MUEST_ T11. 1. MUESTREO DE CONGLOMERADOS CON SUBMUESTREO.

2. ESTIMADORES LINEALES INSESGADOS.

3. VARIANZA de un ESTIMADOR LINEAL.

4. TEOREMA de MADOW. APLICACIONES.

5. MUESTRAS AUTOPONDERADAS.

1- MUESTREO de CONGLOMERADOS con SUBMUESTREO (bietápico)

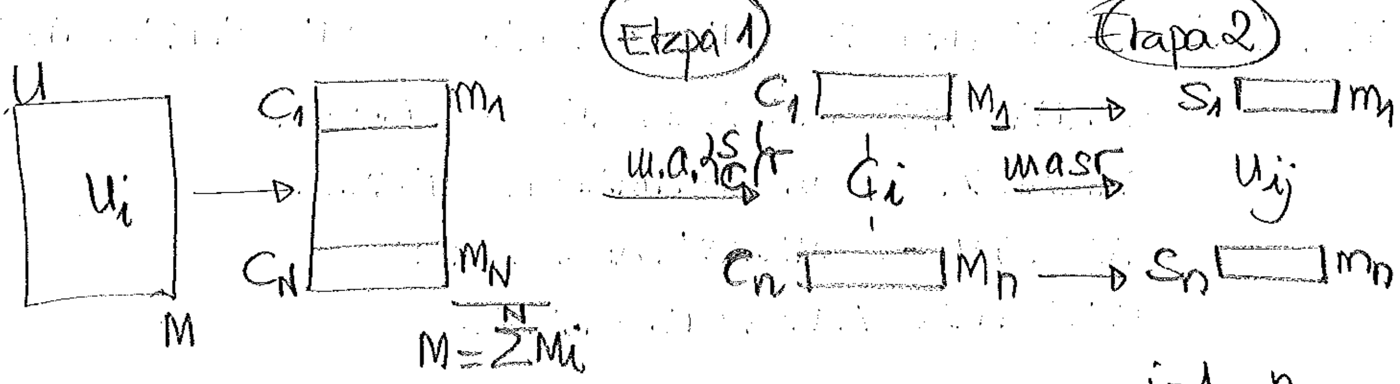
Supongamos que realizamos un muestreo de conglomerados monoetápico, y que las unidades elementales de los conglom. elegidos para la muestra son homogéneas entre sí, entonces no es necesario observar todas las unidades del conglom. muestral para obtener una muestra representativa.
Conviene efectuar un submuestreo en cada conglomerado seleccionado para formar la muestra final → se ahorra tiempo y dinero.

El MCB se lleva a cabo en dos etapas:

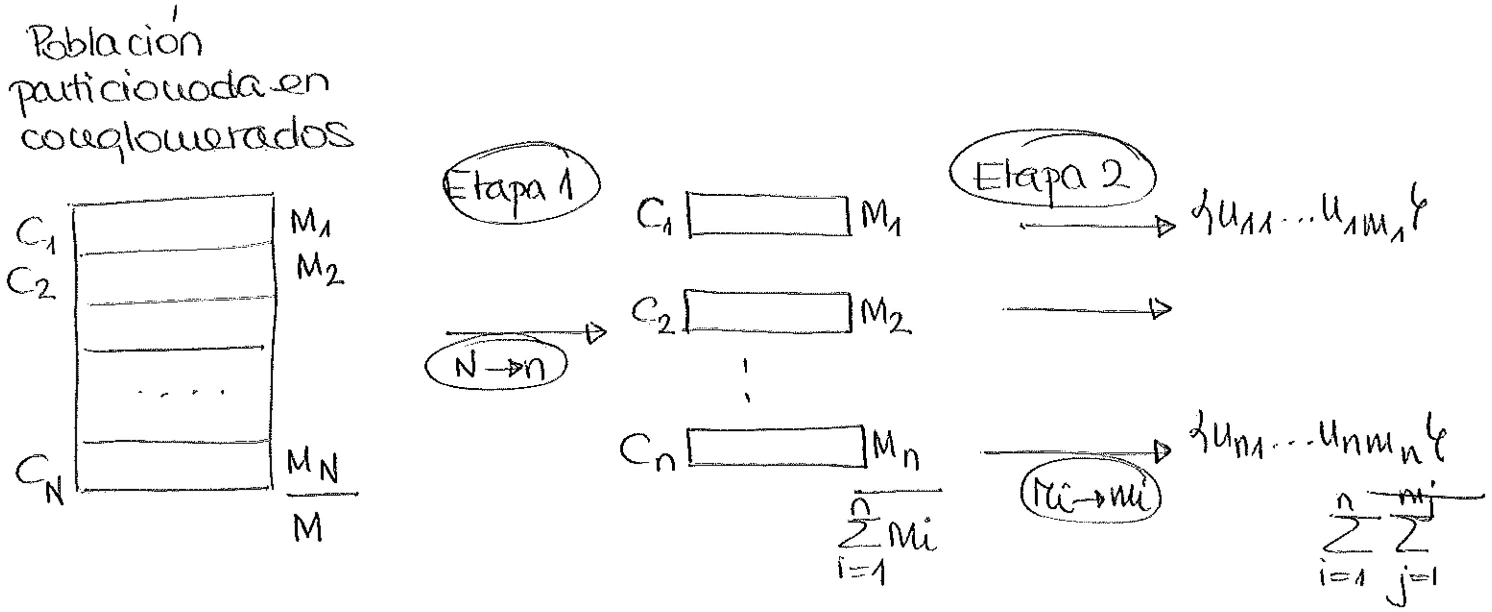
Etapa 1: De una población dividida en N conglomerados de tamaño M_i, $\sum_{i=1}^{N} M_i = M$, se seleccionan aleat. n conglomerados.

