

Población de estudio y muestra

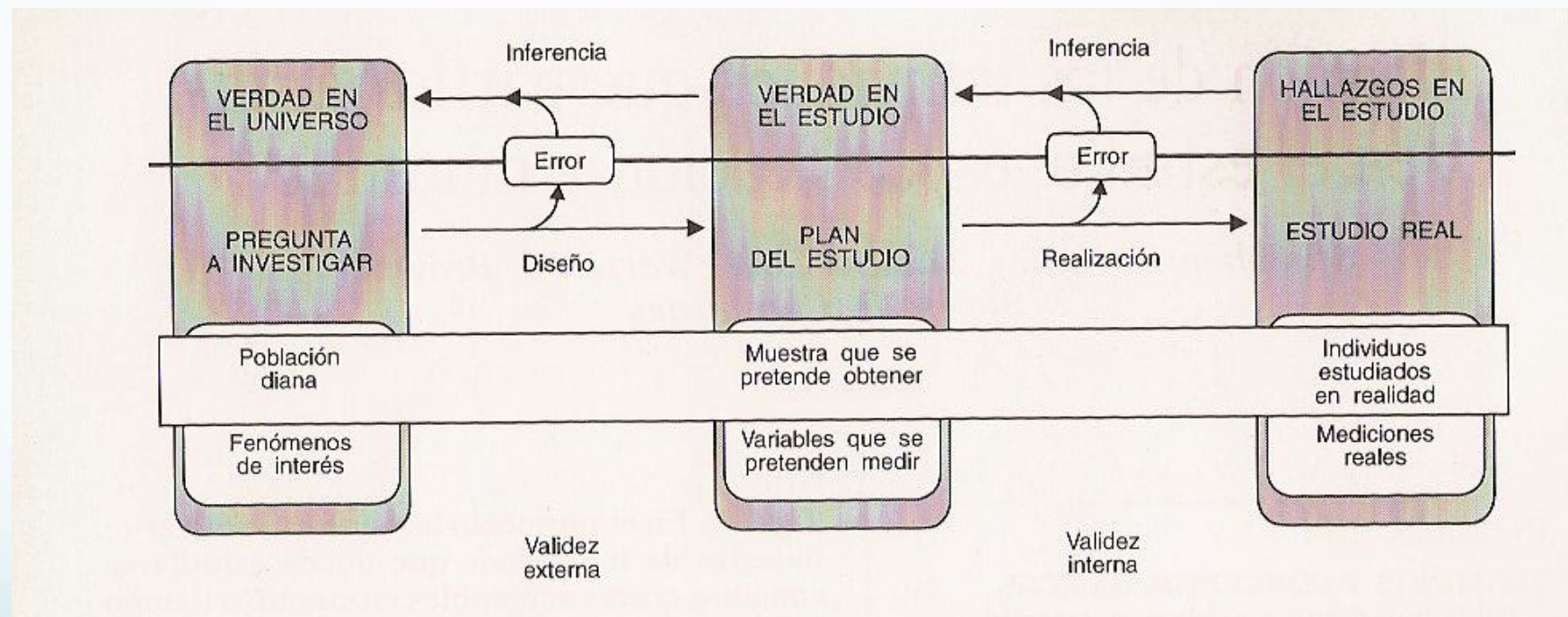
Curso de Metodología de la Investigación
Unidad Docente de MFyC

Patricio Suárez Gil
La Fresneda (Asturias), 2011

Población de estudio

- Conjunto de “individuos” al que se refiere nuestra pregunta de estudio o respecto al cual se pretende concluir algo.
- Las preguntas de estudio nunca remiten a “una muestra”
- Población diana y población accesible

- ***Población diana***: definida por características clínicas, epidemiológicas o demográficas.
- ***Población accesible***: subconjunto de la población diana delimitada geográfica y/o temporalmente disponible para el estudio.
- ***Muestra.***



