



# MUESTREO



# Definición de muestreo

## Algunas definiciones de muestreo

Ibarra (1998)

El muestreo es un proceso mediante el cual es posible determinar en términos aproximados el comportamiento del **universo o población** donde se aplica, partiendo del hecho de inferir sus características en base a los resultados **obtenidos del análisis de una muestra extraída** del mismo.

Stevenson (2006)

El muestreo es un medio que permite el análisis de una pequeña parte de todos los elementos de un determinado grupo, cuyo objetivo es establecer generalizaciones con respecto a grupo total sin tener que examinarlos uno por uno.

Bauce (2004)

El muestro es el procedimiento a través del cual se procede a escoger una muestra de una población.

Seijas (2006)

Consiste en establecer una manera para identificar y escoger un subconjunto de elementos de un conjunto total de unidades que va ser analizado y evaluado con la finalidad de poder describir y estimar caracteres y condiciones de esa totalidad a la cual pertenece

**Muestra**   **Análisis**   **Inferimiento**   **Comportamiento y características de la población**

# Relación entre población- parámetro- muestra- estadístico

Proceso de inferencia estadística



**Población**

*Parámetro*

Magnitud o característica que poseen los elementos de una población

A  
APARTIR DE  
SE ESTIMAN  
PARAMETROS

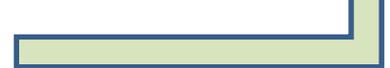
Magnitud o característica que poseen los elementos de una muestra

*Estadístico*

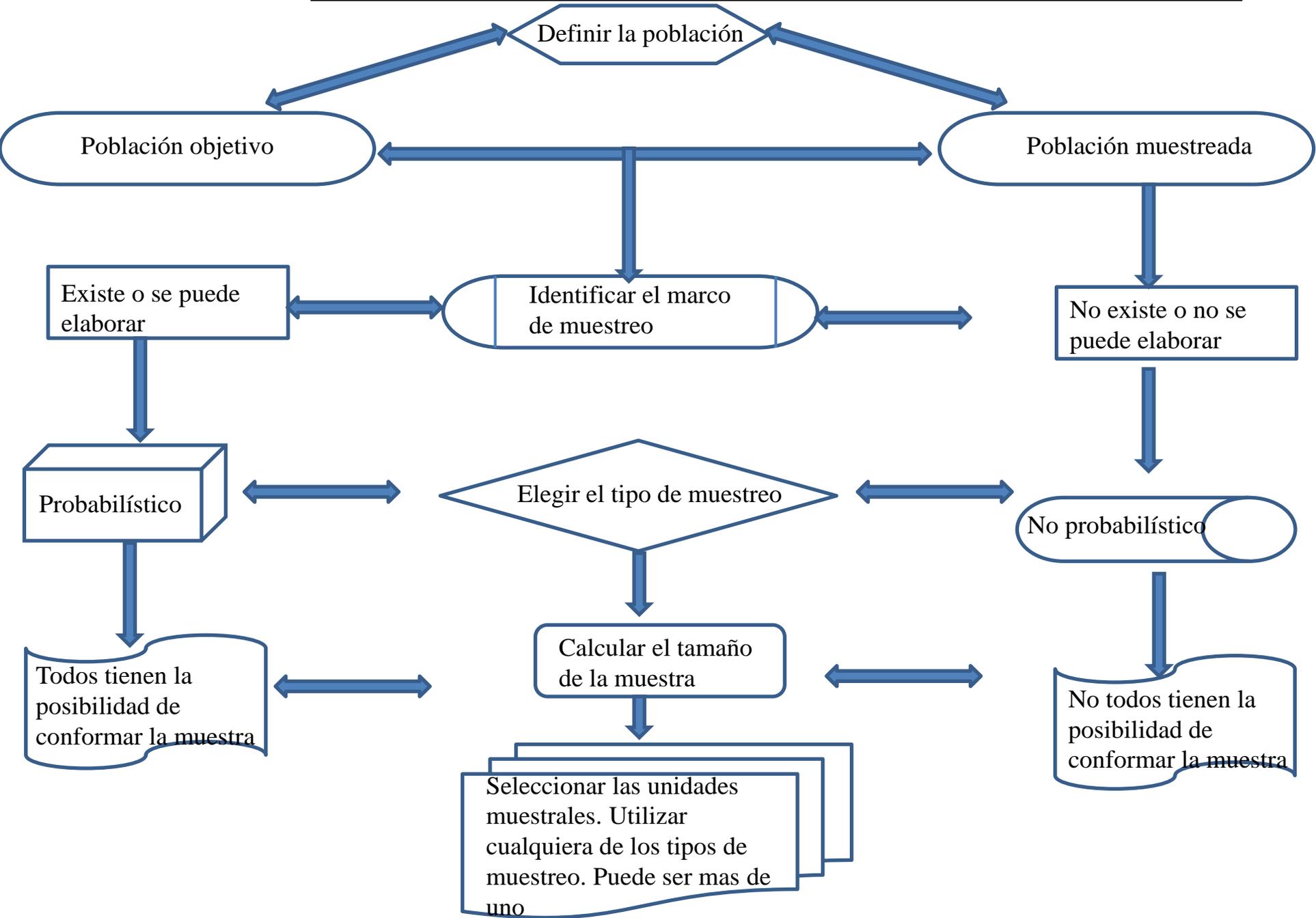
**MUESTRA**



Se extrae una parte de la población



# Procedimiento para realizar un muestreo

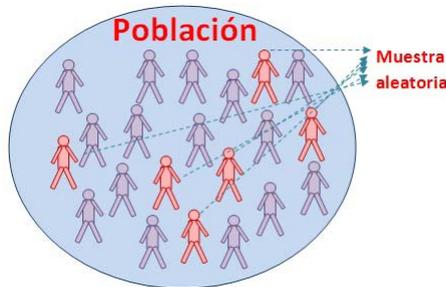


# Tipos de muestreo

De acuerdo a los objetivos e intención de la investigación que se pretende emprender y a los procedimientos a seguir, el muestreo puede clasificarse en dos grupos

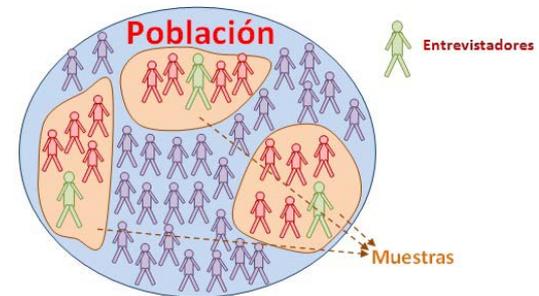
## Muestreo probabilístico

Todos los integrantes de la población tienen la posibilidad igual o diferente de pertenecer a la muestra (pueden ser equiprobables o no equiprobables)



## Muestreo no probabilístico

Cuando no todos los integrantes de la población tienen la posibilidad de conformar la muestra.



