

# Metodología de investigación cuantitativa

.....

## Población, muestra y diseño

Pablo Geraldo Bastías

pdgerald@uc.cl

# Estructura de la presentación

1. Población y marco muestral
2. Problemas de los marcos y error de cobertura
3. Distribución muestral
4. Referencias

# Población

La **población objetivo** corresponde al conjunto de elementos (unidades, personas, etc.) sobre el cual el investigador quiere realizar inferencias.

La población debe encontrarse bien definida a partir de sus características:

- Tipo de unidades (personas, hogares, viviendas, empresas, etc.)
- Restricciones de fecha/tiempo (ej.: prueba PISA)
- Límites geográficos (ej.: zonas extremas)
- Límites de edad

Ejemplo: "Estudiantes chilenos de 15 años cumplidos al 15 de octubre".

## Población a encuestar

La **población a encuestar** es aquella de la cual efectivamente se recolectarán los datos.

Se definen posibles exclusiones:

- Población institucionalizada (ej.: cárceles, SENAME, etc.)
- Población sin residencia habitual
- Áreas remotas o aisladas

Ejemplo: ".Estudiantes chilenos de 15 años cumplidos al 15 de octubre, que asisten a centros educacionales urbanos".

## Marco muestral

El **marco muestral** corresponde a las listas o procedimientos que se utilizan para identificar a la población objetivo (o a encuestar).

Por ejemplo:

- Marco de áreas geográficas (como en el INE: mapas y cartografía de Censos de Población y Vivienda)
- Marco de listas (directorios, miembros de una organización, registro electoral, etc.)

# Estructura de la presentación

1. Población y marco muestral
2. Problemas de los marcos y error de cobertura
3. Distribución muestral
4. Referencias

## Problemas de los marcos muestrales

1. **Cobertura:** ausencia de elementos elegibles en el marco
2. **Elementos no elegibles:** presencia de elementos no elegibles en el marco
3. **Duplicación:** un único elemento de la población objetivo se asocia a múltiples elementos del marco muestral
4. **Conglomeración:** múltiples elementos de la población objetivo es representado por el mismo elemento en el marco muestral

## Error de cobertura

Es la manera en que afectan los problemas previamente mencionados (sub-cobertura, duplicación, conglomeración, etc.) a los estimadores obtenidos a partir de una encuesta.

Es necesario recordar que el error de cobertura es una propiedad de cada estimador y varía por subgrupos (no es una propiedad de la encuesta).

Existen diversas maneras de reducir el error de cobertura: intervalo semi-abierto (en muestreo de áreas geográficas), utilizar marcos múltiples/suplementarios, utilizar marcos más amplios, etc.



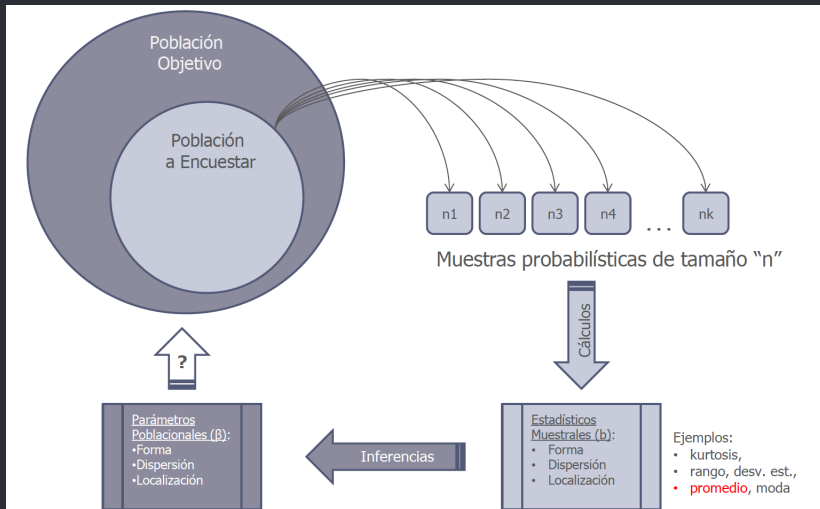
# Estructura de la presentación

1. Población y marco muestral
2. Problemas de los marcos y error de cobertura
3. Distribución muestral
4. Referencias

## Distribución muestral

- **Paradigma del muestreo repetido**: si la selección de la muestra es probabilística, pueden seleccionarse muchas muestras distintas.
- Se llama **realizaciones** a cada set de elementos del mismo tamaño ( $n$ ) que puede seleccionarse bajo un mismo diseño muestral.
- Hipotéticamente, podemos calcular un estadístico (digamos, la media  $\bar{y}$ ) a partir de cada realización.
- La **distribución muestral** de un estadístico (la media, por ejemplo) es la distribución de todas las medias obtenidas de todas las realizaciones.
- Sobre la distribución muestral se construyen los análisis más conocidos: intervalos de confianza, test de diferencia, etc.