Etapa 2: De cada conglomerado seleccionado en 1ª etapa, se selecciona una submuestra de m_i unidades elementales, de manera independiente.

Los conglomerados se denominan unidades primarias, o de primera etapa, y las unidades seleccionadas en el submuestreo unidades secundarias, o de 2ª etapa.


Etapa 1
Etapa 2
U
U_i
M
C_1
C_N
M_1
M_N
$M = \sum^{N} M_i$
C_1
C_i
C_n
M_1
M_n
S_1
S_n
m_1
m_n
U_{ij}

$i = 1 - n$
$j = 1 - m_i$



En ambas etapas, la selección puede ser con o sin reposic., pero en segunda etapa suele utilizarse m.a.s sin r.

Si en segunda etapa, y dentro de cada conglomerado de primera etapa, volvemos a realizar muestreo por conglomerados con submuestreo → muestreo trietápico de conglomerados. De esta forma se podría generalizar al muestreo trietápico.

Aunque el muestreo de conglomerados es menos preciso que el mas, su uso está muy extendido por cuestiones de coste, y porque no requiere un marco de unidades elementales completo.

Además, si los conglomerados son homogéneos dentro, bastará observar un pequeño nº de unid. elementales para obtener una muestra representativa.

Así, se podría observar el mismo nº de unid. elementales que en MCM, pero tomando un mayor nº de conglom. en la etapa 1. La población está más representada en la muestra, aunque se incrementa el coste.

En cualquier caso, un estudio conjunto de la variabilidad y el coste aconsejarán el mt. + oportuno en cada caso.

2_ ESTIMADORES LINEALES INSESGADOS

En el MCB hay dos tipos de unidades en el muestreo, por lo que su selección origina dos fuentes de variación:
- Variación debida al muestreo de unid. primarias (sub. 1)
- Variación debida al submuestreo (sub. 2)

Con esta notación, la esperanza de un estimador sería la esperanza, sobre todas las unid. posibles, de n unid. primarias, de la esperanza condicionada a un conjunto fijo de n unid. primarias, sobre todas las submuestras posibles dentro de dicho conjunto.

MIRAR ATRAS ⟹

cambio sus

$$E[\hat{\theta}] = \sum_{m_2} \hat{\theta}(m_2) P(m_2) = \sum_{m_2} \hat{\theta}(m_2) \sum_{m_1} P(m_2/m_1) P(m_1) =$$
$$= \sum_{m_1} P(m_1) \cdot \sum_{m_2} \hat{\theta}(m_2) \cdot P(m_2/m_1) = \sum_{m_1} P(m_1) E_2(\hat{\theta}) = E_1 E_2(\hat{\theta}).$$

Para construir los estim. lineales insesgados de los distintos parámetros poblacionales, hay que tener en cuenta estas dos etapas.

