

Análisis de las encuestas de cobertura (1)

Andrés Gutiérrez

Comisión Económica para América Latina y el Caribe

2025-02-21

El Sistema de Estimación Dual

Introducción

- ▶ El sistema de estimación dual es una metodología para medir la omisión en censos.
- ▶ Basado en modelos de captura y recaptura, desarrollados por @petersen1896, @lincoln1930, y @schnabel1938.
- ▶ Aplicado a eventos vitales humanos por @sekar1949.
- ▶ Este capítulo establece las condiciones y supuestos para su uso.

Planteamiento del Problema

- ▶ **Población:** U , de tamaño N (fijo pero desconocido).
- ▶ **Censo:** Intenta enumerar a cada individuo, pero algunos no son contados.
- ▶ **Error de cobertura:** Diferencia entre el conteo censal y N .
- ▶ **Encuesta de postenumeración:** Proporciona información adicional para estimar el error de cobertura.

Condiciones y Supuestos

Estabilidad Poblacional

- ▶ La población U es cerrada y de tamaño fijo N .
- ▶ **Período de referencia:** Bien definido y específico.
- ▶ **No incorporaciones:** No hay nacimientos ni inmigraciones.
- ▶ **No pérdidas:** No hay defunciones ni emigraciones.

Estructura Multinomial

- Modelo de probabilidad conjunta para censo y encuesta:

	En la encuesta	Fuera de la encuesta	Total
En el censo	p_{11}	p_{12}	p_{1+}
Fuera del censo	p_{21}	p_{22}	p_{2+}
Total	p_{+1}	p_{+2}	1

- **Probabilidades:**

- p_{11} : En censo y encuesta.
- p_{12} : En censo, no en encuesta.
- p_{21} : No en censo, en encuesta.
- p_{22} : No en censo ni en encuesta.

Independencia Autónoma

- ▶ Censo y encuesta son resultados de N ensayos independientes.
- ▶ **Tabla de contingencia observada:**

	En la encuesta	Fuera de la encuesta	Total
En el censo	N_{11}	N_{12}	N_{1+}
Fuera del censo	N_{21}	N_{22}	N_{2+}
Total	N_{+1}	N_{+2}	$N_{++} = N$

- ▶ **Observables:** N_{1+} , N_{11} , N_{12} , N_{21} .
- ▶ **Desconocidos:** N_{22} , N_{++} .

Independencia Causal

- ▶ Inclusión en el censo es independiente de la inclusión en la encuesta.
- ▶ **Razón de Odds:**

$$\frac{p_{11} \cdot p_{22}}{p_{12} \cdot p_{21}} = 1$$

- ▶ **Dependencia causal:** Puede generar sesgos en las estimaciones.

Emparejamiento

- ▶ **Supuesto:** Emparejamiento preciso entre censo y encuesta.
- ▶ **Identificación:**
 1. Individuos en ambos registros.
 2. Individuos solo en la encuesta.
- ▶ **No respuesta:** Se recopila información auxiliar para ajustar.

Eventos Espurios

- ▶ **Supuesto:** Censo y encuesta están libres de errores espurios.
- ▶ **Ejemplos:**
 1. Duplicaciones en el censo.
 2. Registros de casos inexistentes.
 3. Casos no pertinentes (fuera del período de referencia).

Postestratificación

- ▶ **Técnica:** División de la población en subgrupos homogéneos.
- ▶ **Variables comunes:** Edad, sexo, etnia, nivel educativo, región geográfica.
- ▶ **Objetivo:** Mejorar la precisión de las estimaciones.

Inferencia

- ▶ **Objetivo:** Estimar el tamaño total de la población N_{++} .
- ▶ **Modelo:** Eventos de inclusión como procesos Bernoulli.
- ▶ **Variables aleatorias:**
 - ▶ $N_{1+} \sim \text{Bin}(N_{++}, p_{1+})$
 - ▶ $N_{+1} \sim \text{Bin}(N_{++}, p_{+1})$
 - ▶ $N_{11} \sim \text{Bin}(N_{++}, p_{11})$

Estimadores del Sistema Dual

- ▶ **Estimador de Petersen:**

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

- ▶ **Propiedades:**

- ▶ Insesgado: $E[\tilde{N}_{++}] = N_{++}$.
- ▶ Varianza estimada:

$$\tilde{V}[\tilde{N}_{++}] = \frac{N_{1+} \cdot N_{+1} \cdot N_{12} \cdot N_{21}}{N_{11}^3}$$

Conclusión

- ▶ El sistema de estimación dual permite estimar el tamaño de la población y el error de cobertura.
- ▶ Requiere cumplir con supuestos clave como estabilidad poblacional, independencia causal y emparejamiento preciso.
- ▶ La postestratificación mejora la precisión de las estimaciones.
- ▶ El estimador de Petersen es insesgado y ampliamente utilizado en estudios de captura y recaptura.