En esta clase de muestreo se desconoce la probabilidad a priori que tienen los sujetos u objetos de la población de formar parte de la muestra a seleccionar

# Clasificación del muestreo no probabilístico

## Muestreo intencional



Se apoya en el conocimiento y dominio que tiene un experto o la persona que investiga sobre un tema o hecho en particular, que aunado a la intencionalidad que se tiene sobre la selección de quienes se necesitan en la muestra, determina que muestra seleccionar, en este muestreo.

**Se apoya en la lógica, sentido común o en le sano juicio para seleccionar muestras representativas.**

➤ Las unidades se eligen en forma arbitraria, a juicio del entrevistador.

➤ Se toma cada unidad de observación según las características del informante y que resulten relevantes para el investigador.

➤ La representatividad subjetiva del investigador es sustento para afirmar que la muestra realmente personifica a una población.

## Características

**Ejemplo:** Se plantea una investigación para determinar la influencia de cierta literatura infantil en los valores que desarrollan los niños la comunidad de COLON. Para escoger la muestra de individuos que representaran a su población, no se necesita recurrir a procesos estadísticos matemáticos, sino a la intuición, lógica o al conocimiento que posea el propio investigador sobre el tema de investigación, en este caso la persona encargada de estudio encuestaría niños con hábitos de lectura y que lean este tipo de literatura



## Muestreo Opinático



En este tipo de muestreo la base utilizada para la escogencia de las unidades muestrales **es la opinión del investigador**, de modo que, la persona que selecciona la muestra procura que esta sea representativa .

➤ Al igual que muestreo intencional, la muestra dependerá de la representatividad **subjetiva del entrevistador**; este decide cuales y cuantos elementos serán designados.

### Características

Un gerente de una compañía que presta sus servicios de tv cable en la ciudad de San Cristóbal, quiere emprender una investigación sobre un nuevo producto que próximamente ofertará. Pero a su juicio u opinión personal, considera que las personas que entrevistará serán las que residan en conjuntos residenciales (apartamento), por la percepción que tiene de estas.

Es aquel donde la muestra se obtiene sin ningún plan preconcebido, las unidades de observación elegidas son producto de circunstancias fortuitas

➤ Los sujetos son elegidos en forma arbitraria sin juicios ni criterios previamente elaborados

➤ Las muestras son formadas por individuos que se han reunido casualmente o de acceso fácil a la comodidad del investigador



No permite generalizaciones que vayan más allá de las agrupaciones, por lo tanto, muy rara vez tiene carácter científico

El muestreo que realizan por lo general las corporaciones de resort cuando ubica jóvenes preferiblemente del género femenino en diversas intersecciones con semáforo en determinada ciudad para promocionar y ofertar su producto..

## Muestreo bola de nieve



Este proceso de muestreo es empleado cuando la población objetivo es **difícil de ubicar** y la unidad de análisis resulta ser muy escasa.



- Este muestreo parte del indicio de que las unidades de la población atípica se conocen entre sí.
- El investigador por cualquier medio localiza algún(os) informantes y éste(os) a su vez conducen a otros

Un etnógrafo emprenderá una investigación sobre los hábitos de los masones. Por lo tanto, debe logra identificar a un masón par (otro masón) que éste lo conduzca a otros masones y así completar una pequeña lista que pueda constituir una muestra.

*seis grados de separación*

**Stanley Milgram** psicólogo estadounidense

## Muestreo por cuotas



En este tipo de muestreo, las diversas características de una población (estado civil, raza, clase social, edad, genero, nacionalidad) son muestreadas de acuerdo con la **proporción** porcentaje que ocupan dentro de dicha población.

**Resulta ser relativamente económico y coyuntural en estudios de marketing, dado que es rápido y sencillo.**

- Se deja a discreción del encuestador las formas más convenientes de obtener la muestra.
- No precisa el listado de la población.

Para una investigación en la UNET, se decide seleccionar un 10%, 40%, 20% y 30% de docentes, estudiantes, obreros y administrativos respectivamente

# Clasificación del muestreo probabilístico

## Muestreo aleatorio simple



Todos los integrantes de la población muestreada tienen la misma probabilidad de ser tomados para dicha muestra, además, esta posibilidad de selección es conocida siendo diferente de cero (0) y de uno (1).

El procedimiento se fundamenta en determinar  $n$  elementos de una población de tamaño  $N$ , de manera que, todas las posibles muestras de tamaño  $n$ , es decir,  $\binom{N}{n}$  tengan la misma probabilidad ..  $\frac{1}{\binom{N}{n}}$  de ser elegidas.

### Manera de extraer la muestra

*Con reemplazo de los elementos que se van seleccionando*

Las unidades se pueden incluir mas de una vez en la muestra.

*Sin reemplazar las unidades ya extraídas*

Las unidades tomadas no se pueden repetir en la muestra

### Características

- ✓ Toma cada elemento para la muestra mediante un sorteo o cualquier otro proceso simple de selección como en los juegos de azar.
- ✓ Es el método mas simple y sencillo , pero de poco uso como tal, cuando se muestrea poblaciones grandes
- ✓ Sus propiedades matemáticas son sencillas.
- ✓ Se recomienda cuando las poblaciones son homogéneas (varianza poblacional tienda acero).
- ✓ En una muestra aleatoria simple, la unidad de muestreo y la unidad de observación coinciden.

## MUESTREO ALEATORIO SIMPLE (POBLACIÓN FINITA)

Para seleccionar una muestra aleatoria simple de la población finita de profesores de Ciencias Biológicas de Instituciones Educativas estatales, primero se le asigna a cada profesor un número. Por ejemplo, se les asignan los números del 1 al 2500 en el orden en que aparecen sus nombres en el archivo de personal de la zona educativa estatal.

63271	59986	71744	51102	15141	80714	58683	93108	13554	79945
88547	09896	95436	79115	08303	01041	20030	63754	08459	28364
55957	57243	83865	09911	19761	66535	40102	26646	60147	15702
46276	87453	44790	67122	45573	84358	21625	16999	13385	22782
55363	07449	34835	15290	76616	67191	12777	21861	68689	03263
69393	92785	49902	58447	42048	30378	87618	26933	40640	16281
13186	29431	88190	04588	38733	81290	89541	70290	40113	08243
17726	28652	56836	78351	47327	18518	92222	55201	27340	10493
36520	64465	05550	30157	82242	29520	69753	72602	23756	54935
81628	36100	39254	56835	37636	02421	98063	89641	64953	99337
84649	48968	75215	75498	49539	74240	03466	49292	36401	45525
63291	11618	12613	75055	43915	26488	41116	64531	56827	30825
70502	53225	03655	05915	37140	57051	48393	91322	25653	06543
06426	24771	59935	49801	11082	66762	94477	02494	88215	27191
20711	55609	29430	70165	45406	78484	31639	52009	18873	96927
41990	70538	77191	25860	55204	73417	83920	69468	74972	38712
72452	36618	76298	26678	89334	33938	95567	29380	75906	91807
37042	40318	57099	10528	09925	89773	41335	96244	29002	46453
53766	52875	15987	46962	67342	77592	57651	95508	80033	69828
90585	58955	53122	16025	84299	53310	67380	84249	25348	04332
32001	96293	37203	64516	51530	37069	40261	61374	05815	06714
62606	64324	46354	72157	67248	20135	49804	09226	64419	29457
10078	28073	85389	50324	14500	15562	64165	06125	71353	77669
91561	46145	24177	15294	10061	98124	75732	00815	83452	97355
13091	98112	53959	79607	52244	63303	10413	63839	74762	50289

