



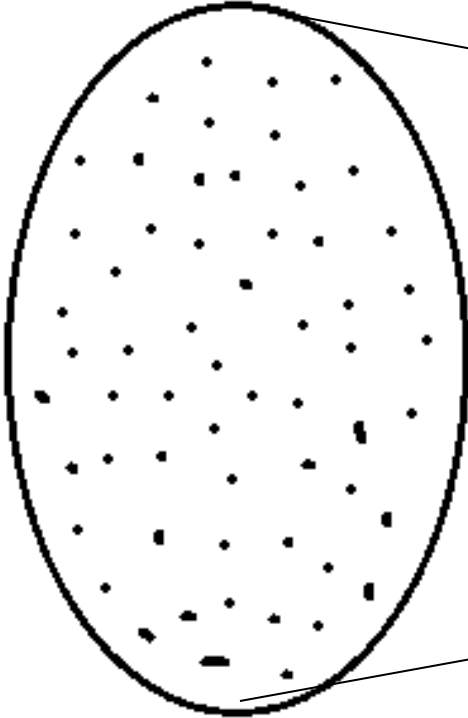
# SEMINARIO DE MUESTREO

Prof. Angel Zambrano

Mayo 2010

Universidad de Los Andes-Escuela de Estadística

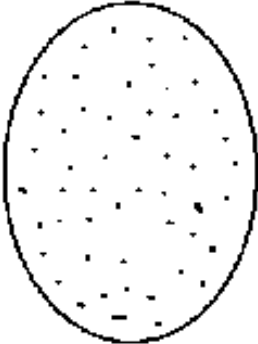
**Población**



**Parámetros**

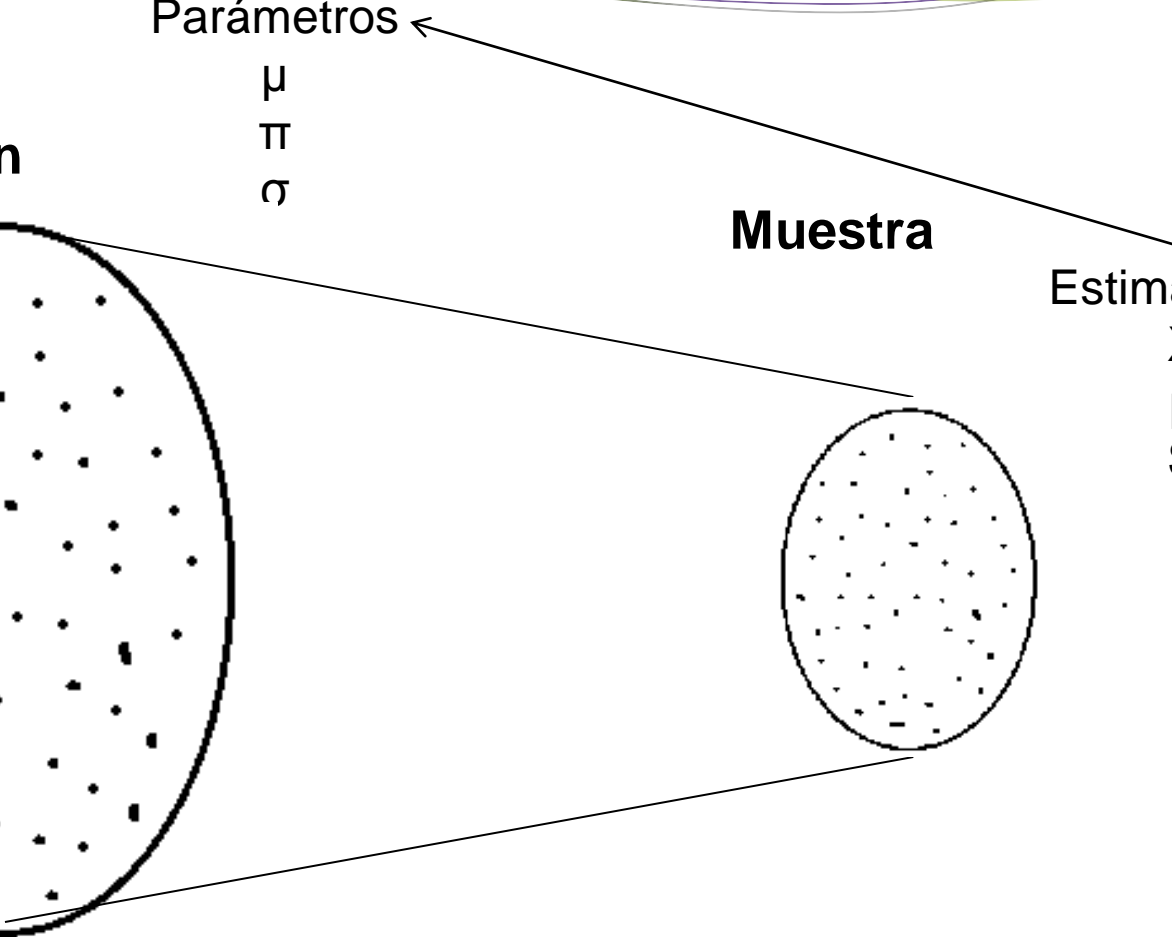
$\mu$   
 $\pi$   
 $\sigma$

**Muestra**



**Estimadores**

X  
p  
S



## **Muestreo:**

Es una metodología que apoyándose en la teoría estadística y de acuerdo a las características del estudio, indica como seleccionar y medir una parte de los elementos de la población (muestra) para hacer inferencia válida sobre el comportamiento global de la población.

## **Población:**

Es una colección de objetos acerca de los cuales deseamos hacer alguna inferencia. Un conjunto finito o infinito de elementos.

## **Elemento o unidad elemental de observación:**

Objeto sobre el cual se realizan las mediciones.

## **Unidad de muestreo:**

Elementos o conjuntos no solapados de elementos los cuales cubren totalmente la Población. Se utilizan para seleccionar la muestra. Pueden ser unidades simples (personas) o complejas (familias, organizaciones).

## **Marco muestral:**

Lista o registro de las unidades de muestreo. Base sobre la cual deben diseñarse los procesos de selección.

## **Muestra:**

Colección de unidades de muestreo seleccionadas de uno o varios marcos. Es un subconjunto de la población obtenido mediante algún procedimiento de selección (Método de Muestreo)

## **Población objetivo:**

Aquella que se pretende cubrir. Definida por los objetivos de la encuesta. Es la colección completa de observaciones que deseamos estudiar.

## **Población muestreada:**

Aquella de la cual se extrae la muestra, es más restringida. Los resultados que se obtengan serán aplicados a la población muestreada.

## **Representatividad de la muestra:**

Grado en el cual la muestra reproduce las características de la población.

## **Métodos de Muestreo:**

Conjunto de técnicas estadísticas que estudian la forma de seleccionar una muestra lo suficientemente representativa de una población. Puede ser: Probabilístico y No Probabilístico