Una vez finalizada la primera etapa (se seleccionan n conglom.), en cada conglomerado se extrae por m.a.s.(*) una muestra de tamaño m_i (segunda etapa). POR COMODIDAD, MAS, pi.

Para el conglomerado i-ésimo se tiene:

Media poblacional:

$$\theta = \overline{X}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} X_{ij} \longrightarrow \hat{\theta} = \overline{x}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ij}$$

Total poblacional:

$$\theta = X_i = \sum_{j=1}^{M_i} X_{ij} = M_i \overline{X}_i \longrightarrow \hat{\theta} = \hat{X}_i = M_i \overline{x}_i = \frac{M_i}{m_i} \sum_{j=1}^{m_i} x_{ij}.$$

(*) Se podría utilizar otro tipo de muestreo (con repos. y prob. ≠, pero puede demasiado largo el tema).

El estimador del parámetro, $\hat{\theta}$, es una v.a. cuyo valor esperado depende de las unidades muestrales de 1ª etapa y de las unid. muestrales de 2ª etapa:

$$E[\hat{\theta}] = \sum_{i=1}^{n} \hat{\theta}(C_i) \cdot P(C_i) = \sum_{i=1}^{n} P(C_i) \cdot \underbrace{\sum_{j=1}^{m_i} \hat{\theta}(u_{ij}) P(u_{ij}/C_i)}_{\text{esperanza condic.}} =$$

$$= \sum_{i=1}^{n} P(C_i) E_2(\hat{\theta}) = E_1 E_2(\hat{\theta}).$$

Proporción poblacional: 1

$$\theta = P_i = \frac{1}{M_i}\sum_{j=1}^{M_i} A_{ij} \longrightarrow \hat{\theta} = \hat{P}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} A_{ij}$$

Total de clase:

$$\theta = A_i = \sum_{j=1}^{M_i} A_{ij} = M_i P_i \longrightarrow \hat{\theta} = \hat{A}_i = M_i \hat{P}_i = \frac{M_i}{m_i}\sum_{j=1}^{m_i} A_{ij}$$

(de 2ª etapa)

A partir de los estimadores para cada conglomerado seleccionado en la primera etapa se construyen los estimadores globales para toda la población, distinguiendo los distintos tipos de muestreo utilizados en la primera etapa.

a) Selección SIN reposición en la 1ª etapa:

Sea $\pi_i = P(C_i \in \text{Muestra})$ = Probab. de que la unidad primaria u_i (congl. i) $\in$ muestra.

$$e_i = \begin{cases} 1 & \text{si } C_i \in \text{Muestra} \\ 0 & \notin \end{cases} \qquad e_i \longrightarrow \text{Bern}(\pi_i)$$

Para el parámetro poblacional $\theta = X = \sum_{i=1}^{N} X_i = \sum_{i=1}^{N}\sum_{j=1}^{M_i} X_{ij}$,

el estimador lineal insesgado es (Horvitz y Thompson)

$$\hat{X} = \sum_{i=1}^{n} \frac{\hat{X}_i}{\pi_i} \underset{\text{prob. iguales}}{\overset{p.i.}{=}} \frac{N}{n}\sum_{i=1}^{n}\hat{X}_i \quad , i = 1 \ldots n \quad = \sum_{i=1}^{n}\frac{M_i}{\pi_i}\cdot \bar{x}_i$$

Es insesgado:

$$E[\hat{X}] = E_1 E_2\left[\sum_{i=1}^{n}\frac{\hat{X}_i}{\pi_i}\right] = E_1\left[\sum_{i=1}^{n}\frac{1}{\pi_i}\underbrace{E_2[\hat{X}_i]}_{=X_i}\right] =$$

$$= E_1\left[\sum_{i=1}^{n}\frac{1}{\pi_i}\cdot X_i\right] = E_1\left[\sum^{N}\frac{1}{\pi_i} X_i e_i\right] = \sum_{i=1}^{N}\frac{1}{\pi_i} X_i \underbrace{E[e_i]}_{=\pi_i} =$$

$$= \sum_{i=1}^{N} X_i = X$$

Las expresiones de los otros 3, igual.

