EJERCICIOS PEDIDOS POR EL TP

EJERCICIOS DE LA GUÍA 2

Ejercicio 7.

a.

$$\hat{\overline{y}}^* = \begin{cases} \overline{y}_\pi - C & y_1 \notin s; y_N \in s \\ \overline{y}_\pi + C & y_1 \in s; y_N \notin s \\ \overline{y}_\pi & y_1 \in s; y_N \in s \end{cases}$$

Demostración. $\hat{\overline{y}}^{*}$[1] es insesgado.

$S_A = \{s \in \Omega / y_1 \notin s; y_N \in s\}$

$S_B = \{s \in \Omega / y_1 \in s; y_N \notin s\}$

$S_C = \overline{S_A \cup S_B} = \{s \in \Omega / (y_1 \in s \wedge y_N \in s) \vee (y_1 \notin s \wedge y_N \notin s)\}$

$\bigcup_{\{A;B;C\}} S_i = \Omega$

$$\mathbb{E}\left(\hat{\overline{y}}^*(S)\right) = \sum_{s \in \Omega} p(s) \cdot \hat{\overline{y}}^*(s)$$

$\Omega = S_A + S_B + S_C$

$$\mathbb{E}\left(\hat{\overline{y}}^*(S)\right) = \sum_{s \in S_A} p(s) \cdot \hat{\overline{y}}^*(s) + \sum_{s \in S_B} p(s) \cdot \hat{\overline{y}}^*(s) + \sum_{s \in S_C} p(s) \cdot \hat{\overline{y}}^*(s)$$

$$\mathbb{E}\left(\hat{\overline{y}}^*(S)\right) = \sum_{s \in S_A} p(s) \cdot (\overline{y}_\pi(s) - C) + \sum_{s \in S_B} p(s) \cdot (\overline{y}_\pi(s) + C) + \sum_{s \in S_C} p(s) \cdot \overline{y}_\pi(s)$$

$$\mathbb{E}\left(\hat{\overline{y}}^*(S)\right) = \sum_{s \in S_A} p(s) \cdot \overline{y}_\pi(s) - \sum_{s \in S_A} p(s) \cdot C + \sum_{s \in S_B} p(s) \cdot \overline{y}_\pi(s) + \sum_{s \in S_B} p(s) \cdot C + \sum_{s \in S_C} p(s) \cdot \overline{y}_\pi(s)$$

$$\mathbb{E}\left(\hat{\overline{y}}^*(S)\right) = \left[\sum_{s \in S_A} p(s) \cdot \overline{y}_\pi(s) + \sum_{s \in S_B} p(s) \cdot \overline{y}_\pi(s) + \sum_{s \in S_C} p(s) \cdot \overline{y}_\pi(s)\right] - \sum_{s \in S_A} p(s) \cdot C + \sum_{s \in S_B} p(s) \cdot C$$

$$\mathbb{E}\left(\hat{\overline{y}}^*(S)\right) = \mathbb{E}(\overline{y}_\pi(s)) - C \cdot \sum_{s \in S_A} p(s) + C \cdot \sum_{s \in S_B} p(s)$$

$\sum_{s \in S_A} p(s) = p(S_A) = p\left(\underset{1}{\mathbb{I}} = 1; \underset{N}{\mathbb{I}} = 0\right) = \pi_1 - \pi_{1N}$

Como estamos en MSA, todos los π_i y π_{ij} son del mismo valor, con $\forall i : \pi_i = \frac{n}{N} = f$ y $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$.

$\sum_{s \in S_B} p(s) = p(S_B) = p\left(\underset{1}{\mathbb{I}} = 0; \underset{N}{\mathbb{I}} = 1\right) = \pi_N - \pi_{1N} = \pi_1 - \pi_{1N}$

$$\mathbb{E}\left(\hat{\overline{y}}^*(S)\right) = \mathbb{E}(\overline{y}_\pi(s)) - C \cdot (\pi_1 - \pi_{1N}) + C \cdot (\pi_1 - \pi_{1N})$$

[1] Favor de leer «Y media sombrero estrella».

$$\mathbb{E}\left(\hat{\overline{y}}^{*}(S)\right)=\mathbb{E}\left(\overline{y}_{\pi}(s)\right)$$

Ya probamos que el π-estimador de la media es insesgado:

$$\mathbb{E}\left(\hat{\overline{y}}^{*}(S)\right)=\overline{y}_{U}$$

$\square$

b Verificar la fórmula de la varianza. Se hace evidente que vamos a necesitar desarrollar de alguna forma la ecuación e intentar aproximarla

$$V\left(\hat{\overline{y}}^{*}(S)\right)=\mathbb{E}\left[\left(\hat{\overline{y}}^{*}\right)^{2}(S)\right]-\mathbb{E}^{2}\left[\hat{\overline{y}}^{*}(S)\right]$$

Buscamos $\mathbb{E}\left[\left(\hat{\overline{y}}^{*}\right)^{2}(S)\right]$.

$$\mathbb{E}\left[\left(\hat{\overline{y}}^{*}\right)^{2}(S)\right]=\sum_{s\in\Omega}p(s)\cdot\left(\hat{\overline{y}}^{*}\right)^{2}(s)$$

$$\mathbb{E}\left(\hat{\overline{y}}^{*}(S)\right)=\sum_{s\in S_A}p(s)\cdot\left(\overline{y}_{\pi}(s)-C\right)^{2}+\sum_{s\in S_B}p(s)\cdot\left(\overline{y}_{\pi}(s)+C\right)^{2}+\sum_{s\in S_C}p(s)\cdot\overline{y}_{\pi}^{2}(s)$$

$$\mathbb{E}\left(\hat{\overline{y}}^{*}(S)\right)=\sum_{s\in S_A}p(s)\cdot\left(\overline{y}_{\pi}^{2}(s)+2C\overline{y}_{\pi}(s)+C^{2}\right)+\sum_{s\in S_B}p(s)\cdot\left(\overline{y}_{\pi}^{2}(s)-2C\overline{y}_{\pi}(s)+C^{2}\right)+\sum_{s\in S_C}p(s)\cdot\overline{y}_{\pi}^{2}(s)$$

$$\begin{aligned}\mathbb{E}\left(\hat{\overline{y}}^{*}(S)\right)=&\sum_{s\in S_A}p(s)\cdot\overline{y}_{\pi}^{2}(s)-2C\cdot\sum_{s\in S_A}p(s)\,\overline{y}_{\pi}(s)+\sum_{s\in S_A}p(s)\cdot C^{2}\\&+\sum_{s\in S_B}p(s)\cdot\overline{y}_{\pi}^{2}(s)+2C\sum_{s\in S_B}p(s)\cdot\overline{y}_{\pi}(s)+C^{2}\sum_{s\in S_B}p(s)+\sum_{s\in S_C}p(s)\cdot\overline{y}_{\pi}^{2}(s)\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left(\hat{\overline{y}}^{*}(S)\right)=&\left[\sum_{s\in S_A}p(s)\cdot\overline{y}_{\pi}^{2}(s)+\sum_{s\in S_B}p(s)\cdot\overline{y}_{\pi}^{2}(s)+\sum_{s\in S_C}p(s)\cdot\overline{y}_{\pi}^{2}(s)\right]\\&+2C\cdot\sum_{s\in S_A}p(s)\,\overline{y}_{\pi}(s)-2C\sum_{s\in S_B}p(s)\cdot\overline{y}_{\pi}(s)+C^{2}\cdot\sum_{s\in S_A}p(s)+C^{2}\cdot\sum_{s\in S_B}p(s)\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left(\hat{\overline{y}}^{*}(S)\right)=&\,\mathbb{E}\left(\overline{y}_{\pi}^{2}\right)\\&-2C\cdot\sum_{s\in S_A}p(s)\,\overline{y}_{\pi}(s)+2C\sum_{s\in S_B}p(s)\cdot\overline{y}_{\pi}(s)+C^{2}\cdot(\pi_1-\pi_{N1})+C^{2}\cdot(\pi_1-\pi_{N1})\end{aligned}$$

$$\mathbb{E}\left(\hat{\overline{y}}^{*}(S)\right)=\mathbb{E}\left(\overline{y}_{\pi}^{2}\right)-2C\cdot\left(\sum_{s\in S_A}p(s)\,\overline{y}_{\pi}(s)-\sum_{s\in S_B}p(s)\cdot\overline{y}_{\pi}(s)\right)+2C^{2}\cdot(\pi_1-\pi_{N1})$$