### **Muestreo Probabilístico:**

Procedimiento que da a cada elemento de la población una probabilidad de ser seleccionado diferente de cero.

### **Muestreo no Probabilístico:**

No involucra ningún elemento probabilístico en el procedimiento de selección. La probabilidad de que un elemento sea seleccionado NO puede estimarse con exactitud. Casual, Por conveniencia, Por cuotas o Intencional

## **Muestreo Aleatorio:**

Consiste en asignar a cada elemento poblacional una probabilidad no nula, de ser seleccionado. Lo que permiten hacer estimaciones de las magnitudes de los errores de muestreo y controlar la precisión de las estimaciones muestrales dentro de ciertos límites fijados con anticipación y con cierto grado de confianza.

## Razones para usar el muestreo:

1. Costo reducido: Una muestra ahorra dinero, pues se estudia una fracción de la población.
2. Mayor rapidez: el muestreo ahorra tiempo.
3. Mayor exactitud: Una muestra da mejores resultados que un censo.
4. Pruebas destructivas
5. Poblaciones infinitas.
6. Más posibilidades: personal calificado, equipo especializado, etc.



## **El muestreo tiene dos aspectos fundamentales:**

- 1) La planificación de la investigación o plan de muestreo
- 2) Diseño muestral

La justa armonía de ambos: un buen plan y un buen diseño muestral garantizan el éxito de la operación

## **El plan del muestreo:**

- 1) Planeamiento de la Investigación
- 2) Elaboración de instrumentos básicos
- 3) Diseño de la muestra
- 4) Organización y ejecución de las operaciones de campo
- 5) Procesamiento de la información
- 6) Análisis de los resultados

## **Diseño muestral:**

- 1) Forma de selección
- 2) Estimadores y propiedades
- 3) Tamaño de la muestra
- 4) Modificaciones del diseño original

## **Características de un buen diseño**

- **Cumplimiento de objetivos**

El diseño muestral (como el plan de la investigación) debe orientarse a cumplir la finalidad y los objetivos declarados. Estas consideraciones deben determinar la definición de la población, de la unidad muestral, del proceso de selección y de medición.

