MUESTREO ESTADÍSTICO

Muestreo Aleatorio Estratificado (M.A.E)

SEMANA-5

Raúl Alberto Pérez

Universidad Nacional de Colombia, Escuela de Estadística, 2021-I

Introducción

En muchas investigaciones no es raro encontrar poblaciones compuestas de subgrupos bien definidos que pueden ser identificados con autoridad.

Cuando dichos subgrupos son disjuntos, reciben el nombre de ***Estratos***.

La estratificación se presenta en algunos casos de manera directa.

En otros casos se construye con base en una o varias variables de interés sobre las cuales se posee suficiente información.

En este caso, se procura que exista una gran homogeneidad dentro de los elementos de cada estrato y una gran heterogeneidad entre dichos grupos o estratos.

TEMAS A DESARROLLAR EN ESTE CAPÍTULO:

1. CASO DE ESTUDIO

2. Definición de MAE.

3. ¿Cómo se selecciona una MAE?

4. Estimación de los parámetros de interés: μ, τ, p, A.

5. Cálculo del Tamaño de muestra- Según diferentes asignaciones.

6. Retomar el CASO DE ESTUDIO

MAE- CASO DE ESTUDIO:

Un Ingeniero está interesado en estudiar la productividad de las máquinas en una multinacional, dicha productividad esta representada por la producción mensual (en toneladas) que dichas máquinas generan del producto. La empresa está dividida en TRES plantas de producción.

Las características de cada planta, así como los resultados del estudio se muestran en la siguiente tabla. Adicionalmente, se registraron las máquinas que necesitan mantenimiento.

TABLA DE DATOS:

	Planta 1	Planta 2	Planta 3
Nro. Máq. por Planta	500	1000	4000
Nro. Máq. en la muestra	50	80	160
Nro. Máq. para mantenimiento	4	6	5
Producción Promedio (ton)	40	60	100
Desv. Est. De la producción	12	17	10

MAE: CASO DE ESTUDIO: PREGUNTAS:

1. Indicar el procedimiento que debió seguir el Ingeniero para garantizar que la muestra seleccionada fuera una MAE.

2. Calcule IC´s del 95 % para la producción promedio mensual en cada planta.

3. Estime la producción media mensual en la multinacional y construya un intervalo de confianza del 95 %.

4. Estime la producción total mensual en la multinacional y construya un intervalo de confianza del 90 %.

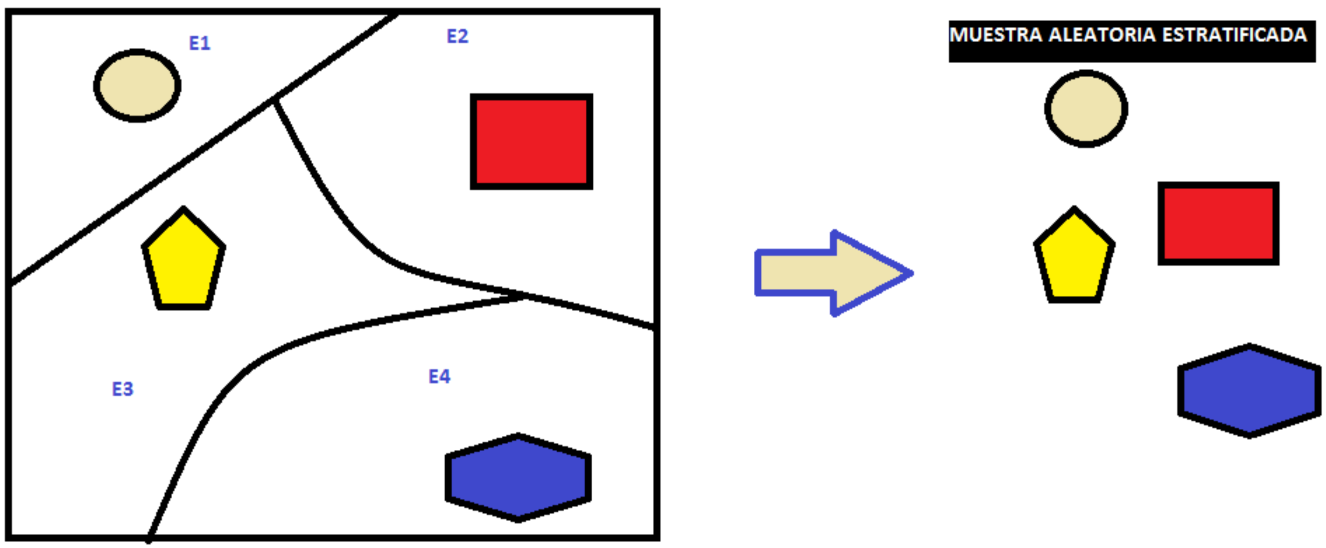
5. Estime la proporción de máquinas para mantenimiento en la empresa con un IC del 95 %.

