MUEST _ T2

1. MUESTREO con PROBAB. IGUALES.
2. ESTIMADORES LINEALES.
3. VARIANZAS Y COVARIANZAS de los ESTIMADORES y sus ESTIMACIONES.
4. COMPARACIÓN entre el MUESTREO CON repos. y SIN repos.
5. CONSIDERACIONES sobre el TAMAÑO de la muestra

---

1 _ MUESTREO con PROBAB. IGUALES.

Muestreo probabilístico en el que cada unidad poblacional u_i, $i = 1 - N$, tiene la misma probabilidad de ser elegida para formar parte de la muestra.

En este tipo de muestreo las unidades son elementales, o siendo compuestas tienen todas la misma importancia.

Podemos clasificarlos atendiendo a la reposición y al orden.

Número muestras posibles	NO orden	SI orden
SIN repos.	$C_{N,n} = \binom{N}{n} = \frac{N!}{(N-n)!\,n!}$	$V_{N,n} = \binom{N}{n} n! = \frac{N!}{(N-n)!}$
CON repos.	$CR_{N,n} = \binom{N+n-1}{n}$	$VR_{N,n} = N^n$

$$C_{N,n} = \frac{V_{N,n}}{P_n = n!}$$

$$CR_{N,n} = \frac{V_{N+n-1,\,n}}{P_n}$$

Probabilidad muestra	NO orden	SI orden
SIN repos.	$\frac{1}{C_{N,n}} = \frac{1}{\binom{N}{n}}$	$\frac{1}{V_{N,n}} = \frac{1}{\binom{N}{n} n!}$
CON repos	$\neq$ para cada muestra	$\frac{1}{VR_{N,n}} = \frac{1}{N^n}$

En la práctica, no se ~~tiene en cuenta~~ distingue por el orden, por lo que solamente nos limitaremos a estudiar el caso SIN reposición y el caso CON reposición.

2_ MUESTREO ALEATORIO SIMPLE SIN REPOSICIÓN, m.a.s.s.r.

(prob= → s, SIN → s)

~~Procedi:~~
También llamado muestreo irrestrictamente aleatorio, ó muestreo aleatorio simple sin más.

Procedimiento de selección de la muestra
- con probab. iguales,
- se seleccionan las unid. ~~una~~ 1 a 1 sin reposición,
- no importa el orden de aparición.

Por lo que las muestras obtenidas mediante este procedimiento se caracterizan por:
- muestras con elementos repetidos son imposibles,
- muestras equiprobables (pq todas las unid. tienen la misma probab. de ∈ muestra)

PROBABILIDADES

$\# S = C_{N,n} = \binom{N}{n} = \frac{N!}{n!(N-n)!}$ muestras posibles (sr, no orden)

$\mathbb{P}(\text{muestra}) = P(u_1 \ldots u_n) = \frac{CF}{CP} = \frac{1}{\binom{N}{n}}$ ← muestras equiprobables

(tb. se puede hacer con $P_n \cdot P(\{u_1, \ldots u_n\}) = n! \frac{1}{N} \cdot \frac{1}{N-1} \cdots \frac{1}{N-(n-1)}$).

$$\boxed{\pi_i = P(u_i \in \text{muestra}) = \frac{CF}{CP} = \frac{\text{nº mu. con } u_i}{\text{nº tot. mu}} = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}}$$

(tb. se puede hacer utilizando c. probab ⇒

$\pi_i = P(u_i = u_1) + P(u_i \neq u_1 \cap u_i = u_2) + \ldots + P(u_i \neq u_1 \cap u_i \neq u_2 \cap \ldots \cap u_i \neq u_{n-1} \cap u_i = u_n)$

$= \frac{1}{N} + \ldots + \frac{1}{N} = \frac{n}{N}$.

$$\boxed{\pi_{ij} = P((u_i, u_j) \in \text{Muestra}) = \frac{CF}{CP} = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}}$$

con probab. $P(u_i = u_1 \cap u_j = u_2) = \frac{1}{N} \cdot \frac{1}{N-1}$ $\xrightarrow[u_j\ (n-1)\ \text{posic.}]{u_i \text{ fijo}}$ $\frac{(n-1)}{N(N-1)}$ ⇒

u_i n posic. ⇒ $\frac{n(n-1)}{N(N-1)}$.

ESTIMADOR lineal insesgado

El estimador lineal insesgado en el caso de m.a.s.s.r. es el estimador de Horvitz y Thompson, $\hat{\theta}_{HT}$.

En general, para un parámetro θ resultado de sumar las características poblacionales, el estimador HT es una suma ponderada de las mismas características muestrales, donde las ponderaciones son las inversas de las probab. de pertenecer a la muestra.

$$\theta = \sum_{i=1}^{N} Y_i \xrightarrow{\text{SIN r.}} \hat{\theta}_{HT} = \sum_{i=1}^{n} \frac{Y_i}{\pi_i} \leftarrow \text{General}$$

El estimador Horvitz y Thompson es un estimador insesgado de θ, $E[\hat{\theta}_{HT}] = \theta$

$$E[\hat{\theta}_{HT}] = E\left[\sum^{n} \frac{Y_i}{\pi_i}\right] = E\left[\sum^{N} \frac{1}{\pi_i} \cdot Y_i e_i\right] = \sum_{i=1}^{N} \frac{Y_i}{\pi_i} E[e_i] =$$

$$= \sum_{i=1}^{N} \frac{Y_i}{\pi_i} (1 \cdot \pi_i + 0(1-\pi_i)) = \sum^{N} \frac{Y_i}{\pi_i} \pi_i = \sum^{N} Y_i = \theta.$$

Al ser en el m.a.s.s.r. todas las probab. de pertenecer a la muestra iguales, $\pi_i = \frac{n}{N}$, podemos especificar los estimadores lineales insesgados para los parámetros más comunes: $\hat{\theta}_{HT} = N \sum_{i=1}^{n} \frac{Y_i}{n}$.

Total poblacional	$\theta = X = \sum_{i=1}^{N} X_i$	$\hat{\theta}_{HT} = \frac{\sum^{n} X_i}{n/N}$	$\hat{X}_{HT} = N\bar{x}$
Media poblacional	$\theta = \bar{X} = \sum_{i=1}^{N} \frac{X_i}{N}$	$\hat{\theta}_{HT} = \frac{\sum^{n} X_i/N}{n/N} = \frac{\sum^{n} X_i}{n}$	$\hat{\bar{X}}_{HT} = \bar{x}$
Total de clase	$\theta = A = \sum_{i=1}^{N} A_i$	$\hat{\theta}_{HT} = \frac{\sum^{n} A_i}{n/N} = N \cdot \frac{\sum^{n} A_i}{n}$	$\hat{A}_{HT} = N\hat{P}$
Proporción de clase	$\theta = P = \sum_{i=1}^{N} \frac{A_i}{N}$	$\hat{\theta}_{HT} = \frac{\sum^{n} A_i/N}{n/N} = \frac{\sum^{n} A_i}{n}$	$\hat{P}_{HT} = \hat{P}$

VARIANZA del estimador $\hat{\theta}_{HT}$

$e_i = \{1, 0\}$

var(suma) = suma (var) + suma (cov)

En general, $V(\hat{\theta}_{HT}) = V\left(\sum^{n} \frac{y_i}{\pi_i}\right) = V\left(\sum^{N} \frac{y_i}{\pi_i} e_i\right) =$

$$= \sum_{i=1}^{N} V\left(\frac{y_i}{\pi_i} e_i\right) + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} cov\left(\frac{y_i}{\pi_i} e_i, \frac{y_j}{\pi_j} e_j\right) =$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{\pi_i^2} V(e_i) + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{y_i}{\pi_i} \cdot \frac{y_j}{\pi_j} cov(e_i, e_j) =$$

$V(e_i) = \pi_i(1-\pi_i)$

$cov(e_i, e_j) = \pi_{ij} - \pi_i \pi_j$

$$= \sum_{i=1}^{N} \frac{y_i^2}{\pi_i^2} \pi_i (1-\pi_i) + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{y_i}{\pi_i} \cdot \frac{y_j}{\pi_j} (\pi_{ij} - \pi_i \pi_j) =$$

$$V(\hat{\theta}_{HT}) = \sum_{i=1}^{N} \frac{y_i^2 (1-\pi_i)\cancel{\pi_i}}{\cancel{\pi_i^2}\,\pi_i} + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{y_i}{\pi_i} \cdot \frac{y_j}{\pi_j} (\pi_{ij} - \pi_i \pi_j)$$

