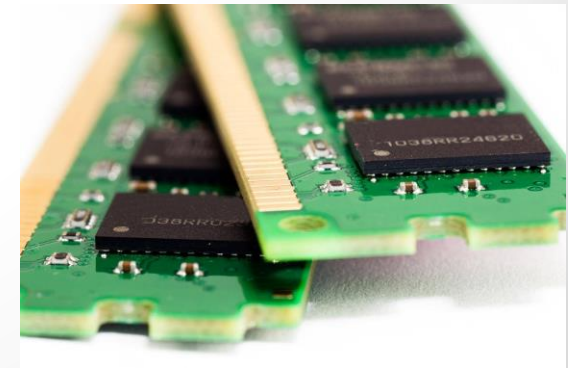
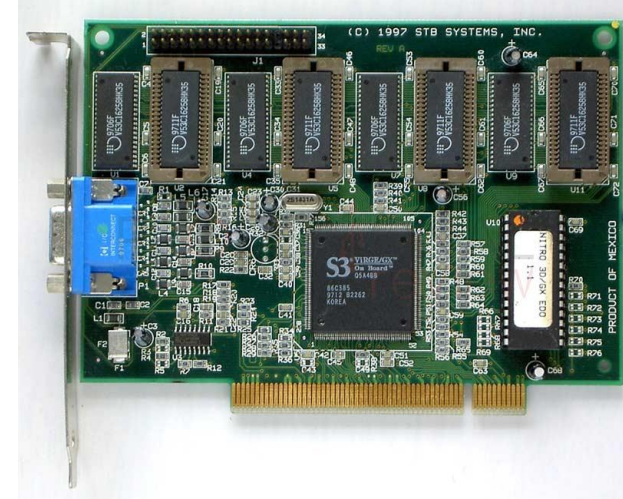
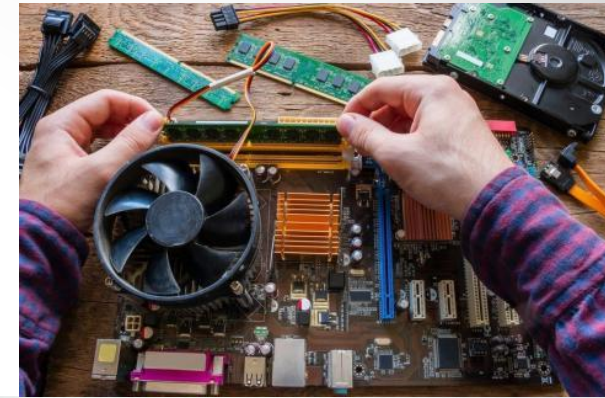