6. Asuma como valores para σ_i^2, los correspondientes valores estimados en la tabla anterior. Determine el tamaño de muestra necesario para estimar μ con una confianza del 95 % y un límite para el error de estimación de TRES toneladas.

MUESTREO ALEATORIO ESTRATIFICADO-MAE

Definición: Un M.A.E es una técnica de muestreo mediante la cual se obtiene una muestra a partir de la separación de los elementos de una población en grupos no traslapados, llamados **estratos**, y la selección posterior de una muestra aleatoria simple, independiente, en cada estrato.

Esquemáticamente:



Observaciones en-MAE:

1. El objetivo de un diseño de encuesta es MAXIMIZAR la información obtenida (o MINIMIZAR el valor de B) para un presupuesto fijo, teniendo en cuenta la variabilidad de la población cuando ésta NO es HOMOGÉNEA.

2. El costo de obtener las observaciones varía según el diseño de la encuesta: costo de seleccionar los adultos que se van a muestrear, costo tiempo-traslado del encuestador y costo de administración de todo el proceso de muestreo; entre otros.

Los costos mencionados se pueden minimizar mediante la selección de una muestra aleatoria estratificada (M.A.E.), además se pueden obtener estimaciones de los parámetros poblacionales de interés para CADA UNO de los estratos individualmente.

Motivos para el uso de MAE:

Motivos principales para seleccionar una MAE:

1. Un MAE puede producir un B más pequeño que el generado por un MAS (con el mismo tamaño) , esto es cierto si la medición dentro de los estratos es homogénea.

2. El costo por observación se puede reducir mediante un MAE en grupos convenientes.

3. Se pueden obtener estimaciones de parámetros para subgrupos de la población.

SELECCIÓN DE UNA MAE:

Proceso de Selección MAE:

1. Especificar claramente los estratos, así cada unidad muestral se ubica en el estrato adecuado.

2. Se selecciona, en forma **independiente**, una MAS de cada uno de los estratos.

3. Se debe asegurar independencia en la selección de las MAS de cada estrato.

Notaciones:

1. Para la POBLACIÓN: H : # de estratos. N_i :# de unidades muestrales en el i-ésimo estrato. N : # de unidades muestrales en la población. $N = \sum_{i=1}^{H} N_i$.

2. Para la muestra:

   Tamaño de cada muestra n_i # de unidades muestrales seleccionadas del estrato i, $i = 1, 2, \ldots, H$.

   Medidas de resumen: $\bar{y}_i$, s_i^2: Media y varianza muestral de la MAS seleccionada del estrato i, $i = 1, 2, \ldots, H$.

Estimación en Muestreo Aleatorio Estratificado (MAE)

El procedimiento de estimación más frecuente en el M.A.E, consiste en considerar cada estrato como una subpoblación y realizar, en primera instancia, estimaciones acerca de los parámetros correspondiente a cada una de esas subpoblaciones.

Luego las estimaciones realizadas para cada subgrupo se combinan para obtener las estimaciones globales de los parámetros de interés.

Las estimaciones dentro de cada estrato se pueden llevar a cabo de acuerdo a diferentes procedimientos, lo importante es que las muestras seleccionadas en cada uno de los estratos sean independientes para poder obtener expresiones directa de estimación para los parámetros poblacionales.

Estimación de la Media Poblacional en MAES

Un estimador insesgado de la media poblacional μ está dado por:

$$\widehat{\mu} = \overline{y}_{est} = \sum_{h=1}^{H} W_h \overline{y}_h = \sum_{h=1}^{H} \left(\frac{N_h}{N}\right) \overline{y}_h = \frac{1}{N} \sum_{h=1}^{H} N_h \overline{y}_h$$

donde, $\overline{y}_h$-es un estimador insesgado de la media verdadera en el estrato h y W_h-es la proporción de la población que pertenecen a dicho estrato.