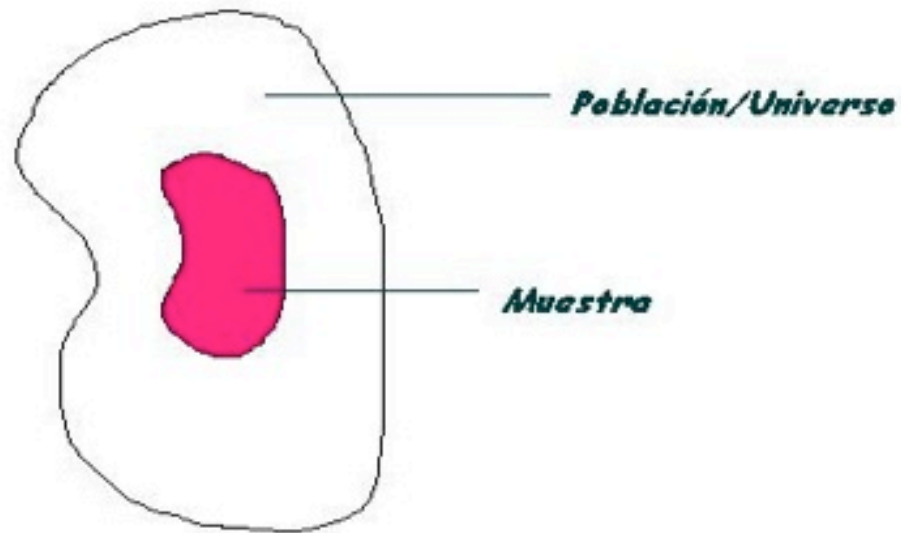
Definición población de estudio

- Definición operativa: geográfica y temporal.
- Fuente de información
- Criterios de exclusión y/o inclusión
- Definición de pérdidas

¿Qué es muestreo?

- Herramienta de la investigación científica (otros usos)
- Determinar qué parte de una realidad debe examinarse para hacer inferencias sobre el todo del que procede
- Siempre lleva asociado un error (ERROR DE MUESTREO)

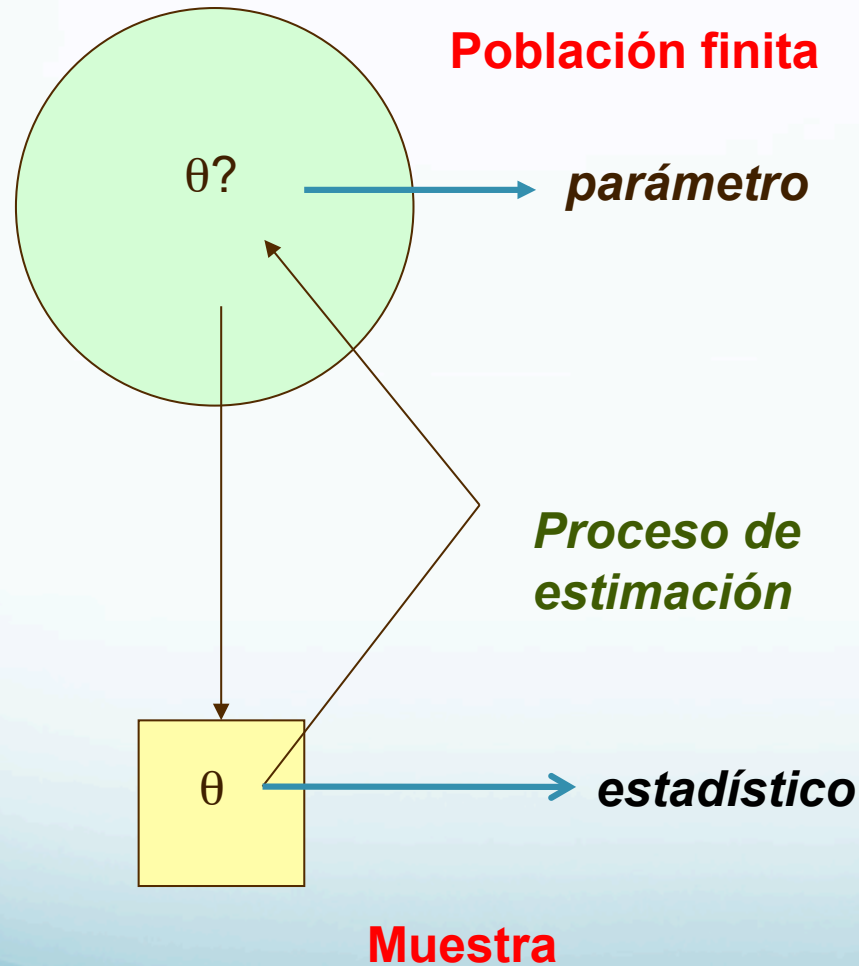
(esquema de población y muestra)



¿Cuándo realizar muestreo?