Tabla: NÚMEROS ALEATORIOS

Como el número mayor en la lista de la población de profesores, 2500, tiene cuatro dígitos, se seleccionarán números aleatorios de la tabla en conjuntos o grupos de cuatro dígitos

Los primeros 7 números aleatorios de cuatro dígitos son: 6327 **1599** 8671 7445 **1102 1514 1807**  
 El primer número, 6327, es mayor que 2500. No corresponde a ninguno de los profesores numerados que forman la población y por tanto se descarta. Igualmente se ignoran 8671 y 7445

# MUESTRA ALEATORIA SIMPLE (POBLACIÓN INFINITA)

Una muestra aleatoria simple de una población infinita es una muestra seleccionada de manera que se satisfagan las condiciones siguientes.

- 1. Cada uno de los elementos seleccionados proviene de la población.**
- 2. Cada elemento se selecciona independientemente.**

población infinita

- ✓ Partes fabricadas en una línea de producción
- ✓ Transacciones en un banco
- ✓ Llamadas que llegan a un centro de asesoría técnica
- ✓ Clientes que entran en las tiendas.

El número de muestras aleatorias simples distintas de tamaño  $n$  que pueden seleccionarse de una población finita de tamaño  $N$  es

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Los datos del problema de los profesores, en el que  $N=2500$  y  $n=30$ ,

$$\binom{N}{n} = \frac{2500!}{30!(2500-30)!} = \frac{2500!}{30 \times 2470!} = \frac{2500 \times 2499 \times 2498 \times \dots \times 2470!}{30 \times 2470!} \approx 10^{69}$$

se pueden tomar  $\approx 10^{69}$  muestras aleatorias simples distintas de 30 profesores

## Media poblacional

$$\bar{x} = \frac{1}{n} \sum_{i \in S} x_i$$

Desviación estándar de la distribución muestral de

*Población finita*

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}} \left( \frac{\sigma}{\sqrt{n}} \right)$$

$\sigma_{\bar{x}}$  desviación estándar de  $\bar{x}$

$\sigma$  desviación estándar de la población

$n$  tamaño de la muestra

$N$  tamaño de la población

*Población infinita*

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

siempre que

**1. La población sea infinita; o**

**2. La población sea finita y el tamaño de la muestra sea menor o igual a 5% del**

**tamaño de la población; es decir,  $n/N \leq 0.05$ .**

Se procede a la selección de los integrantes de la muestra. Se puede utilizar cualquier recurso o procedimiento aleatorio simple o complejo.

**Situaciones  
para un  
muestreo  
aleatorio  
simple**

**Favorables**

- ❖ Cuando sea una exigencia el uso de ecuaciones
- ❖ Cuando se disponga de poca información sobre el área de interés para diseñar la encuesta.
- ❖ Cuando el eje de interés del estudio son relaciones multivariadas o regresiones

**Desfavorables**

- ❖ **Cuando las encuestas no constituyan la mayor herramienta para el análisis de los datos de una investigación.**
- ❖ Cuando no se disponga de un listado completo de las unidades de observación para el estudio, o cuando resulte demasiado costoso obtenerlo.
- ❖ Cuando resulte más beneficioso y se ahorre en inversión al dividir la población en subgrupos

## Distribución muestral de $\bar{p}$

La proporción muestral  $\bar{p}$  es el estimador puntual de la proporción poblacional  $p$ . *La fórmula para calcular la proporción muestral es*

$$\bar{p} = \frac{x}{n}$$

$x$  número de elementos de la muestra que poseen la característica de interés

$n$  tamaño de la muestra

## Desviación estándar de $\bar{p}$

*Población finita*

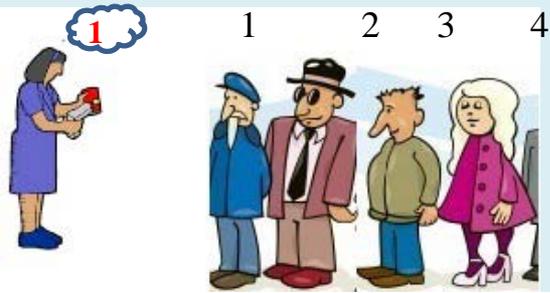
*Población infinita*

$$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}} \left( \sqrt{\frac{p(1-p)}{n}} \right)$$

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

La distribución muestral de  $\bar{p}$  se aproxima mediante una distribución normal siempre que  $np \geq 5$  y  $n(1-p) \geq 5$ .

# Muestreo sistemático



Consiste en seleccionar muestras mediante un procedimiento organizado a través de una lista ordenada (creciente o decreciente) de los integrantes de una población ( $K$ ), para que de este modo se puedan escoger unidades de la población en intervalos fijos.

*En este muestreo, las unidades de la población se toman en intervalos de  $K$  en  $K$ , siendo  $K$ , una constante calculada a partir del tamaño de la población y el tamaño de la muestra. El procedimiento consiste en numerar los elementos de la población desde uno (1) hasta “N”, en cualquier orden, para luego dividirla en cierta cantidad de partes de tamaño  $K$  (**intervalo de selección sistemático**).*

➤ *Los elementos que integran la muestra se toman de acuerdo a cierto orden (de  $K$  en  $K$ ).*

➤ Los elementos se seleccionan de forma regular y ninguno de ellos se puede tomar más de una vez.

➤ *Este muestreo resulta preciso si las unidades que integran la muestra son heterogéneas.*

➤ Generalmente, un muestreo sistemático no proporciona una muestra representativa si, la lista de las unidades integrantes de una población poseen algún tipo de orden periódico o cíclico.

➤ *No requiere el uso de tablas de números aleatorios.*

➤ A veces se puede usaren sustitución del muestreo aleatorio simple, pero siempre y cuando, no se disponga de una lista de la población, o cuando dicha lista tenga un orden más o menos aleatorio.

➤ *El muestreo sistemático no es igual al muestreo aleatorio simple, puesto que no posee la propiedad de que cada posible colección “n” unidades tenga la misma probabilidad de formar parte de la muestra (equiprobables).*

## Características

## Procedimiento para el muestreo Sistemático

1. Se elabora un listado de  $N$  elementos o integrantes de la población. Pueden enumerarse desde el 1 hasta el  $N$ .
2. Se determina el tamaño de la muestra siguiendo los postulados matemáticos en los que se apoya la estadística, o mediante un procedimiento sencillo.
3. Se establece el intervalo fijo de selección ( $K$ ) mediante la expresión

$$\frac{N}{n}$$

$N \rightarrow$  Es el tamaño de la población

$n \rightarrow$  Es el tamaño de la muestra

1. Luego se procede a definir y determinar en qué elemento se iniciará el proceso de selección. A este valor se le denota por  $R$ .
2. Por último, se configura la muestra, la cual quedará definida por el siguiente conjunto :

$$n = \{R, R + K, R + 2K, \dots, R + nK\}$$

Para ciertos muestreos, en especial en aquellos con poblaciones grandes, se necesita mucho tiempo para tomar una muestra aleatoria simple (hallando primero los números aleatorios y después contando y recorriendo toda una lista de la población hasta encontrar los elementos correspondientes). Una alternativa al muestreo aleatorio simple es el **muestreo sistemático**.