Si las muestras de cada estrato son independientes, la varianza del estimador de la media poblacional está dada por:

$$Var[\overline{y}_{est}] = \frac{1}{N^2} \sum_{h=1}^{H} N_h^2 Var[\overline{y}_h] \ , \quad \text{con:}$$

$$Var[\overline{y}_h] = \left(1 - \frac{n_h}{N_h}\right) \frac{\sigma_h^2}{n_h}$$

I. C PARA LA MEDIA POBLACIONAL:

B dependiendo del tamaño de la MAE

El **L. E. E** para la estimación de la **MEDIA poblacional** será:

	$n \leq 30$	$n > 30$
B	$t_{\alpha/2,n-H}\sqrt{\widehat{Var}(\overline{y}_{est})}$	$Z_{\alpha/2}\sqrt{\widehat{Var}(\overline{y}_{est})}$

I. C del $(1-\alpha)\times 100\,\%$ para μ:

$$\overline{y}_{est} \;\pm\; B$$

Estimación del Total Poblacional en MAES

Un estimador insesgado del total poblacional τ está dado por:

$$\widehat{\tau}_{est} = N\overline{y}_{est} = N\left(\frac{1}{N}\sum_{h=1}^{H} N_h\overline{y}_h\right) = \sum_{h=1}^{H} N_h\overline{y}_h = \sum_{h=1}^{H}\widehat{\tau}_h$$

La varianza de $\widehat{\tau}_{est}$ esta dada por:

$$Var[\widehat{\tau}_{est}] = N^2 Var[\overline{y}_{est}]$$

$$= N^2\left(\frac{1}{N^2}\sum_{h=1}^{H} N_h^2 Var[\overline{y}_h]\right)$$

$$Var[\widehat{\tau}_{est}] = \sum_{h=1}^{H} N_h^2 Var[\overline{y}_h]$$

I. C PARA EL TOTAL POBLACIONAL:

B dependiendo del tamaño de la MAE:

El **L. E. E** para la estimación del **TOTAL poblacional** será:

	$n \leq 30$	$n > 30$
B	$t_{\alpha/2,n-H}\sqrt{\widehat{Var}(\hat{\tau}_{est})}$	$Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\tau}_{est})}$

I. C del $(1-\alpha) \times 100\,\%$ para τ:

$$\hat{\tau}_{est} \ \pm \ B$$

EJEMPLO-1: En la siguiente tabla se presenta la información correspondiente al gasto mensual en servicios públicos (en salarios mínimos) de una muestra aleatoria estratificada de 120 familias en una ciudad geográficamente dividida en tres estratos: Norte, Centro y Sur.

	Estratos		
	Norte	Centro	Sur
N_h	4000	6000	10000
n_h	36	40	44
$\overline{y}_h$	2.4	1.2	0.6
$\widehat{\tau}_h$	9600	7200	6000
S_h^2	1.21	0.36	0.04
$\widehat{Var}[\overline{y}_h]$	0.0333	0.00894	0.0009051

Con base en estos datos, se desea estimar el gasto promedio en servicios públicos de toda la población,el gasto total en servicios públicos y los correspondientes intervalos de confianza. Suponga que el tamaño N de la población es de 20000 familias.

Para la media se tiene que:

$$\widehat{\mu}_{est} = \overline{y}_{est} = \sum_{h=1}^{H} W_h \overline{y}_h = \frac{1}{N} \sum_{h=1}^{H} N_h \overline{y}_h$$

$$= \frac{1}{20000}\Big[4000(2{,}4) + 6000(1{,}2) + 10000(0{,}6)\Big] = 1.14 \text{ Salarios mínimos por familia.}$$