MIRAR ATRÁS $\Longrightarrow$

Para la media:

$$\hat{X} = \sum_{i=1}^{n} \frac{\hat{X}_i}{\pi_i} = \sum_{i=1}^{n} \frac{\overline{x}_i}{\pi_i} \xrightarrow{p.i.} \frac{N}{n} \sum_{i=1}^{n} \overline{x}_i$$

Para la proporción:

$$\hat{P} = \sum_{i=1}^{n} \frac{\hat{P}_i}{\pi_i}$$

Para el total de clase:

$$\hat{A} = \sum_{i=1}^{n}$$

b) Selección CON reposición en la 1ª etapa:

Sea $P_i = P(\mathcal{C}_i \in \text{Muestra}) \equiv$ Probab. de seleccionar el congl. i en una extracción.

$e_i \equiv$ nº veces que aparece $\mathcal{C}_i$ en la muestra

$e_i \longrightarrow B(n, P_i)$

El estimador lineal insesgado del total poblacional es

$$\hat{X} = \sum_{i=1}^{n} \frac{\hat{X}_i}{nP_i},$$

Es insesgado de X:

$$E[\hat{X}] = E_1E_2\left[\sum_{i=1}^{n} \frac{\hat{X}_i}{nP_i}\right] = E_1\left[\sum_{i=1}^{n} \frac{1}{nP_i} \underbrace{E_2[\hat{X}_i]}_{X_i}\right] = E_1\left[\sum_{i=1}^{n} \frac{X_i}{nP_i}\right] =$$

$$= E_1\left[\sum_{i=1}^{N} \frac{X_i}{nP_i} e_i\right] = \sum_{i=1}^{N} \frac{X_i}{nP_i} \underbrace{E[e_i]}_{nP_i} = \sum^{N} X_i = X.$$

- Si los conglomerados se seleccionan con probab. proporcionales a su tamaño en cada extracción:

$$P_i = \frac{M_i}{M}, \; i = 1 \dots N \implies \hat{X} = \sum_{i=1}^{n} \frac{M\hat{X}_i}{nM_i} = \frac{M}{n}\sum_{i=1} \bar{X}_i = M\bar{\bar{x}}$$

- Si el muestreo es con probab. iguales sobre los conglomerados:

$$P_i = \frac{1}{N}, \; \forall i = 1 \dots N \implies \hat{X} = \sum_{i=1}^{n} \frac{\hat{X}_i}{n/N} = N\sum_{i=1}^{n} \frac{\hat{X}_i}{n} = \frac{N}{n}\sum_{i=1}^{n} \hat{X}_i$$

* Los estimadores de la média poblacional por conglomerado

$$\bar{X} = \frac{X}{N} \longrightarrow \hat{\bar{X}} = \bar{x} = \frac{\hat{X}}{N}$$

y de la media poblacional por unidad elemental.

$$\bar{\bar{X}} = \frac{X}{M} \longrightarrow \hat{\bar{\bar{X}}} = \bar{\bar{x}} = \frac{\hat{X}}{M}$$

3_ VARIANZA de un ESTIMADOR LINEAL.

Al igual que ocurría con el cálculo de la esperanza de un estimador lineal en MCB, para calcular la varianza de los estimadores hay que tener en cuenta los dos tipos de variación presentes en este muestreo:

- variación debida al muestreo en la etapa 1
- variación debida al submuestreo en la etapa 2.

Si $\hat{\theta}$ es un estimador insesgado de θ, $E[\hat{\theta}]=\theta$, su varianza se puede escribir como

$$V(\hat{\theta}) = E[(\hat{\theta}-\theta)^2] = E_1E_2[(\hat{\theta}-\theta)^2]$$

El teorema de Madow facilita el cálculo de las varianzas de los estimadores, proporcionando una expresión para la varianza de los estimadores insesgados en el muestreo bietápico.