Otro Ejemplo

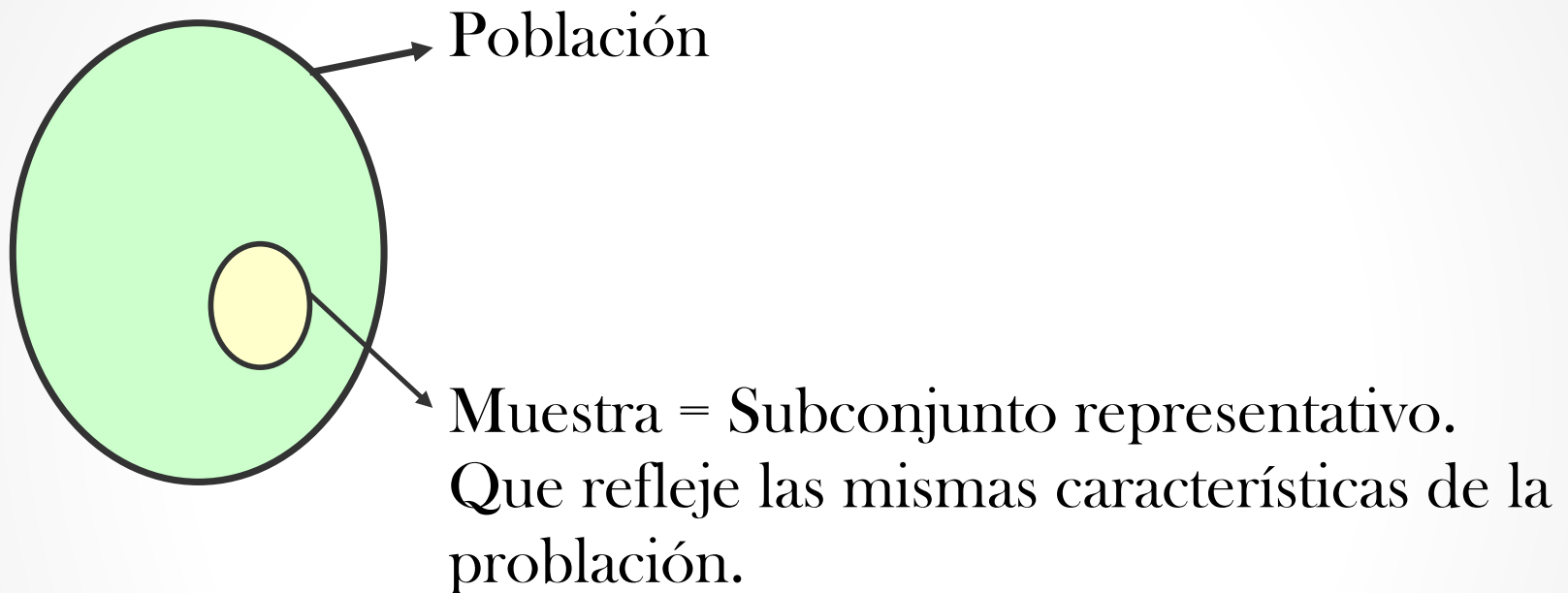
Por ejemplo, un fabricante de tarjetas para computadora podría desear eliminar defectos.

Un proceso de muestreo implicaría recolectar información de 50 tarjetas de computadora tomadas aleatoriamente durante el proceso.

En este caso la población sería representada por todas las tarjetas de computadora producidas por la empresa en un periodo específico.



Muestra representativa



Esa es la **idea del muestreo**: disponemos de cierta información, la que nos da la muestra, y queremos conocer lo que pasa en toda población de interés.

MUESTREO

Es la técnica para la selección de una muestra a partir de una población.

Es la acción por la cual se seleccionan en una población algunos elementos, a partir de los cuales se van tomar decisiones.

A través del muestreo se puede hacer análisis de situaciones de una población biológica, una empresa o algún campo de la sociedad.

Al elegir una muestra se espera que sus propiedades sean extrapolables a la población. Este proceso permite ahorrar recursos, obteniendo resultados similares a los que se obtendrían si se realizase un estudio de toda la población.

Teoría del muestreo: estudio de las relaciones existentes entre la distribución de un carácter en una población y en todas sus muestras.

Ventajas del Muestreo

↗ Costo reducido.