Volvemos a la fórmula. Considerando que conocemos $\mathbb{E}\left(\overline{y}_\pi\right) = \overline{y}_U$.

$$V\left(\hat{\overline{y}}^*(S)\right) = \mathbb{E}\left[\left(\hat{\overline{y}}^*\right)^2(S)\right] - \mathbb{E}^2\left[\hat{\overline{y}}^*(S)\right]$$

$$V\left(\hat{\overline{y}}^*(S)\right) = \left[\mathbb{E}\left(\overline{y}_\pi^2\right) + 2C\cdot\left(\sum_{s\in S_A} p(s)\,\overline{y}_\pi(s) - \sum_{s\in S_B} p(s)\cdot\overline{y}_\pi(s)\right) + 2C^2\cdot\left(\pi_1 - \pi_{N1}\right)\right] - \mathbb{E}^2\left[\hat{\overline{y}}^*(S)\right]$$

$$V\left(\hat{\overline{y}}^*(S)\right) = \mathbb{E}\left(\overline{y}_\pi^2\right) - \mathbb{E}^2\left[\hat{\overline{y}}^*(S)\right] - 2C\cdot\left[\left(\sum_{s\in S_A} p(s)\,\overline{y}_\pi(s) - \sum_{s\in S_B} p(s)\cdot\overline{y}_\pi(s)\right) - C\cdot\left(\pi_1 - \pi_{N1}\right)\right]$$

Bajo MSA, la probabilidad de cada muestra es la misma: $\forall s\in\Omega : p(s) = 1/\binom{N}{n}$.

$$V\left(\hat{\overline{y}}^*(S)\right) = V\left(\overline{y}_\pi^2\right) - 2C\left[p(s)\cdot\left(\sum_{s\in S_A}\overline{y}_\pi(s) - \sum_{s\in S_B}\overline{y}_\pi(s)\right) - C\cdot\left(\pi_1 - \pi_{N1}\right)\right]$$

$\overline{y}_\pi(s) = \sum_{i=1}^N e_i\cdot s$
$\forall s\in S_A : \exists j/e_j\cdot s = N$
$\forall s\in S_B : \exists j/e_j\cdot s = 1$

$N_A = N_B = \binom{N-2}{n-1}$, ya que son las colas que se arman cuando se fija uno de los elementos atípicos y se excluye el otro.

En un punto aparte vamos a desarrollar la idea de que $\sum_{s\in S_A}\overline{y}_\pi(s) - \sum_{s\in S_B}\overline{y}_\pi(s) = N_A\cdot\left(\frac{y_N}{n} - \frac{y_1}{n}\right)$.

$$V\left(\hat{\overline{y}}^*(S)\right) = V\left(\overline{y}_\pi^2\right) - 2C\left[\frac{p(s)}{n}\cdot\left(\sum_{s\in S_A}\sum_{i=1}^{N_A} e_i\cdot s - \sum_{s\in S_B}\sum_{i=1}^{N_B} e_i\cdot s\right) - C\cdot\left(\pi_1 - \pi_{N1}\right)\right]$$

$$V\left(\hat{\overline{y}}^*(S)\right) = V\left(\overline{y}_\pi^2\right) - 2C\cdot\left[\frac{p(s)}{n}\cdot\left(N_A\cdot y_N - N_A\cdot y_1\right) - C\cdot\left(\pi_1 - \pi_{N1}\right)\right]$$

$$V\left(\hat{\overline{y}}^*(S)\right) = V\left(\overline{y}_\pi^2\right) - 2C\left[\frac{1}{\binom{N}{n}}\cdot\binom{N-2}{n-1}\left(y_N - y_1\right) - C\cdot\left(\pi_1 - \pi_{N1}\right)\right]$$

$\frac{1}{\binom{N}{n}}\cdot\binom{N-2}{n-1} = \frac{(N-2)!}{(n-1)!((N-2)-(n-1))!}\cdot\frac{n!(N-n)!}{N!} = \frac{(N-n)!}{(N-n-1)!}\cdot\frac{n!}{(n-1)!}\cdot\frac{(N-2)!}{N!}$
$\frac{1}{\binom{N}{n}}\cdot\binom{N-2}{n-1} = \frac{(N-n)\cdot n}{N(N-1)}$

$$V\left(\hat{\overline{y}}^*(S)\right) = V\left(\overline{y}_\pi^2\right) - 2C\cdot\left[\frac{1}{n}\cdot\frac{(N-n)\cdot n}{N(N-1)}\cdot\left(y_N - y_1\right) - C\cdot\left(\pi_1 - \pi_{N1}\right)\right]$$

Además, $\pi_1 - \pi_{N1} = \frac{n}{N} - \frac{n(n-1)}{N(N-1)} = \frac{n(N-1)-n(n-1)}{N(N-1)} = \frac{nN-n^2}{N(N-1)} = \frac{n(N-n)}{N(N-1)}$

$$V\left(\hat{\overline{y}}^*(S)\right) = V\left(\overline{y}_\pi^2\right) - 2C\cdot\left[\frac{(N-n)}{N(N-1)}\cdot\left(y_N - y_1\right) - C\cdot\frac{n(N-n)}{N(N-1)}\right]$$

$$V\left(\hat{\overline{y}}^*(S)\right) = V\left(\overline{y}_\pi^2\right) - 2C\frac{(N-n)}{N(N-1)}\cdot\left[\left(y_N - y_1\right) - nC\right]$$

$$V\left(\hat{\overline{y}}^{*}(S)\right) = V\left(\overline{y}_{\pi}^{2}\right) - 2C \cdot \frac{N-n}{N} \cdot \frac{1}{N-1} \cdot \left[(y_N - y_1) - nC\right]$$

$$V\left(\hat{\overline{y}}^{*}(S)\right) = V\left(\overline{y}_{\pi}^{2}\right) - (1-f) \cdot \frac{2C}{N-1} \cdot \left[(y_N - y_1) - nC\right]$$

$V\left(\overline{y}_{\pi}^{2}\right)$ es conocida.

$$V\left(\hat{\overline{y}}^{*}(S)\right) = \frac{1-f}{n} \cdot S_{yU}^{2} - (1-f) \cdot \frac{2C}{N-1} \cdot \left[(y_N - y_1) - nC\right]$$

$$V\left(\hat{\overline{y}}^{*}(S)\right) = (1-f) \cdot \left[\frac{S_{yU}^{2}}{n} - \frac{2C}{N-1} \cdot (y_N - y_1 - nC)\right]$$

Identidad usada: $\sum_{s \in S_A} \overline{y}_{\pi}(s) - \sum_{s \in S_B} \overline{y}_{\pi}(s) = N_A \cdot \left(\frac{y_N}{n} - \frac{y_1}{n}\right)$

La idea coloquial a la que acudimos acá es que cada conjunto tiene todas las muestras donde está su elemento característico pero no está el otro y por lo demás sus «colas» son iguales.