- Es un problema “económico”
- Muchas investigaciones conllevan un problema de muestreo, pero no en todas es clave
- Más importante en **estudios descriptivos** que en comparativos: importa más la **representatividad** que la **homogeneidad** (*Encuestas de Salud vs. Estudio de casos y controles*)

Formalmente: problema de muestreo de poblaciones finitas



¿Muestra representativa o probabilística?

- La noción de representatividad sólo tiene un alcance intuitivo, carece de definición formal
- Una muestra será representativa cuando exhiba internamente el mismo grado de diversidad que la población
- El azar no garantiza la representatividad ante una muestra concreta, pero...

- ...la confianza que pueda depositarse en una muestra depende de la que merezca **el método** que la produjo
- El azar es un seguro contra distorsiones sistemáticas (sesgos)

¿Qué es una muestra?

Cualquier subconjunto de una población

Procedimientos de extracción de muestras

- MUESTREOS PROBABILÍSTICOS
- MUESTREOS NO PROBABILÍSTICOS

Muestreos no probabilísticos

- *Selección sin método*
- *Criterio de experto/autoridad*
- *Muestreo semiprobabilístico*
- *Muestreo por cuotas*
- *Muestreos consecutivos*

¡No permiten calcular errores de muestreo, pero “pueden” producir muestras representativas!

Muestreo probabilístico

1. Todo elemento de la población tiene probabilidad no nula de ser elegido
2. La probabilidad de selección debe ser conocida (de antemano)
3. Si además, la probabilidad es igual para todos:
Muestreo equiprobabilístico

¡Se pueden calcular errores de muestreo e intervalos de confianza!

Tipos de muestreo

- Muestreo Simple Aleatorio (MSA)
- Muestreo Aleatorio Estratificado (MAE)
- Muestreo Sistemático (MS)

Técnicas básicas de selección

- Muestreos por conglomerados y diseños complejos (polietápicos)

Muestreo Simple Aleatorio (MSA)

- Cualquier subconjunto de tamaño n tiene la misma probabilidad de selección
- De una población finita de tamaño N se pueden obtener k muestras de tamaño n , según

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

- Cuando las k muestras tienen la misma probabilidad de ser elegidas estamos ante un MSA
- Es equiprobabilístico
- La probabilidad de selección de cada uno de los elementos es

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

Ojo: ¡No todo equiprobabilístico es MSA!

¿Cómo se hace?

- Mecanismo físico (urna)
- Tabla de números aleatorios
- Algoritmos de generación de números pseudo-aleatorios

Error e intervalo de confianza (proporción)

$$Var(p) = \frac{pq}{n-1} (1-f) \quad \text{siendo} \quad q = 1-p$$

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

$$e = z_{1-\alpha} \sqrt{Var(p)} = 1,96 \sqrt{Var(p)}$$

$$IC_{1-\alpha} = [p - e; p + e]$$

Error e intervalo de confianza (media)

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

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$e = z_{1-\alpha} \sqrt{\frac{s^2}{n} (1-f)}$$

$$IC_{1-\alpha} = [\bar{x} - e; \bar{x} + e]$$

Error absoluto vs. relativo

- *Error absoluto* es la diferencia entre la estimación y el verdadero parámetro poblacional

$$e_0 = \hat{\theta} - \theta \quad e_0 = p - \pi$$

$$e_0 = \bar{x} - \mu$$

- *Error relativo* es la razón entre la magnitud del error (absoluto) y la magnitud medida

$$e_r = \frac{e_0}{\hat{\theta}} = \frac{e_0}{p} = \frac{e_0}{\bar{x}}$$

Tamaño de muestra

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

→ *fracción de muestreo*

- Para estimar una proporción (%)

$$n_0 = \frac{z_{1-\alpha}^2 pq}{e_0^2}$$

- Para estimar una media

$$n_0 = \frac{z_{1-\alpha}^2 s^2}{e_0^2}$$

Predicción de pérdidas

- Calcular el tamaño muestral corregido según fracción esperada de no-respuesta (β)

$$n_c = \frac{n}{1 - \beta}$$

Algunos problemas...