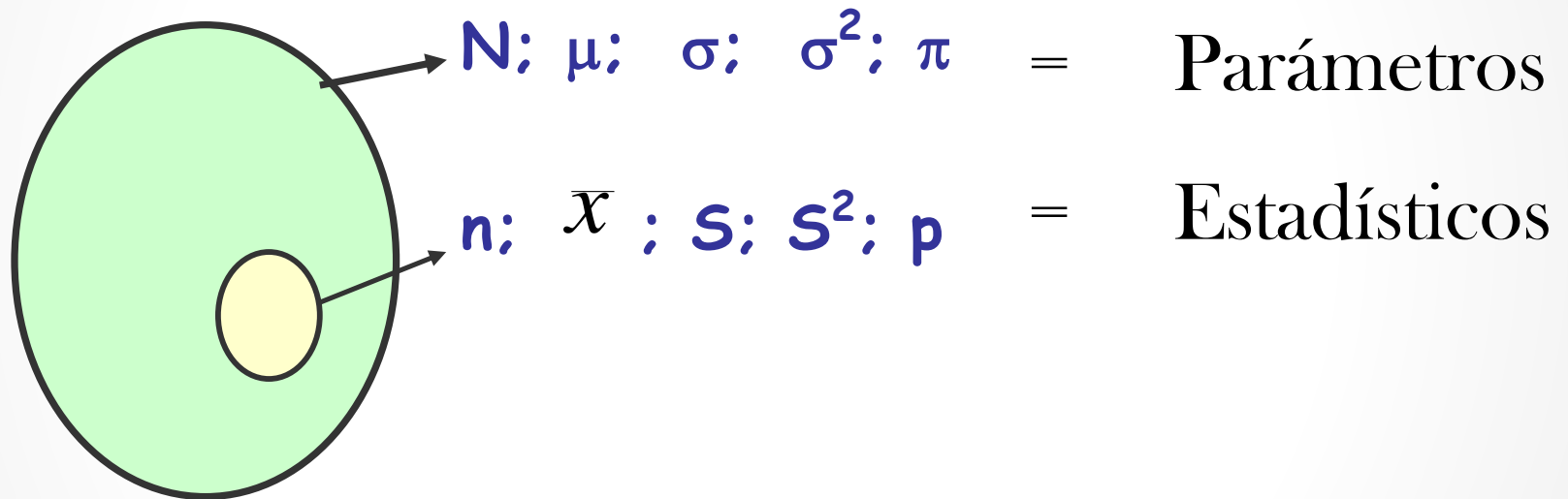
↗ Mayor rapidez.

↗ Más posibilidades.

Al hacer estadística inferencial debemos enfrentarnos con dos problemas:

- Elección de la muestra (*muestreo*).
- Extrapolación de las conclusiones obtenidas sobre la muestra, al resto de la población (*inferencia*).

Población y Muestra



Pasos para el muestreo

Definir:

 Objetivos

 Variables a estudiar

 Población objeto

 Unidad de muestreo

 Marco

 Tamaño de muestra

 Método de muestreo

Ejemplo

Se elige una muestra para contestar diferentes preguntas respecto a **la calidad del insumo principal para la realización de un producto**. Se toma una muestra de **50 unidades** de ese insumo. Suponga que se utiliza un **muestreo aleatorio** para seleccionar las unidades de muestreo.

Se analizará la **duración, resistencia, sabor**.

Población objeto proveedores del insumo.

Unidad de muestreo: paquete de kg?

Marco: listado de todos los proveedores

Definir:

 **Objetivos**

 **Variables a estudiar**

 **Población objeto**

 **Unidad de muestreo**

 **Marco**

 **Tamaño de muestra**

 **Método de muestreo**

Error muestral o error de muestreo

La diferencia entre el resultado obtenido de una muestra (un estadístico) y el resultado el cual deberíamos haber obtenido de la población (el parámetro correspondiente) se llama el error muestral o error de muestreo.

Un error de muestreo ocurre cuando no se lleva a cabo el estudio completo de la población, sino que se toma una muestra para estimar las características de la población.

El error muestral es medido por el error estadístico, en términos de probabilidad. El resultado de la media indica la precisión de la estimación de la población basada en el estudio de la muestra.

MUESTREO

Métodos de Muestreo:

Características de la Población

- Azar Simple
- Azar Sistemático
- Azar Estratificado
- Azar por Conglomerados

Tamaño de Muestra:

- Variabilidad de la Población
- Nivel de Confianza de las estimaciones
- Error de estimación permitido

Selección de muestras

Una muestra debe ser representativa si va a ser usada para estimar las características de la población. Los métodos para seleccionar una muestra representativa son numerosos, dependiendo del tiempo, esfuerzo, dinero y habilidad disponibles para tomar una muestra y la naturaleza de los elementos individuales de la población.

Los métodos de selección de muestras pueden ser clasificados de acuerdo a la forma de seleccionar los elementos que serán incluidos en la muestra.