(cq prob)
GENERAL

En el caso particular del m.a.s.s.r, $\begin{cases} \pi_i = \frac{n}{N} \\ \pi_{ij} = \frac{n(n-1)}{N(N-1)} \end{cases}$

por lo que

$$V(\hat{\theta}_{HT}) = \sum_{i=1}^{N} y_i^2 \underbrace{\frac{(1 - n/N)}{n/N}}_{\frac{\frac{N-n}{\cancel{N}}}{\frac{n}{\cancel{N}}} \to \frac{N-n}{n}} + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \underbrace{\frac{y_i}{n/N} \cdot \frac{y_j}{n/N}}_{\frac{1}{\frac{n}{N} \cdot \frac{n}{N}}} \underbrace{\left(\frac{n(n-1)}{N(N-1)} - \frac{n}{N} \cdot \frac{n}{N}\right)}_{\frac{n}{N}\left[\frac{n-1}{N-1} - \frac{n}{N}\right]} =$$

$$\frac{n}{N}\underbrace{\left[\frac{n-1}{N-1} - \frac{n}{N}\right]}_{\frac{N(n-1) - n(N-1)}{N(N-1)}}$$

$$\frac{N(n-1) - n(N-1)}{N(N-1)} \to \frac{\cancel{Nn} - N \cancel{- nN} + n}{N(N-1)} \to -\frac{(N-n)}{N(N-1)}$$

$$\underbrace{\frac{1}{\frac{n}{\cancel{N}}} \left[-\frac{(N-n)}{\cancel{N}(N-1)}\right]}_{-\left(\frac{N-n}{n}\right) \cdot \frac{1}{N-1}}$$

sacando factor común:

$$V(\hat{\theta}_{HT}) = \frac{N-n}{n} \left[\sum_{i=1}^{N} y_i^2 - 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{y_i y_j}{N-1}\right]$$

(prob =)
massr

y podemos especificar la varianza del estimador de Horvitz y Thompson para los parámetros poblacionales más comunes:

Total poblacional, ⊗ $\theta = X = \sum_{i=1}^{N} X_i$

$$V(\hat{X}_{HT}) = \frac{N-n}{n}\left[\sum_{i=1}^{N} X_i^2 - 2\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{X_i X_j}{N-1}\right]$$

Como la varianza es invariante ante cambios de origen

$$V(\hat{X}_{HT}) = \frac{N-n}{n}\left[\sum_{i=1}^{N}(X_i-\bar{X})^2 - 2\underbrace{\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{(X_i-\bar{X})(X_j-\bar{X})}{N-1}}_{\frac{\sum_{i=1}^{N}(X_i-\bar{X})^2}{N-1}}\right] =$$

(*) mirar atrás →

$$= \frac{N-n}{n}\left[\sum_{i=1}^{N}(X_i-\bar{X})^2 + \frac{\sum_{i=1}^{N}(X_i-\bar{X})^2}{N-1}\right] =$$

CUASIVARIANZA pobl. cuasi S^2

$$= \frac{N-n}{n}\cdot\frac{N}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})^2 = \frac{N(N-n)}{N}\cdot\frac{\sum_{i=1}^{N}(X_i-\bar{X})^2}{N-1} =$$

$$= \frac{N(N-n)}{n}\cdot S^2_{cuasiv} \overset{\text{multipl. y divido por N}}{=} \frac{N^2(N-n)}{N}\cdot\frac{S^2_{cuasiv}}{n} =$$

$$= N^2\left(1-\frac{n}{N}\right)\cdot\frac{S^2_{cuasiv}}{n} = N^2(1-f)\,\frac{S^2_{cuasiv}}{n}$$

de donde se obtiene que la varianza del estimador se ~~coincide~~ puede obtener a partir de la cuasivarianza poblacional ~~muestral~~, el tamaño muestral y el tamaño poblacional.

$$V(\hat{X}_{HT}) = N^2\cdot(1-f)\cdot\frac{S^2_{cuasiv}}{n} \overset{(N-1)S^2 = N\sigma^2}{=} N^3(1-f)\cdot\frac{\sigma^2}{n(N-1)}$$

$f = \frac{n}{N}$ = fracción de muestreo

Mirar atrás →

→ Varianza del estimador del total ES el cuadrado del tamaño poblacional por (1 - fracción de muestreo) por la cuasivarianza poblacional partido por el tamaño muestral

A partir de la varianza del estimador del total poblac. es fácil obtener la varianza del estimador de la media poblacional.

$$\hat{X}_{HT} = N\bar{x} = N\hat{\bar{X}} \Rightarrow \hat{\bar{X}} = \frac{\hat{X}}{N}$$

luego

$$V(\hat{\bar{X}}) = V\left(\frac{\hat{X}}{N}\right) = \frac{1}{N^2} V(\hat{X}) = \frac{1}{\cancel{N^2}} \cdot \cancel{N^2}(1-f) \cdot \frac{S^2_{cuasiv}}{n}$$

$$V(\hat{\bar{X}}_{HT}) = (1-f) \cdot \frac{S^2_{cuasiv}}{n} = (1-f) \cdot \frac{1}{n(N-1)} \sum_{i=1}^{N} (X_i - \bar{X})^2$$

Para obtener la varianza del estimador de la proporción poblacional, hay que calcular primero la expresión de la cuasivarianza poblacional para la característica dicotómica $A_i = \{1, 0\}$.

$$S^2_{cuasiv} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2 = \frac{1}{N-1}\sum_{i=1}^{N}(A_i - P)^2 = \frac{1}{N-1}\sum_{i=1}^{N}(A_i^2 - 2PA_i + P^2) =$$

$$= \frac{1}{N-1}\Big(\underbrace{\sum_{i=1}^{N} A_i^2}_{\sum A_i = NP \; \{pq\ 1, 0\}} - 2P\underbrace{\sum_{i=1}^{N} A_i}_{NP} + NP^2\Big) = \frac{1}{N-1}\left(NP - 2PNP + NP^2\right) =$$

$$= \frac{1}{N-1}\left(NP - NP^2\right) = \frac{1}{N-1} NP(1-P) = \frac{1}{N-1} \cdot NPQ$$

Ahora ya podemos obtener la varianza del estimador de la proporción poblacional, a partir de la de $\bar{X}$.

$$V(\hat{P}_{HT}) \underset{V(\hat{\bar{X}})}{=} (1-f) \cdot \frac{S^2_{cuasiv}}{n} \overset{A_i}{=} (1-f) \cdot \frac{\frac{N}{N-1} PQ}{n} = \frac{N}{N-1} \cdot \frac{1}{n}(1-f) PQ$$

$$V(\hat{P}_{HT}) = \frac{N}{N-1} \cdot \frac{1}{n} \cdot (1-f) PQ$$

Análogamente, la varianza del estimador del total de clase será $(\hat{A} = N\hat{P})$

$$V(\hat{A}) = V(N\hat{P}) = N^2 \cdot V(\hat{P}) = N^2 \cdot \frac{N}{N-1} \cdot \frac{1}{n} (1-f) PQ$$

$$\boxed{V(\hat{A}) = \frac{N^3}{N-1} \cdot \frac{1}{n} \cdot (1-f) PQ}$$

Parámetro	Estimador HT	Varianza del estimador
Total poblacional	$\hat{X} = N\bar{X}$	$V(\hat{X}) = N^2 \cdot (1-f) \cdot \frac{S^2}{n}$
Media poblacional	$\hat{\bar{X}} = \bar{X}$	$V(\hat{\bar{X}}) = (1-f) \cdot \frac{S^2}{n}$
Total de clase	$\hat{A} = N\hat{P}$	$V(\hat{A}) = \frac{N^3}{N-1} \cdot \frac{1}{n}(1-f) \cdot PQ$
Proporción de clase	$\hat{P} = \hat{P}$	$V(\hat{P}) = \frac{N}{N-1} \cdot \frac{1}{n} \cdot (1-f) PQ$

donde $S^2 \equiv$ cuasiV. poblacional, $f = \frac{n}{N} \equiv$ fracción muestreo

ESTIMACIÓN de la varianza del estimador $\hat{\Theta}_{HT}$

Como la varianza del estimador de Horvitz y Thompson depende de las unidades poblacionales $(i = 1 \dots N)$ y sólo disponemos de n unidades, necesitamos estimar dicha varianza con los valores muestrales.

Además, dicha estimación será insesgada.