La media está fuertemente emparentada con la suma de los totales. Es claro que cada conjunto de S_A tiene un análogo en S_B que es igual en todos sus elementos excepto en $1/N$. Entonces, es fácil ver que la diferencia entre estos dos es $y_N - y_1$. Podemos agrupar así los $N_A = \binom{N-2}{n-1}$ conjuntos obteniendo siempre la misma diferencia.

$$\sum_{s \in S_A} \overline{y}_{\pi}(s) - \sum_{s \in S_B} \overline{y}_{\pi}(s) = \sum_{a_i \in S_A} \sum_{j=1}^{n} \frac{a_{ij}}{n} - \sum_{b_i \in S_B} \sum_{j=1}^{n} \frac{b_{ij}}{n}$$

Reordenamos S_B para agrupar los conjuntos similares:
$S'_B = \left\{ b'_{ij} / b'_{ij} \in S_B \wedge b'_{ij} \setminus a_{ij} = y_1 \wedge a_{ij} \setminus b'_{ij} = y_N \right\}$
$S'_B = S_B$

$$\sum_{s \in S_A} \overline{y}_{\pi}(s) - \sum_{s \in S_B} \overline{y}_{\pi}(s) = \frac{1}{n}\left[\sum_{s_i \in S_A} \sum_{j=1}^{n} a_{ij} - \sum_{s_i \in S'_B} \sum_{j=1}^{n} b'_{ij}\right]$$

$\#S_B = \#S_A = \binom{N-2}{n-1} = N_A = N_B$

$$\sum_{s \in S_A} \overline{y}_{\pi}(s) - \sum_{s \in S_B} \overline{y}_{\pi}(s) = \frac{1}{n}\left[\sum_{i=1}^{N_A} \sum_{j=1}^{n} a_{ij} - \sum_{i=1}^{N_A} \sum_{j=1}^{n} b'_{ij}\right]$$

$$\sum_{s \in S_A} \overline{y}_{\pi}(s) - \sum_{s \in S_B} \overline{y}_{\pi}(s) = \frac{1}{n}\left[\sum_{i=1}^{N_A} \sum_{j=1}^{n} \left(a_{ij} - b'_{ij}\right)\right]$$

Estamos comparando siempre los conjuntos que sólo se diferencian en los elementos y_N e y_1, ya que $b'_{ij} \setminus a_{ij} = y_1 \wedge a_{ij} \setminus b'_{ij} = y_N$.

$$\sum_{s \in S_A} \overline{y}_{\pi}(s) - \sum_{s \in S_B} \overline{y}_{\pi}(s) = \frac{1}{n} \sum_{i=1}^{N_A} (y_N - y_1)$$

$$\sum_{s \in S_A} \overline{y}_{\pi}(s) - \sum_{s \in S_B} \overline{y}_{\pi}(s) = \frac{N_A}{n} (y_N - y_1)$$

c ¿Para qué valores de C el nuevo estimador es más preciso que el π-estimador?

$$\frac{V\left(\hat{\overline{y}}^{*}(S)\right)}{V\left(\overline{y}_{\pi}(S)\right)}=\frac{(1-f)\cdot\left[\frac{S_{yU}^{2}}{n}-\frac{2C}{N-1}\cdot\left(y_{N}-y_{1}-nC\right)\right]}{(1-f)\cdot\frac{S_{yU}^{2}}{n}}$$

$$\frac{V\left(\hat{\overline{y}}^{*}(S)\right)}{V\left(\overline{y}_{\pi}(S)\right)}=1-\frac{2C\cdot\left({}^{y_N}\!/\!_{n}-{}^{y_1}\!/\!_{n}-C\right)}{(N-1)\,S_{yU}^{2}}$$

$\frac{V\left(\hat{\overline{y}}^{*}(S)\right)}{V\left(\overline{y}_{\pi}(S)\right)}<1\Rightarrow\frac{2C\cdot\left({}^{y_N}\!/\!_{n}-{}^{y_1}\!/\!_{n}+C\right)}{(N-1)S_{yU}^{2}}>0$

$$\frac{2C\cdot\left({}^{y_N}\!/\!_{n}-{}^{y_1}\!/\!_{n}-C\right)}{(N-1)\,S_{yU}^{2}}>0$$

$\forall N;\Omega:(N-1)\,S_{yU}^{2}>0$
Busquemos el caso límite:

$$2C\cdot\left({}^{y_N}\!/\!_{n}-{}^{y_1}\!/\!_{n}-C\right)>0$$

$$2C\cdot\left({}^{y_N}\!/\!_{n}-{}^{y_1}\!/\!_{n}-C\right)>0$$

$C>0$
Si $2C=0$, estamos ante el estimador $\overline{y}_{\pi}$. No analizamos este caso.

$${}^{y_N}\!/\!_{n}-{}^{y_1}\!/\!_{n}-C>0$$

$${}^{y_N}\!/\!_{n}-{}^{y_1}\!/\!_{n}>C$$

$$y_N-y_1>nC$$

$\blacksquare$

$$C<\frac{y_N-y_1}{n}$$

Esto condice con lo que sería intuitivamente esperable: El indicador disminuye la varianza respecto al tradicional siempre y cuando C esté «sustrayendo» la diferencia introducida por los atípicos y_1 e y_N, es decir si es menor que el aporte que hace esta diferencia.

13 Suponer que $\forall k\in U:y_k=k$. Demostrar que $V_{SIST}\leq V_{MSA}$ del estimador de la media para una muestra de tamaño n del total N de elementos de U con $a=\frac{N}{n}$ entero. Es inmediato que hay una alta heterogeneidad entre muestras.

Probar lo pedido equivale a mostrar que el efecto de diseño tiene un valor mayor a 1. Este indicador tiene una forma conocida que se desarrolló detaladadamente en Särndall.

$$\text{efdis}\left(MS;\hat{t}_{\pi}\right)=\frac{V_{SIST}}{V_{MSA}}=1+\frac{n-1}{1-f}\cdot\delta$$

Para comprobar lo pedido hay que observar si $\delta<0$:

$$\delta=1-\frac{N-1}{N-a}\cdot\frac{SCD}{SCT}$$

$$1-\frac{N-1}{N-a}\cdot\frac{SCD}{SCT}<0$$
$$-\frac{N-1}{N-a}\cdot\frac{SCD}{SCT}<-1$$
$$\frac{N-1}{N-a}\cdot\frac{SCD}{SCT}>1$$

$SCD=\sum_{r=1}^{a}\sum_{s_r}\left(y_k-\overline{y}_{sr}\right)^2$. Se observa en este caso que las muestras s_r son de la forma $\{s_{ir}\}_{i=0}^{n}/s_{ir}=ia+r$ con r una constante fija para cada muestra, $1<r<a$.

$\overline{y}_{sr}=\frac{1}{n}\cdot\sum_{i=0}^{n-1}(i\cdot a+r)$
$\overline{y}_{sr}=\frac{n}{n}r+\frac{a}{n}\cdot\sum_{i=1}^{n-1}i$
$\overline{y}_{sr}=r+\frac{a}{\cancel{n}}\cdot\frac{\cancel{n}(n-1)}{2}i$
La suma de esta serie finita es conocida.
$\overline{y}_{sr}=r+a\frac{n-1}{2}$

- $$\overline{y}_{sr}=r+a\cdot\frac{n-1}{2}$$
$$SCD=\sum_{r=1}^{a}\sum_{s_r}\left(y_k-\overline{y}_{sr}\right)^2$$
$$SCD=\sum_{r=1}^{a}\sum_{i=0}^{n-1}\left((i\cdot a+r)-r-a\cdot\frac{n-1}{2}\right)^2$$
$$SCD=\frac{a^2}{4}\sum_{r=1}^{a}\sum_{i=0}^{n-1}\left(2i-(n-1)\right)^2$$
$$SCD=\frac{a^2}{4}\sum_{r=1}^{a}\sum_{i=0}^{n-1}\left[4i^2-4i\,(n-1)+(n-1)^2\right]$$

$\sum_{k=1}^{N}k=\frac{n\cdot(n+1)}{2}$
$\sum_{k=1}^{N}k^2=\frac{n\cdot(n+1)(2n+1)}{6}$

$$SCD=\frac{a^2}{4}\sum_{r=1}^{a}\left[4\frac{(n-1)\cdot n\,(2n-1)}{6}-4\frac{n\cdot(n-1)}{2}(n-1)+n\,(n-1)^2\right]$$
$$SCD=\frac{a^2}{12}n\cdot(n-1)\sum_{r=1}^{a}\left[2\,(2n-1)-6\,(n-1)+3\,(n-1)\right]$$
$$SCD=\frac{a^2}{12}n\cdot(n-1)\sum_{r=1}^{a}(4n-2-6n+6+3n-3)$$
$$SCD=\frac{a^2}{12}n\cdot(n-1)\sum_{r=1}^{a}(n+1)$$
$$SCD=\frac{a^2}{12}n\cdot(n-1)\,a\,(n+1)$$

$N=an$

$$SCD = \frac{1}{12} N \cdot (N-a)(N+a)$$

$SCT = \sum_U (y_k - \overline{y}_U)^2$. $\overline{y}_U = \frac{1}{N} \sum_{i=1}^{N} i$

Esta es la misma serie finita que nos cruzamos en $\overline{y}_{sr}$.