# Inferencia en base a muestras



Fuente: Casas-Cordero (2013)

## Distribución muestral de la media

La **distribución muestral** es la distribución a través de todas las muestras posibles (de tamaño  $n$  a partir de una población  $N$ )

Si la distribución muestral se relaciona de algún modo con la población, podemos hacer conjeturas sobre la población a partir de **una sola muestra**

# Teorema Central del Límite

El **teorema central del límite** nos permite conocer la relación entre la distribución muestral y la población, pues establece que:

- El promedio de la distribución muestral de las medias es la media de los elementos de la población ( $\bar{x}$  es insesgado para  $\bar{X}$ )
- La desviación estándar de la distribución muestral de las medias ( $\sigma_{\bar{x}}$ ) se relaciona con la desviación estándar de los elementos ( $\sigma_X$ ) a través de:  $\sigma_{\bar{x}} = \frac{\sigma_X}{\sqrt{n}}$
- La forma de la distribución muestral de las medias de **cualquier distribución** poblacional converge aproximadamente a una distribución **Normal** a medida que el tamaño de la muestra aumenta.

## Fórmulas para la inferencia

- **Valor esperado:** El valor esperado de la distribución muestral de la media es igual a la media poblacional.
- **Dispersión:** La dispersión de la distribución muestral de la media puede medirse por su varianza ( $\frac{\sum (y_i - \bar{y})^2}{(n-1)}$ ), o su error estándar ( $\sqrt{\text{var}(\bar{y})}$ )

## Intervalo de confianza

Si seleccionamos muchas muestras y calculamos la media en cada una de ellas, generamos la distribución muestral de la media.

El intervalo de confianza de la media muestral, al  $100(1 - \alpha) \%$  de confianza, está dado por la fórmula  $\bar{y} \pm se(\bar{y}) \times Z_{1-\alpha/2}$

Ejemplo: Sea  $\bar{y} = 250$  el promedio del SIMCE en Chile, con  $se(\bar{y}) = 50$ . ¿Cual es el intervalo de confianza al 95 % para el puntaje SIMCE?

## Interpretación del IC

En el contexto de muestreo de poblaciones finitar, la interpretación del intervalo de confianza es frecuentista, no probabilístico:

### Correcto

Intervalo que contiene (cubre) el valor verdadero de la media población el 95 % de las veces, bajo el diseño muestral dado"

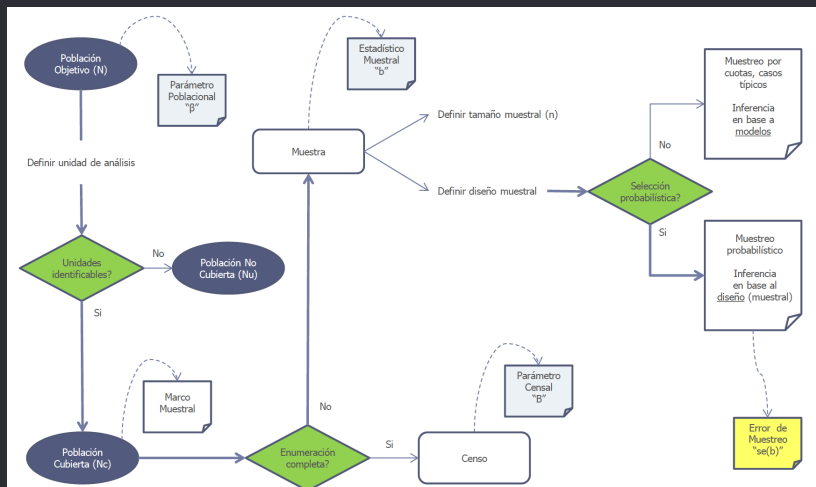
### Incorrecto

"95 % de probabilidad de que el valor verdadero de la media poblacional es  $\bar{y}$ "

"95 % de confianza de que el valor verdadero de la media poblacional es  $\bar{y}$ "



# Resumen



Fuente: Casas-Cordero (2013)

# Estructura de la presentación

1. Población y marco muestral
2. Problemas de los marcos y error de cobertura
3. Distribución muestral
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.