En general, la estimación de la varianza que se utiliza es:

$$\hat{V}(\hat{\Theta}_{HT}) = \sum_{i=1}^{n} \frac{y_i^2}{\pi_i^2}(1-\pi_i)\frac{1}{\pi_i} + 2\sum_{i=1}^{n}\sum_{j>i}^{n} \frac{y_i}{\pi_i} \cdot \frac{y_j}{\pi_j} \cdot \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}}$$

Esta expresión es muy parecida a la $V(\hat{\Theta}_{HT})$, pero

- sumatorio → hasta n
- 1.er sumando → dividido por π_i
- 2.º sumando → dividido por π_{ij}

Es una estimación insesgada, $E[\hat{V}(\hat{\theta}_{HT})] = V(\hat{\theta}_{HT})$.

$$E[\hat{V}(\hat{\theta}_{HT})] = E\left[\sum^{n} \frac{y_i^2}{\pi_i^2}(1-\pi_i)\right] + 2E\left[\sum^{n}\sum_{j>i}^{n} \frac{y_i}{\pi_i}\cdot\frac{y_j}{\pi_j}\,\frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}}\right] =$$

$$= E\left[\sum_{i=1}^{N} \frac{y_i^2}{\pi_i^2}(1-\pi_i)e_i\right] + 2E\left[\sum_{i=1}^{N}\sum_{j>i}^{N} \frac{y_i}{\pi_i}\cdot\frac{y_j}{\pi_j}\cdot\frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}}\cdot e_ie_j\right] =$$

$$= \sum^{N} \frac{y_i^2}{\pi_i^2}(1-\pi_i)\underbrace{E[e_i]}_{\pi_i} + 2\sum\sum \frac{y_i}{\pi_i}\,\frac{y_j}{\pi_j}\,\frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}}\underbrace{E[e_ie_j]}_{\pi_{ij}} =$$

$$= \sum^{N} \frac{y_i^2}{\pi_i^2}\pi_i(1-\pi_i) + 2\sum^{N}\sum_{j>i}^{N} \frac{y_i}{\pi_i}\,\frac{y_j}{\pi_j}(\pi_{ij}-\pi_i\pi_j) = V(\hat{\theta}_{HT}).$$

En el caso de m.a.s.s.r. $\begin{cases} \pi_i = \dfrac{n}{N} \\ \pi_{ij} = \dfrac{n(n-1)}{N(N-1)} \end{cases}$

$$\hat{V}(\hat{\theta}_{HT}) = \sum_{i=1}^{n} \frac{y_i^2}{\pi_i^2}(1-\pi_i) + 2\sum_{i=1}^{n}\sum_{j>i}^{n} \frac{y_i}{\pi_i}\cdot\frac{y_j}{\pi_j}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}}\right) =$$

$$= \sum_{i=1}^{n} \frac{y_i^2}{n^2/N^2}\underbrace{\left(1-\frac{n}{N}\right)}_{\frac{N-n}{N}} + 2\sum_{i=1}^{n}\sum_{j>i}^{n} \frac{y_i}{n/N}\cdot\frac{y_j}{n/N}\underbrace{\left(\frac{\frac{n(n-1)}{N(N-1)}-\frac{n}{N}\cdot\frac{n}{N}}{\frac{n(n-1)}{N(N-1)}}\right)}$$

$$\frac{N-n}{N} \rightarrow \frac{N^2-Nn}{N^2}$$

$$\underbrace{\frac{1}{\frac{n^2}{\cancel{N^2}}}\cdot\frac{N^2-Nn}{\cancel{N^2}}}$$

$$\sum_{i=1}^{n} y_i^2\cdot\frac{N(N-n)}{n^2}$$

$$\frac{1}{\frac{n^2}{N^2}}\cdot\underbrace{\frac{\cancel{\frac{n}{N}}\left[\frac{n-1}{N-1}-\frac{n}{N}\right]}{\cancel{\frac{n}{N}}\cdot\frac{n-1}{N-1}}}_{\left(1-\frac{n(N-1)}{N(n-1)}\right)}$$

$$\underbrace{\qquad}_{\frac{N^2}{n^2}-\frac{N}{n}\cdot\frac{N-1}{n-1}}$$

$$\downarrow$$

$$\frac{N^2(n-1)-Nn(N-1)}{n^2(n-1)}$$

$$\frac{\cancel{N^2n}-N^2-\cancel{N^2n}+Nn}{n^2(n-1)}$$

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

luego
$$\hat{V}(\hat{\theta}_{HT}) \overset{\text{massr}}{=} \frac{N(N-n)}{n^2}\left[\sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n}\sum_{j>i}^{n} \frac{y_i \cdot y_j}{n-1}\right]$$

Una vez simplificada esta expresión, podemos particularizarla para los parámetros poblacionales más utilizados:

Para el total poblacional, $\theta = \sum^{N} X_i$, aprovechando que la varianza es invariante frente a cambios de origen:

$$\hat{V}(\hat{X}_{HT}) = \frac{N(N-n)}{n^2}\left[\sum (X_i-\bar{x})^2 - \underbrace{2\sum_{i=1}^{n}\sum_{j>i}^{n}\frac{(X_i-\bar{x})\cdot(X_j-\bar{x})}{n-1}}_{=\frac{\sum^{n}(X_i-\bar{X})^2}{n-1}}\right] =$$

media muestral

(mirar atrás pág. 5)

$$= \frac{N(N-n)}{n^{\cancel{2}}} \cdot \frac{\cancel{n}}{n-1}\sum_{i=1}^{n}(X_i-\bar{x})^2$$

← Cuasivarianza muestral $\hat{S}^2 = \frac{\sum^{n}(X_i-\bar{X})^2}{n-1}$

$$= \frac{N(N-n)}{n}\cdot \hat{S}^2 = \frac{N\cdot N(N-n)}{n\cdot N}\hat{S}^2 = \frac{N^2}{n}\left(1-\frac{n}{N}\right)\cdot \hat{S}^2 =$$

$$\hat{V}(\hat{X}_{HT}) = N^2 \cdot (1-f)\cdot \frac{\hat{S}^2}{n}$$

Aprovechando que el estimador insesgado de la media poblacional es función del estimador del total poblacional

$$\hat{V}(\hat{\bar{X}}) = \hat{V}\left(\frac{\hat{X}}{N}\right) = \frac{1}{N^2}\hat{V}(\hat{X}) = \frac{1}{\cancel{N^2}}\cdot \cancel{N^2}(1-f)\frac{\hat{S}^2}{n} = (1-f)\frac{\hat{S}^2}{n}$$

Al igual que en el caso de la varianza del estimador de la proporción poblacional, para calcular el estimador insesgado de la varianza de la proporción poblacional, hay que calcular primero la expresión de la cuasivarianza muestral en poblac. dicotómicas $(A_i=\{1,0\})$.

Cuasivarianza muestral:

$$\hat{S}^2 = \frac{1}{n-1}\sum^n (X_i - \bar{X})^2 \overset{X_i = A_i}{=} \frac{1}{n-1}\sum^n (A_i - \hat{P})^2 = \frac{1}{n-1}\sum^n (A_i^2 - 2\hat{P}A_i + \hat{P}^2) =$$

$$= \frac{1}{n-1}\Big(\underbrace{\sum^n A_i^2}_{=\sum A_i = n\hat{P}} - 2\hat{P}\sum^n A_i + n\hat{P}^2\Big) = \frac{1}{n-1}\left(n\hat{P} - 2n\hat{P}^2 + n\hat{P}^2\right) =$$

$$= \frac{1}{n-1}\left(n\hat{P} - n\hat{P}^2\right) = \frac{1}{n-1}\, n\hat{P}(1-\hat{P}) = \frac{n}{n-1}\cdot \hat{P}\hat{Q}$$

Para A_i, $\hat{S}^2 = \frac{n}{n-1}\hat{P}\hat{Q}$

Utilizando la expresión para el estimador de la media poblacional,

$$\hat{V}(\hat{P}) = (1-f)\cdot\frac{\hat{S}^2}{n} \overset{A_i}{=} (1-f)\cdot\frac{\frac{\cancel{n}}{n-1}\hat{P}\hat{Q}}{\cancel{n}} = (1-f)\cdot\frac{\hat{P}\hat{Q}}{n-1}$$

y en el caso del total de clase:

$$\hat{V}(\hat{A}) = N^2\cdot(1-f)\cdot\frac{\hat{S}^2}{n} = \dots\dots = N^2(1-f)\cdot\frac{\hat{P}\hat{Q}}{n-1}$$

Parámetro, θ	$\hat{\theta}_{HT}$	$V(\hat{\theta}_{HT})$	$\hat{V}(\hat{\theta}_{HT})$
Total poblac.	$\hat{X}_{HT} = N\bar{x}$	$V(\hat{X}) = N^2(1-f)\frac{S^2}{n}$	$\hat{V}(\hat{X}) = N^2(1-f)\cdot\frac{\hat{S}^2}{n}$
Media poblac.	$\hat{\bar{X}}_{HT} = \bar{x}$	$V(\hat{\bar{X}}) = (1-f)\frac{S^2}{n}$	$\hat{V}(\hat{\bar{X}}) = (1-f)\cdot\frac{\hat{S}^2}{n}$
Total de clase	$\hat{A}_{HT} = N\hat{P}$	$V(\hat{A}) = \frac{N^3}{N-1}(1-f)\frac{PQ}{n}$	$\hat{V}(\hat{A}) = N^2(1-f)\frac{\hat{P}\hat{Q}}{n-1}$
Proporción poblacional	$\hat{P}_{HT} = \hat{P}$	$V(\hat{P}) = \frac{N}{N-1}\frac{1}{n}(1-f)PQ$	$\hat{V}(\hat{P}) = (1-f)\frac{\hat{P}\hat{Q}}{n-1}$

$S^2 \equiv$ cuasivarianza poblacional

$\hat{S}^2 \equiv$ cuasivarianza muestral

$f \equiv$ fracción de muestreo

3_ MUESTREO ALEATORIO SIMPLE CON REPOSICIÓN

Procedimiento de selección de la muestra

- con probab. iguales,
- las muestras con elementos repetidos son posibles,
- cada elemento poblacional puede estar $0, 1, \dots n$ veces en la muestra.

