

Metodología para el calculo de ponderadores

Muestreo II

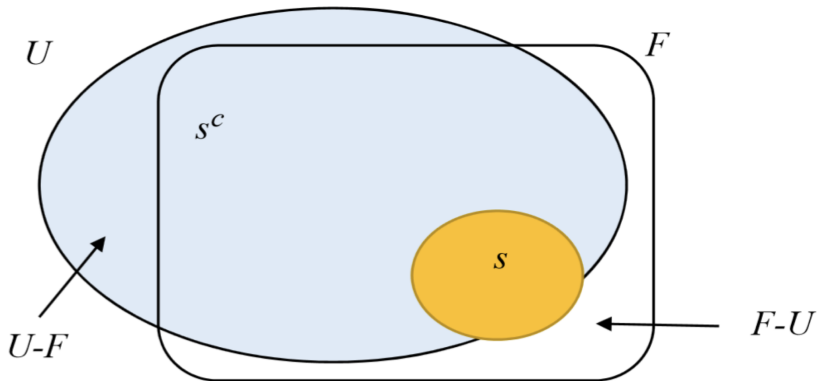
Licenciatura en Estadística

2023

objetivos de la ponderación

- ▶ expandir la muestra a la población total (U)
- ▶ corregir problemas de cobertura en el marco de muestreo (F)
- ▶ reducir sesgos ocasionados por la no respuesta
- ▶ utilizar información auxiliar para conseguir estimadores insesgados y más precisos (menor SE)

objetivos de la ponderación



- ▶ un total puede ser escrito como:

$$Y = \sum_{i \in s} y_i - \sum_{i \in U-s} y_i$$

- ▶ estimador de un total de la población:

$$\hat{Y} = \sum_{i \in s_{ER}} w_i y_i$$

donde

- ▶ s_{ER} es la muestra efectiva (i.e. elegibles respondentes)
- ▶ w_i indica la cantidad de individuos que representa el elemento i

características del diseño muestral

- ▶ estratificación
- ▶ conglomerados
- ▶ probabilidad de inclusión distintas (e.g. tasas de muestreo variadas por estratos)

estratificación

- ▶ una buena construcción y asignación de muestra en los estratos puede reducir los SE para la estimación cantidades poblacionales.
- ▶ utilizando para controlar el tamaño de muestra y precisiones en las estimaciones a nivel de estrato

conglomeración

usualmente incrementa los SEs.

los efectos pueden ser diferentes para:

- ▶ estimación a nivel de toda la población y subgrupos.
- ▶ diferentes variables
- ▶ diferentes parámetros: totales, medias, parámetros de modelos.

- ▶ ajustes de los ponderadores pueden incrementar o reducir los SEs
- ▶ ajustes por NR usualmente incrementan los SEs
 - ▶ objetivo: reducir los sesgos ocasionados por la NR
- ▶ calibración a totales poblacionales pueden reducir los SEs
 - ▶ variables de control deben estar correlacionadas con las variables interés

Población de Interés

No incluidos en la población marco

Inelegibles

s_{ER}

s_{ENR}

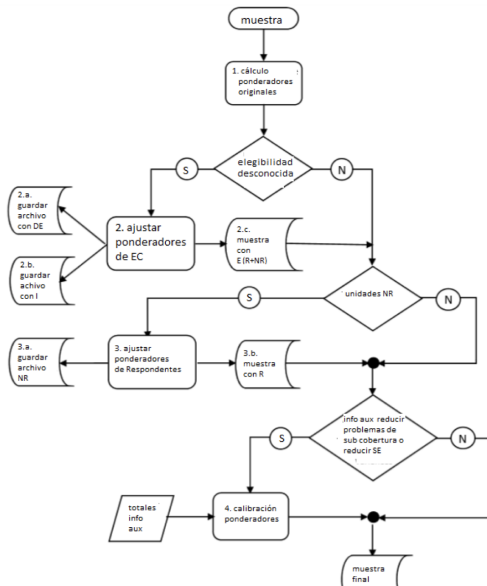
s_{ED}

s_I

Población marco

cuatro pasos para la ponderación

- ① computar ponderadores originales
- ② ajuste teniendo en cuenta las unidades con elegibilidad desconocida (si es que hay)
- ③ ajuste por no respuesta
- ④ calibración a totales (conteos) poblacionales



paso 1 - ponderadores originales

- ▶ los ponderadores originales dependen del diseño muestral
- ▶ se computan para toda la muestra (s)
- ▶ se definen como el inverso de la probabilidad de selección

$$w_{i0} = \frac{1}{\text{Prob}(i \in s)} = \frac{1}{\pi_i}$$

- ▶ pueden ser distintos (e.g. tasas de muestreo distintas por estrato, dos etapas, PPS, etc.)

paso 2 - ajuste elegibilidad desconocida

- ▶ distribución de los ponderadores originales de las unidades con ED entre las unidades con EC
- ▶ se **crean clases** (celdas) con información auxiliar a nivel del marco F (estratos, UPM, etc). Se buscan patrones.