$\overline{y}_U = \frac{1}{N} \frac{N(N+1)}{2}$

$$\overline{y}_U = \frac{N+1}{2}$$

$$\sum_U (y_k - \overline{y}_U)^2 = \sum_{k=1}^{N} (k - \overline{y}_U)^2$$

$\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2$

$$\sum_U (y_k - \overline{y}_U)^2 = \left(\sum_{k=1}^{N} k^2\right) - N\overline{y}_U^2$$

$\sum_{k=1}^{N} k^2 = \frac{n \cdot (n+1)(2n+1)}{6}$

$$\sum_U (y_k - \overline{y}_U)^2 = \frac{N \cdot (N+1)(2N+1)}{6} - N\overline{y}_U^2$$

$\overline{y}_U = \frac{N+1}{2}$

$$\sum_U (y_k - \overline{y}_U)^2 = \frac{N \cdot (N+1)(2N+1)}{6} - N\left(\frac{N+1}{2}\right)^2$$

$$\sum_U (y_k - \overline{y}_U)^2 = (N+1)\left[\frac{N \cdot (2N+1)}{6} - N\frac{N+1}{4}\right]$$

$$\sum_U (y_k - \overline{y}_U)^2 = \frac{1}{12} N \cdot (N+1)(N-1)$$

$$\sum_U (y_k - \overline{y}_U)^2 = \frac{1}{12} N \cdot (N+1)(N-1)$$

Podemos armar el cociente

$$\frac{SCD}{SCT} \cdot \frac{N-1}{N-a} = \frac{\frac{1}{12} N \cdot (N-a)(N+a)}{\frac{1}{12} N \cdot (N+1)(N-1)} \cdot \frac{N-1}{N-a}$$

$$\frac{SCD}{SCT} \cdot \frac{N-1}{N-a} = \frac{N+a}{N+1}$$

$a > 1$

Ya tenemos lo que necesitamos. Esta condición es razonable ya que si $a = 1$ estamos en realidad ante un MSA.

-

$$\delta \leq 1$$

Empleando las relaciones antes citadas, es evidente que

$$\frac{V_{MS}}{V_{MSA}} \leq 1$$

$$V_{MS} \leq V_{MSA}$$

Ejercicios de la Guía 3

3 Considerando el ejercicio 13 de la práctica 2, para la población de N unidades dividida en $H = n$ estratos de cada uno de los cuales se selecciona un elemento, probar que $V_{EST} \leq V_{SIS}$.

> 13-2: «Suponer que $\forall k \in U : y_k = k$. Demostrar que $V_{SIST} \leq V_{MSA}$ del estimador de la media para una muestra de tamaño n del total N de elementos de U con $a = \frac{N}{n}$ entero.»

En este caso, no desarrollamos una fórmula especial para el efecto diseño, así que conviene comparar directamente las dos expresiones de varianza.

$$V_{SIS}\left(\hat{t}_\pi\right) = a \sum_{r=1}^{a} \left(t_{s_r} - \overline{t}\right)^2$$

$$V_{EST}\left(\hat{t}_\pi\right) = \sum_{h=0}^{H-1} N_h^2 \frac{1 - f_h}{n_h} S_{yU_h}^2$$

- Se aclaró que los N_h son todos iguales. Además, se toma un elemento por estrato.

$H = n = \frac{N}{a}$
$\forall 1 < h < H : (N_h; n_h) = (a; 1)$

$$V_{EST}\left(\hat{t}_\pi\right) = \sum_{h=0}^{H-1} N_h^2 \frac{1 - \frac{1}{N_h}}{1} S_{yU_h}^2$$

$$V_{EST}\left(\hat{t}_\pi\right) = \sum_{h=0}^{H-1} N_h \frac{N_h - 1}{1} S_{yU_h}^2$$

$$V_{EST}\left(\hat{t}_\pi\right) = \sum_{h=0}^{n-1} a\,(a-1)\, S_{yU_h}^2$$

$$V_{EST}\left(\hat{t}_\pi\right) = a\,(a-1) \sum_{h=0}^{n-1} S_{yU_h}^2$$

Buscamos $S_{yU_h}^2$ *y* t_{s_r} *y* $\overline{t}$. Cada uno de los n estratos tiene a elementos. Por la expresión elegida, $y_{hi} = ah+i$, los n estratos están numerados con etiquetas $\{h\}_0^{n-1}$.

$$S_{yU_h}^2 = \frac{\sum_{i=1}^{a} \left(y_{i+h} - \overline{y}_h\right)^2}{a-1}$$

$\overline{y}_{hEST} = \frac{\sum_{i=1}^{a} ah+i}{a}$
$\overline{y}_{hEST} = ah + \frac{\frac{a(a+1)}{2}}{a}$
$\overline{y}_{hEST} = ah + \frac{a+1}{2}$

$$S_{yU_h}^2 = \frac{\sum_{i=1}^{a}\left[ah+i-\left(ah+\frac{a+1}{2}\right)\right]^2}{a-1}$$

$$S_{yU_h}^2 = \frac{\sum_{i=1}^{a}\left(i-\frac{a+1}{2}\right)^2}{a-1}$$

$$S_{yU_h}^2 = \frac{1}{4}\frac{\sum_{i=1}^{a}\left[2i-(a+1)\right]^2}{a-1}$$

$$S_{yU_h}^2 = \frac{1}{4}\frac{\sum_{i=1}^{a}\left[4i^2-4i\left(a+1\right)+\left(a+1\right)^2\right]}{a-1}$$

$\sum_{k=1}^{N} k = \frac{n\cdot(n+1)}{2}$
$\sum_{k=1}^{N} k^2 = \frac{n\cdot(n+1)(2n+1)}{6}$

$$S_{yU_h}^2 = \frac{1}{4}\frac{4\frac{a\cdot(a+1)(2a+1)}{6}-4\frac{a\cdot(a+1)}{2}\left(a+1\right)+a\left(a+1\right)^2}{a-1}$$

$$S_{yU_h}^2 = \frac{a\left(a+1\right)}{12}\frac{2\cdot(2a+1)-6\cdot(a+1)+3\left(a+1\right)}{a-1}$$

$$S_{yU_h}^2 = \frac{a\left(a+1\right)}{12}\frac{4a+2-6a-6+3a+3}{a-1}$$

$$S_{yU_h}^2 = \frac{a\left(a+1\right)}{12}\frac{a-1}{a-1}$$

- $$S_{yU_h}^2 = \frac{a\left(a+1\right)}{12}$$

$\overline{y}_{sr} = r + a\frac{n-1}{2}$

$$t_{s_r} = n\overline{y}_{sr}$$

$$t_{s_r} = nr + a\frac{\left(n-1\right)n}{2}$$

- $$t_{s_r} = n\left(r+\frac{N-a}{2}\right)$$

$$\bar{t} = \frac{t}{a}$$

$t = \sum_{i=1}^{N} i = \frac{N(N+1)}{2}$

- $$\bar{t} = \frac{N\left(N+1\right)}{2a}$$

Determinamos ambas varianzas con los datos hallados.

