

Tópicos especiales - calibración

Muestreo II

Licenciatura en Estadística

2023

- ① calibración cuando la UUM es un cluster
- ② prevención de ponderadores extremos o influyentes

- ▶ en algunos casos los ponderadores finales w_i deben ser iguales para todas las unidades del conglomerado/cluster cuando la misma es la UUM
- ▶ **ejemplo:** UUM = vivienda y la unidad de análisis son los hogares y todos sus integrantes (e.g. ECH)
- ▶ tener ponderadores iguales produce consistencia en las estimaciones entre los microdatos de personas y hogares

- ▶ exigir ponderadores calibrados w_i^* para todas las unidades dentro del conglomerado implica una pérdida de "eficiencia".
- ▶ esto se debe (como veremos más adelante) a que el aumento en la variabilidad o spread de los w_i^* suele implicar un aumento en los SE de los estimadores

- ▶ **método de integración:** el set/vector de variables auxiliares \mathbf{x}_i pueden promediarse para cada UUM previo a la realización de la calibración
- ▶ **forma iterativa:**
 - ① se realiza la calibración de los ponderadores utilizando información a nivel del elemento i (i.e \mathbf{x}_i)
 - ② los ponderadores calibrados w_i^* son promediados posteriormente a nivel de cada cluster
 - ③ se repite la calibración utilizando como "ponderadores originales" los ponderadores promediados en el punto 2.
 - ④ se repiten los puntos anteriores hasta lograr la convergencia, i.e. se cumplan las ecuaciones de calibración y los ponderadores calibrados sean iguales para cada elemento del cluster

cuando se calibran los ponderadores puede ocurrir:

- ▶ obtener ponderadores negativos (i.e. factores de ajuste $g_i < 0$) si utilizamos la función de distancia lineal (i.e. el estimador de regresión)
 - ▶ es poco probable que pase si las probabilidades de inclusión son pequeñas (π_i) y la cantidad de variables de control no son excesivas para lo que la muestra puede soportar.

cuando se calibran los ponderadores puede ocurrir:

- ▶ Obtener ponderadores muy grandes
 - ▶ esto puede ocurrir si las tasas de repuesta o elegibilidad obtenidas en algunos dominios son pequeñas y si la cantidad de variables de control son excesivas para lo que la muestra puede soportar.

Podemos agregar una restricción en el cálculo para evitar ponderadores negativos (o pequeños) y muy grandes

$$L \leq \frac{w_i^*}{w_i} \leq U$$

tengamos en cuenta que estamos fijando los cambios relativos, es decir, los factores de ajustes g_i

utilizamos el script llamado *calibración truncada.r*

- ▶ trabajamos con U =ocupados
- ▶ seleccionamos una muestra bajo un SI de tamaño $n = 500$
- ▶ Calibramos utilizando las funciones:
 - 1 lineal
 - 2 raking
 - 3 logit

elegimos una cantidad excesiva de variables de control:

- ▶ educación (6 categorías)
- ▶ sexo (2 categorías)
- ▶ edad (7 categorías)
- ▶ estrato socioeconómico (5 categorías)
- ▶ salud (2 categorías)
- ▶ pobreza (2 categorías)

existen tres argumentos útiles:

- ❶ **maxit**= numero de iteraciones permitidas. Por defecto 50
- ❷ **epsilon**= tolerancia entre las estimaciones y los totales poblacionales. Por defecto 10^{-7}
- ❸ **force**= Devuelve una respuesta si los niveles requeridos no son alcanzados. Por defecto es FALSE

pasos a seguir (jerárquicos o no):

- ① si la convergencia no es alcanzada con los valores por defecto, aumentar el **maxit**.
- ② aumentar el nivel de tolerancia (**epsilon**)
- ③ forzar el cálculo de los ponderadores. **force=TRUE**
- ④ si **force=TRUE** chequear que tan lejos están de cumplirse las ecuaciones de calibración.

recorte (trimming) de los ponderadores

- ▶ generalmente procedimientos ad-hoc.
- ▶ los límites inferior y superior (L , U) son arbitrarios y dependen de la persona
- ▶ básicamente tienen en cuenta la distribución de los ponderadores en encuestas anteriores.

- 1 indicar L y U de forma arbitraria.
- 2 cualquier ponderador fuera de los límites es recortado, es decir,

$$w_{i,\text{trim}}^* = \begin{cases} U & \text{si } w_i^* \geq U \\ w_i^* & \text{si } L \leq w_i^* < U \\ L & \text{si } w_i^* < L \end{cases} \quad (1)$$

- 3 Determinar la suma $K = \sum_{i \in S} |w_i^* - w_{i,\text{trim}}^*|$, es decir, la cantidad de ponderadores perdida por el recorte
- 4 distribuir K entre los ponderadores no recortados.
- 5 repetir los pasos del 2 al 4 hasta que no haya más ponderadores por fuera de los límites.

trimWeights(cal,upper=U,lower=L, strict=TRUE)

si **strict=TRUE** la función se llama así misma de forma recursiva hasta que los límites son satisfechos

- ▶ cuando aplicamos **trimWeights** los ponderadores dejan de estar calibrados a los conteos poblacionales.
- ▶ debemos chequear que “tan lejos” están de cumplir las ecuaciones de calibración