

Repaso - diseños muestrales

Muestreo II

Licenciatura en Estadística

2023

muestreo simple

- ▶ probabilidad de selección $\pi_i = n/N \quad \forall i \in U$
- ▶ ponderador original $w_i = \pi_i^{-1} = N/n \quad \forall i \in s$
- ▶ estimador del total de la variable y , $Y = \sum_{i \in U} y_i$

$$\hat{Y} = \sum_{i \in s} w_i \times y_i = N \times \bar{y}$$

donde $\bar{y} = n^{-1} \times \sum_{i \in s} y_i$

- ▶ **estiman sin error** el total de la población (i.e. los ponderadores se encuentran calibrados al tamaño de la población)

$$\hat{N} = \sum_{i \in s} w_i = N$$

estimación del SE

Un estimador insesgado viene dado como:

$$\widehat{SE}^2(\hat{Y}) = \widehat{\text{var}}(\hat{Y}) = N^2 \times \text{fpc} \times \text{var}_s[y]/n$$

donde

- ▶ $\text{fpc} = (1 - n/N)$ = es el factor de corrección por población finita
- ▶ $\text{var}_s[y] = \sum_{i \in s} (y_i - \bar{y})^2 / (n - 1)$ = varianza de la variable y en la muestra

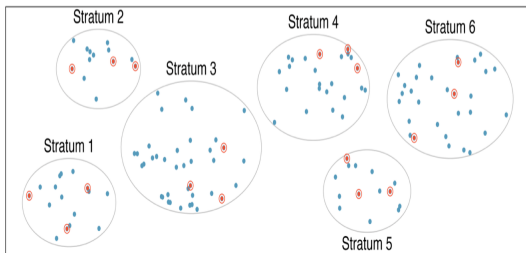
sistemático

- ▶ útil para seleccionar muestras en campo
- ▶ si la variable presenta un orden que explica las características de los individuos es más eficiente
- ▶ los individuos pueden ser seleccionados con igual probabilidad de selección (EPSEM)

muestreo estratificado

La población es agrupada/separada en estratos ($H = 6$).

Dentro de cada estrato se selecciona una muestra independiente s_h de tamaño n_h bajo un diseño muestral cualquiera.



estratificación

variables de estratificación en encuestas a hogares y personas:

- ▶ regionales (departamentos, ciudades etc.)
- ▶ variables demográficas (sexo, tramo de edad)
- ▶ socio-económicas (tramo de ingreso, etc.)

estratificación

variables de estratificación en encuestas a empresas:

- ▶ tamaño de la empresa (cantidad de empleados, monto de ventas, etc)
- ▶ actividad económica de la empresa (industria, comercio, etc)

Importante: las variables x deben estar contenidas en el marco muestral (F)

por qué estratificar?

- ▶ **protegernos** de la posibilidad de tener una muestra realmente mala
- ▶ definir **precisiones específicas** para distintas subpoblaciones (dominios)
- ▶ obtener **estimaciones más precisas** para el total de la población
- ▶ definir **ponderadores distintos** (tasas de muestreo por estrato)

muestreo estratificado simple

Dentro de cada estrato seleccionamos una muestra aleatoria bajo un diseño simple

- ▶ N_h = tamaño del estrato h
- ▶ n_h = tamaño de muestra en el estrato h
- ▶ $\pi_{ih} = n_h/N_h$ = probabilidad de selección del individuo i en el estrato h
- ▶ $w_{ih} = N_h/n_h$ = ponderador original del del individuo i perteneciente al estrato h

$$\hat{Y} = \sum_{i \in s} w_i \times y_i = \sum_{h=1}^H \hat{Y}_h = \sum_{h=1}^H N_h \times \bar{y}_h$$

ejemplo

| h | sector | N_h | n_h | w_{hi} |
|-------|-----------------------|-------|-------|----------|
| 1 | Industria | 600 | 50 | 12 |
| 2 | Comercio al por menor | 1200 | 50 | 24 |
| 3 | Comercio al por mayor | 400 | 50 | 8 |
| 4 | Servicios | 2300 | 50 | 46 |
| 5 | Salud | 500 | 50 | 10 |
| Total | | 5000 | 250 | |

estimación del SE

$$\widehat{SE}^2(\hat{Y}) = \widehat{var}(\hat{Y}) = \sum_{h=1}^H N_h^2 \times fpc_h \times var_{s_h}[y]/n_h$$

donde

- ▶ $fpc_h = (1 - n_h/N_h)$ = es el factor de corrección por población finita en el estrato h
- ▶ $var_{s_h}[y] = \sum_{i \in s_h} (y_i - \bar{y}_h)^2 / (n_h - 1)$ = varianza de la variable y en la muestra del estrato h

Asignación de muestra por estrato

- **proporcional**: réplica en miniatura de la población y autoponderada (EPSEM)

$$n_h = n \times \text{peso}_h = n \times \frac{N_h}{N}$$

- **óptima**: minimiza el error estándar de la estimación de un total.

$$n_h = n \times \frac{N_h \times \text{sd}_h[x]}{\sum_{h=1}^H N_h \times \text{sd}_h[x]}$$

- **precisiones aprox constantes**

$$n_h = n \times \frac{\sqrt{N_h}}{\sum_{h=1}^H \sqrt{N_h}}$$