$$V_{SIS}\left(\hat{t}_\pi\right)=a\sum_{r=1}^{a}\left(t_{s_r}-\bar{t}\right)^2$$

$$V_{SIS}\left(\hat{t}_\pi\right)=a\sum_{r=1}^{a}\left(n\left(r+\frac{N-a}{2}\right)-\frac{N\left(N+1\right)}{2a}\right)^2$$

$na=N$

$$V_{SIS}\left(\hat{t}_\pi\right)=an^2\sum_{r=1}^{a}\left(r+\frac{N-a}{2}-\frac{N+1}{2}\right)^2$$

$$V_{SIS}\left(\hat{t}_\pi\right)=an^2\sum_{r=1}^{a}\left(r-\frac{a+1}{2}\right)^2$$

$$V_{SIS}\left(\hat{t}_\pi\right)=n^2\frac{a}{4}\sum_{r=1}^{a}\left[2r-\left(a+1\right)\right]^2$$

$$V_{SIS}\left(\hat{t}_\pi\right)=n^2\frac{a}{4}\sum_{r=1}^{a}\left[4r^2-4r\left(a+1\right)+\left(a+1\right)^2\right]$$

$\sum_{k=1}^{N}k=\frac{n\cdot(n+1)}{2}$
$\sum_{k=1}^{N}k^2=\frac{n\cdot(n+1)(2n+1)}{6}$

$$V_{SIS}\left(\hat{t}_\pi\right)=n^2\frac{a}{4}\left[4\frac{a\cdot\left(a+1\right)\left(2a+1\right)}{6}-4\frac{a\cdot\left(a+1\right)}{2}\left(a+1\right)+a\left(a+1\right)^2\right]$$

$$V_{SIS}\left(\hat{t}_\pi\right)=n^2a^2\frac{a+1}{4}\left[\frac{2}{3}\left(2a+1\right)-2\cdot\left(a+1\right)+\left(a+1\right)\right]$$

$$V_{SIS}\left(\hat{t}_\pi\right)=n^2a^2\frac{a+1}{12}\left[4a+2-6\cdot\left(a+1\right)+3\left(a+1\right)\right]$$

- $$V_{SIS}\left(\hat{t}_\pi\right)=n^2a^2\frac{a+1}{12}\left(a-1\right)$$

$$V_{EST}\left(\hat{t}_\pi\right)=a\left(a-1\right)\sum_{h=0}^{n-1}S^2_{yU_h}$$

$$V_{EST}\left(\hat{t}_\pi\right)=a\left(a-1\right)\sum_{h=0}^{n-1}\frac{a\left(a+1\right)}{12}$$

- $$V_{EST}\left(\hat{t}_\pi\right)=a\left(a-1\right)n\frac{a\left(a+1\right)}{12}$$

$$\frac{V_{EST}\left(\hat{t}_\pi\right)}{V_{SIS}\left(\hat{t}_\pi\right)}=\frac{a\left(a-1\right)n\frac{a(a+1)}{12}}{n^2a^2\frac{a+1}{12}\left(a-1\right)}$$

$$\frac{V_{EST}\left(\hat{t}_\pi\right)}{V_{SIS}\left(\hat{t}_\pi\right)}=\frac{ana}{n^2a^2}$$

$$\frac{V_{EST}\left(\hat{t}_\pi\right)}{V_{SIS}\left(\hat{t}_\pi\right)}=\frac{1}{n}$$

$$V_{EST}\left(\hat{t}_\pi\right)\cdot n = V_{SIS}\left(\hat{t}_\pi\right)$$

$n \geq 1$

$$V_{EST}\left(\hat{t}_\pi\right) < V_{SIS}\left(\hat{t}_\pi\right)$$

Ejercicio 4.

a Sea una población dividida en dos estratos. Se define ϕ a la razón entre $^{n_1}/_{n_2}$ y $^{n_{1Ney}}/_{n_{2Ney}}$ que surjen de la adjudicación de Neyman. El objetivo es estimar el total usando un diseño MESA. Probar que la varianza obtenida con la adjudicación de Neyman es menor.

$$V_{EST}\left(\hat{t}_\pi\right) = N_1^2\frac{1-f_1}{n_1}S_{yU_1}^2 + N_2^2\frac{1-f_2}{n_2}S_{yU_2}^2$$

$N_h^2\frac{1-f_h}{n_h}S_{yU_h}^2 = N_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)S_{yU_h}^2$
$N_h \gg 0 \Rightarrow \frac{1}{N_h} \approx 0$
$N_h^2\frac{1-f_h}{n_h}S_{yU_h}^2 = \frac{N_h^2}{n_h}S_{yU_h}^2$
■

$$V_{EST}\left(\hat{t}_\pi\right) \approx \frac{N_1^2}{n_1}S_{yU_1}^2 + \frac{N_2^2}{n_2}S_{yU_2}^2$$

$$V_{EST,Ney}\left(\hat{t}_\pi\right) = \frac{N_1^2}{n_{1Ney}}S_{yU_1}^2 + \frac{N_2^2}{n_{2Ney}}S_{yU_2}^2$$

$$V_{EST,Ney}\left(\hat{t}_\pi\right) = \frac{N_1^2}{n\frac{N_1S_{yU_1}}{N_1S_{yU_1}+N_2S_{yU_2}}}S_{yU_1}^2 + \frac{N_2^2}{n\frac{N_2S_{yU_2}}{N_1S_{yU_1}+N_2S_{yU_2}}}S_{yU_2}^2$$

$$V_{EST,Ney}\left(\hat{t}_\pi\right) = \frac{1}{n}\left(N_1S_{yU_1}+N_2S_{yU_2}\right)\frac{1}{n}\left(N_1S_{yU_1}+N_2S_{yU_2}\right)$$

$$V_{EST,Ney}\left(\hat{t}_\pi\right) = \frac{1}{n^2}\left(N_1S_{yU_1}+N_2S_{yU_2}\right)^2$$

Planteamos la diferencia:

$$V_{EST}\left(\hat{t}_\pi\right) - V_{EST,Ney}\left(\hat{t}_\pi\right) = \frac{N_1^2}{n_1}S_{yU_1}^2 + \frac{N_2^2}{n_2}S_{yU_2}^2 - \frac{1}{n^2}\left(N_1S_{yU_1}+N_2S_{yU_2}\right)^2$$

$$V_{EST}\left(\hat{t}_\pi\right) - V_{EST,Ney}\left(\hat{t}_\pi\right) = \frac{N_1^2}{n_1}S_{yU_1}^2 + \frac{N_2^2}{n_2}S_{yU_2}^2 - \frac{1}{n^2}\left(N_1^2S_{yU_1}^2 + N_2^2S_{yU_2}^2 + 2N_1S_{yU_1}N_2S_{yU_2}\right)$$