Ahora,

$$\widehat{Var}[\overline{y}_{est}] = \frac{1}{N^2} \sum_{h=1}^{H} N_h^2 \widehat{Var}[\overline{y}_h]$$

$$= \frac{1}{(20000)^2}\Big[(4000)^2(0{,}0333) + (6000)^2(0{,}00894) + (10000)^2(0{,}0009051)\Big] = 0.002363$$

y el error estándar de estimación es:

$$e.e[\overline{y}_{est}] = \sqrt{\widehat{Var}[\overline{y}_{est}]} = \sqrt{0{,}002363} = 0.0486$$

$$B = 2e.e[\overline{y}_{est}] = 2(0.0486) = 0.0972 \text{ , salarios mínimos}$$

Un I.C de aproximadamente el 95 % para μ es:

$$\overline{y}_{est} \pm B \Longleftrightarrow 1.14 \pm 0.0972 \Longleftrightarrow (\ 1.0428\ ,\ 1.2392\ )$$

Ahora para el total se tiene que:

$$\widehat{\tau}_{est} = N\overline{y}_{est} = 20000(1.14) = 22800 \text{ Salarios mínimos totales}$$

$$e.e[\widehat{\tau}_{est}] = Ne.e[\overline{y}_{est}] = 20000(0.0486) = 972 \text{ Salarios mínimos}$$

$$B = 2e.e[\widehat{\tau}_{est}] = 2(972) = 2Ne.e[\overline{y}_{est}] = 1944, \quad \text{salarios mínimos}$$

Un I.C de aproximadamente el 95 % para τ es:

$$\widehat{\tau}_{est} \pm B \Longleftrightarrow 22800 \pm 972 \Longleftrightarrow (\, 21828 \,,\, 23772 \,)$$

EJEMPLO-2: Una empresa publicitaria está interesada en determinar cuánto debe enfatizar la publicidad televisiva en una determinada región, y decide realizar una encuesta por muestreo para estimar el número promedio de horas por semana que se ve la televisión en los hogares de la región. La región comprende dos pueblos, pueblo A y pueblo B, y un área rural. El pueblo A está ubicado en torno a una fábrica, y la mayoría de los hogares son de trabajadores industriales con niños de edad escolar. El pueblo B es un suburbio exclusivo de una ciudad vecina y consta de habitantes más viejos con pocos niños en casa. Existen 155 hogares en el pueblo A, 62 hogares en el pueblo B y 93 hogares en el área rural.

Suponga que se llevó a cabo una encuesta planificada. La empresa publicitaria tenia tiempo y dinero suficientes para entrevistar a $n = 40$-hogares de la región y decide m.a.s de tamaños $n_1 = 20$-hogares del pueblo-A, $n_2 = 8$-hogares del pueblo-B y $n_3 = 12$-hogares del área rural.

Se seleccionan las m.a.s en cada pueblo y en el área rural y se realizan las encuestas. Los resultados con las mediciones del tiempo que se ve la televisión en horas por semana, se dan en la siguiente tabla:

Pueblo-A: 35, 43, 36, 39, 28, 28, 29, 25, 38, 27, 26, 32, 29, 40, 35, 41, 37, 31, 45, 34.

Pueblo-B: 27, 15, 4, 41, 49, 25, 10, 30.

Área Rural: 8, 14, 12, 15, 30, 32, 21, 20, 37, 7, 11, 24.

Estimar el tiempo promedio que se ve TV, en horas por semana, para:
a) Todos los hogares de la región, b) Todos los hogares del pueblo-B

En ambos casos establezca un límite para el error de estimación.

Solución: En la siguiente tabla se presenta un resumen de los datos muestrales:

	N_h	n_h	$\overline{y}_h$	s_h
Pueblo-A	155	20	33.90	5.95
Pueblo-B	62	8	25.12	15.25
Áreas Rural	93	12	19.00	9.36

(a) Para toda la Región:

$$\widehat{\mu}_{est} = \overline{y}_{est} = \frac{1}{N}\sum_{h=1}^{H} N_h \overline{y}_h$$

$$= \frac{1}{310}\Big[155(33{,}90) + 62(25{,}12) + 93(19{,}00)\Big]$$

$\overline{y}_{est} = 27.7$ horas por semana.

