



# Metodología de investigación cuantitativa

.....

## Diseños muestrales

Pablo Geraldo Bastías

pdgerald@uc.cl

# Estructura de la presentación

1. Muestreo Aleatorio Simple (MAS)
2. Muestreo Estratificado (ST)
3. Muestreo por Conglomerados (CG)
4. Referencias

## Muestreo Aleatorio Simple (MAS)

- Constituye la **base teórica** de los demás diseños muestrales; sin embargo, no se utiliza en muestras de gran escala por ser poco eficiente.
- Para obtener una muestra a través de MAS se pueden utilizar tablas de números aleatorios, algoritmos de generación, etc.
- Cada elemento de la población tiene la misma probabilidad de ser seleccionado en la muestra (**EPSEM: equal probability of selection**), y cada combinación de elementos de tamaño  $n$  (muestra) tiene la misma probabilidad de selección.

## Fórmulas para MAS

- Media muestral:  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$
- Varianza:  $\text{var}(\bar{y}) = (1-f) \frac{s^2}{n}$
- Donde:  $s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{(n-1)}$
- Corrección para población finita:  $(1-f)$
- Fracción de muestreo:  $f = \frac{n}{N}$
- Cuando  $f$  es pequeña:  $\text{var}(\bar{y}) \approx \frac{s^2}{n}$
- Para una proporción:  $s^2 \approx p(1-p)$
- Luego, su varianza:  $\text{var}(p) = \frac{(1-f)}{(n-1)} p(1-p)$

## Ejercicio MAS

En la Universidad de Talca, sede Santiago, seleccione una muestra de  $n = 40$  de un total de 400 estudiantes. Considerando que de la muestra de 40, hay 28 que desapruban la calidad de los completos del kiosko, calcule:

- Porcentaje que desapruaba los completos
- Error estándar de la proporción
- Intervalo de confianza al 95 %

# Estructura de la presentación

1. Muestreo Aleatorio Simple (MAS)
- 2. Muestreo Estratificado (ST)**
3. Muestreo por Conglomerados (CG)
4. Referencias

## Muestreo Estratificado (ST)

- Se divide a la población en *estratos*, es decir, subconjuntos **mutuamente exclusivos y exhaustivos**, y se selecciona una muestra de forma independiente al interior de cada uno.
- En cada estrato pueden utilizarse diferentes diseños muestrales y fracciones de muestreo.
- Si se utiliza la misma fracción de muestreo en cada estrato, se conoce como **estratificación proporcional**
- En cambio, cuando las fracciones varían, se conoce como **estratificación desproporcionada**. En este escenario, se requiere realizar ajustes al analizar los datos.
- Es importante notar que la estratificación puede traer ganancias de eficiencia, siempre que la estratificación sea por variables asociadas al fenómeno en estudio.

# Fórmulas para ST

## Nivel poblacional

- Tamaño poblacional en estrato  $h$ :  $N_h$
- Tamaño poblacional:  $N = \sum N_h$
- Proporción de población en estrato  $h$ :  $W_h = \frac{N_h}{N}$

## Nivel muestral

- Muestra en estrato  $h$ :  $n_h$
- Tamaño muestral:  $n = \sum n_h$
- Ponderador muestral:  $w_h = \frac{n_h}{N_h}$



## Fórmulas para ST

- Media muestral:  $\bar{y}_{ST} = \sum W_h \bar{y}_h = \frac{1}{n} \sum \sum y_{hi}$
- Varianza de la media muestral:  $var(\bar{y})_{ST} = \sum W_h^2 (1 - f_h) \frac{s_h^2}{n_h}$
- Si hay afijación proporcional:  $f_h = \frac{n_h}{N_h} = \frac{n}{N} = f$
- Entonces:  $var(\bar{y})_{ST} = (1 - f) \sum \frac{W_h s_h^2}{n} = (1 - f) \frac{s_w^2}{n}$
- Donde:  $s_w^2 = \sum W_h s_h^2$

## Ejercicio ST

En una empresa, se selecciona una muestra estratificada de  $n = 480$  empleados de un total de  $N = 8000$  trabajadores. Se busca estimar el promedio de visitas al doctor durante el año anterior ()

- Los empleados fueron agrupados en tres estratos ( $h=1,2,3$ ), de acuerdo a su edad.
- Se seleccionó al mismo número de empleados al interior de cada estrato ( $n_1 = n_2 = n_3$ ), utilizando MAS al interior de cada estrato.