Al ser un muestreo ~~con probab.~~ iguales:

$P(u_i \in \text{muestra}) = P_i = \frac{1}{N}, \ \forall i.$

en cada extracción

Al ser todas las muestras equiprobables:

$$P(\text{Muestra}) = P(u_1 \dots u_n) \overset{\text{con repos.}}{=} P(u_1)P(u_2)\dots P(u_n) = \underbrace{\frac{1}{N} \cdot \dots \cdot \frac{1}{N}}_{n \text{ veces}} =$$

$$= \frac{1}{N^n} \leftarrow \text{sí importa el orden}$$

Para cada elemento poblacional u_i, $i = 1 \dots N$, definimos

$e_i \equiv$ nº veces que aparece u_i en la muestra

$e_i = 0, 1, 2 \dots n \qquad e_i \to B(n, p = P_i = \frac{1}{N})$ por lo que:

$$E[e_i] = np = \frac{n}{N}$$

$$V[e_i] = npq = n \cdot \frac{1}{N} \underbrace{\left(1 - \frac{1}{N}\right)}_{\frac{N-1}{N}} = \frac{n(N-1)}{N^2}$$

$$\text{cov}[e_i e_j] = E[e_i e_j] - E[e_i]E[e_j] = n(n-1)P_iP_j - nP_i nP_j = -nP_iP_j$$

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

Recordar que:

$E[e_i e_j] = n(n-1)P_iP_j$

$\text{cov}[e_i e_j] = -nP_iP_j$

← (general)

Estimador lineal insesgado

El estimador lineal insesgado en el caso de m.a.s.c.r. es el estimador de Hansen y Hurwitz, $\hat{\Theta}_{HH}$.

En general, para un parámetro poblacional $\Theta = \sum_{i=1}^{N} Y_i$, resultado de sumar las características de todas las unidades poblacionales, el estimador HH es una suma ponderada de las mismas características observadas sobre las unidades muestrales, donde las ponderaciones son $\frac{1}{nP_i}$.

$$\Theta = \sum_{i=1}^{N} Y_i \xrightarrow{\text{CON r.}} \hat{\Theta}_{HH} = \sum_{i=1}^{n} \frac{y_i}{nP_i} \quad \leftarrow \text{General}$$

coincide con $\hat{\Theta}_{HT}$ (en variante uno)

Es un estimador insesgado:

$$e_i \to B(n, P_i)$$

$$E[\hat{\Theta}_{HH}] = E\left[\sum^{n} \frac{y_i}{nP_i}\right] = E\left[\sum^{N} \frac{Y_i}{nP_i} e_i\right] = \sum^{N} \frac{Y_i}{nP_i} E[e_i] = \sum^{N} \frac{Y_i}{nP_i} \cdot nP_i =$$

$$= \sum^{N} Y_i = \Theta.$$

En el caso de probabilidades iguales, el estimador lineal insesgado de Hansen y Hurwitz queda:

$$\hat{\Theta}_{HH} = \sum_{i=1}^{n} \frac{y_i}{nP_i} = \sum_{i=1}^{n} \frac{y_i}{n \cdot \frac{1}{N}} = \frac{N}{n} \sum_{i=1}^{n} y_i$$

(coincide con $\hat{\Theta}_{HT}$, aunque [illegible] ≠)

cuya expresión se puede particularizar para los parám. más comunes:

Parámetro	Θ	Estimador	$\hat{\Theta}_{HH}$
Total poblacional	$X = \sum_{i=1}^{N} X_i$	$\hat{X}_{HH} = \sum_{i=1}^{n} \frac{x_i}{n \cdot 1/N} = N \sum_{i=1}^{n} \frac{x_i}{n}$	$\hat{X}_{HH} = N\bar{x}$
Media poblacional	$\bar{X} = \sum_{i=1}^{N} \frac{X_i}{N}$	$\hat{\bar{X}}_{HH} = \sum_{i=1}^{n} \frac{x_i/\cancel{N}}{n/\cancel{N}} = \sum_{i=1}^{n} \frac{x_i}{n}$	$\hat{\bar{X}}_{HH} = \bar{x}$
Total de clase	$A = \sum_{i=1}^{N} A_i$	$\hat{A} = \sum_{i=1}^{n} \frac{A_i}{n \cdot 1/N} = N \sum_{i=1}^{n} \frac{A_i}{n}$	$\hat{A}_{HH} = N\hat{P}$
Proporción poblacional	$P = \sum_{i=1}^{N} \frac{A_i}{N}$	$\hat{P} = \sum_{i=1}^{n} \frac{A_i/N}{n/N} = \sum_{i=1}^{n} \frac{A_i}{n}$	$\hat{P}_{HH} = \hat{P}$

$\bar{X}$ ≡ media poblacional, $\bar{x}$ ≡ media muestral

VARIANZAS de los estimadores $\hat{\theta}_{HH}$

En general, $V(\hat{\theta}_{HH}) = V\left(\sum_{i=1}^{n} \frac{y_i}{nP_i}\right) = V\left(\sum_{i=1}^{N} \frac{y_i}{nP_i} e_i\right) =$

$$= \sum_{i=1}^{N} V\left(\frac{y_i}{nP_i} e_i\right) + \sum_{i\neq j}^{N} \operatorname{cov}\left(\frac{y_i}{nP_i} e_i, \frac{y_j}{nP_j} e_j\right) =$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{n^2 P_i^2} V(e_i) + \sum_{i\neq j}^{N} \frac{y_i}{nP_i}\cdot\frac{y_j}{nP_j} \operatorname{cov}(e_i, e_j) =$$

mirar atrás

$$= \sum_{i=1}^{N} \frac{y_i^2}{n^2 P_i^2} \cancel{nP_i}(1-P_i) + \sum_{i\neq j}^{N} \frac{y_i}{\cancel{nP_i}}\cdot\frac{y_j}{\cancel{nP_j}}(-\cancel{nP_iP_j}) =$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{nP_i}(1-P_i) - \sum_{i=1}^{N} \frac{y_i\cdot y_j}{n} = \sum_{i=1}^{N} \frac{y_i^2}{nP_i} - \sum_{i=1}^{N} \frac{y_i^2\cdot \cancel{P_i}}{n\cancel{P_i}} - \sum_{i=1}^{N} \frac{y_i y_j}{n} \overset{(*)}{=}$$

$$= \frac{1}{n}\sum_{i=1}^{N} \frac{y_i^2}{P_i} - \cancel{\frac{1}{n}\sum_{i=1}^{N} y_i^2} + \cancel{\frac{1}{n}\sum_{i=1}^{N} y_i^2} - \frac{\theta^2}{n} = \frac{1}{n}\left[\sum_{i=1}^{N} \frac{y_i^2}{P_i} - \theta^2\right]$$

Otra manera de escribir la varianza del estimador de Hansen y Hurvitz es:

$$V(\hat{\theta}_{HH}) = \frac{1}{n}\sum_{i=1}^{N}\left(\frac{y_i}{P_i} - \theta\right)^2 P_i \, ,$$

ya que:

$$\sum^{N}\left(\frac{y_i}{P_i} - \theta\right)^2 P_i = \sum^{N}\left(\frac{y_i^2}{P_i^2} - 2\theta\frac{y_i}{P_i} + \theta^2\right)P_i = \sum^{N} \frac{y_i^2}{P_i^2}\cdot P_i -$$

$$- 2\theta \underbrace{\sum_{i=1}^{N} \frac{y_i}{\cancel{P_i}}\cdot \cancel{P_i}}_{\theta} + \theta^2 \underbrace{\sum_{i=1}^{N} P_i}_{1} = \sum_{i=1}^{N} \frac{y_i^2}{P_i} - 2\theta^2 + \theta^2 = \sum_{i=1}^{N} \frac{y_i^2}{P_i} - \theta^2 \text{ cqd.}$$