Por ejemplo, si se quiere una muestra de tamaño *10* de una población que tiene *50* elementos

se muestrea uno de cada  $\frac{N}{n} = \frac{50}{10} = 5$  elementos de la población

Un muestreo sistemático consiste en seleccionar en forma aleatoria uno de los primeros elementos de la lista de la población. Se selecciona o se sortea el inicio “*R*”, puede ser cualquier valor aleatorio entre *1* y *K*. *Por ejemplo R=2*

**Muestra:** 2, 2+5, 2+10, 2+15, 2+20, 2+25, 2+30, 2+35, 2+40, 2+45

*n=2,7,12,17,22,27,32,37,42,47*

De esta manera es más fácil de identificar la muestra de *50* que si se usara el muestreo aleatorio simple

# Muestreo estratificado



Este tipo de muestreo es aplicable cuando una población se encuentra dividida en subconjuntos pero con la particularidad de que sus integrantes poseen **características comunes**, es decir, existe una partición en el grupo con estratos homogéneos en su interior .

*El procedimiento en este tipo de muestreo consiste en dividir el conjunto total de elementos  $N$  en subconjuntos o estratos utilizando ciertas variables de control llamadas **variables claves de estratificación**, y estas a su vez se relacionan con las variables objeto de estudio.*

**Para algunos investigadores consideran a este tipo de muestreo como una variación de muestreo aleatorio simple**

## Características

➤ **Es aplicable cuando la variable objeto de estudio asume diferentes valores promedio en diferentes subpoblaciones o estratos.**

➤ Se utiliza cuando quien investiga o realiza el estudio se quiere resguardar de la posibilidad de obtener muestras malas (no representativas).

➤ Su aplicación es ideal cuando se quiere obtener datos de precisión conocida de los subgrupos.

➤ Este muestreo puede administrarse de manera mas conveniente ya un menor costo, ya que se pueden utilizar diversos patrones de muestreo para los diferentes estratos.

➤ Si el muestreo estratificado se realiza en forma correcta, permitirá obtener estimaciones e inferencias sobre la población más precisas, es decir, los datos presentan menor variabilidad pues su varianza se hace pequeña.

➤ **Es aplicable a estudios socioeconómicos y demográficos.**

## Procedimiento para el muestreo estratificado

1. Se clasifica la población en subgrupos o estratos que no se traslapan, es decir, los estratos no se cruzan entre sí. Una misma unidad de observación (sujeto) no se puede ubicar en más de un (1) subgrupo.
2. Se procede a calcular tamaño de la muestra  $n$ , y se construye una tabla para organizar los estratos con sus respectivos porcentajes y cantidad de elementos que aportará a la muestra.

Población:  $N$

muestra:  $n$  10% de  $N$

$$\text{peso } (W) = \frac{\text{\# de observaciones en el estrato}}{\text{Total de la población}}$$

$$\text{porcentaje} = \text{peso } (W) \cdot 100$$

$$\text{Muestra} = n[\text{peso } (W)]$$

3. Ahora corresponde seleccionar en cada estrato los individuos que integraran la muestra. A tal efecto se puede recurrir a un muestreo sistemático o aleatorio simple, según los recursos y tiempo del que se disponga.

# Aspectos matemáticos del muestreo estratificado

En una población de  $N$  unidades de muestreo en  $H$  estratos, cada uno con  $N_h$  unidades muestreadas en cada estrato  $h$ , se dice que se realiza un muestreo estratificado cuando se determinan los valores  $N_1, N_2, \dots, N_h$  y se establece que:

$N_1 + N_2, \dots, + N_h = N$        $N$  Es el número total de unidades de toda la población que se le aplica el muestreo

**Total de la población en el estrato  $h$**   $t_h = \sum_{j=1}^{N_h} y_{hj}$

**Total de la población**  $t = \sum_{h=1}^H t_h$

**Media de la población en el estrato  $h$**   $\bar{y}_h U = \frac{\sum_{j=1}^{N_h} y_{hj}}{N_h}$

**Media global de la población**

$$YU = \frac{t}{N} = \frac{\sum_{h=1}^H \sum_{j=1}^{N_h} y_{hj}}{N}$$

**Varianza de la población en el estrato  $h$**   $S_h^2 = \sum_{j=1}^{N_h} \frac{(y_{hj} - \bar{y}_h)^2}{N_h - 1}$

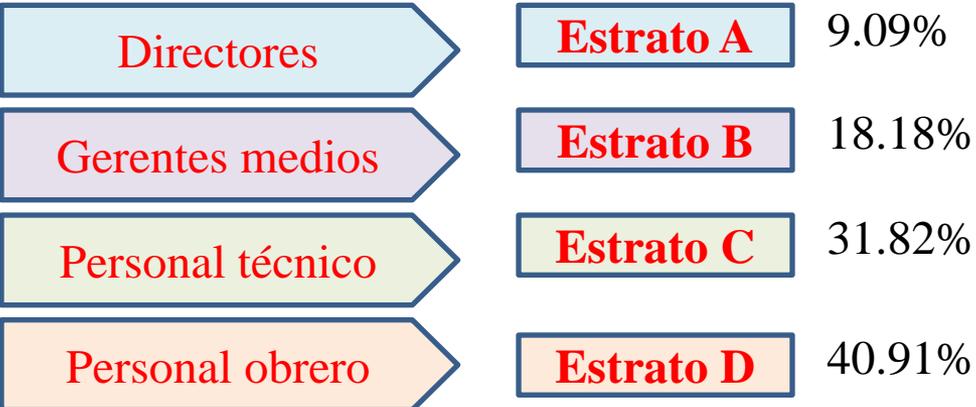
**Cantidades de la muestra para estimar la muestra aleatoria de cada uno de los estratos**

$$y_h = \frac{\sum_{j \in S_h} y_{hj}}{n_h}$$
$$\hat{t}_h = \frac{N_h}{n_h} \sum_{h=1}^H \bar{y}_{hj}$$
$$S_h^2 = \sum_{j \in S_h} \frac{(y_{hj} - \bar{y}_h)^2}{n_h - 1}$$

**Total de la población y la media aritmética estimada**

$$t = \sum_{h=1}^H t_h$$
$$\hat{t}_{estimado} = \sum_{h=1}^H t_h = \sum_{h=1}^H N_h \bar{Y}_h$$
$$\bar{y}_{estimada} = \frac{\hat{t}_{estimado}}{N} = \sum_{h=1}^H \frac{N_h}{N} \bar{Y}_h$$

Suponga una gran empresa corporativa que cuenta con personal directivo, gerentes a nivel medio, personal técnico o profesional y personal obrero. Suponga que se desea obtener una muestra donde estén representados estos cuatro renglones: Suponga, que la corporación la conforman cinco (5) directores, veinticinco (25) gerentes medio, ciento setenta (170) personas entre personal técnico y profesional y 300 personas del sector obrero. Se pide un muestreo para seleccionar el 10% de quienes laboran en la corporación para realizar un estudio. De modo que se cuentan con los siguientes estratos.