Según la forma de seleccionar los elementos de una muestra

Una muestra es extraída al azar cuando la manera de selección es tal, que cada elemento de la población tiene igual oportunidad de ser seleccionado.

Una muestra aleatoria, también llamada muestra probabilística es preferida por los estadísticos porque la selección de las muestras es objetiva y el error muestral puede ser medido en términos de probabilidad.

Los tipos comunes de muestreo aleatorio son el muestreo aleatorio simple, muestreo sistemático, muestreo estratificado y muestreo de

- conglomerados

Muestreo aleatorio simple

- Consideremos una población finita, de la que deseamos extraer una muestra. Cuando el proceso de extracción es tal que garantiza a cada uno de los elementos de la población la misma oportunidad de ser incluidos en dicha muestra, denominamos al proceso de selección muestreo aleatorio.
- Por conveniencia, este método puede ser realizado con una tabla de números aleatorios.

La muestra aleatoria simple puede ser extraída con
• reemplazo o sin reemplazo. •

Muestreo aleatorio con reemplazo

Sobre una población de tamaño N podemos realizar extracciones de n elementos, pero de modo que cada vez el elemento extraído es repuesto al total de la población. De esta forma un elemento puede ser extraído varias veces.

Se caracteriza por que:

- Cada elemento de la población tiene la misma probabilidad de ser elegido, y
- Como las observaciones se realizan con reemplazamiento, cada observación es realizada sobre la misma población (no disminuye con las extracciones sucesivas).

La probabilidad de cualquier individuo de una población de tamaño N , de ser seleccionado para participar de la muestra es $p = 1/N$

La probabilidad de cualquier muestra de tamaño n de ser seleccionada de una población de tamaño N es $p = n/N$

Muestreo aleatorio sin reemplazo

En algunos casos no es posible o recomendable realizar un muestreo con reposición, en esos casos cada individuo que es seleccionado para participar de la muestra no es devuelto a la población.

El tamaño poblacional disminuye con cada extracción y la probabilidad de cada individuo será:

$p = 1/N$ en la primer extracción, $p=1/(N-1)$ en la segunda y así sucesivamente.

En cada extracción todos los individuos tienen la misma probabilidad.

MUESTREO ALEATORIO SIMPLE

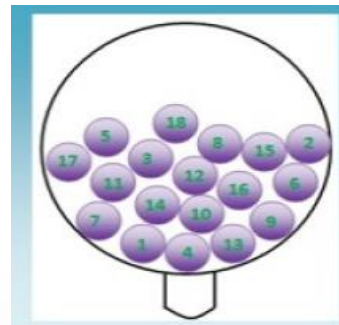
Ejemplo:

Supongamos que nos interesa elegir una muestra aleatoria de 5 estudiantes en un grupo de estadística de 20 alumnos.



1. Escribir los 20 nombres en pedazos separados de papel, colocarlos en un recipiente, revolverlos y luego extraer cinco papeles al mismo tiempo.
2. Listar los estudiantes y utilizar una tabla de números aleatorios, para escoger los 5 alumnos de acuerdo al orden de lista.
3. Excel tiene una opción para extraer una muestra aleatoria de una lista.