Resumiendo, expresiones alternativas de la varianza del estimador de Hansen y Hurvitz son:

$$V(\hat{\theta}_{HH}) = \frac{1}{n}\left(\sum_{i=1}^{N} \frac{y_i^2}{P_i} - \theta^2\right)$$

$$V(\hat{\theta}_{HH}) = \frac{1}{n}\sum_{i=1}^{N}\left(\frac{y_i}{P_i} - \theta\right)^2 P_i \quad (*) \quad \leftarrow \text{General}$$

$$V(\hat{\theta}_{HH}) = \frac{1}{n}\sum_{i=1}^{N}\sum_{j>i}^{N}\left(\frac{y_i}{P_i} - \frac{y_j}{P_j}\right)^2 P_i P_j$$

Al ser en m.a.s.c.r. las probabilidades $P_i = \frac{1}{N}$, el estimador de Hansen y Hurvitz queda para la siguiente expresión

$$V(\hat{\theta}_{HH}) = \frac{1}{n}\sum_{i=1}^{N}\left(\frac{Y_i}{P_i}-\theta\right)^2 P_i = \frac{1}{n}\sum_{i=1}^{N}\left(\frac{Y_i}{1/N}-\theta\right)^2 \cdot \frac{1}{N} =$$

$$V(\hat{\theta}_{HH}) = \frac{1}{n}\sum_{i=1}^{N}(NY_i-\theta)^2 \cdot \frac{1}{N} \longleftarrow \text{m.a.s.c.r.}$$

que se puede particularizar para los parámetros poblacionales más comunes:

Para el total poblacional, $\theta = X = \sum_{i=1}^{N} X_i$; $\hat{X} = N\bar{x}$

$$V(\hat{X}_{HH}) = \frac{1}{n}\sum_{i=1}^{N}(NX_i - X)^2 \cdot \frac{1}{N} = \frac{1}{n}\sum_{i=1}^{N}(NX_i - N\bar{X})^2 \cdot \frac{1}{N} =$$

$$= \frac{1}{n} N^2 \underbrace{\sum_{i=1}^{N}(X_i-\bar{X})^2 \cdot \frac{1}{N}}_{\text{varianza poblac.}} = \frac{N^2}{n} \cdot \sigma^2$$

Para la media poblacional, $\theta = \bar{X} = \sum_{i=1}^{N}\frac{X_i}{N}$; $\hat{\bar{X}} = \bar{x}$

$$V(\hat{\bar{X}}) = V\left(\frac{\hat{X}}{N}\right) = \frac{1}{N^2}V(\hat{X}) = \frac{1}{\cancel{N^2}} \cdot \frac{\cancel{N^2}}{n}\sigma^2 = \frac{\sigma^2}{n}$$

Para la proporción poblacional, $\theta = P$; $\hat{\theta} = \hat{P}$

hay que tener en cuenta la relación entre la varianza poblacional σ^2 y la cuasivarianza poblacional S^2:

$$N\sigma^2 = (N-1)S^2 \longrightarrow \sigma^2 = \frac{N-1}{N}S^2 \qquad (\sigma^2 = PQ)$$

y la expresión de la cuasivarianza poblacional para poblaciones dicotómicas, $A_i = 1$ ó 0. (mirar pág 6)

$$S^2 = \frac{N}{N-1}PQ$$

$$V(\hat{P}) = \frac{\sigma^2}{n} = \frac{\frac{N-1}{N}S^2}{n} = \frac{\frac{N-1}{N}\cdot\frac{N}{N-1}PQ}{n} = \frac{PQ}{n}$$

y para el total de clase, $A = \sum A_i$, $\hat{A} = N\hat{P}$

$$V(\hat{A}) = V(N\hat{P}) = N^2 V(\hat{P}) = N^2\frac{PQ}{n}$$

ESTIMACIÓN de las varianzas de $\hat{\theta}_{HH}$

Al igual que ocurría con $\hat{\theta}_{HT}$, las varianzas de los estimadores lineales insesgados de Hansen y Hurwitz dependen de parámetros poblacionales, desconocidos, por lo que es necesario estimarlas.

Una estimación insesgada para la varianza del estimador de Hansen y Hurwitz puede expresarse como:

$$\hat{V}(\hat{\theta}_{HH}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{Y_i}{P_i} - \hat{\theta}_{HH}\right)^2 \leftarrow \text{General}$$

Efectivamente, es una estimación insesgada de $V(\hat{\theta}_{HH})$, ya que: (1º cuadrado, 2º esperanzas).

$$E[\hat{V}(\hat{\theta}_{HH})] = E\left[\frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{Y_i}{P_i} - \hat{\theta}_{HH}\right)^2\right] = \cancel{E\left[\frac{1}{n(n-1)}\right.}$$

$$= E\left[\frac{1}{n(n-1)}\left(\sum_{i=1}^{n} \frac{Y_i^2}{P_i^2} - 2\hat{\theta}_{HH} \cdot \underbrace{\sum^{n} \frac{Y_i}{P_i}}_{n\cdot\hat{\theta}_{HH}} + n\hat{\theta}_{HH}^2\right)\right] =$$

$$= E\left[\frac{1}{n(n-1)}\left(\sum_{i=1}^{n} \frac{Y_i^2}{P_i^2} - n\hat{\theta}_{HH}^2\right)\right] \overset{\text{tomando esperanzas}}{=}$$

$$= \frac{1}{n(n-1)}\left[E\left(\sum^{n}\left(\frac{Y_i}{P_i}\right)^2\right) - n\,\underbrace{E[\hat{\theta}_{HH}^2]}\right] =$$

$$E\left[\sum^{N}\left(\frac{Y_i}{P_i}\right)^2 \cdot e_i\right] \qquad V[\hat{\theta}_{HH}] + \underbrace{E[\hat{\theta}_{HH}]}_{\theta}{}^2,$$

$$\sum^{N}\left(\frac{Y_i}{P_i}\right)^2 \cdot \underbrace{E[e_i]}_{nP_i} \qquad V[\hat{\theta}_{HH}] + \theta^2$$

$$n \cdot \sum^{N} \frac{Y_i^2}{P_i}$$

$$= \frac{1}{\cancel{n}(n-1)}\left[\cancel{n}\sum^{N} \frac{Y_i^2}{P_i} - \cancel{n}\left(V[\hat{\theta}_{HH}] + \theta^2\right)\right] =$$

$$= \frac{1}{n-1}\left[\underbrace{\sum^{N} \frac{Y_i^2}{P_i} - \theta^2}_{n\,V(\hat{\theta}_{HH})} - V[\hat{\theta}_{HH}]\right] = \frac{\cancel{n-1}}{\cancel{n-1}} V[\hat{\theta}_{HH}] \text{ cqd.}$$

Al ser las probabilidades iguales en el m.a.s.c.r. $P_i = \frac{1}{N}, \ \forall i$, la estimación de la varianza del estimador de Hansen y Hurwitz queda:

$$\hat{V}(\hat{\theta}_{HH}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left[\frac{Y_i}{P_i} - \hat{\theta}_{HH} \right]^2 =$$

$$P_i = \frac{1}{N} \qquad \hat{\theta}_{HH} = \frac{N}{n} \sum_{i=1}^{n} Y_i$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(N Y_i - \frac{N}{n} \sum_{i=1}^{n} Y_i \right)^2 = \frac{N^2}{n(n-1)} \sum_{i=1}^{n} \left(Y_i - \bar{y} \right)^2 =$$

$$= \frac{N^2}{n} \sum_{i=1}^{n} \frac{(Y_i - \bar{y})^2}{n-1}$$

~~que~~ cuya expresión se puede particularizar para las características poblacionales más comunes:

Para el total poblacional: media muestral

$$\hat{V}(\hat{X}) = \frac{1}{n(n-1)} \left(\sum_{i=1}^{n} \frac{X_i}{1/N} - N\bar{x} \right)^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} (N X_i - N\bar{x})^2 =$$

$$= \frac{N^2}{n(n-1)} \sum_{i=1}^{n} (X_i - \bar{x})^2 = \frac{N^2}{n(n-1)} \cdot (n-1) \hat{S}^2 = \frac{N^2}{n} \hat{S}^2$$

donde $\hat{S}^2 \equiv$ cuasivarianza muestral $= \frac{\sum^{n} (X_i - \bar{x})^2}{n-1}$.