4_ TEOREMA de MADOW

La varianza de un estimador insesgado en MCB se puede expresar de la siguiente forma:

$$V(\hat{\theta}) = E_1[V_2(\hat{\theta})] + V_1[E_2(\hat{\theta})]$$

Dem:

$$V(\hat{\theta}) = E[(\hat{\theta}-\theta)^2] = E_1E_2[(\hat{\theta}-\theta)^2] = E_1E_2[\hat{\theta}^2+\theta^2-2\hat{\theta}\theta] =$$

$$= E_1[E_2(\hat{\theta}^2)+\theta^2-2\theta E_2(\hat{\theta})] \underset{\uparrow}{=} E_1[V_2(\hat{\theta})+(E_2(\hat{\theta}))^2+\theta^2-2\theta E_2(\hat{\theta})]$$

$$E_2(\hat{\theta}^2) = V_2(\hat{\theta})+(E_2(\hat{\theta}))^2$$

$$= E_1V_2(\hat{\theta}) + E_1(E_2(\hat{\theta}))^2+\theta^2-2\theta E_1E_2(\hat{\theta}) =$$

como $\hat{\theta}$ es insesgado de θ $\rightarrow$ $E_1E_2(\hat{\theta})=\theta$

$$= E_1V_2(\hat{\theta}) + E_1(E_2(\hat{\theta}))^2 - \theta^2 = E_1V_2(\hat{\theta}) + \underbrace{E_1(E_2(\hat{\theta}))^2}_{\text{momento de orden 2}} - \underbrace{(E_1(E_2(\hat{\theta})))^2}_{\text{momento de orden 1 al 2}} =$$

$$= E_1V_2(\hat{\theta}) + V_1E_2(\hat{\theta})$$

El tma de Madow se puede generalizar a tres ó más etapas:

$$V(\hat{\theta}) = E_1E_2V_3(\hat{\theta}) + E_1V_2E_3(\hat{\theta}) + V_1E_2E_3(\hat{\theta})$$

5_ APLICACIONES del THA. de MADOW (INVENTADO)

El tma de Madow permite desglosar la varianza de los estimadores, por lo que ~~permi~~ facilita su cálculo.
Por comodidad, suponemos m.a.s.s.r. en segunda etapa y distinguimos dos posibilidades en la primera etapa.
Para cada C_i, $i = 1 \ldots n$

$E_2(\hat{X}_i) = X_i$, insesgado

$V_2(\hat{X}_i) = M_i^2 \cdot (1 - f_{2i}) \cdot \frac{S_i^2}{m_i}$ (SR)

donde $f_{2i} = \frac{m_i}{M_i} \rightarrow$ fr. muestreo E2.

$S_i^2 = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} (X_{ij} - \bar{X}_i)^2 \rightarrow$ cuasiv. pobl. conglom. i.

En el caso del total poblacional:

$$V(\hat{X}) = E_1 V_2(\hat{X}) + V_1 E_2(\hat{X}).$$

a) E1 SR ó E2 $\rightarrow$ la varianza del estimador del total se puede escribir como la var. del estim. monoetápico de HT + un término de penalización debido al submuestreo.

$$V(\hat{\theta}) = V_{1E}(\hat{\theta}_{HT}) + \text{PENALIZACIÓN}_{\text{subm.}}$$

$\hat{X} = \sum_{i=1}^{n} \frac{\hat{X}_i}{\pi_i} \longrightarrow \hat{X}_{HT}$ $\begin{cases} \text{insesgado} \\ V = \Sigma V + \Sigma COV \end{cases}$

$\circ\ E_2(\hat{X}) = E_2\left[\sum^n \frac{\hat{X}_i}{\pi_i}\right] = \sum^n \frac{X_i}{\pi_i} \longleftarrow \hat{X}_{HT}$

$\hookrightarrow V_1(E_2(\hat{X})) = V_1(\hat{X}_{HT}) = \sum_i^N \frac{X_i^2}{\pi_i^2} \cdot \pi_i(1-\pi_i) + \sum_{j \neq i} \frac{X_i}{\pi_i} \cdot \frac{X_j}{\pi_j} (\pi_{ij} - \pi_i \pi_j)$

$\circ\ V_2(\hat{X}) = V_2\left(\sum^n \frac{\hat{X}_i}{\pi_i}\right) = \sum^n \frac{1}{\pi_i^2} V_2(\hat{X}_i) = \sum^n \frac{1}{\pi_i^2} \cdot M_i^2 \cdot (1 - f_{2i}) \cdot \frac{S_i^2}{m_i}$