Población (N): 500

Muestra (n)=10%=0.1x500=50

$$\text{peso (W)} = \frac{\text{\# de observaciones en el estrato}}{\text{Total de la población}}$$

$$\text{porcentaje} = \text{peso (W)} \cdot 100$$

Estrato	Población (N)	Porcentaje	Muestra (n)
Directores	5	1	0.5=1.0
Gerentes medios	25	5	2.5=3.0
Personal técnico	170	34	17
Personal obrero	300	60	30
<b>Total:</b>	500	100	50

$$\text{Muestra} = n [\text{peso (W)}]$$

Corresponde seleccionar en cada estrato los individuos que integran la muestra.

# Muestreo por conglomerado



Parte del principio de dividir la población objeto de estudio en módulos y submódulos menores pero con idénticas características a dicha población, para luego, seleccionar aleatoriamente algunos de esos módulos y así, poder obtener la muestra. El muestreo por conglomerados agrupa una población en pequeño aglomerados. A diferencia del muestreo por estratos, los conglomerados son muy idénticos, pero internamente las características de sus integrantes son muy heterogéneas dentro de cada subgrupo.

**En este caso la unidad primaria no es el individuo sino el grupo.**

¿Por qué el muestreo por conglomerado?

- El Listado completo de las unidades es difícil, imposible o costoso de obtener.
- La Población se encuentra muy dispersa geográficamente, o esta formada en cúmulos naturales.

## Características

➤ *Los elementos individuales de una población pueden formar parte de la muestra si y sólo si pertenecen a unos de los conglomerados que se ha tomado para la muestra.*

➤ *Facilita la reducción de costos*

➤ *Resulta ideal cuando la población se encuentra dispersa geográficamente.*

➤ Los elementos contenidos en los conglomerados seleccionados para la muestra son encuestados en su totalidad

➤ Por lo general, es precedido por un muestreo estratificado

➤ Para el observador externo, los conglomerados se perciben homogéneos pero internamente presentan varianza grande.

## Ventajas del muestreo por conglomerados

- ❖ Sus principales beneficios redundan en el trabajo de campo, en la aplicación y manejo de las encuestas.
- ❖ **No hacen falta las engorrosas listas o listados de sujetos de la población a muestrear**
- ❖ Se puede concentrar la aplicación de las entrevistas o instrumentos de recabar información
- ❖ **Los costos disminuyen de manera muy sustancial al disponer de la información en cúmulos**
- ❖ Se logra evitar los traslados continuos de sujeto en sujeto para encuestar la muestra
- ❖ La información se obtiene de manera mas rápida, además se logra supervisar mejor
- ❖ Facilita la construcción de unidades estadísticas artificiales para lugares de alto riesgo
- ❖ El equipo de encuestadores puede trabajar en conjunto y simultáneamente cubrir todos los conglomerados

## Procedimiento para el muestreo por conglomerados

Por la naturaleza de la investigación se determinan y enumeran los subgrupos o conglomerados que se van a muestrear, una vez se tenga el listado de conglomerados (unidad muestral), se procede a seleccionar aleatoriamente los que integraran la muestra a analizar.

**Se encuestan todos y cada uno de los elementos o sujetos que integran los conglomerados seleccionados para la muestra**

**Ejemplo:** Un investigador social, desea recolectar información entre los integrantes de la comunidad educativa de la ULA-Táchira. Desde fuera todos se ven como integrantes de la comunidad, pero internamente existen varios subgrupos diferentes entre si. Por lo tanto, se decide tomar la muestra dividiendo la población en carreras y menciones, para seleccionar al azar ciertas cantidades de grupos para entrevistar a todos los que pertenezcan a los grupos escogidos.

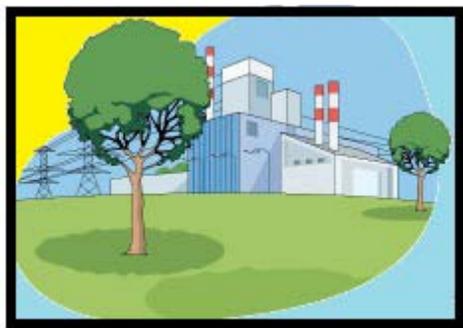
## Modalidades del muestreo por conglomerados

**Una etapa:** La población se divide en grupos y se selecciona aleatoriamente uno como muestra. Se obtienen unidades de muestreo primarias

**Bietapa:** Se considera como un submuestreo. Las unidades primarias se subdividen para seleccionar una entre ellas. Se obtienen unidades de muestreo secundarias

**Polietapico:** Muestreo que se realiza en mas de dos etapas. Partiendo de una unidad secundaria, ésta se vuelve a subdividir para obtener unidades terciarias, ... ..

## Muestreo por conglomerado: Una etapa



Recibe este nombre, el muestreo en el cual la población se divide en grupos y se seleccionan algunos de ellos como unidades muestrales. Existe cierta similitud con el muestreo estratificado, con la diferencia de que los estratos son conjuntos mas o menos homogéneos en su estructura interna, mientras que en los conglomerados esa estructura interna presenta mayor heterogeneidad, con el fin de que cada conglomerado sea representativo de la población a la cual pertenece.

### Características

Quienes lo conforman

Todos o ninguno de los elementos que integran un conglomerado (unidad primaria de muestreo).

Cuando se utiliza

En encuestas donde el costo del muestreo con unidades secundarias resulta despreciable en comparación con un muestreo de unidades primarias

En el muestreo por conglomerados de una etapa, se pueden presentar dos casos

## Conglomerados del mismo tamaño

Aplicable áreas agrícolas e industriales, Cuando se estima la media o totales de los aglomerados y se ignoran los elementos individuales. Entonces en una muestra aleatoria simple de  $n$  observaciones  $\{t_i, i \in S\}$ ; donde  $t_i$  es el total para todos los elementos de la unidad primaria  $I$ , entonces,  $t_s$  estima el promedio de los totales por cumulo

Estimadores

**Promedio**

**Varianza**

$$\hat{t} = \frac{N}{n} \sum_{i \in S} t_i \quad V(\hat{t}) = N^2 \left(1 - \frac{n}{N}\right) \frac{S_t^2}{n}$$

Para estimar proporciones

$$P_{st} = \frac{\sum M_h P_h}{\sum m_h} \quad P_h = \frac{\sum a_{hi}}{\sum m_{hi}} \quad V(\hat{t}) = (NM)^2 \left(1 - \frac{n}{N}\right) \left(\frac{1}{n(n-1)}\right) \sum_{i=1}^N (p_i - \hat{P})^2 \quad \hat{t} = NM\hat{P} \quad \hat{P} = \frac{1}{n} \sum_{i=1}^n P_i$$

Error absoluto del muestreo  $\hat{C}_v(\hat{P}) = \frac{\sqrt{\hat{V}(\hat{t})}}{\hat{t}}$

## **Conglomerados de distinto tamaño**

En estudios de naturaleza social es muy raro que los conglomerados tengan el mismo tamaño. Los procedimientos de estimación son prácticamente iguales a los del muestreo en una etapa, las diferencias que se pueden encontrar se deben a que la variación entre los totales de los conglomerados sean grandes cuando los conglomerados tienen distintos tamaños.