Para la media poblacional:

$$\hat{V}(\hat{\bar{X}}) = \hat{V}\left(\frac{\hat{X}}{N}\right) = \frac{1}{N^2} \hat{V}(\hat{X}) = \frac{1}{N^2} \cdot \frac{N^2}{n} \hat{S}^2 = \frac{\hat{S}^2}{n}$$

Para la proporción poblacional, hay que tener en cuenta la expresión de la cuasivarianza muestral en poblaciones dicotómicas, $A_i = \{0, 1\}$.

$$\hat{S}^2 = \frac{\sum^{n} (X_i - \bar{X})^2}{n-1} = \frac{\sum^{n} (A_i - \hat{P})^2}{n-1} = \frac{1}{n-1} \cdot n \hat{P} \hat{Q} \quad \text{con} \quad \hat{P} = \frac{\sum^{n} A_i}{n}, \quad \hat{Q} = 1 - \hat{P}$$

luego

$$\hat{V}(\hat{P}) = \frac{\hat{S}^2}{n} = \frac{\frac{n}{n-1}\hat{P}\hat{Q}}{n} = \frac{\hat{P}\hat{Q}}{n-1}$$

Para el total de clase:

$$\hat{V}(\hat{A}) = \hat{V}(N\hat{P}) = N^2\hat{V}(\hat{P}) = N^2 \cdot \frac{\hat{P}\hat{Q}}{n-1}$$

θ	$\hat{\theta}_{HH}$	$V(\hat{\theta}_{HH})$	$\hat{V}(\hat{\theta}_{HH})$
Total poblacional	$\hat{X}_{HH} = N\bar{x}$	$V(\hat{X}) = \frac{N^2}{n}\sigma^2$	$\hat{V}(\hat{X}) = N^2 \cdot \frac{\hat{S}^2}{n}$
Media poblacional	$\hat{\bar{X}}_{HH} = \bar{x}$	$V(\hat{\bar{X}}) = \frac{\sigma^2}{n}$	$\hat{V}(\hat{\bar{X}}) = \frac{\hat{S}^2}{n}$
Total de clase	$\hat{A}_{HH} = N\hat{P}$	$V(\hat{A}) = N^2\frac{PQ}{n}$	$\hat{V}(\hat{A}) = N^2 \cdot \frac{\hat{P}\hat{Q}}{n-1}$
Proporción poblacional	$\hat{P}_{HH} = \hat{P}$	$V(\hat{P}) = \frac{PQ}{n}$	$\hat{V}(\hat{P}) = \frac{\hat{P}\hat{Q}}{n-1}$

Nótese que las estimaciones de las varianzas de los estimadores de Hansen y Hurvitz (m.a.s.c.r) coinciden con las estimaciones de las varianzas de los estimadores de Horvitz y Thompson (m.a.s.s.r), salvo en el factor de corrección $(1-f)$.

4. CONSIDERACIONES SOBRE EL TAMAÑO de la MUESTRA

En el muestreo, a la hora de seleccionar la muestra, es estrictamente necesario conocer su tamaño.

Al aproximar las características poblacionales desconocidas mediante estimadores basados en la muestra, se comete un error, error que mide la representatividad de la muestra.

Dependiendo de la accesibilidad y disponibilidad del marco, del coste de la entrevista de las unidades encuestados, del presupuesto disponible y de muchos otros factores, fijaremos un error de muestreo.

Error de muestreo
- absoluto, $e = \sigma(\hat{\theta}) \equiv$ desv. típica
- relativo, $e_r = CV(\hat{\theta}) = \dfrac{\sigma(\hat{\theta})}{E[\hat{\theta}]}$.

NO En este caso, consideramos únicamente un error de muestreo absoluto, y distinguiremos entre muestreo con reposición y sin reposición.

a) Muestreo SIN reposición

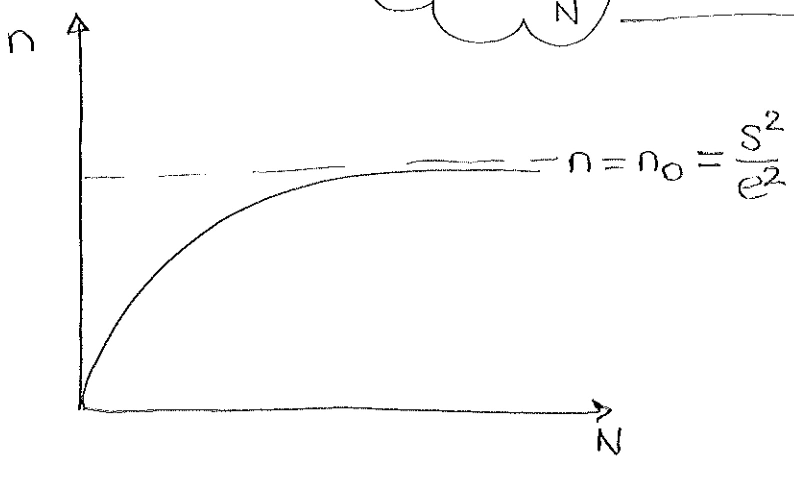
a.1. $\theta = \overline{X} \equiv$ media poblacional $\longrightarrow n = \dfrac{NS^2}{Ne^2+S^2}$

$$e = \sigma(\hat{\theta}) = \sqrt{(1-f)\frac{S^2}{n}} \longrightarrow e^2 = (1-f)\frac{S^2}{n} = \left(1-\frac{n}{N}\right)\frac{S^2}{n}$$

divid. por N

$$e^2 = \frac{S^2}{n} - \frac{S^2}{N} \Rightarrow \frac{S^2}{n} = e^2 + \frac{S^2}{N} \Rightarrow n = \frac{S^2}{e^2 + S^2/N} = \frac{NS^2}{Ne^2+S^2}$$

Gráficamente: $n = \dfrac{n_0}{1+\dfrac{n_0}{N}}$ $\quad n_0 = \dfrac{S^2}{e^2}$ (tamaño muestral inversamente proporcional al error de muestreo)



- Siempre creciente
- Sin máximos ni mínimos
- Pasa por el origen
- Asíntota horizontal en $\dfrac{S^2}{e^2}$ (tamaño de la muestra en poblac. infinitas)
- Sin punto de inflexión
- a partir de cierto N no aumenta n.

a.2. Total poblacional, $\theta = X$ $\longrightarrow$ $n = \frac{N^2 S^2}{e^2 + N S^2}$

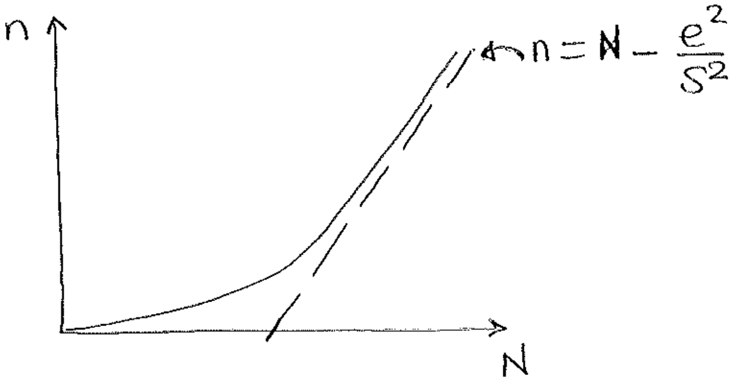
$$e = \sigma(\hat{\theta}) = \sqrt{N^2(1-f)\frac{S^2}{n}} \rightarrow e^2 = N^2\left(1-\frac{n}{N}\right)\frac{S^2}{n} = N^2 \cdot \frac{S^2}{n} - \frac{N^2}{N}S^2 =$$

$$= \frac{N^2 S^2}{n} - NS^2 \Rightarrow \frac{N^2 S^2}{n} = e^2 + NS^2$$

Despejando, Divido por e^2

$$n = \frac{N^2 S^2}{e^2 + NS^2} \stackrel{\swarrow}{=} \frac{N^2 \cdot S^2/e^2}{e^2/e^2 + N \cdot S^2/e^2} = \frac{N^2 \cdot n_1}{1 + N \cdot n_1} \quad \left(n_1 = S^2/e^2\right) \quad \left(n \xrightarrow[N \to \infty]{} \infty\right)$$

Gráficamente,



- Siempre creciente
- Pasa por el origen
- Asíntota oblicua en $n = N - e^2/S^2$
- Sin máximos ni mínimos
- Convexa
- Al aumentar N tb aumenta n

En poblaciones dicotómicas, ocurre prácticamente lo mismo con algunas salvedades.

a.3. Proporción poblacional, $\theta = P$

Teniendo en cuenta que $S^2 = \frac{N}{N-1} PQ$ para $A_i = \{0,1\}$, sustituyendo en la fórmula de $\bar{X}$

$$n = \frac{NS^2}{Ne^2 + S^2} = \frac{N \cdot \frac{N}{N-1} PQ}{Ne^2 + \frac{N}{N-1} PQ} = \ldots = \frac{NPQ}{(N-1)e^2 + PQ} \quad \left(n \xrightarrow[N \to \infty]{} \frac{PQ}{e^2}\right).$$

Sigue teniendo una asíntota horizontal, $n = \frac{PQ}{e^2}$

El tamaño ~~p~~ muestral es inversamente proporcional al cuadrado del error de muestreo y directamente proporcional a P.

Para poblaciones grandes o fracción de muestreo pequeña, el valor máximo de n se obtiene para $P = Q = 1/2$.