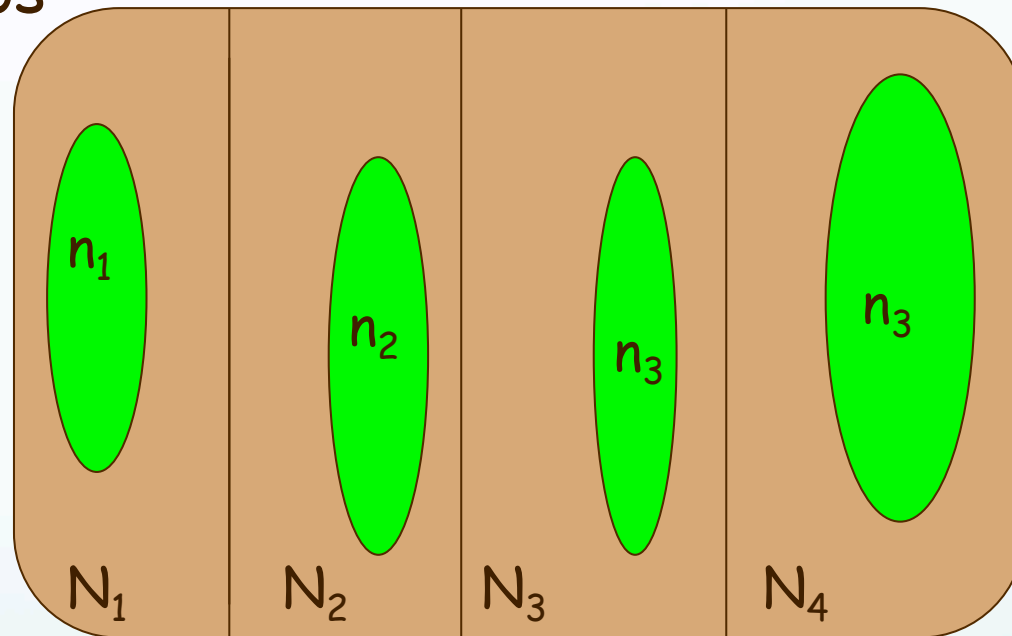
- En rigor, sólo sirve para MAS
- Hay que fijar error absoluto
- Exige parámetros a estimar
- Definir un nivel de confianza
- En la práctica se quieren estimar muchos parámetros
- ...

En resumen, el MAS

- Importancia teórica
- Poco usado como método “puro o aislado”
- Forma parte de diseños más complejos
- **¡Exige listado completo de las unidades de análisis (universo o población a muestrear)!**

Muestreo Aleatorio Estratificado

L=4 estratos



$$N_1 + N_2 + N_3 + N_4 + \dots = N \text{ (población)}$$

$$n_1 + n_2 + n_3 + n_4 + \dots = n \text{ (muestra)}$$

Muestreo Aleatorio Estratificado

- Cuando podemos identificar diferentes grupos cuya representación quiera asegurarse, podemos hacer listados separados de cada uno de ellos
- Ideal: grupos homogéneos internamente y diferentes entre sí (estratos)
- En general, cuando se quiere hacer estimaciones en los estratos
- Suele dar errores menores (mayor precisión)

Muestreo Aleatorio Estratificado

- Asignación proporcional

$$\frac{n_i}{N_i} = \frac{n}{N}$$

La fracción de muestreo es constante en todos los estratos: ¡es equiprobabilístico!

- Asignación fija
- Asignación óptima (de Neyman)

Muestreo Sistemático (MS)