- **Medibilidad**

Esta característica de los diseños permite calcular, a partir de la propia muestra, estimaciones válidas y aproximaciones de su variabilidad en el muestreo. Esta es la base de la inferencia estadística, y sirve como puente científico y objetivo, entre el resultado de la muestra y el valor desconocido de la población. Las muestras aleatorias replicadas son medibles.

## Características de un buen diseño

- **Practicidad**

Se refiere a los problemas que deben resolverse para llevar a cabo el diseño muestral como se propuso. Es decir traducir el modelo teórico a un conjunto de instrucciones que sean: simples, claras, prácticas y completas.

- **Economía (eficiencia)**

Se refiere a cumplir los objetivos de la encuesta con un costo (dinero, tiempo y esfuerzo) mínimo. Los objetivos suelen expresarse en términos de la precisión y de la exactitud.

La precisión es el número de varianzas de las estimaciones

La exactitud es el inverso del error total, incluye el sesgo y la varianza

# Taxonomía de los métodos probabilísticos de selección (5 alternativas de combinación)

I.	<b>Igual probabilidad</b> para todos los elementos	<b>Probabilidades desiguales</b> para elementos diferentes; compensadas generalmente con ponderaciones inversas.
II.	<b>Muestreo de elementos:</b> en una sola etapa, la unidad de muestreo contiene solamente un elemento	<b>Muestreo de conglomerados:</b> las unidades de muestreo son conglomerados de elementos, en una o varias etapas
III.	<b>Selección no estratificada:</b> las unidades de muestreo son seleccionadas de la población completa	<b>Muestreo estratificado:</b> selecciones separadas de particiones, o estratos de la población
IV.	<b>Selección aleatoria</b> de unidades de muestreo del estrato completo o de la población	<b>Selección sistemática</b> de unidades de muestreo con un intervalo de selección aplicado al marco
V.	<b>Muestreo en una fase:</b> la muestra final seleccionada directamente de la población completa	<b>Muestreo de dos fases:</b> la muestra final seleccionada de una muestra de primera fase, que obtiene información para estratificación o para estimación

## **Muestreo Aleatorio Simple.**

Consiste en seleccionar una muestra de tamaño “n” de una población de tamaño “N” de manera que cada una de las muestras tenga la misma probabilidad de ser seleccionada.

M.A.S. sin Reemplazo. Muestras con la misma probabilidad de ser seleccionadas. Cada elemento puede ser seleccionado solo una vez.

M.A.S. con Reemplazo. Muestras con la misma probabilidad de ser seleccionadas. Cada elemento puede ser seleccionado más de una vez.

## **Muestreo Sistemático Aleatorio:**

Los elementos de la muestra son seleccionados a intervalos fijos. Se reparte más uniformemente en la población. *Es más preciso que el muestreo aleatorio simple.*

## **Muestreo Estratificado Aleatorio Simple:**

La población se divide en estratos (subpoblaciones). Luego se toma una muestra aleatoria simple de cada estrato. Los estratos deben ser mutuamente excluyentes y exhaustivos. Los estratos deben ser internamente homogéneos



## **Muestreo por Conglomerados:**

La población se divide en grupos y luego se elige aleatoriamente una muestra de conglomerados. Los conglomerados son grupos de elementos cercanos desde algún punto de vista (Administrativa o físicamente).

## **Muestreo Polietápico:**

Diseño muestral que consta de varias etapas en las cuales se usa un diseño simple (diseños que comprenden una sola etapa de clasificación, por ejemplo; muestreo aleatorio simple, etc.).

## **¿Cómo seleccionar los elementos de la muestra?**

1. Usar un Bingo
2. Usar una caja o bolsa con elementos enumerados
3. Tabla de números aleatorios
4. Números pseudo-aleatorios
5. Generadores de números aleatorios
6. Posibilidades de los paquetes estadísticos



**En una investigación por muestreo se producen dos tipos de errores:**

- Errores de muestreo
- Errores ajenos al muestreo

## **Error de muestreo $e$ :**

se define como error de muestreo o error de estimación a la diferencia absoluta entre el valor del parámetro desconocido que se desea estimar  $\theta$  (la característica poblacional) y el valor observado del estimador  $\hat{\theta}$ , es decir,  $e = |\theta - \hat{\theta}|$

El tamaño de este error es fijado de antemano por el muestrista de acuerdo a criterios de precisión y costo. El error de estimación decrece al aumentar el tamaño de la muestra y desaparece si se observa toda la población, pues en este caso,  $\hat{\theta} \equiv \theta$

## **Errores ajenos al muestreo:**

Este tipo de error se debe a factores muy diversos inherentes a las diversas operaciones que se realizan en el proceso de la encuesta o los instrumentos, los mas frecuentes son:

- a) Errores de cobertura, marco muestral defectuoso o sesgo en la selección.
- b) Errores en los instrumentos de recolección:
  - variables mal definidas
  - preguntas mal formuladas
  - inadecuado orden en el cuestionario
  - grado de dificultad en las preguntas
  - encuestadores mal entrenados y capacitados
- c) No respuesta (total o parcial)
- d) Errores en el procesamiento (codificación, transcripción y calculo)

## ¿Cómo determinar n?

1. Fijar el grado de precisión deseado en términos del error de estimación de acuerdo a los objetivos de la encuesta.
2. Fijar la probabilidad de o grado de confianza.
3. Encontrar una ecuación que relacione a n con la precisión deseada
4. Pre-estimar los valores de los parámetros de la ecuación que contiene a n.
5. Si se desean resultados válidos por dominio debe fijarse el error para cada uno y calcular el tamaño de muestra como suma de los tamaños de los dominios.
6. Cuando se miden varios atributos y se precisa el error para cada uno de ellos, esto origina varios tamaños de n, hay que reconciliarlos en uno sólo.
7. Buscar un compromiso entre la precisión (tamaño de n) y el costo.

# Tamaño de muestra para estimar P la proporción poblacional

Sea:

$e$  el margen máximo de error admisible en la estimación de P.

$\alpha$  el riesgo de que el error sea mayor a  $e$

$N$  el tamaño de la población y

$\sigma_p$  la varianza de  $p$  (o su estimación)

$$\Pr(|p - P| \geq e) = \alpha$$

conocemos que: 
$$\sigma_p^2 = \frac{N - n}{N - 1} \frac{PQ}{n}$$

la formula que liga a **n** con la precisión deseada

$$e = t_{\alpha/2} \sqrt{\frac{N-n}{N-1} \frac{PQ}{n}} \quad \text{donde } t_{\alpha/2} \text{ absisa de la normal.}$$

al resolver para **n**

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

o equivalente

$$n = \frac{\frac{t^2 PQ}{e^2}}{1 + \frac{1}{N} \left( \frac{t^2 PQ}{e^2} - 1 \right)} = \frac{n_o}{1 + \frac{1}{N} (n_o - 1)}$$

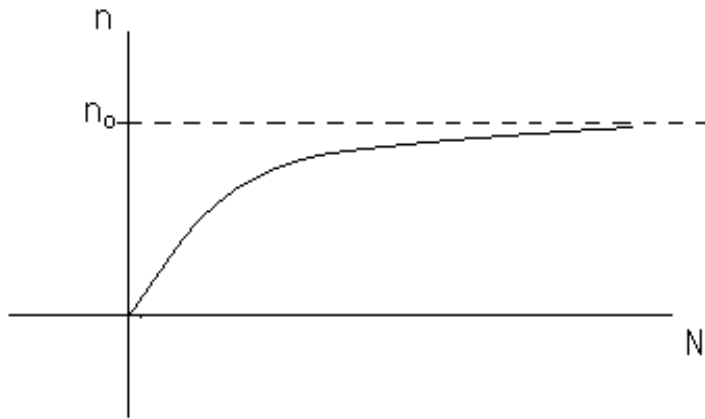
Como **P** es desconocida se sustituye por un estimación anticipada o se usa el criterio de máxima varianza con  $P=0.5$



## Comportamiento de N:

Es de interés ver como es el comportamiento del tamaño de la muestra  $n$  en función de la población  $N$ . La función:

$n = \frac{n_0}{1 + \frac{n_0}{N}}$  representa una hipérbola equilátera que pasa por el origen y tiene una asíntota en  $n_0$  paralela al eje de las abscisas



Este resultado es importante pues nos dice que la misma precisión puede dar una muestra de tamaño  $n$  para una población de por ejemplo  $N_1=5000$  elementos que para otro de  $N_2=100000$ . (siempre que verifique que  $N_1 > n_0(n_0-1)$ )

Es de interés ver que  $n$  es inversamente proporcional al cuadrado del error absoluto y por lo tanto para, por ejemplo aumentar la precisión en el doble (error en la mitad) haría necesario aumentar la muestra en 4 veces.

# Bibliografía

1. Cochran, W. (1980) **Técnicas de Muestreo**. Editorial CECSA. España.
2. Lohr, Sharon. (2000) **Muestreo: Diseño y Análisis**. International Thomson Editores, México.
3. Pérez, César. (2000) **Muestreo con aplicaciones informáticas**. Editorial Alfaomega, Madrid, España.
4. Scheaffe, R., Mendenhall, W., y Ott, L. (1991) **Elementos de Muestreo**. Duxbury Press, Boston.