Ahora,

$$\widehat{Var}[\overline{y}_{est}] = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2\left(\frac{N_h - n_h}{N_h}\right)\frac{S_h^2}{n_h}$$

$$= \frac{1}{(310)^2}\left[\frac{(155)^2(0{,}871)(5{,}95)^2}{20} + \frac{(62)^2(0{,}871)(15{,}25)^2}{8} + \frac{(93)^2(0{,}871)(9{,}36)^2}{12}\right]$$

$\widehat{Var}[\overline{y}_{est}] = 1.97.$

y el error estándar de estimación es:

$$e.e[\overline{y}_{est}] = \sqrt{\widehat{Var}[\overline{y}_{est}]} = \sqrt{1{,}97} = 1.4035.$$

El límite para el error de estimación está dado por:

$$B = 2e.e[\overline{y}_{est}] = 2(1.4035) = 2.807 \text{ horas.}$$

Un I.C de aproximadamente el $95\,\%$ para μ es:

$$\overline{y}_{est} \pm B \Longleftrightarrow 27.7 \pm 2.807 \Longleftrightarrow (\ 24.893\ ,\ 30.507\ )$$

(b) Para el pueblo-B $\hat{\mu}_2 = \overline{y}_2 = 25.1$ Ahora,

$$\widehat{Var}[\overline{y}_2] = \left(\frac{N_2 - n_2}{N_2}\right)\left(\frac{S_2^2}{n_2}\right)$$

$$= \left(\frac{62-8}{62}\right)\left(\frac{(15.25)^2}{8}\right)$$

$$\widehat{Var}[\overline{y}_2] = 25.32$$

y el error estándar de estimación es:

$$e.e[\overline{y}_2] = \sqrt{\widehat{Var}[\overline{y}_2]} = \sqrt{25{,}3193} = 5.0318$$

El límite para el error de estimación está dado por:

$$B = 2e.e[\overline{y}_{est}] = 2(5.0318) = 10.0636 \text{ horas}$$

Un I.C de aproximadamente el 95 % para μ_2 es:

$$\overline{y}_2 \pm B \iff 25.1 \pm 10.0636 \iff (\ 15.0364\ ,\ 35.1636\ )$$

Estimación de la Proporción y del Total de elementos de la Población con un Atributo Específico en M.A.E.S

Sea A_h-el número de elementos en el estrato h con el atributo deseado.

a_h-el número de elementos en la muestra del estrato h con el atributo deseado.

$P_h = \frac{A_h}{N_h}$-es la proporción de elementos de la población en el estrato h con el atributo deseado.

$p_h = \frac{a_h}{n_h}$-es la proporción de elementos en la muestra del estrato h con el atributo deseado.

Los estimadores insesgados para P_h y A_h son respectivamente:

$$\widehat{P}_h = p_h \quad \text{y} \quad \widehat{A}_h = N_h p_h$$

Estimadores de P y A Globales

Los estimadores de la proporción poblacional P y del número total de elementos en la población con el atributo deseado A, son:

$$\widehat{P}_{est} = p_{est} = \sum_{h=1}^{H} W_h p_h = \sum_{h=1}^{H} \left(\frac{N_h}{N}\right) p_h = \frac{1}{N} \sum_{h=1}^{H} N_h p_h = \frac{1}{N} \sum_{h=1}^{H} \widehat{A}_h$$

$$\widehat{A}_{est} = N p_{est}$$

La varianza de estos estimadores son:

$$Var[p_{est}] = \frac{1}{N^2} \sum_{h=1}^{H} N_h^2 Var[p_h] \; , \qquad Var[\widehat{A}_{est}] = N^2 Var[p_{est}]$$

EJEMPLO: Continuando con el ejemplo de la ciudad mencionada anteriormente, ahora se tiene como objetivo estimar el porcentaje y el número total de familias que poseen al menos un automóvil.

La información resumen de los datos aparece en la siguiente tabla:

	Estratos		
	Norte	Centro	Sur
N_h	4000	6000	10000
n_h	36	40	44
p_h	0.9	0.6	0.2
$\widehat{A}_h$	3600	3600	2000
q_h	0.1	0.4	0.8
$\widehat{Var}[p_h]$	0.002548	0.006113	0.003704

A partir de la información anterior, se tienen las siguientes estimaciones.