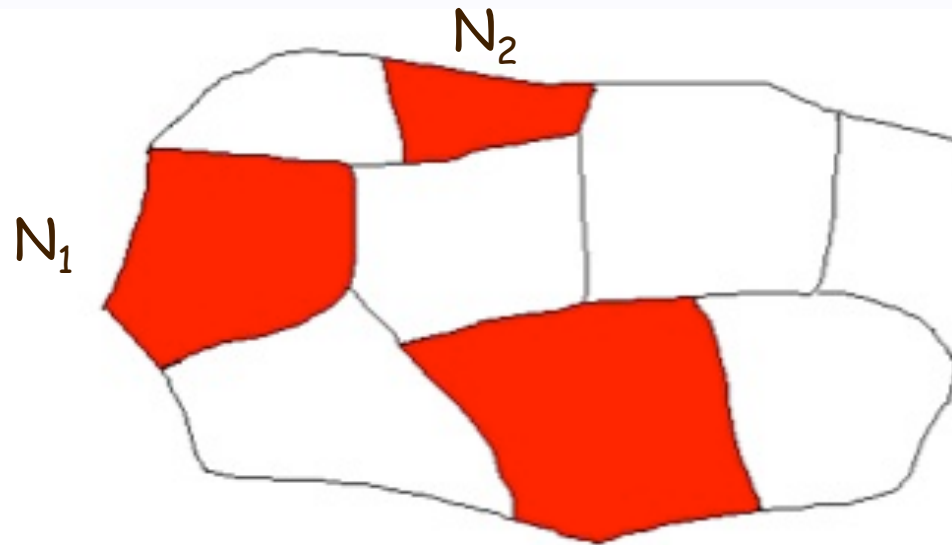
- Procedimiento equiprobabilístico que exige listado de las unidades de análisis
- Equivalente al MAS, se utilizan las mismas fórmulas de cálculo de errores e intervalos de confianza

- A partir de una población N se desea extraer una muestra de tamaño n ; se calcula $k=N/n$ (intervalo de selección sistemática) y se seleccionan los n elementos a partir de r (arranque aleatorio $0 \leq r \leq k$) tomando $r, r+k, r+2k, \dots$
- El azar sólo interviene en el arranque, pero siempre es equiprobabilístico $f=1/k$

Muestreo por conglomerados

- La población se divide en partes o conglomerados (*Unidades de Primera Etapa, UPE*); cada UPE puede ser subdividida en *Unidades de Segunda Etapa (USE)* y así sucesivamente hasta las unidades de análisis

M. Conglomerados monoetápico

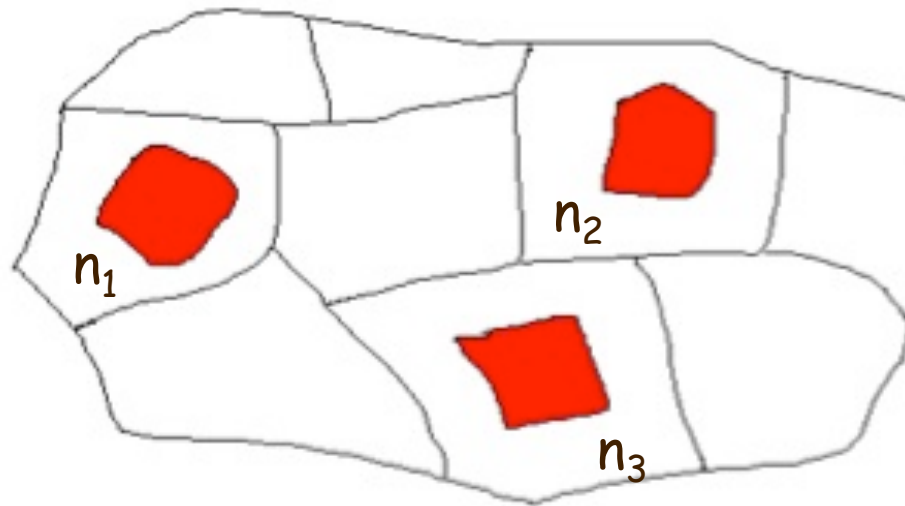


$M = 9$ conglomerados (UPE) N_3

$m = 3$

$n = N_1 + N_2 + N_3$

M. Conglomerados bietápico



$M = 9$ congl (UPE)

$M = 3$

$n = n_1 + n_2 + n_3$

Software

- **Epidat 3.1** (<http://dxsp.sergas.es>)
- *Stata*
- *R*

Bibliografía

- Silva Ayçaguer LC (1993). ***Muestreo para la investigación en Ciencias de la salud.*** Ed Díaz de Santos, Madrid.
- Silva Ayçaguer LC (2000). ***Diseño razonado de muestras y captación de datos para la investigación sanitaria.*** Ed Díaz de Santos, Madrid.
- Lwanga SK & Lemeshow S. (1991). ***Determinación del tamaño de las muestras en los estudios sanitarios. Manual práctico.*** OMS, Ginebra.