Adecuamos el primer término para poder usar un factor común.
$\frac{N_1^2}{n_1}S_{yU_1}^2 + \frac{N_2^2}{n_2}S_{yU_2}^2 = \frac{1}{n}\left(\frac{N_1^2}{\frac{n_1}{n_1+n_2}}S_{yU_1}^2 + \frac{N_2^2}{\frac{n_2}{n_1+n_2}}S_{yU_2}^2\right)$
$\frac{N_1^2}{n_1}S_{yU_1}^2 + \frac{N_2^2}{n_2}S_{yU_2}^2 = \frac{1}{n}\left[N_1^2\left(n_2+1\right)S_{yU_1}^2 + N_2^2\left(n_1+1\right)S_{yU_2}^2\right]$
$\frac{N_1^2}{n_1}S_{yU_1}^2 + \frac{N_2^2}{n_2}S_{yU_2}^2 = \frac{1}{n^2}\left[N_1^2n\left(n_2+1\right)S_{yU_1}^2 + N_2^2n\left(n_1+1\right)S_{yU_2}^2\right]$

$$\begin{aligned} V_{EST}\left(\hat{t}_\pi\right) - V_{EST,Ney}\left(\hat{t}_\pi\right) &= \frac{1}{n^2}\left[N_1^2n\left(n_2\cancel{+1}\right)S_{yU_1}^2 + N_2^2n\left(n_1\cancel{+1}\right)S_{yU_2}^2\right] \\ &- \frac{1}{n^2}\left(\cancel{N_1^2S_{yU_1}^2}\cancel{+N_2^2S_{yU_2}^2} + 2N_1S_{yU_1}N_2S_{yU_2}\right) \end{aligned}$$

$$
\begin{aligned}
V_{EST}\left(\hat{t}_\pi\right) - V_{EST,Ney}\left(\hat{t}_\pi\right) &= \frac{1}{n^2}\left[N_1^2 n n_2 S_{yU_1}^2 + N_2^2 n n_1 S_{yU_2}^2\right] \\
&- \frac{1}{n^2}\left(2N_1 S_{yU_1} N_2 S_{yU_2}\right)
\end{aligned}
$$

$$V_{EST}\left(\hat{t}_\pi\right) - V_{EST,Ney}\left(\hat{t}_\pi\right) = \frac{1}{n^2}\left[N_1^2 n n_2 S_{yU_1}^2 + N_2^2 n n_1 S_{yU_2}^2 - 2N_1 S_{yU_1} N_2 S_{yU_2}\right]$$

Dada la existencia de estratos, es razonable asumir
$n \geq 2$
$n_h > 1$
$nn_h > 1$

$$V_{EST}\left(\hat{t}_\pi\right) - V_{EST,Ney}\left(\hat{t}_\pi\right) > \frac{1}{n^2}\left[N_1^2 S_{yU_1}^2 + N_2^2 S_{yU_2}^2 - 2N_1 S_{yU_1} N_2 S_{yU_2}\right]$$

$$V_{EST}\left(\hat{t}_\pi\right) - V_{EST,Ney}\left(\hat{t}_\pi\right) > \frac{1}{n^2}\left(N_1^2 S_{yU_1}^2 - N_2^2 S_{yU_2}^2\right)^2$$

$$V_{EST}\left(\hat{t}_\pi\right) - V_{EST,Ney}\left(\hat{t}_\pi\right) > 0$$

$$V_{EST}\left(\hat{t}_\pi\right) > V_{EST,Ney}\left(\hat{t}_\pi\right)$$

b Probar que $\frac{V_{Ney}}{V_{EST}} \geq \frac{4\phi}{(1+\phi)^2}$[2]. $V_{EST}\left(\hat{t}_\pi\right) \approx \frac{N_1^2}{n_1} S_{yU_1}^2 + \frac{N_2^2}{n_2} S_{yU_2}^2$
$V_{EST,Ney}\left(\hat{t}_\pi\right) = \frac{N_1^2}{n_{1Ney}} S_{yU_1}^2 + \frac{N_2^2}{n_{2Ney}} S_{yU_2}^2$

Bucamos alguna relación entre los n_{hNey}.

$$n_{hNey} = n\frac{N_h S_{yU_h}}{\sum_{h=1}^{H} N_h S_{yU_h}}$$

$$n_{hNey} = n\frac{N_h S_{yU_h}}{N_1 S_{yU_1} + N_2 S_{yU_2}}$$

$$n_{1Ney}/n_{2Ney} = \frac{N_1 S_{yU_1}}{N_2 S_{yU_2}}$$

Buscamos Expresiones para Ambas Varianzas. Recordando la hipótesis de que los N_h son «muy grandes».

$$V_{MESA}\left(\hat{t}_\pi\right) = \sum N_h^2 \frac{(1-\cancel{f})}{n_h} S_h^2$$

$$V_{MESA}\left(\hat{t}_\pi\right) = \frac{N_1^2 S_1^2}{n_1} + \frac{N_2^2 S_2^2}{n_2}$$

Aplicamos la asignación neyman para encontrar la otra expresión:

[2]**Ayuda:** $\frac{x+a}{x+b} \geq \frac{a}{b}$ si $x \geq 0$ y $a \geq b$.

$$V_{MESA_{Ney}}\left(\hat{t}_\pi\right) = \frac{N_1^2 S_1^2}{n\left(\frac{N_1S_1}{N_1S_1+N_2S_2}\right)} + \frac{N_2^2 S_2^2}{n\left(\frac{N_2S_2}{N_1S_1+N_2S_2}\right)}$$

$$V_{MESA_{Ney}}\left(\hat{t}_\pi\right) = (N_1S_1 + N_2S_2)\left(\frac{N_1S_1}{n} + \frac{N_2S_2}{n}\right)$$

$$V_{MESA_{Ney}}\left(\hat{t}_\pi\right) = \frac{(N_1S_1 + N_2S_2)^2}{n}$$

Buscamos Algunas identidades para ϕ. $\phi = \frac{n_1/n_2}{n_{1Ney}/n_{2Ney}}$

- $\phi = \frac{n_1 n_{2Ney}}{n_2 n_{1Ney}}$

$n_{2Ney}/n_{1Ney} = \frac{n\left(\frac{N_2S_2}{N_1S_1+N_2S_2}\right)}{n\left(\frac{N_1S_1}{N_1S_1+N_2S_2}\right)} = \frac{N_2S_{yU_2}}{N_1S_{yU_1}}$

$\phi = \frac{n_1}{n_2}\frac{N_2S_{yU_2}}{N_1S_{yU_1}}$

$\phi\frac{n_2}{n_1} = \frac{N_2S_{yU_2}}{N_1S_{yU_1}}$

Trabajamos con la Razón pedida entre las varianzas.

$$V_{MESA_{Ney}}\left(\hat{t}_\pi\right)/V_{EST}\left(\hat{t}_\pi\right) = \frac{\frac{(N_1S_1+N_2S_2)^2}{n}}{\frac{N_1^2S_1^2}{n_1} + \frac{N_2^2S_2^2}{n_2}}$$

$$V_{MESA_{Ney}}\left(\hat{t}_\pi\right)/V_{EST}\left(\hat{t}_\pi\right) = \frac{1}{n}\frac{N_1^2S_1^2\left(1+\frac{N_2S_2}{N_1S_1}\right)^2}{N_1^2S_1^2\left[\frac{1}{n_1} + \frac{1}{n_2}\left(\frac{N_2S_2}{N_1S_1}\right)^2\right]}$$

$$V_{MESA_{Ney}}\left(\hat{t}_\pi\right)/V_{EST}\left(\hat{t}_\pi\right) = \frac{1}{n}\frac{\left(1+\frac{N_2S_2}{N_1S_1}\right)^2}{\frac{1}{n_1} + \frac{1}{n_2}\left(\frac{N_2S_2}{N_1S_1}\right)^2}$$

$$V_{MESA_{Ney}}\left(\hat{t}_\pi\right)/V_{EST}\left(\hat{t}_\pi\right) = \frac{\left(1+\frac{N_2S_2}{N_1S_1}\right)^2}{(n_1+n_2)\left[\frac{1}{n_1} + \frac{1}{n_2}\left(\frac{N_2S_2}{N_1S_1}\right)^2\right]}$$

$\phi\frac{n_2}{n_1} = \frac{N_2S_{yU_2}}{N_1S_{yU_1}}$

$$V_{MESA_{Ney}}\left(\hat{t}_\pi\right)/V_{EST}\left(\hat{t}_\pi\right) = \frac{1}{n}\frac{\left(1+\phi\frac{n_2}{n_1}\right)^2}{\left[\frac{1}{n_1} + \frac{1}{n_2}\left(\phi\frac{n_2}{n_1}\right)^2\right]}$$

$$V_{MESA_{Ney}}\left(\hat{t}_\pi\right)/V_{EST}\left(\hat{t}_\pi\right) = \frac{1}{n}\frac{\left(\frac{n_1+\phi n_2}{n_1}\right)^2}{\frac{n_2+n_1\left(\phi\frac{n_2}{n_1}\right)^2}{n_1n_2}}$$

$$V_{MESA_{Ney}}\left(\hat{t}_\pi\right)/V_{EST}\left(\hat{t}_\pi\right) = \frac{1}{n}\cdot\frac{n_1n_2}{n_1^2}\cdot\frac{(n_1+\phi n_2)^2}{n_2+n_1\left(\phi\frac{n_2}{n_1}\right)^2}$$