4. Enumerarlos y luego extraer 5 bolillas de un bolillero



Muestreo al azar simple

Población



Muestra

1	2	6	8	11	13	14	17	20	21
---	---	---	---	----	----	----	----	----	----



Muestreo al azar sistemático

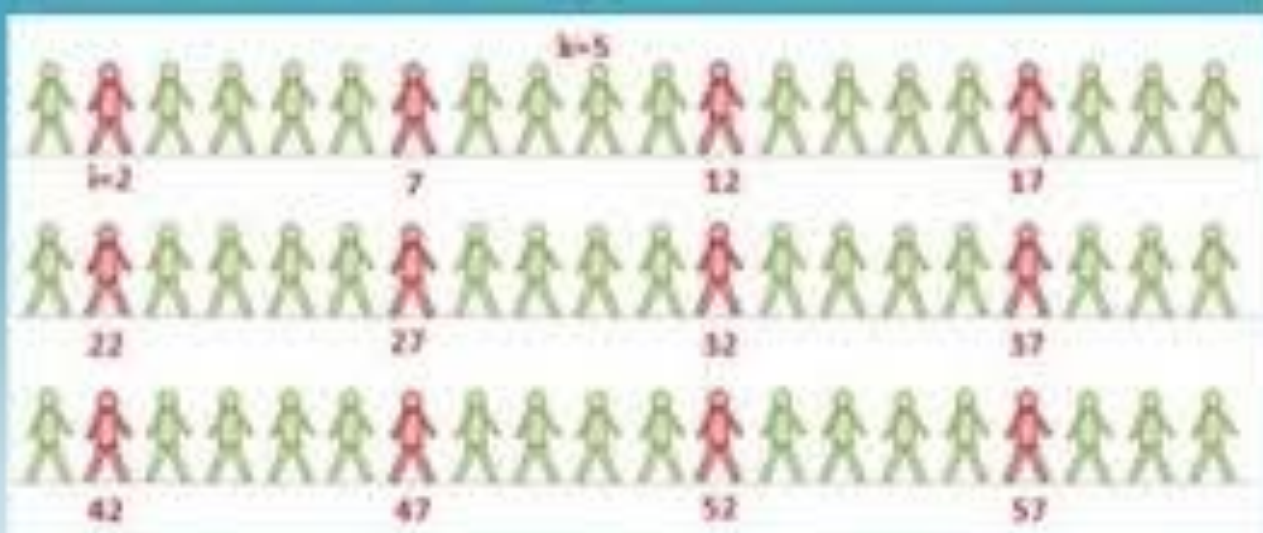
Una muestra sistemática es obtenida cuando los elementos son seleccionados en una manera ordenada. La manera de la selección depende del número de elementos incluidos en la población y el tamaño de la muestra.

El número de elementos en la población es, primero, dividido por el número deseado en la muestra. El cociente indicará si cada k-ésimo elemento en la población va a ser seleccionado.

$$k = \frac{N}{n}$$

El primer elemento de la muestra es seleccionado al azar. Por lo tanto, una muestra sistemática puede dar la misma precisión de estimación acerca de la población, que una muestra aleatoria simple cuando los elementos en la población están ordenados al azar.

Muestreo al azar sistemático



Muestreo sistemático

Muestreo al azar estratificado

Una muestra aleatoria estratificada se obtiene dividiendo o estratificando la población en subgrupos mutuamente excluyentes (estratos), tomando una muestra aleatoria de cada estrato. Las unidades tomadas de cada estrato se combinan para formar la muestra completa.

Los estratos son más homogéneos que la población como un todo.

El número de elementos seleccionados de cada estrato (asignación), puede ser fijo, proporcional al tamaño del estrato en relación con la población u óptimo.