$\hookrightarrow E_1(V_2(\hat{X})) = E_1\left[\sum^N \frac{1}{\pi_i^2} \cdot M_i^2 \cdot (1 - f_{2i}) \cdot \frac{S_i^2}{m_i} \cdot e_i\right] \underset{E_1(e_i) = \pi_i}{=}$

$$= \underbrace{\sum_{i=1}^{N} \frac{1}{\pi_i} \cdot M_i^2 (1 - f_{2i}) \cdot \frac{S_i^2}{m_i}}_{\text{penalización subm.}}$$

Por lo que la varianza del estim. del total poblacional queda:

$$V(\hat{X}) = \sum_{i=1}^{N} \frac{X_i^2}{\pi_i^2}\pi_i(1-\pi_i) + \sum_{j\neq i}^{N} \frac{X_i}{\pi_i}\cdot\frac{X_j}{\pi_j}(\pi_{ij}-\pi_i\pi_j) + \sum_{i=1}^{N}\frac{1}{\pi_i}M_i^2(1-f_{2i})\cdot\frac{S_i^2}{m_i}$$

*Casos particulares:

Probab. primera etapa iguales $\longrightarrow \pi_i = \frac{n}{N},\ \forall i=1 \_ n$

$$\pi_{ij} = \frac{n(n-1)}{N(N-1)}$$

$$\hat{X} = \frac{N}{n}\sum_{i=1}^{n}\hat{X}_i = N\cdot\bar{\hat{X}}$$

$$V(\hat{X}) = N^2(1-f_1)\cdot\frac{S_c^2}{n} + \frac{N}{n}\sum_{i=1}^{N}M_i^2\cdot(1-f_{2i})\frac{S_i^2}{m_i}$$

donde: $f_1 = n/N \longrightarrow$ fracción muestreo 1ª etapa.

$S_c^2 = \frac{\sum(X_i-\bar{X})}{N-1} \longrightarrow$ cuasiv. poblacional 1ª etapa.

b) E1 CR $\longrightarrow$ La varianza del estimador del total poblacional en MCB es igual a la varianza del estimador monoetápico de Hansen y Hurwitz + un término de penalización debido al submuestreo.

$$\hat{X} = \sum_{i=1}^{n}\frac{\hat{X}_i}{nP_i} \longrightarrow \hat{X}_{HH}\begin{cases}\text{insesgado}\\ V = \frac{1}{n}\sum^{N}\left(\frac{X_i}{P_i}-X\right)^2\cdot P_i\end{cases}$$

Tma Madow $\Rightarrow V(\hat{X}) = E_1V_2(\hat{X}) + V_1E_2(\hat{X})$

- $E_2(\hat{X}) = E_2\left(\sum\frac{\hat{X}_i}{nP_i}\right) = \sum_{i=1}^{n}\frac{1}{nP_i}E_2(\hat{X}_i) = \sum_{i=1}^{n}\frac{X_i}{nP_i} \longleftarrow \hat{X}_{HH}$

  $\hookrightarrow V_1[E_2(\hat{X})] = V_1(\hat{X}_{HH}) = \frac{1}{n}\sum_{i=1}^{N}\left(\frac{X_i}{P_i}-X\right)^2\cdot P_i$

- $V_2(\hat{X}) = V_2\left(\sum^{n}\frac{\hat{X}_i}{nP_i}\right) = \sum^{n}\frac{1}{n^2P_i^2}V(\hat{X}_i) = \sum_{i=1}^{n}\frac{1}{n^2P_i^2}\cdot M_i^2\cdot(1-f_{2i})\frac{S_i^2}{m_i}$

  $\hookrightarrow E_1(V_2(\hat{X})) = E_1\left[\sum^{N}\frac{1}{n^2P_i^2}\cdot M_i^2\cdot(1-f_{2i})\cdot\frac{S_i^2}{m_i}e_i\right] = \longleftarrow E[e_i] = nP_i$

$$= \underbrace{\sum^{N}\frac{1}{nP_i}\cdot M_i^2\cdot(1-f_{2i})\cdot\frac{S_i^2}{m_i}}_{\text{penalit. subm.}}$$

luego la varianza del estimador bietápico del total queda:

$$V(\hat{X}) = \frac{1}{n}\sum_{i=1}^{N}\left(\frac{X_i}{P_i} - X\right)^2 P_i + \frac{1}{n}\sum^{N}\frac{1}{P_i} M_i^2 \cdot (1-f_{2i}) \cdot \frac{S_i^2}{m_i}$$