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{1}{n}\cdot\frac{n_2}{n_1}\cdot\frac{(n_1+\phi n_2)^2}{n_2+\phi^2\frac{n_2^2}{n_1}}$$

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{1}{n}\cdot\frac{n_2}{n_1}\cdot\frac{(n_1+\phi n_2)^2}{n_2\left(1+\phi^2\frac{n_2}{n_1}\right)}$$

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{1}{n}\cdot\frac{1}{n_1}\cdot\frac{(n_1+\phi n_2)^2}{1+\phi^2\frac{n_2}{n_1}}$$

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{1}{n}\cdot\frac{(n_1+\phi n_2)^2}{n_1+\phi^2 n_2}$$

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{(n_1+\phi n_2)^2}{(n_1+n_2)(n_1+\phi^2 n_2)}$$

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{(n_1+\phi n_2)^2}{n_1(n_1+\phi^2 n_2)+n_2(n_1+\phi^2 n_2)}$$

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{(n_1+\phi n_2)^2}{n_1^2+\phi^2 n_1 n_2+n_1 n_2+\phi^2 n_2^2}$$

$(n_1+\phi n_2)^2 = (n_1-\phi n_2)^2 + 4\phi n_1 n_2$

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{(n_1-\phi n_2)^2+4\phi n_1 n_2}{n_1^2+\phi^2 n_1 n_2+n_1 n_2+\phi^2 n_2^2+2\phi n_1 n_2-2\phi n_1 n_2}$$

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{(n_1-\phi n_2)^2+4\phi n_1 n_2}{n_1^2-2\phi n_1 n_2+\phi^2 n_2^2+\phi^2 n_1 n_2+n_1 n_2+2\phi n_1 n_2}$$

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{(n_1-\phi n_2)^2+4\phi n_1 n_2}{(n_1-\phi n_2)^2+\phi^2 n_1 n_2+n_1 n_2+2\phi n_1 n_2}$$

Aplicamos la identidad de la consigna.

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{4\phi n_1 n_2}{\phi^2 n_1 n_2+n_1 n_2+2\phi n_1 n_2}$$

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{4\phi}{\phi^2+1+2\phi}$$

$$V_{MESA_{Ney}}(\hat{t}_\pi)/V_{EST}(\hat{t}_\pi) = \frac{4\phi}{(1+\phi)^2}$$

Punto i del Ejercicio de Estratificación

i Construir el estimador de HT $\hat{d}_{i\pi}$ siendo $d_i = \overline{y}_{U_i} - \overline{y}_U$.

$$d_i = \overline{y}_{U_i} - \overline{y}_U$$

$W_h = {}^{N_h}/_{N} \Rightarrow \overline{y}_U = \sum_{h=1}^{H} W_h \overline{y}_{U_h}$

$$d_i = \overline{y}_{U_i} - \sum_{h=1} W_h \overline{y}_{U_h}$$

$$d_i = (1 - W_i)\, \overline{y}_{U_i} - \sum_{h \neq i} W_h \overline{y}_{U_h}$$

En Särndall, probamos que siempre se cumple $\mathbb{E}\left(\sum_s \check{a}_k\right) = \sum_U a_k$.
Es decir, el estimador de HT es insertable.
Sean $\hat{\overline{y}}_{s_h} = \frac{1}{N_h} \sum_{s \cap U_h} \frac{y_h}{\pi_h} = \frac{1}{N_h} \sum_{s \cap U_h} \check{y}_h$
$\pi_h = \frac{n_h}{N_h}$

- $$\hat{d}_{i\pi} = (1 - W_i)\, \hat{\overline{y}}_{s_i} - \sum_{h \neq i} W_h \hat{\overline{y}}_{s_h}$$

Probamos que es un estimador de la variable deseada (e insesgado).

$$\mathbb{E}\left(\hat{d}_{i\pi}\right) = \mathbb{E}\left[(1 - W_i)\, \hat{\overline{y}}_{s_i} - \sum_{h \neq i} W_h \hat{\overline{y}}_{s_h}\right]$$

$$\mathbb{E}\left(\hat{d}_{i\pi}\right) = (1 - W_i)\, \mathbb{E}\left(\hat{\overline{y}}_{s_i}\right) - \sum_{h \neq i} W_h \mathbb{E}\left(\hat{\overline{y}}_{s_h}\right)$$

Por cómo están definidos los $\hat{\overline{y}}_{s_h}$, el resultado de Särndall aplica inmediatamente.

- $$\mathbb{E}\left(\hat{d}_{i\pi}\right) = (1 - W_i)\, \overline{y}_{U_i} - \sum_{h \neq i} W_h \overline{y}_{U_h}$$

ii Desarollar la varianza de $\hat{d}_{i\pi}$.

$$V\left(\hat{d}_{i\pi}\right) = V\left[(1 - W_i)\, \hat{\overline{y}}_{s_i} - \sum_{h \neq i}^{H} W_h \hat{\overline{y}}_{s_h}\right]$$

$$V\left(\hat{d}_{i\pi}\right) = (1 - W_i)^2\, V\left(\hat{\overline{y}}_{s_i}\right) + \sum_{h \neq i}^{H} W_h^2 V\left(\hat{\overline{y}}_{s_h}\right)$$

- Este punto está entrocado con el resto del los ejercicios sobre muestreo estratificado y por lo tanto asumimos que se extiende a él la elección del MESA, siendo «cualquier adjudicación» referido a variaciones en la cantidad de estratos y métodos para determinar los n_h siempre dentro de este esquema.

$V_{MAS}\left(\hat{\overline{y}}_{s_h}\right) = \frac{1 - f_h}{n_h} \cdot S_{yU_h}^2 =$

$$V\left(\hat{d}_{i\pi}\right) = (1 - W_i)^2\, \frac{1 - f_i}{n_i} \cdot S_{yU_i}^2 + \sum_{h \neq i}^{H} W_h^2 \frac{1 - f_h}{n_h} \cdot S_{yU_h}^2$$

iii ¿Cuál es la adjudicación de Neyman para minimizar la varianza de $\hat{d}_{i\pi}$ para un tamaño de muestra fijo? Para no tener que optimizar desde cero la adjudicación sobre los mismos supuestos, intentamos lograr amoldar el problema a las hipótesis del caso conocido con un cambio de variable. Necesitamos una variable cuya varianza vamos a evaluar estrato por estrato.

$$\hat{d}_{i\pi} = (1 - W_i)\,\hat{\bar{y}}_{s_i} - \sum_{h\neq i} W_h \hat{\bar{y}}_{s_h}$$

$$\hat{d}_{i\pi} = (1 - W_i)\left(\frac{1}{N_i}\sum_{s\cap U_i} \check{y}_k\right) - \sum_{h\neq i} W_h \left(\frac{1}{N_h}\sum_{s\cap U_h} \check{y}_k\right)$$