Muestreo al azar estratificado

- Asignación fija

$$n1 = n2 = n3 = 20$$

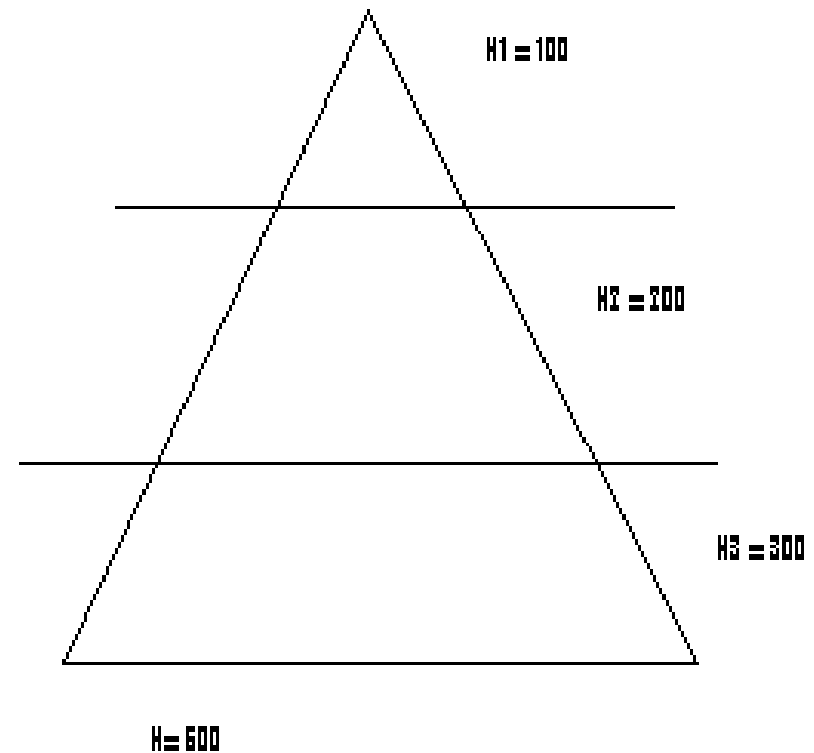
- Asignación proporcional

$$n1 = 10$$

$$n2 = 20$$

$$n3 = 30$$

- Asignación óptima



Muestreo Aleatorio Estratificado

- Una muestra aleatoria estratificada es la obtenida mediante la separación de los elementos de la población en grupos que no presenten traslapes, llamados estratos, y la selección posterior de una muestra aleatoria simple de cada estrato

Procedimiento

- Una población se divide en grupos llamados estratos.
- Hay homogeneidad dentro de los estratos.
- De cada estrato se selecciona el tamaño de muestra

Ventajas

- Aumento de la precisión sobre el MAS.
- Estimaciones separadas para cada estrato.
- Bajos costos de muestreo.

Desventajas

- Es necesario un marco para cada estrato.
- La división en estratos en algunas poblaciones puede no ser sencilla