* Casos particulares:

- Si los conglomerados se eligen con probab. proporcionales a su tamaño:

$$P_i = \frac{M_i}{M}, \forall i = 1 \_ N \longrightarrow \hat{X} = \sum_{i=1}^{n}\frac{\hat{X}_i}{\frac{n}{M}\cdot M_i} = \frac{M}{n}\sum_{i=1}^{n}\bar{x}_i \quad (\hat{X}_i = M_i \bar{x}_i)$$

$$\rightarrow V(\hat{X}) = \frac{M}{n}\sum_{i=1}^{N} M_i(\bar{X}_i - \bar{\bar{X}})^2 + \frac{M}{n}\sum^{N} M_i(1-f_{2i})\cdot\frac{S_i^2}{m_i}$$

- Si el muestreo de la primera etapa se realiza con probab. iguales:

$$P_i = \frac{1}{N}, \forall i = 1 \_ N \longrightarrow \hat{X} = \frac{N}{n}\sum_{i=1}^{n}\hat{X}_i$$

$$\rightarrow V(\hat{X}) = \frac{N}{n}\sum^{N}(X_i - \bar{X})^2 + \frac{N}{n}\sum^{N} M_i^2(1-f_{2i})\cdot\frac{S_i^2}{m_i}$$

CONCLUSIÓN: si E2 es m.a.s.s.r.:

$$V(\hat{X}) = \text{Var. estim. monoetápico}\begin{Bmatrix} HT \\ HH \end{Bmatrix} + \text{Pérdid. submuestreo}$$

6_ MUESTRAS AUTOPONDERADAS

Una muestra bietápica se dice autoponderada, o más correctamente se dice que el tipo de muestreo genera muestras autoponderadas cuando se cumple que la var. aleatoria e_{ij} cte número de veces que la unidad secundaria u_{ij} aparece en la muestra, toma valor constante $\forall i,j \Rightarrow$ todas las observaciones X_{ij} realizadas sobre las unidades secundarias de la muestra, contribuyen por igual a la formación del estimador.

Si el muestreo es sin reposición (E2SR), la condición es equivalente a exigir que todas las unidades de segunda etapa tengan la misma probabilidad de aparecer en la muestra.

Por ejemplo, el estimador del total, dependiendo del tipo de muestreo en la 1ª etapa es:

$$\left.\begin{array}{ll} (SR) \longrightarrow & \hat{X} = \sum_{i=1}^{n} \frac{\hat{X}_i}{\pi_i} \\ (CR) \longrightarrow & \hat{X} = \sum_{i=1}^{n} \frac{\hat{X}_i}{nP_i} \end{array}\right\} \text{ En general, } \hat{X} = \sum_{i=1}^{n} \frac{\hat{X}_i}{\alpha_i}$$

Si hacemos intervenir la 2ª etapa:

$$\hat{X} = \sum_{i=1}^{n} \frac{1}{\alpha_i} \sum_{j=1}^{m_i} \frac{M_i}{m_i} X_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{M_i}{\alpha_i m_i} X_{ij}$$

cuya expresión no varía si se multiplica por $e_{ij} \equiv$ nº veces que aparece en la muestra la unidad secundaria u_{ij}.

$$\hat{X} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{M_i}{m_i \alpha_i} X_{ij} e_{ij} \qquad ¿\sum^{N} \sum^{M_i} (\quad) e_{ij}? \quad e_{ij} = \begin{cases} 1 \\ 0 \end{cases}$$

Para que el estimador sea insesgado:

$$E[\hat{X}] = X \iff \frac{M_i}{\alpha_i m_i} = \frac{1}{E[e_{ij}]} \equiv \text{factor de elevación}$$

La importancia de las muestras autoponderadas reside en que en ellas el factor de elevación es constante, y tb $E[e_{ij}]$, de modo que los estimadores insesgados del ~~la población~~ total de la población admiten la forma:

$$\hat{X} = K_0 \cdot \mathcal{X}$$

donde $K_0 \equiv$ factor de elevación ~~cte~~ , $K_0 = \frac{M_i}{\alpha_i m_i}$