Por ser un MESA: $\forall y_h \in U_h : \pi_k = \frac{n_h}{N_h}$

$$\hat{d}_{i\pi} = (1 - W_i)\left(\frac{1}{N_i}\sum_{s\cap U_i} \frac{y_k}{{n_i}/{N_i}}\right) - \sum_{h\neq i} W_h \left(\frac{1}{N_h}\sum_{s\cap U_h} \frac{y_k}{{n_h}/{N_h}}\right)$$

$$\hat{d}_{i\pi} = (1 - W_i)\left(\frac{1}{n_i}\sum_{s\cap U_i} y_k\right) - \sum_{h\neq i} W_h \left(\frac{1}{n_h}\sum_{s\cap U_h} y_k\right)$$

$$\hat{d}_{i\pi} = W_i \sum_{s\cap U_i} \frac{\left({1}/{W_i} - 1\right)}{n_i} y_k + \sum_{h\neq i}\left(\sum_{s\cap U_h}\left(-\frac{W_h}{n_h}\right) y_k\right)$$

$$\hat{d}_{i\pi} = \sum_{s\cap U_i} \frac{W_i}{n_i}\left[\left({1}/{W_i} - 1\right) y_k\right] + \sum_{h\neq i}\left(\sum_{s\cap U_h}\left(-\frac{W_h}{n_h}\right) y_k\right)$$

$$z_k = \begin{cases} \left({1}/{W_i} - 1\right) y_k & k \in h_i \\ -y_k & k \notin h_i \end{cases}$$

$$\hat{d}_{i\pi} = \sum_{s\cap U_i} z_k$$

Ahora la ecuación tiene una forma que nos deja aplicar la fórmula ya conocida para la asignación óptima de Neyman:

$$n_h = n \cdot \frac{N_h S_{zU_h}}{\sum_k N_k S_{zU_k}}$$

Buscamos la varianza de esta nueva variable:

$$V\left(z_{kh}\right) = \begin{cases} V\left[\left({1}/{W_i} - 1\right) y_k\right] & k \in h_i \\ V\left(-y_k\right) & k \notin h_i \end{cases}$$

$$S_{zU_h} = \begin{cases} \left({1}/{W_i} - 1\right)^2 S_{yU_i} & h = i \\ S_{yU_h} & h \neq i \end{cases}$$

Especializamos con la varianza obtenida

$$n_h = \begin{cases} n \cdot \frac{N_i \cdot \left({1}/{W_i} - 1\right)^2 S_{yU_i}}{\sum_{k\neq i} N_k S_{yU_k} + N_i \cdot \left({1}/{W_i} - 1\right)^2 S_{yU_i}} & h = i \\ n \cdot \frac{N_h S_{yU_h}}{\sum_{k\neq i} N_k S_{yU_k} + N_i \cdot \left({1}/{W_i} - 1\right)^2 S_{yU_i}} & h \neq i \end{cases}$$

iv ¿Cuándo existirán entre esta adjudicación y la de Neyman considerando la variable de estratificación? Recordamos la adjudicación de Neyman:

$$n_{h_{Ney}} = n \cdot \frac{N_h \cdot S_{yU_h}}{\sum_{k=1}^{H} N_k S_{yU_k}}$$

- Efecto sobre n_i:

$$n_i = n_{i_{Ney}}$$

$$n \cdot \frac{N_i \cdot \left({}^1\!/W_i - 1\right)^2 S_{yU_i}}{\sum_{k\neq i} N_k S_{yU_k} + N_i \cdot \left({}^1\!/W_i - 1\right)^2 S_{yU_i}} = n \cdot \frac{N_i \cdot S_{yU_i}}{\sum_{k=1}^{H} N_k S_{yU_k}}$$

$$\frac{1 \cdot \left({}^1\!/W_i - 1\right)^2}{\sum_{k\neq i} N_k S_{yU_k} + N_i \cdot \left({}^1\!/W_i - 1\right)^2 S_{yU_i}} = \frac{1}{\sum_{k=1}^{H} N_k S_{yU_k}}$$

$$\frac{1}{\frac{\sum_{k\neq i} N_k S_{yU_k}}{\left({}^1\!/W_i - 1\right)^2} + N_i \cdot S_{yU_i}} = \frac{1}{\sum_{k\neq i} N_k S_{yU_k} + N_i \cdot S_{yU_i}}$$

Como la función $\frac{1}{x}$ es biyectiva cuando restringimos los valores de x a $\mathbb{R}^+$ y este es el caso de nuestro denominador, podemos asegurar que esta igualdad sólo se da cuando los denominadores son iguales:

$$\frac{1}{\frac{\sum_{k\neq i} N_k S_{yU_k}}{\left({}^1\!/W_i - 1\right)^2} + N_i \cdot S_{yU_i}} = \frac{1}{\sum_{k\neq i} N_k S_{yU_k} + N_i \cdot S_{yU_i}} \Rightarrow \frac{\sum_{k\neq i} N_k S_{yU_k}}{\left({}^1\!/W_i - 1\right)^2} + N_i \cdot S_{yU_i} = \sum_{k\neq i} N_k S_{yU_k} + N_i \cdot S_{yU_i}$$

$$\frac{\sum_{k\neq i} N_k S_{yU_k}}{\left({}^1\!/W_i - 1\right)^2} + N_i \cdot S_{yU_i} = \sum_{k\neq i} N_k S_{yU_k} + N_i \cdot S_{yU_i}$$

$$\frac{\sum_{k\neq i} N_k S_{yU_k}}{\left({}^1\!/W_i - 1\right)^2} = \sum_{k\neq i} N_k S_{yU_k}$$

$$\frac{\sum_{k\neq i} N_k S_{yU_k}}{\sum_{k\neq i} N_k S_{yU_k}} = \left({}^1\!/W_i - 1\right)^2$$

$$\left({}^1\!/W_i - 1\right)^2 = 1$$

$W_i \leq 1 \Rightarrow ({}^1\!/W_i - 1) > 0$

$${}^1\!/W_i - 1 = 1$$

$${}^1\!/W_i = 2$$

- $$W_i = \frac{1}{2}$$

El caso de igualdad es cuando el estrato a considerar representa la mitad de la muestra.

Observando la igualdad $\frac{1}{\frac{\sum_{k\neq i} N_k S_{yU_k}}{\left({}^1\!/W_i - 1\right)^2} + N_i \cdot S_{yU_i}} = \frac{1}{\sum_{k\neq i} N_k S_{yU_k} + N_i \cdot S_{yU_i}}$ vemos que estaremos cerca de ella cuando el factor $\left({}^1\!/W_i - 1\right)^2$ cumple $\left({}^1\!/W_i - 1\right)^2 \approx 1$.

Si el estrato es pequeño, su peso W_i lo es también y la expresión $1/W_i$ tiende a crecer así como lo hace su cuadrado. El numerador entonces disminuye y obtenemos un n_i mayor.

- Para los estratos $h \neq i$:

$$n_h = n_{h_{Ney}}$$

$$n \cdot \frac{N_h \cdot S_{yU_h}}{\sum_{k \neq i} N_k S_{yU_k} + N_i \cdot \left(1/W_i - 1\right)^2 S_{yU_i}} = n \cdot \frac{N_h \cdot S_{yU_h}}{\sum_{k=1}^{H} N_k S_{yU_k}}$$

$$\frac{1}{\sum_{k \neq i} N_k S_{yU_k} + N_i \cdot \left(1/W_i - 1\right)^2 S_{yU_i}} = \frac{1}{\sum_{k=1}^{H} N_k S_{yU_k}}$$

Otra vez, tenemos valores similares cuando $\left(1/W_i - 1\right)^2 \approx 1$.

En este caso, en la medida en que el estrato i tenga un peso W_i pequeño, el factor $\left(1/W_i - 1\right)^2$ se va a ver incrementado y por lo tanto vamos a asignarle menos muestras a los demás estratos.

- Como era de esperar, conocer más información sobre una de las variables siempre nos obliga a asignarle más muestras, en este caso hallamos la ley de acuerdo a la cual eso sucede.
- Los valores que adopta la variable en el universo de estudio no son drásticamente importantes, el efecto se da de acuerdo al peso W_i del estrato cuya diferencia respecto a la media nos importa.
- En nuestro marco el marco construido con nuestra variable auxiliar, las empresas clasificadas como pequeñas representan un estrato muy poco significativo y por lo tanto la diferencia en la asignación va a ser apreciable.