Ejemplo de Muestreo Aleatorio Estratificado



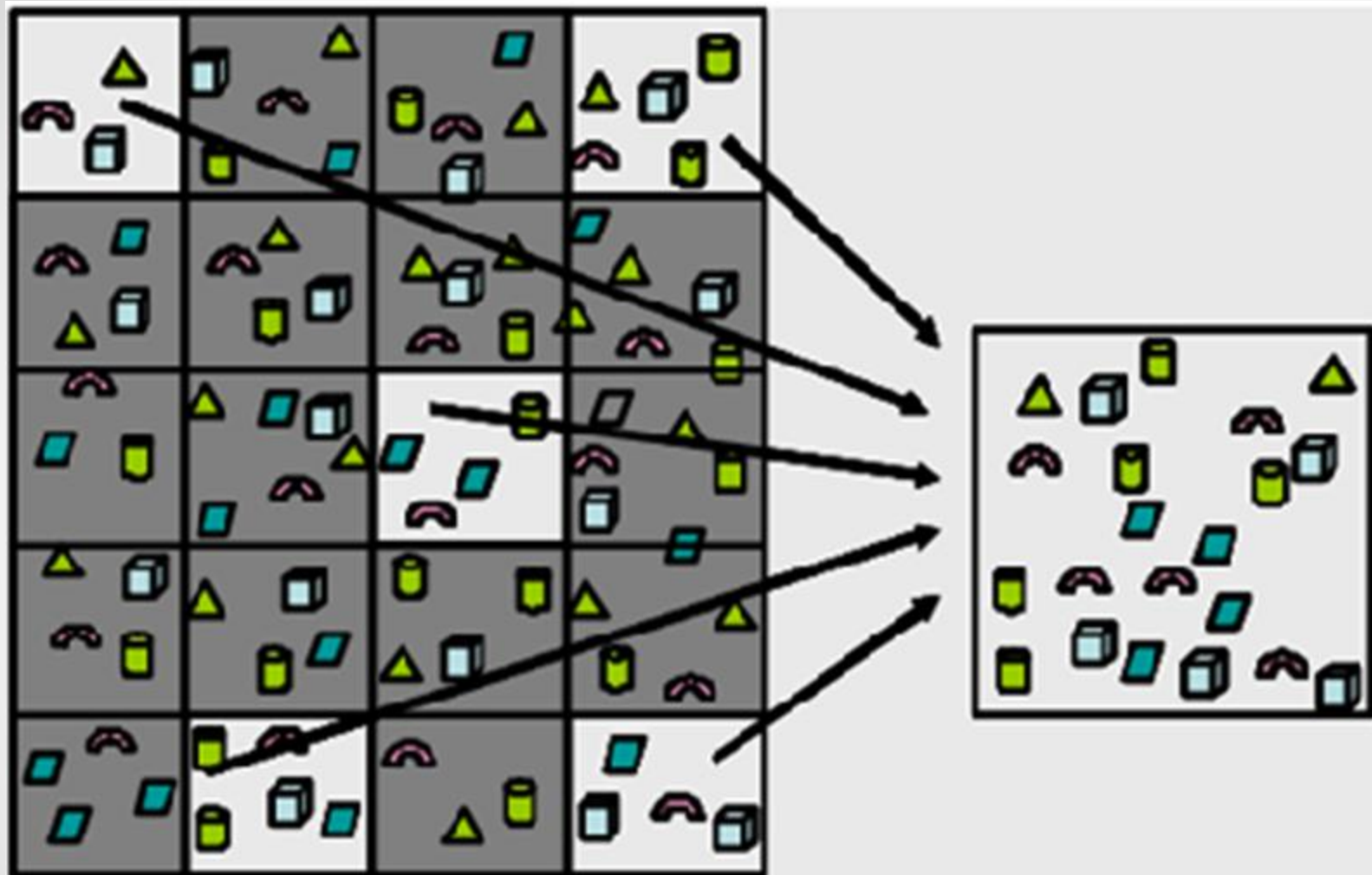
Muestra = 25% de la población

Muestreo por conglomerados

El muestreo por conglomerado consiste en agrupar las unidades de la población en conglomerados. Luego uno o mas conglomerados son seleccionados al azar y todas las unidades de esos conglomerados compondrán la muestra.

Bajo este método, aunque no todos los grupos son muestreados, cada grupo tiene igual probabilidad de ser seleccionado. Por lo tanto la muestra es aleatoria.

Una muestra de conglomerados, produce igual o menor error muestral (por lo tanto mayor precisión de las estimaciones acerca de la población) que una muestra aleatoria simple del mismo tamaño si los elementos individuales dentro de cada "conglomerado" conservan la mayor variabilidad de la población.



Objetivos de la determinación del tamaño adecuado de una muestra

- ✚ Estimar un parámetro determinado con el nivel de confianza deseado.
- ✚ Detectar una determinada diferencia, si realmente existe, entre los grupos de estudio con un mínimo de garantía.
- ✚ Reducir costos o aumentar la rapidez del estudio.
- ✚ La determinación de un tamaño adecuado de la muestra tendría como objetivo su factibilidad:
 - ✧ Los estudios con tamaños muestrales insuficientes, no son capaces de detectar diferencias entre grupos, llegando a la conclusión errónea de que no existe tal diferencia.
 - ✧ Si el número de sujetos es excesivo, el estudio se encarece desde el punto de vista económico y de recursos humanos.

Cálculo de n

Depende del parámetro a estimar y del método de muestreo a utilizar. Existen diversas fórmulas para el cálculo del tamaño mínimo de muestra, pero en general se basan en:

- ✎ La variabilidad existente en la población (σ).
- ✎ El nivel de confianza que se requiere de las estimaciones que se harán (α).
- ✎ El grado de precisión exigida a los estimadores que se calcularán ($EM=\theta-\theta$).
Error de muestreo.

$$n = \left(\frac{2z_{(1-\alpha/2)} * Sx}{d} \right)^2$$

Sx = es la desviación estándar estimada
 z = es el valor tabular de la distribución
“ z ”, con α nivel de probabilidad;
 d = es el margen de error permitido en la
estimación del parámetro.