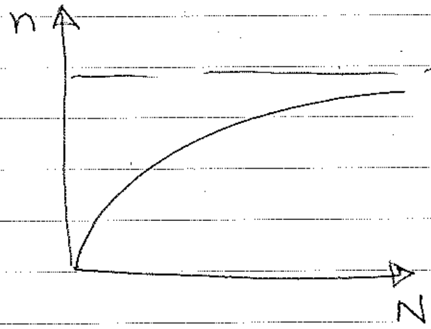
Por lo tanto, para un error de muestreo prefijado, se necesitarán tamaños muestrales más pequeños para P próximo a 0 ó a 1.

Tamaño de la muestra

SR

Para la media:

$$e = \sigma(\hat{\bar{X}}) = \sqrt{V(\hat{\bar{X}})} = \sqrt{(1-f)\frac{S^2}{n}} \Rightarrow e^2 = \left(1 - \frac{n}{N}\right)\frac{S^2}{n}$$



$$n_0 = \frac{S^2}{e^2}$$

$$= \frac{S^2}{n} - \frac{S^2}{N}$$

$$n = \frac{N S^2}{N e^2 + S^2}$$

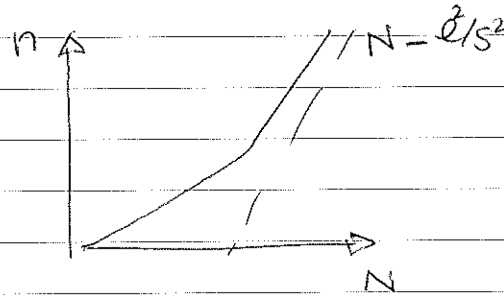
$$e_r = \frac{\sigma(\hat{\bar{X}})}{E(\hat{\bar{X}})} = \frac{\sqrt{(1-f)S^2/n}}{\bar{X}} \Rightarrow e_r^2 = \frac{\frac{S^2}{n} - \frac{S^2}{N}}{\bar{X}^2} =$$

$$e_r^2 = \frac{1}{n} C_{1,X}^2 - \frac{1}{N} C_{1,X}^2$$

$$n = \frac{N C_{1,X}^2}{N e_r^2 + C_{1,X}^2}$$

Para el total:

$$e = \sigma(\hat{X}) = \sqrt{V(\hat{X})} = \sqrt{N^2(1-f)S^2/n} \Rightarrow e^2 = N^2\left(1 - \frac{n}{N}\right)\frac{S^2}{n}$$



$$N - e^2/S^2$$

$$e^2 = N^2 \cdot \frac{S^2}{n} - N S^2$$

$$n = \frac{N^2 S^2}{e^2 + N S^2}$$

$$e_r = \frac{\sigma(\hat{X})}{E(\hat{X})} = \frac{N\sigma(\hat{\bar{X}})}{N E(\hat{\bar{X}})} = e_r(\hat{\bar{X}}) \Rightarrow \text{Coincide para la media y para el total}$$

En la práctica, para estimar la proporción poblacional se utiliza una encuesta piloto, cuyo tamaño se determina con $P=\frac{1}{2}$, (máximo tamaño muestral para el error fijado).

a.4. Total de clase

$$n=\frac{N^2S^2}{e^2+NS^2}=\frac{N^2\frac{N}{N-1}PQ}{e^2+N\frac{N}{N-1}PQ}=\ldots=\frac{N^3PQ}{(N-1)e^2+N^2PQ} \quad \left(n \xrightarrow[N\to\infty]{} \infty\right).$$

b) Muestreo CON reposición

De la misma manera que en el muestreo SIN reposición, a partir del error de muestreo se puede despejar el tamaño muestral.

b.1. Media poblacional, $\cancel{\ }\ \overline{X}$

$$e^2=V(\hat{\bar{X}})=\frac{\sigma^2}{n} \Rightarrow n=\frac{\sigma^2}{e^2} \quad \text{(no depende de N)}$$

b.2. Total poblacional, X

$$e^2=V(\hat{X})=N^2\cdot\frac{\sigma^2}{n} \Rightarrow n=\frac{N^2\sigma^2}{e^2} \quad (N\to\infty, n\to\infty)$$

b.3. Proporción poblacional, P

$$e^2=V(\hat{P})=\frac{PQ}{n} \Rightarrow n=\frac{PQ}{e^2} \quad \text{(no depende de N)}$$

b.4. Total de clase, A

$$e^2=V(\hat{A})=N^2\cdot\frac{PQ}{n} \Rightarrow n=\frac{N^2PQ}{e^2} \quad \left(n \xrightarrow[N\to\infty]{} \infty\right)$$

5. COMPARACIÓN entre muestreo CON y SIN

Se pueden realizar comparaciones de precisión entre el muestreo aleatorio simple sin reposición (m.a.s.s.r.) y el muestreo aleatorio simple con reposición (m.a.s.c.r.).

Estas comparaciones podrán hacerse a través de las varianzas de los estimadores o a través del tamaño muestral necesario para cometer un error de muestreo dado.

a) Criterio: error del muestreo

$e = \sqrt{V(\hat{\theta})}$ = desviación típica del estimador → concentrac. de los valores del estimador alrededor de su valor medio.

Será más preciso el método de selección con un error de muestreo menor ⟹ menor varianza de los estim.

Si acudimos a las expresiones generales de las varianzas de los estimadores CR y SR, éstas no son fáciles de comparar, por lo que lo haremos para cada una de las características poblacionales más usuales

$(N-1)S^2 = N\sigma^2$

~~$\hat{X} \to \{ V_{SR}(\hat{X}) = (1-f)\cdot\frac{S^2}{n} = (1-$~~

Para la media poblacional:

$$V_{SR}(\hat{X}) = (1-f)\cdot\frac{S^2}{n} = \left(1-\frac{n}{N}\right)\cdot\frac{\frac{N}{N-1}\sigma^2}{n} = \left(\frac{N-n}{\cancel{N}}\right)\left(\frac{\cancel{N}}{N-1}\right)\cdot\frac{\sigma^2}{n} =$$

$$= \frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}.$$

$$V_{CR}(\hat{X}) = \frac{\sigma^2}{n}$$

$\frac{N-n}{N-1} < 1$ para $n>1$

luego $V_{SR}(\hat{X}) < V_{CR}(\hat{X})$

Para el total poblacional:

$$V_{SR}(\hat{X}) = V_{SR}(N\hat{\bar{X}}) = N^2 V_{SR}(\hat{\bar{X}}) = N^2 \cdot \frac{N-1}{N-n} \cdot \frac{\sigma^2}{n}$$

$$V_{CR}(\hat{X}) = V_{CR}(N\hat{\bar{X}}) = N^2 V_{CR}(\hat{\bar{X}}) = N^2 \cdot \frac{\sigma^2}{n}$$

Por la misma razón, $V_{SR}(\hat{X}) < V_{CR}(\hat{X})$.

Para la proporción poblacional:

$$V_{SR}(\hat{P}) = \frac{N}{N-1} \cdot \frac{1}{n} \cdot (1-f)PQ = \frac{N}{N-1} \cdot \left(1 - \frac{n}{N}\right) \cdot \frac{PQ}{n} =$$

$$= \frac{\cancel{N}}{N-1} \cdot \frac{N-n}{\cancel{N}} \cdot \frac{PQ}{n} = \frac{N-n}{N-1} \cdot \frac{PQ}{n}$$

$$V_{CR}(\hat{P}) = \frac{PQ}{n}$$

luego $V_{SR}(\hat{P}) < V_{CR}(\hat{P})$.

Análogamente,

$$V_{SR}(\hat{A}) = V_{SR}(N\hat{P}) = N^2 V_{SR}(\hat{P}) = N^2 \cdot \frac{N-n}{N-1} \cdot \frac{PQ}{n}$$

$$V_{CR}(\hat{A}) = V_{CR}(N\hat{P}) = N^2 V_{CR}(\hat{P}) = N^2 \cdot \frac{PQ}{n}$$

luego $V_{SR}(\hat{A}) < V_{CR}(\hat{A})$.

Resumiendo, para $\theta = X, \bar{X}, A, P$ se verifica que

$$V_{SR}(\hat{\theta}) = \underbrace{\frac{N-n}{N-1}}_{<1} \cdot V_{CR}(\hat{\theta}) \Rightarrow V_{SR}(\hat{\theta}) < V_{CR}(\hat{\theta}),$$

por lo que en todos los casos el muestreo SIN reposición es más preciso que el muestreo <u>con</u> reposición.

b) Criterio: tamaño muestral

Atendiendo a este criterio, será mejor el método con menor tamaño muestral para un error de muestreo dado.

En todas las situaciones (error absoluto, error relativo y error de muestreo y coef. de confianza), el muestreo SIN reposición necesita menos tamaño muestral para cometer el mismo error que el muestreo CON reposición, por lo que SIN mejor que CON.