ajuste por elegibilidad desconocida en una clase b

$$ed_i = \frac{\sum_{i \in s_b} w_{i0}}{\sum_{i \in s_b \cap s_{b,EC}} w_{i0}}$$

El ponderador ajustado por ED es:

$$w_{i1} = w_{i0} \times ed_i \quad \forall i \in s_{EC}$$

- ▶ **MCAR**: no respuesta totalmente al azar
- ▶ **MAR**: no respuesta al azar
- ▶ **NINR**: no respuesta no ignorable

bajo estas tres enfoques, la respuesta o no de una unidad es tratada como un evento aleatorio (estocástica).

la no respuesta puede categorizarse como determinística, i.e. una unidad está garantizada de que responda o no.

para el ajuste por NR se asume que la NR es estocástica.

- ▶ suponemos que la probabilidad o propensión de que la unidad i responda es ϕ_i
- ▶ si podemos estimar ϕ_i , entonces puede utilizarse como insumo para ponderar de la muestra de elegibles respondentes (ER).
- ▶ el ajuste por no respuesta puede pensarse como un problema de **predicción** de ϕ_i . Distintas estrategias posibles.

$$w_{i2} = w_{i1} \times nr_i \quad \forall i \in s_{ER}$$

donde $nr_i = \hat{\phi}_i^{-1}$

Paso 3 - ajuste por no respuesta

MCAR: todas las unidades tienen la misma probabilidad de responder .

- ▶ no es necesario ajuste por no respuesta para estimar medias.
- ▶ Ajuste global para totales (nivel)

MAR: probabilidad de responder depende de variables auxiliares (x)

- ▶ es posible calcular los ajustes si las variables auxiliares son conocidas para los ER y ENR.
- ▶ En la práctica poco sabemos de los ENR.

NINR: probabilidad de responder depende de las variables que intenta relevar la encuesta (y). Ajuste por NR difícil o imposible

Paso 3 - ajuste por NR - grupos de respuesta homogéneos

- ▶ se crean grupos o celdas para las cuales se suponen que las probabilidades de responder son similares

$$\text{Prob}(i \in s_{\text{ER}|s}) = \phi_c \quad \forall i \in s_c$$

- ▶ los grupos los construimos de la mejor forma (que podemos) para que describa el estatus de respondente.
- ▶ la idea es que los individuos dentro de los grupos sean lo más parecidos entre si respecto a las variables de interés (desconocida para NR)
- ▶ en la práctica son contruidos en base a las tasas de respuesta.

modificar los ponderadores w_{i2} para construir nuevos ponderadores:

$$w_{i3} = w_{i2} \times g_i$$

de forma tal que se cumpla:

$$\sum_{i \in s_{ER}} w_{i3} \times \mathbf{x}_i = \sum_{i \in U} \mathbf{x}_i$$

i.e. la muestra expandida coincida con conteos poblacionales conocidos.

paso 4 - calibración a conteos poblacionales

- ▶ utilizar variables auxiliares para corregir errores de sub cobertura y reducir los SE .
- ▶ también reduce el sesgo de la no respuesta (e.g. sub o sobre representación de ciertos grupos).
- ▶ variables auxiliares solo deben ser conocidas para los ER.
- ▶ Permite utilizar mayor cantidad de variables que en el ajuste por NR

Post-estratificación

- ▶ clasifica a los ER en grupos (post-estratos).
- ▶ los ponderadores son ajustados para que total estimado del grupo coincida con el total conocido.
- ▶ e.g tramos de edad \times sexo \times región

Raking

- ▶ Similiar a la postestratificación pero solo se necesita los totales de las marginales.
- ▶ e.g tramo de edad, sexo y región.
- ▶ permite mayor cantidad de variables auxiliares

GREG

- ▶ variables cualitativas y cuantitativas pueden usarse
- ▶ e.g tramos de edad \times sexo y edad simples

Calibración truncada

- ▶ se restringen los ajustes de la calibración.
- ▶ se fija un limite inferior o superior para los ajustes
- ▶ las ecuaciones de calibración pueden no cumplirse (pero estar muy cerca)

- ▶ evalúa la variabilidad de los ponderadores
- ▶ representa el incremento en la varianza de los ponderadores por usar ponderadores distintos respecto a un diseño autoponderado

$$\text{deff}_w = 1 + \text{cv}(w_i)^2$$

- ▶ debe ser calculado en cada etapa de ajuste
- ▶ en la práctica mayores a 1.5 indican que los ajustes realizados tengan que ser revisados. sesgo vs varianza
- ▶ e.g. pasar a la calibración truncada o utilizar celdas de ajuste por no respuesta más amplias.
- ▶ si lo anterior no reduce la variabilidad de los ponderadores, los mismos pueden ser recortados (trimming)