Por lo tanto, con frecuencia es más grande en un muestreo por conglomerado cuando las unidades primarias tienen distintos tamaños que cuando las unidades primarias tienen el mismo número de unidades secundarias.

**Población estudiantil de la Universidad de Los Andes “Pedro Rincón Gutiérrez”-Táchira sobre la lectura de literatura especializada que hace el estudiantado de esta casa de estudios superiores. Esta base de datos fue tomada de una muestra de 200 personas, pero, para la explicación se asumirá que es una población de 200 estudiantes.**

Numero de conglomerados

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

$N$ : Tamaño de la población

$n$ : Tamaño del conglomerado

El tamaño del conglomerado es a criterio del investigador. Para nuestro ejemplo asumamos que fue de  $n=25$ .

$$I = \frac{200}{25} \Rightarrow I = 8$$

Son ocho (8) conglomerados con 25 estudiantes cada uno. Para su estructuración se puede recurrir a la tabla de números aleatorios, de manera que aleatoriamente se conformen los conglomerados.

Conglomerado 1

c	Sorteo	Gasto	Apoyo
1	58	210	Pa
2	77	200	Pa
3	73	30	Pa
4	41	150	Pa
5	186	200	Be
6	162	60	Pa
7	26	100	Pa
8	48	150	Pa
9	118	0	Pa
10	194	0	Pa
11	12	500	Pa
12	68	50	Pa
13	32	0	Pa
14	19	0	Pa
15	1	100	Pa
16	128	0	Pa
17	95	120	Pa
18	35	0	Pa
19	113	50	Pa
20	33	50	Pa
21	94	70	Pa
22	11	500	Pa
23	81	50	Pa
24	63	120	Pa
25	54	100	pa

Conglomerado 2

c	Sorteo	Gasto	Apoyo
1	189	100	Pa
2	37	300	Mp
3	21	150	Pa
4	148	0	Pa
5	78	30	Pa
6	25	100	Pa
7	53	150	Pa
8	199	0	Pa
9	146	0	Pa
10	183	0	Be
11	182	0	Pa
12	186	100	Pa
13	111	0	Be
14	57	210	Pa
15	126	0	Pa
16	75	0	Pa
17	93	0	Pa
18	43	176	Pa
19	91	60	Mp
20	127	100	Pa
21	176	50	Pa
22	55	39	Mp
23	158	20	Pa
24	8	150	Pa
25	31	200	pa

Conglomerado 3

c	Sorteo	Gasto	Apoyo
1	14	100	Pa
2	121	0	Pa
3	123	42	Pa
4	86	200	Pa
5	96	150	Pa
6	87	16	Pa
7	16	150	Pa
8	59	100	Pa
9	82	40	Pa
10	88	78	Pa
11	152	100	Be
12	22	100	Be
13	62	150	Pa
14	108	0	Pa
15	49	111	Pa
16	163	105	Pa
17	76	100	Pa
18	99	100	Pa
19	17	0	Pa
20	104	0	Pa
21	97	30	Pa
22	69	50	Pa
23	6	0	Pa
24	13	80	Pa
25	117	0	Pa

Conglomerado 4

c	Sorteo	Gasto	Apoyo
1	74	100	Pa
2	165	35	Pa
3	159	50	Pa
4	149	0	Pa
5	29	200	Pa
6	65	210	Pa
7	3	300	Pa
8	84	50	Pa
9	114	0	Pa
10	142	0	Pa
11	129	0	Pa
12	18	0	Pa
13	24	100	Pa
14	47	111	Pa
15	85	1200	Pa
16	143	0	Pa
17	155	50	Pa
18	34	100	Pa
19	5	0	Pa
20	133	50	Pa
21	66	0	Be
22	174	0	Pa
23	2	80	Pa
24	67	30	Be
25	40	200	pa

Conglomerado 5

c	Sorteo	Gasto	Apoyo
1	184	400	Be
2	52	30	Pa
3	36	100	Pa
4	15	600	Pa
5	153	108	Be
6	168	0	Pa
7	122	40	Pa
8	61	150	Pa
9	51	50	Pa
10	180	0	Pa
11	161	300	Pa
12	192	0	Pa
13	195	0	Pa
14	134	15	Pa
15	191	100	Pa
16	138	20	Pa
17	131	50	Pa
18	38	200	Pa
19	200	0	Pa
20	193	0	Pa
21	56	220	Pa
22	10	600	Pa
23	107	0	Pa
24	140	25	Pa
25	83	50	Mp

Conglomerado 6

c	Sorteo	Gasto	Apoyo
1	120	0	Pa
2	71	200	Pa
3	132	100	Pa
4	125	0	Mp
5	101	0	Pa
6	119	0	Mp
7	139	200	Pa
8	181	0	Pa
9	72	40	Pa
10	115	0	Pa
11	102	100	Be
12	7	150	Be
13	141	0	Pa
14	20	0	Pa
15	198	80	Pa
16	89	70	Pa
17	135	0	Pa
18	124	0	Be
19	30	150	Pa
20	157	50	Be
21	50	50	Pa
22	167	0	Pa
23	151	0	Pa
24	116	0	Pa
25	144	0	Pa

Conglomerado 7

c	Sorteo	Gasto	Apoyo
1	188	0	Pa
2	117	0	Pa
3	9	150	Pa
4	154	200	Pa
5	169	100	Pa
6	92	0	Pa
7	150	0	Pa
8	103	0	Pa
9	197	0	Pa
10	105	0	Pa
11	4	0	Pa
12	110	0	Pa
13	112	0	Pa
14	179	0	Pa
15	147	0	Pa
16	45	100	Pa
17	98	600	Pa
18	164	0	Pa
19	42	150	Pa
20	79	20	Pa
21	136	0	Pa
22	100	0	Pa
23	39	100	Pa
24	130	0	Pa
25	90	50	Pa

Conglomerado 8

c	Sorteo	Gasto	Apoyo
1	23	80	Pa
2	27	0	Pa
3	28	100	Pa
4	44	120	Pa
5	46	0	Be
6	60	80	Pa
7	64	200	Pa
8	70	50	Pa
9	80	32	Pa
10	106	160	Be
11	109	0	Pa
12	137	70	Pa
13	145	0	Pa
14	160	0	Pa
15	166	90	Pa
16	170	50	Be
17	171	30	Pa
18	172	100	Pa
19	173	80	Pa
20	175	0	Mp
21	178	100	Pa
22	185	0	Pa
23	187	70	Pa
24	190	0	Pa
25	196	50	Be

## Muestreo por conglomerado: Dos etapas

El muestreo en dos etapas (bietápico) de conglomerados es un tipo de muestreo en el que en una primera etapa se selecciona una muestra de  $n$  conglomerados de tamaños  $M_i$ , y en una segunda etapa se selecciona, independientemente en cada conglomerado de la primera etapa, una submuestra de  $m_i$  unidades elementales de entre las  $M_i$  del conglomerado

- Características**
- En la segunda etapa se puede usar cualquiera de los tipos de muestreo ya estudiados.
  - No es necesario utilizar todas las unidades elementales de los conglomerados seleccionados en la primera etapa.
  - No es necesario un marco de unidades elementales completo.
  - La precisión es menor, ya que aparecen fuentes de variación que complican los cálculos algebraicos.