~~En el epí~~
Cuando se estudió el tamaño muestral se demostró que para poblaciones grandes o fracciones de muestreo pequeñas (que es lo común):

$$n_{SR} = \frac{n_0}{1+n_0/N} \quad , \quad n_0 = \frac{S^2}{e^2}$$

tanto en el caso de estimaciones de medias y proporciones para e_a como en el caso de estimaciones de medias, totales, proporciones y total de clase para e_r con o sin coef. de confianza.

En los mismos casos, se observa

$$n_{CR} = n_0 \qquad \left(n_0 = \frac{\sigma^2}{e^2} \text{, pero suponemos que para tamaños poblac. grandes } \sigma^2 \sim S^2 \right)$$

Por lo que:

$$n_{SR} = \frac{n_0}{1+n_0/N} = \frac{n_{CR}}{\underbrace{1+n_{CR}/N}_{>1 \text{, siempre}}} < n_{CR} \Rightarrow n_{SR} < n_{CR}$$

Para totales: $\left(n_{CR} = N^2 \cdot \frac{\sigma^2}{e^2} \sim N^2 \cdot n_1 \right)$

$$n_{SR} = \frac{N^2 n_1}{1+N n_1} \simeq \frac{n_{CR}}{\underbrace{1+\frac{n_{CR}}{N}}_{>1 \text{ siempre}}} < n_{CR} \Rightarrow n_{SR} < n_{CR}$$

PROBABILIDADES IGUALES

PROBAB

$SR \rightarrow P(\text{muestra}) = \frac{CF}{CP} = \frac{1}{\binom{N}{n}} =$

$P(u_i \in \text{muestra}) = \pi_i = \frac{CF}{CP} = \frac{\text{nº muestras con } u_i}{\text{nº total muestras}} = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$

$P((u_i, u_j) \in \text{Muestra}) = \pi_{ij} = \frac{n(n-1)}{N(N-1)}$

$CR \rightarrow P(\text{muestra}) = \frac{1}{N^n}$

$P_i = \frac{1}{N}, \forall i$

$e_i \rightarrow B(n, \frac{1}{N}) \Rightarrow E = \frac{n}{N}, V = \frac{n}{N}(1-\frac{1}{N}) = \frac{n(N-1)}{N^2}$

$COV = -\frac{n}{N^2}$

ESTIMADOR

$SR \rightarrow \hat{\theta}_{HT} = \sum_{i=1}^{n} \frac{Y_i}{\pi_i} \overset{\pi_i = \frac{n}{N}}{=} N \sum_{i=1}^{n} \frac{Y_i}{n}$ $\left(\hat{X} = N\bar{x};\ \hat{\bar{X}} = \bar{x};\ \hat{A} = N\hat{p};\ \hat{P} = \sum \frac{A_i}{n} \right)$

$CR \rightarrow \hat{\theta}_{HH} = \sum_{i=1}^{n} \frac{Y_i}{nP_i} \overset{P_i = \frac{1}{N}}{=} \sum_{i=1}^{n} \frac{Y_i}{n/N} = \hat{\theta}_{HT}$ (sólo el estimador)

VARIANZA

$SR \rightarrow V(\hat{\theta}_{HT}) = \frac{N-n}{n}\left[\sum_{i=1}^{N} Y_i^2 - 2\sum_{i=1}^{N}\sum_{j>i}^{N} \frac{Y_i Y_j}{N-1}\right]$ (prob. iguales)

Para los casos particulares, memorizar

$V(\hat{\bar{X}}_{HT}) = (1-f) \cdot \frac{1}{n(N-1)} \underbrace{\sum_{i=1}^{N} (X_i - \bar{X})^2}_{= NPQ \text{ para dicotómicas}}$; $V(\hat{\bar{X}}_{HT}) = (1-f) \cdot \frac{S^2}{n}$

$CR \rightarrow V(\hat{\theta}_{HH}) = \frac{1}{n} \cdot \sum_{i=1}^{N} (NY_i - \theta)^2 \cdot \frac{1}{N}$

Recordar que $V(\hat{\bar{X}}_{HH}) = \frac{\sigma^2}{n}$, etc.

$V(\hat{\bar{X}}_{HH}) = \frac{\sigma^2}{n}$

ESTIMACIÓN

SR $\longrightarrow \hat{V}(\hat{\Theta}_{HT}) = \frac{N(N-n)}{n^2}\left[\sum_{i=1}^{n} Y_i^2 - 2\sum_{i=1}^{n}\sum_{j>i}^{n} \frac{Y_i Y_j}{n-1}\right]$.

Las estimaciones particulares:

$\hat{V}(\hat{\bar{X}}) = (1-f)\cdot\frac{\hat{S}^2}{n}$ y $\hat{S}^2 = \frac{n}{n-1}\hat{P}\hat{Q}$ en dicotómicas

CR $\longrightarrow \hat{V}(\hat{\Theta}_{HH}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(NY_i - \hat{\Theta}_{HH})^2$

$\hat{V}(\hat{\bar{X}}) = \frac{\hat{S}^2}{n}$ y $\hat{S}^2 = \frac{n\,\hat{P}\hat{Q}}{n-1}$ en dicotómicas

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

(CR) $P_i = \frac{1}{N}$; $E[e_i] = nP_i = \frac{n}{N}$, $V[e_i] = \frac{nP_i}{N}(1-\frac{1}{N}) = \frac{n(N-1)}{N^2}$; $Cov[e_i e_j] = -nP_iP_j = -\frac{n}{N^2}$

$\hat{\theta}$ (SR) $\hat{\theta}_{HT} = \sum_{i=1}^{n} \frac{Y_i}{\pi_i} \overset{P_i}{=} N \sum_{i=1}^{n} \frac{Y_i}{n}$ $(\hat{\bar{X}} = \bar{x}$; $\hat{P} = \hat{p})$.

(CR) $\hat{\theta}_{HH} = \sum_{i=1}^{n} \frac{Y_i}{nP_i} = N \sum_{i=1}^{n} \frac{Y_i}{n}$ $(\hat{\bar{X}} = \bar{x})$

(SR) $V(\hat{\theta}_{HT}) = \left(\sum_{i=1}^{N} \frac{Y_i^2}{\pi_i^2} \pi_i(1-\pi_i) + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{Y_i}{\pi_i} \frac{Y_j}{\pi_j} (\pi_{ij} - \pi_i\pi_j) \right)$.

$V(\hat{\bar{X}}_{HT}) = (1-f) \cdot \frac{S^2}{n} \to N^2 (1-f) \frac{S^2}{n}$.

$(1-f) \cdot \frac{N}{N-1} \cdot \frac{PQ}{n} \to N^3$

(CR) $V(\hat{\theta}_{HH}) = \frac{1}{n}\left(\sum_{i=1}^{N} \frac{Y_i^2}{P_i} - \theta^2 \right) = \frac{1}{n} \sum_{i=1}^{N} \left(\frac{Y_i}{P_i} - \theta \right)^2 P_i$

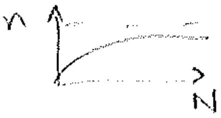
$V(\hat{\bar{X}}_{HH}) = \frac{\sigma^2}{n} \to \frac{PQ}{n}$.

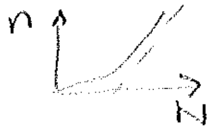
(SR) $\hat{V}(\hat{\theta}_{HT}) = \sum_{i=1}^{n} \frac{Y_i^2}{\pi_i^2}(1-\pi_i) + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{Y_i}{\pi_i} \cdot \frac{Y_j}{\pi_j} \frac{(\pi_{ij} - \pi_i\pi_j)}{\pi_{ij}}$

$\hat{V}(\hat{\bar{X}}_{HT}) = (1-f) \cdot \frac{\hat{S}^2}{n} \longrightarrow (1-f) \cdot \frac{\hat{P}\hat{Q}}{n-1}$

(CR) $\hat{V}(\hat{\theta}_{HH}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{Y_i}{P_i} - \hat{\theta}_{HH} \right)^2$

$\hat{V}(\hat{\bar{X}}_{HH}) = \frac{\hat{S}^2}{n} \to \frac{\hat{P}\hat{Q}}{n-1}$

(SR) $\theta = \bar{X} \longrightarrow n \xrightarrow[N \to \infty]{} \frac{S^2}{e^2}$  $(P = Q = 1/2)$

$\theta = X \longrightarrow n \xrightarrow[N \to \infty]{} \infty$ 

(CR) $\theta = \bar{X} \to n = \frac{\sigma^2}{e^2}$ (no depende de N)

$\theta = X \to n \xrightarrow[N \to \infty]{} \infty$

SIN repos. más preciso que CON repos.

$V(\hat{\theta}) <<$

$n <<$