$\mathcal{X} \equiv$ total muestral , $\mathcal{X} = \sum_{i=1}^{n} \sum_{j=1}^{m_j} X_{ij}$

Otra ventaja de las muestras autoponderadas es la sencillez del cálculo práctico.

a) SIN reposición en la 1ª etapa:

• Si π_i son proporcionales a los tamaños M_i, $\pi_i = n \cdot \frac{M_i}{M} \Rightarrow$

$$\hat{X} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{M_i}{m_i} X_{ij} \Big/ \frac{n M_i}{M} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{M}{n m_i} X_{ij}$$

la muestra será autoponderada si m_i igual para todos los conglom.

$\frac{M}{n m_i}$ cte $\Rightarrow m_i = \bar{m}$, $i = 1 \ldots n$ $\longrightarrow K_0 = \frac{N}{n \bar{m}}$

por lo que $\hat{X} = \frac{M}{n \bar{m}} \mathcal{X} = \frac{\mathcal{X}}{f}$

donde $f = \frac{n \bar{m}}{M} \equiv$ fracción global de muestreo

• Si π_i son iguales , $\pi_i = \frac{n}{N}, \forall i \Rightarrow$ $\left(M_i = \frac{M}{N}\right)$.

$$\hat{X} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{M_i}{m_i} X_{ij} \Big/ \frac{n}{N} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{N}{n} \cdot \frac{M_i}{m_i} X_{ij}$$

la muestra será autoponderada si la fracción de muestreo en la segunda etapa es igual para todos los conglom:

$\frac{N}{n} \cdot \frac{M_i}{m_i}$ cte $\Leftrightarrow \frac{M_i}{m_i} = \frac{1}{f_{2i}} =$ cte $\Rightarrow f_{2i} = \frac{m_i}{M_i} = f_2$ cte.

luego $\hat{X} = \frac{N}{n} \cdot \frac{1}{f_2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} X_{ij} = \frac{1}{f_1 \cdot f_2} \mathcal{X}$

b) CON reposición en la 1ª etapa

- Para probab. P_i proporcionales a los tamaños M_i:

$$P_i = \frac{M_i}{M} \Rightarrow \hat{X} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{M_i}{m_i} X_{ij} \Big/ n \cdot \frac{M_i}{M} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{M}{n \cdot m_i} X_{ij}$$

$\frac{M}{n m_i}$ cte $\Rightarrow m_i = \overline{m}$, cte $\forall i = 1 \ldots n$

La muestra será autoponderada si de cada conglomerado se seleccionan en 2ª etapa el mismo nº de unid. elementales.

$$\hat{X} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{M}{n\overline{m}} X_{ij} = \frac{M}{n \cdot \overline{m}} X = \frac{1}{f} X \quad (\text{tq } f = \frac{n \cdot \overline{m}}{M}).$$

donde $f \equiv$ fracción global de muestreo

- Para probab. P_i iguales:

$$P_i = \frac{1}{N} \Rightarrow \hat{X} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{M_i}{m_i} X_{ij} \Big/ n \frac{1}{N} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{N}{n} \cdot \frac{M_i}{m_i} X_{ij}$$

$\frac{N}{n} \cdot \frac{M_i}{m_i}$ cte $\Rightarrow \frac{M_i}{m_i} = \frac{1}{f_{2i}}$ cte $\forall i \Rightarrow f_{2i} = f_2$ $\quad f_{2i} = \frac{m_i}{M_i}$ cte

La muestra será autoponderada cuando la fracción de muestreo de segunda etapa sea igual para todas las unidades primarias. Luego:

$$\hat{X} = \frac{1}{f_1 \cdot f_2} X.$$

Conclusión: Da igual $\left\{ \begin{array}{l} E1SR \\ E1CR \end{array} \right\}$.

si a) Probab. proporcionales al tamaño, entonces autoponderada cuando

a) $m_i = \overline{m}, \forall i$

$\hat{X} = \frac{1}{f} \cdot X$.

b) Prob. iguales, b) $f_2 =$, $\forall$ unid. pri

$\hat{X} = \frac{1}{f_1 \cdot f_2} \cdot X$.