

Análisis de las encuestas de cobertura (2)

Andrés Gutiérrez

Comisión Económica para América Latina y el Caribe

2025-02-21

Estimación Dual con la Muestra de la Encuesta

Introducción

- ▶ En el capítulo anterior, se asumió que todos los individuos de la población podían ser incluidos en el censo y la encuesta.
- ▶ En la práctica, solo una muestra de la población es incluida en la encuesta.
- ▶ Este capítulo ajusta el enfoque para situaciones más realistas, donde el censo cubre a toda la población, pero la encuesta solo cubre una muestra.

Planteamiento del Problema

- ▶ **Población:** U , de tamaño N (fijo pero desconocido).
- ▶ **Censo:** Intenta enumerar a todos los individuos, pero algunos no son contados.
- ▶ **Encuesta:** Solo una muestra de la población es incluida.
- ▶ **Error de cobertura:** Diferencia entre el conteo censal y N .

Estructura de los Datos

► Tabla de contingencia estimada:

	En la encuesta	Fuera de la encuesta	Total
En el censo	\hat{N}_{11}	$\hat{N}_{12} = \hat{N}_{1+} - \hat{N}_{11}$	\hat{N}_{1+}
Fuera del censo	$\hat{N}_{21} = \hat{N}_{+1} - \hat{N}_{11}$		
Total	\hat{N}_{+1}		$\hat{N}_{++} =$

► Cantidades estimadas:

- \hat{N}_{11} : Individuos en censo y encuesta.
- \hat{N}_{12} : Individuos en censo, no en encuesta.
- \hat{N}_{21} : Individuos en encuesta, no en censo.
- \hat{N}_{1+} : Individuos en censo.
- \hat{N}_{+1} : Individuos en encuesta.
- \hat{N}_{++} : Tamaño total de la población.

Diseño de Muestreo

- ▶ **Estratificación:** División de la población en grupos homogéneos (geográficos).
- ▶ **Conglomerados:** Selección de unidades primarias de muestreo (UPM), como manzanas o sectores censales.
- ▶ **Efecto de diseño:** Aumenta el error estándar debido a la correlación intra-clase.

Limitaciones del Diseño de Muestreo

1. **Población institucionalizada:** Excluye personas en cárceles, hospitales, residencias de ancianos, etc.
2. **Desfase temporal:** Viviendas ocupadas durante el censo pueden estar deshabitadas durante la encuesta.
3. **Población móvil:** Individuos sin vivienda fija no son cubiertos.

Muestra E y Muestra P

- ▶ **Muestra P:** Áreas enumeradas después del censo. Objetivo: estimar N_{11} y N_{+1} .
- ▶ **Muestra E:** Muestra de registros del censo. Objetivo: estimar N_{1+} .
- ▶ **Roles:**
 1. Muestra E: Corrige inclusiones erróneas en el censo.
 2. Muestra P: Estima coincidencias entre censo y encuesta.

Estimadores de Muestreo

- ▶ **Pesos de muestreo:** $w_k = \pi_k^{-1}$, donde π_k es la probabilidad de inclusión del elemento k .
- ▶ **Estimadores:**
 - ▶ $\hat{N}_{+1} = \sum_{k \in s_P} w_k$
 - ▶ $\hat{N}_{11} = \sum_{k \in s_P} w_k x_{k,11}$
 - ▶ $\hat{N}_{1+} = N_{1+}^0 - \sum_{k \in s_E} w_k (1 - z_k)$
 - ▶ $\hat{N}_{1+} = \sum_{k \in s_E} w_k z_k$ (estimador alternativo)

Explicación de las Variables

- ▶ z_k : Variable dicotómica que toma el valor 1 si el individuo k fue **correctamente enumerado en el censo**, y 0 en caso contrario.
- ▶ $x_{k,11}$: Variable dicotómica que toma el valor 1 si el individuo k fue **encontrado tanto en la encuesta como en el censo**, y 0 en caso contrario.

Estimador del Tamaño Poblacional

- ▶ **Estimador de Petersen:**

$$\hat{N}_{++} = \frac{\hat{N}_{1+} \cdot \hat{N}_{+1}}{\hat{N}_{11}}$$

- ▶ **Estimadores adicionales:**

- ▶ $\hat{N}_{12} = \hat{N}_{1+} - \hat{N}_{11}$

- ▶ $\hat{N}_{21} = \hat{N}_{+1} - \hat{N}_{11}$

Varianza del Estimador

► **Varianza total:**

$$\tilde{V}(\hat{N}) = \tilde{V}_m(\tilde{N}) + \tilde{V}_p(\hat{N})$$

► **Varianza del modelo:**

$$\tilde{V}_m(\tilde{N}) = \frac{\hat{N}_{1+} \cdot \hat{N}_{+1} \cdot (\hat{N}_{1+} - \hat{N}_{11}) \cdot (\hat{N}_{+1} - \hat{N}_{11})}{\hat{N}_{11}^3}$$

► **Varianza del muestreo:**

$$\tilde{V}_p(\hat{N}) \approx \frac{M^2}{m} (1 - f) S_d^2$$

Estimadores de Razón y Postestratificación

- ▶ **Estimador de razón:**

$$\hat{N}_{++}^{ratio} = \frac{N_{1+}^0}{\hat{N}_{1+}^0} \frac{\hat{N}_{1+} \cdot \hat{N}_{+1}}{\hat{N}_{11}}$$

- ▶ **Estimador postestratificado:**

$$\hat{N}_{++}^{post} = \sum_{g=1}^G \left[\frac{N_{g1+}^0}{\hat{N}_{g1+}^0} \frac{\hat{N}_{g1+} \cdot \hat{N}_{g+1}}{\hat{N}_{g11}} \right]$$

Conclusión

- ▶ La estimación dual con muestreo requiere ajustar los métodos para considerar la variabilidad del muestreo y el modelo.
- ▶ Los estimadores de razón y postestratificación mejoran la precisión de las estimaciones.
- ▶ La muestra E y la muestra P son fundamentales para corregir errores y estimar la cobertura poblacional.