**Un estimador insesgado de la media aritmética será**

$$\bar{y} = \frac{1}{n\bar{m}} \sum_{i=1}^n \sum_{j=1}^{\bar{m}} Y_{ij} = \frac{1}{n} \sum_{i=1}^n \bar{y}_i$$

**Para el total poblacional, proporción y total de clase, los estimadores insesgados son**

$$\hat{Y} = N\bar{M}_y = \frac{N\bar{M}}{n} \sum_{i=1}^n \bar{y}_i \quad \hat{p} = \frac{1}{n} \sum_{i=1}^n p_i \quad \hat{t} = N\bar{M}P = \frac{N\bar{M}}{n} \sum_{i=1}^n \hat{p}_i$$

# Modalidades del muestreo por conglomerados.

## Muestreo en una etapa

- ✓ La población se divide en grupos
- ✓ Se selecciona aleatoriamente alguna de ellas para conformar la muestra

## Muestreo bietápico

Se debe realizar la etapa anterior

## Muestreo polietápico

Se deben cumplir las etapas anteriores



## Muestreo por áreas o geográfico



Se aplica cuando **no se dispone de una población objetivo completa**, y además existe mucha dispersión entre los integrantes de la misma. El muestreo por áreas es aplicable a poblaciones nacionales o regionales que poseen características físicas, políticas o naturales similares.

**Su aplicación surge por la ausencia de listas completas u ordenadas**

*Se delimitan áreas a partir de mapas:* Regiones, municipios, poblados o caseríos

No aplicable a áreas pequeñas, en cuyo caso se recomienda **el censo**

**Para realizar un muestreo por áreas o geográfico se procede de la siguiente manera**

1. Se divide el área total en varias áreas para ser muestreadas.
2. Cada división es subdividida y numerada en áreas mas pequeñas.
3. De ser necesario se continua subdividiendo.
4. Este procedimiento da origen a un muestreo por fase o etapas.

**Ejemplo:** Un estudio de mercado pretende determinar las costumbres y hábitos de diversión de ciudadanos hindúes que residen en Venezuela. Puesto que llegar a este tipo de personas es bastante difícil, y además se tiene como limitante las distancias geográficas, en este caso conviene realizar un muestreo por áreas o geográfico

## Muestreo por Racimos



En algunas investigaciones se dan casos en las que el investigador se le presenta limitaciones e inconvenientes debidas a factores económicos, al factor tiempo, a las distancias geográficas o simplemente a una combinación de éstos y otros obstáculos.

Si una población se va a muestrear por racimos, trae como consecuencia establecer las diferencias entre lo que es la unidad de análisis y lo que es la unidad muestral.

### Unidad de análisis

Será quien va a ser medido. Por ejemplo, individuo o individuos a quienes se les aplicara el instrumento de recolección de información.

### Unidad muestral

Esta representada en el racimo a través del cual se conseguirá el acceso a la unidad de análisis.

Un muestreo por racimos deriva en una selección de unidades en dos etapas, donde en cada una de ellas se obtienen mediante procedimientos probabilísticos

**Primera etapa** Se procede a seleccionar los racimos, de acuerdo a los pasos estipulados para obtener una muestra probabilística simple o estratificada .

**Segunda etapa** Se consideran los racimos a su interior para poder seleccionara los individuos o elementos que se van a medir.

**Ejemplo:** Cierta empresa procesadora de productos lácteos aborda una investigación sobre la calidad de alimentación que se le proporciona al ganado vacuno. En este caso favorece un muestreo por racimos, donde las unidades de análisis serán los encargados de alimentar a dicho ganado y los posibles racimos serán las fincas y parcelas.

## Comparación entre muestreo probabilístico y no probabilístico

muestreo probabilístico	muestreo no probabilístico
Se considera objetivo, científico y representativo	Se considera simple y rápido
Resulta mas costoso	Resulta de bajo costo
Depende de postulados y teorías matemáticas	Depende del juicio del investigador
Su fundamento es la estadística matemática	No tiene fundamento teórico que lo sustente

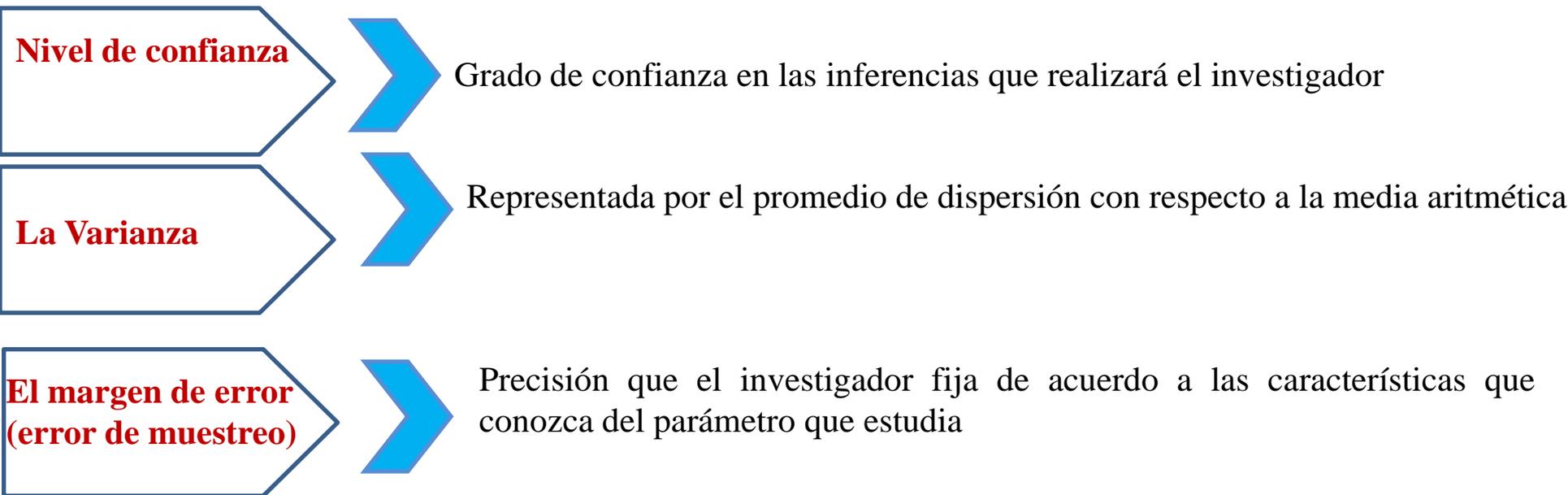
## Tamaño de la muestra

Utilizando los datos que aporta una muestra de la población se pueden obtener conclusiones o se puede inferir sobre las características poblacionales.