## Ejercicio ST

El siguiente esquema sintetiza el muestreo:

Estrato	$N_h$	$W_h$	$n_h$	$f_h$	$y_h$	$\bar{y}_h$	$s_h^2$
Menores de 30 años	3.200	0,4	160	0,05	960	6	5
Entre 30-50 años	4.000	0,5	160	0,04	800	5	4
Mayores de 50 años	800	0,1	160	0,20	1.280	8	7
Total	8.000	1,0	480		3.040	$\bar{y}$	

## Ejercicio ST

- La expresión  $\bar{y} = \frac{\sum y_{hi}}{\sum n_{hi}} = \frac{3,040}{480} = 6,33$  es un **estimador sesgado** del promedio de visitas de los empleados. ¿Por qué?
- Estime insesgadamente el promedio de visitas al doctor del año anterior
- Estime la varianza del promedio de visitas
- Estime el error estándar del promedio de visitas

# Estructura de la presentación

1. Muestreo Aleatorio Simple (MAS)
2. Muestreo Estratificado (ST)
3. Muestreo por Conglomerados (CG)
4. Referencias

## Muestreo por Conglomerados (CG)

- Los conglomerados son utilizados cuando no se dispone de un marco muestral que identifique directamente a los miembros de la población.
- Los elementos de la población pueden ser identificados sólo a través de ciertos grupos, y no directamente (direcciones y no personas, carreras y no de titulados)
- Es importante tener en cuenta que hacer una lista de grupos es menos costoso que hacer una lista de elementos.
- Asimismo, incluso teniendo una lista de elementos puede ser muy costoso muestrear aleatoriamente a partir de ellos (dispersión geográfica, por ejemplo)
- Una vez que se visita a un elemento en terreno, resulta costo-efectivo entrevistar a otros del mismo conglomerado.

## Muestreo por Conglomerados (CG)

El tamaño del conglomerado depende de distintos factores:

- En aplicaciones industriales suelen tener el mismo tamaño (packs, resmas, docenas, etc.).
- En poblaciones naturales, los conglomerados suelen tener tamaños variables (colmenas, jaurías, aldeas, etc.)

Se distinguen dos tipos de muestreo por conglomerados:

- **Monoetápico (CG1E):** Al seleccionar un conglomerado, se incluyen todos sus elementos en la muestra.
- **Multietápico (CG2E):** Al seleccionar un conglomerado, se incluye un subconjunto de sus elementos en la muestra.

## Notación muestreo por Conglomerados (CG)

Conglomerados:

- $A$ : número de conglomerados en la población
- $\alpha$ : número de conglomerados en la muestra

Elementos:

- $B_\alpha$ : número de elementos del conglomerado
- $b_\alpha$ : número de elementos del conglomerado en la muestra

Pribabilidades de selección:

- $P_\alpha$ : probabilidad de que un conglomerado sea seleccionado
- $P_\beta$ : probabilidad de que un elemento sea seleccionado en el conglomerado
- $f = P_\alpha P_\beta$ : probabilidad de que un elemento sea seleccionado en la muestra



## Notación muestreo por Conglomerados

Considere un muestreo por conglomerados en una población con 30 elementos constituida por 3 conglomerados de 10 unidades cada uno, de los cuales se seleccionan dos.

Ejemplo en muestreo monoetápico (CG1E), y conglomerados de igual tamaño:

- $B_{\alpha} = 10, b_{\alpha} = 10$
- $P_{\alpha} = \frac{\alpha}{A} = \frac{2}{3}$
- $P_{\beta} = \frac{B}{B} = \frac{10}{10}$
- $f = P_{\alpha}P_{\beta} = \frac{2}{3}$

## Notación en muestreo por Conglomerados

Ejemplo en muestreo bietápico (CG2E), y conglomerados de igual tamaño:

- $B_\alpha = 10, b_\alpha = 5$
- $P_\alpha = \frac{\alpha}{A} = \frac{2}{3}$
- $P_\beta = \frac{b}{B} = \frac{5}{10}$
- $f = P_\alpha P_\beta = \frac{2}{3} \times \frac{5}{10} = \frac{1}{3}$

# Estructura de la presentación

1. Muestreo Aleatorio Simple (MAS)
2. Muestreo Estratificado (ST)
3. Muestreo por Conglomerados (CG)
4. Referencias

## Referencias

Las sesiones sobre métodos de encuestas son una adaptación de las clases de Carolina Casas-Cordero:

- Casas-Cordero, C. (2013). Encuestas para al investigación social. Apuntes de Clase, Magister en Sociología UC.

Una exposición más detallada puede encontrarse en el libro Groves et al. (2009). Survey Methodology. Wiley, 2da edición.