



FACULTAD DE MATEMÁTICAS  
PONTIFICIA UNIVERSIDAD  
CATÓLICA DE CHILE

Pontificia Universidad Católica de Chile

Facultad de Matemáticas

EYP2417 - Muestreo

# Proyecto de Análisis

Encuesta CASEN 2022

Métodos y Diseño Muestral

Grupo 4

Alexander Pinto      Esteban Román

Julián Vargas      Francisca Sepúlveda

- 1 Introducción
- 2 Diseño Muestral
- 3 Plan de Análisis
- 4 Metodología Detallada



## Encuesta CASEN

La Encuesta de Caracterización Socioeconómica Nacional (CASEN) tiene como objetivo medir las condiciones de vida de los hogares y la población en el territorio chileno.

- ✓ Permite estimar **indicadores de pobreza**, desigualdad e inclusión social a nivel nacional
- ✓ Representa a la población que reside en **viviendas particulares ocupadas** en todo el territorio nacional
- ✓ Excluye las denominadas **áreas especiales** (zonas de acceso restringido, alto costo o condiciones climáticas adversas)



## Características principales

El diseño muestral de la Encuesta CASEN se caracteriza por ser:

- 1. Probabilístico:** Todas las unidades de la población tienen una probabilidad conocida y distinta de cero de ser seleccionadas
- 2. Estratificado:** El territorio se divide en **estratos geográficos y socioeconómicos**
  - Mejora la precisión de las estimaciones
  - Asegura representatividad en distintos niveles (nacional, regional y urbano/rural)



**3. Bietápico:** La selección se realiza en dos etapas:

## Primera etapa

Selección sistemática con probabilidad proporcional al tamaño (**PPT**) de las **unidades primarias de muestreo (UPM)** - conglomerados de viviendas

## Segunda etapa

Selección aleatoria simple (**MAS**) de **viviendas** dentro de cada unidad primaria



Consideraciones para el tamaño de muestra óptimo

CASEN considera los siguientes criterios:

1. Representatividad del **territorio nacional**, regiones y zonas urbanas/rurales
2. **Simulaciones** para optimizar precisión



# Tamaño de Muestra CASEN 2022

Nivel	Tamaño Objetivo	Error Absoluto	Error Relativo	Tamaño con Sobremuestreo
<b>País</b>	71.028	0,4 %	3,3 %	106.856
Urbano	56.905	0,5 %	4,6 %	87.252
Rural	14.123	1,3 %	9,2 %	19.604

Fuente: *Manual Metodológico CASEN 2022*, p. 35



## Construcción del Marco

A partir del **MMV 2020** se elabora el marco de selección para CASEN 2022

- Conformado por **335 comunas** definidas para el nuevo diseño de la Encuesta Nacional de Empleo (ENE 2020)
- Las UPM del marco muestral están estratificadas por:
  - **Geografía** (comunas)
  - **Áreas** (urbana–rural)
  - **Nivel socioeconómico (NSE)** - construida a partir de la clasificación del MMV 2020



## Objetivo Analítico

Producir inferencias hacia la población que reside en **viviendas particulares ocupadas** (elegibles)

## Exclusiones

No se consideran para fines analíticos las viviendas **no elegibles**:

- Oficinas de empresas
- Viviendas abandonadas
- Viviendas de veraneo
- Viviendas demolidas



# Objetivo 1: Brecha Salarial de Género

## Objetivo

Analizar diferencias por género en variables socioeconómicas, laborales y educativas utilizando CASEN 2022

### Variables principales:

- **sexo**: Sexo de la persona
- **ytrabajocorh**: Ingreso del trabajo principal corregido
- **esc**: Años de educación formal

### Variables de control:

- **edad**: Edad de la persona
- **oficio4\_08**: Ocupación
- **tot\_per\_h**: Total personas en el hogar



## Métodos Propuestos

Análisis de la brecha salarial de género

**1. Análisis descriptivo:** Cálculo de ingresos promedio del trabajo principal según sexo

**2. Modelo de regresión lineal ponderado:**

- Variable dependiente: `ytrabajocorh` (Ingreso)
- Variables explicativas: sexo, educación, edad y ocupación
- Utiliza factores de expansión del diseño muestral



# Objetivo 2: Distribución de la Pobreza

## Objetivo

Explorar la distribución de la pobreza y diferencias entre zonas rural y urbana

### Variables principales:

- **pobreza**: Condición de pobreza
- **ytotalcorh**: Ingreso total corregido
- **zona**: Rural/Urbana

### Variables de control:

- **esc**: Años de educación
- **región**: Región del país
- **edad**: Edad de la persona
- **expr**: Factor de expansión



## Métodos Propuestos

### Análisis multidimensional de la pobreza

1. Calcular el **porcentaje de hogares en pobreza** por región y zona (rural/urbana)
2. Comparar el **ingreso promedio** según nivel educativo y tamaño del hogar
3. **Pruebas de hipótesis ponderadas** para determinar si las diferencias entre zonas son estadísticamente significativas
4. **Modelo de regresión logística** para explorar condiciones de pobreza con variables: educación, tamaño del hogar, edad y zona



# Estimador de Horvitz-Thompson

## Estimador de la media ponderada

Para estimar medias y proporciones bajo un diseño muestral complejo, utilizamos el estimador de Horvitz-Thompson ajustado por el factor de expansión:

$$\hat{Y} = \frac{\sum_{i \in s} \frac{y_i}{\pi_i}}{\sum_{i \in s} \frac{1}{\pi_i}} \approx \frac{\sum_{i \in s} expr_i y_i}{\sum_{i \in s} expr_i}$$

donde:

- $y_i$  corresponde a la **variable de interés**
- $expr_i$  es el **factor de expansión** provisto por CASEN 2022
- $\pi_i$  es la probabilidad de inclusión de la unidad  $i$



# Varianza del Estimador

## Varianza e intervalos de confianza

La varianza del estimador se calcula considerando el **efecto del diseño** (estratificación y conglomeración):

$$V(\hat{Y}) = \sum_h \left(1 - \frac{n_h}{N_h}\right) \frac{s_h^2}{n_h}$$

## Ponderadores

Se usará el factor de expansión `expr` para corregir las **probabilidades desiguales de selección** en el diseño muestral



## Implementación en R

Los métodos se implementarán utilizando los siguientes paquetes:

✓ **survey** y **srvyr**

- Estimaciones bajo diseño complejo

✓ **ggplot2**

- Visualización de datos

✓ **dplyr**

- Procesamiento de datos



# ¿Preguntas?

Gracias por su atención

Alexander Pinto | Esteban Román  
Julián Vargas | Francisca Sepúlveda

**Pontificia Universidad Católica de Chile**