La muestra debe ser representativa

Solo se trabaja con toda la población cuando se quiere realizar un censo o que la población sea muy pequeña y manejable..

*Para determinar el tamaño de una muestra que se utilizara para los respectivos análisis es necesario considerar aspectos como:*



Cuando el parámetro que se quiere analizar de la población es

**Naturaleza cuantitativa** Media aritmética      **Naturaleza cualitativa** proporción

## Formas de calcular el tamaño de la muestra

### Tamaño de la muestra para la media aritmética de poblaciones infinitas

$$n = \left( \frac{Z\sigma}{e} \right)^2$$

$Z$ : Valor tipificado

$\sigma$ : Desviación típica

$e$ : Error de muestreo

### Tamaño de la muestra para la media aritmética de poblaciones finitas

$$n = \frac{N}{1 + Ne^2}$$

$N$ : Tamaño de la población

$e$ : Error de muestreo

### Tamaño de la muestra para proporciones de poblaciones infinitas

$$n = \frac{Z^2 PQ}{e^2}$$

$Z$ : Valor tipificado

$P$ : Proporción de poblaciones

$e$ : Error de muestreo

$$Q = 1 - P$$

### Tamaño de la muestra para proporciones de poblaciones finitas

$$n = \frac{Z^2 PQN}{e^2(N-1) + Z^2 PQ}$$

$Z$ : Valor tipificado

$P$ : Proporción de poblaciones

$e$ : Error de muestreo

$$Q = 1 - P$$

$N$ : Tamaño de la población

**Existen algunas tablas con valores específicos para determinar el tamaño de la muestra.**

## Tabla1

TABLA PARA LA DETERMINACIÓN DE UNA MUESTRA SACADA DE UNA POBLACIÓN FINITA, PARA MÁRGENES DE ERROR DEL 1, 2, 3, 4 Y 5 POR 100, EN LA HIPÓTESIS DE  $p = 50\%$

Margen de confianza del 99,7 %

AMPLITUD DE LA POBLACIÓN	Amplitud de la muestra para márgenes de error abajo indicados				
	±1 %	±2 %	±3 %	±4 %	±5 %
500					
1.000					474
1.500				726	563
2.000				826	621
2.500				900	662
3.000			1.364	958	692
3.500			1.458	1.003	716
4.000			1.539	1.041	735
4.500			1.607	1.071	750
5.000			1.667	1.098	763
6.000		2.093	1.765	1.139	783
7.000		3.119	1.842	1.171	798
8.000		3.303	1.905	1.196	809
9.000		3.462	1.957	1.216	818
10.000		3.600	2.000	1.233	826
15.000		4.091	2.143	1.286	849
20.000		4.390	2.222	1.314	861
25.000	11.842	4.592	2.273	1.331	869
50.000	15.517	5.056	2.381	1.368	884
100.000	18.367	5.325	2.439	1.387	892
∞	22.500	5.625	2.500	1.406	900

## Tabla2

TABLA PARA LA DETERMINACIÓN DE UNA MUESTRA SACADA DE UNA POBLACIÓN FINITA, PARA MÁRGENES DE ERROR DEL 1, 2, 3, 4 Y 5 POR 100, EN LA HIPÓTESIS DE  $p = 50\%$

Margen de confianza del 95,5 %

AMPLITUD DE LA POBLACIÓN	Amplitud de la muestra para márgenes de error abajo indicados					de error	
	±1 %	±2 %	±3 %	±4 %	±5 %	±10 %	
500						222	83
1.000					385	286	91
1.500				638	441	316	94
2.000				714	476	333	95
2.500		1.250		769	500	345	96
3.000		1.364		811	517	353	97
3.500		1.458		843	530	359	97
4.000		1.538		870	541	364	98
4.500		1.607		891	549	367	98
5.000		1.667		909	556	370	98
6.000		1.765		938	566	375	98
7.000		1.842		949	574	378	99
8.000		1.905		976	580	381	99
9.000		1.957		989	584	383	99
10.000		2.000		1.000	588	385	99
15.000		2.143		1.034	600	390	99
20.000		2.222		1.053	606	392	100
25.000	7.143	2.273		1.064	610	394	100
50.000	8.333	2.381	1.087		617	397	100
100.000	9.091	2.439	1.099		621	398	100
∞	10.000	2.500	1.111		625	400	100

**Tablas** con valores específicos para determinar el tamaño de una muestra.

**Colton Arkin**

De una población de **5000** de una gran corporación se desea tomar una muestra para estudiar su funcionamiento. Con un error de 3%. ¿Cuál debe ser el tamaño de la muestra?

$$n = \frac{N}{1 + Ne^2} \qquad n = \frac{5000}{1 + 5000 \times 0.03^2} \Rightarrow n = 909$$

El tamaño de la muestra debe ser de 909 sucursales

Cuántos sujetos debe tener una muestra, si se desea estimar la proporción de estudiantes que tienen una opinión determinada sobre el servicio de comedor, a nivel de confianza 99.74% y un error de estimación del 5%, si se sabe que:  $P=25\%$

$$n = \frac{Z^2 PQ}{e^2}$$

$$Z : 3$$

$$P : 0.25$$

$$e : 0.05$$

$$Q = 1 - P = 0.75$$

$$n = \frac{3^2 \cdot 0.25 \times 0.75}{0.05^2} \Rightarrow n = 675$$

La muestra debe tener 675 estudiantes

En un poblado de 60000 habitantes, se desea calcular el número de personas necesarias para realizar una investigación sobre la opinión de crear un centro para la educación. ¿ de que tamaño debe ser la muestra con una confianza de 95.5% y un error de 3%. Estudios anteriores estiman que el 50% por ciento favorece esta idea

$$n = \frac{Z^2 PQN}{e^2(N-1) + Z^2 PQ}$$

- Z : 2
- P : 0.5
- e : 0.03
- Q = 1 - P = 0.5
- N : 60000

$$n = \frac{2^2 \times 0.5 \times .05 \times 60000}{0.03^2(60000 - 1) + 2^2 0.5 \times 0.5} \quad n = 1.091$$

Se necesitan 1091 personas para la muestra. Coincide con el valor reportado por la Tabla 2 (rectángulo azul).

# Características deseables en una muestra

**Representativas**

**Aleatoria**

**Confiabilidad**

**Óptima**

## **Muestras optimas**

**Eficiente**

**Suficiente**

**Segura**

**Flexible**