Para la proporción de familias con más de un automóvil:

$$p_{est} = \frac{1}{N}\sum_{h=1}^{H} N_h p_h = \frac{1}{20000}\Big[4000(0{,}9) + 6000(0{,}6) + 10000(0{,}2)\Big] = 0.46(46\,\%),$$

$$\widehat{Var}[p_{est}] = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \widehat{Var}[p_h] = \frac{1}{(20000)^2}\Big[(4000)^2(0{,}002548) + (6000)^2(0{,}006113)$$
$$+ (10000)^2(0{,}003704)\Big]$$

$$\widehat{Var}[p_{est}] = 0.001578$$

$$e.e[p_{est}] = \sqrt{\widehat{Var}[p_{est}]} = \sqrt{0.001465} = 0.0397$$

$$B = 2 \times e.e[p_{est}] = 2(0.0397) = 0.0794$$

Un I.C de aproximadamente el $95\,\%$ para p es:

$$p_{est} \ \pm \ B \Longleftrightarrow \ 0.46 \ \pm \ 0.0794 \ \Longleftrightarrow \ (\ 0.3806\ ,\ 0.5394\ )$$

Ahora, para el total de familias con más de un automóvil:

$$\widehat{A}_{est} = Np_{est} = 20000(0{,}46) = 9200 \quad \text{familias,}$$

$$\widehat{Var}[\widehat{A}_{est}] = N^2\widehat{Var}[p_{est}] = (20000)^2(0.001578) = 631200$$

$$e.e[\widehat{A}_{est}] = \sqrt{\widehat{Var}[\widehat{A}_{est}]} = \sqrt{631200} = 795 \quad \text{familias}$$

$$B = 2 \times e.e[\widehat{A}_{est}] = 2 \times (795) = 1589$$

Un I.C de aproximadamente el 95 % para A es:

$$A_{est} \;\pm\; B \Longleftrightarrow \;\; 9200 \;\pm\; 1589 \;\Longleftrightarrow\; (\,7611\;,\;10789\,)$$

Tamaño de Muestra n

En el M.A.S el único procedimiento para mejorar la eficiencia del estimador de la media (y de otros parámetros relacionados con la media) consiste en aumentar el tamaño de muestra n, ya que los demás términos en la expresión correspondiente están dados.

En el M.A.E, es diferente ya que la varianza del estimador de la media es un función no sólo de los tamaños (N_h) y varianzas verdaderas por estrato (σ_h^2), sino también de los tamaños muestrales asignados a cada uno de dichos estratos, ie. de los n_h

Nuevamente para MAE se hace:

$$D = Var[\overline{y}_{est}] = \frac{B^2}{Z^2}, \quad \text{donde,} \quad B = Z\sqrt{Var[\overline{y}_{est}]} = ZE.E(\overline{y}_{est})$$

pero, en este caso no es posible despejar a n-directamente de ninguna de las siguientes expresiones:

$$Var[\overline{y}_{est}] = D = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \left(\frac{N_h - n_h}{N_h}\right)\frac{\sigma_h^2}{n_h}$$

a no ser que se conozca algo de la relación entre las n_h y n

Existen varios procedimientos para asignar el tamaño de la muestra n entre los diferentes estratos.

En cada caso el número de observaciones n_h asignado al h-ésimo estrato es una fracción del tamaño total de la muestra n.

La fracción de afijación será denotada por w_h, es decir:

$$n_h = nw_h \quad \text{o} \quad w_h = \frac{n_h}{n}, \quad \text{para} \;\; h = 1, 2, \ldots, H.$$

De esta manera, reemplazando las n_h por nw_h en la expreción anterior:

$$D = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \left(\frac{N_h - n_h}{N_h}\right)\frac{\sigma_h^2}{n_h} = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \left(\frac{N_h - nw_h}{N_h}\right)\frac{\sigma_h^2}{nw_h}$$

$$= \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \left(1 - \frac{nw_h}{N_h}\right)\frac{\sigma_h^2}{nw_h}$$

$$D = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \frac{\sigma_h^2}{nw_h} - \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \frac{\sigma_h^2}{N_h},$$

es decir que,

$$N^2D + \sum_{h=1}^{H} N_h^2 \frac{\sigma_h^2}{N_h} = \frac{1}{n} \sum_{h=1}^{H} N_h^2 \frac{\sigma_h^2}{w_h}$$

de donde se obtiene lo siguiente:

El tamaño de muestra aproximado para estimar a μ con un límite para el error de estimación de B es:

$$n = \frac{\sum_{h=1}^{H} \left(N_h^2 \sigma_h^2 / w_h\right)}{N^2D + \sum_{h=1}^{H} N_h \sigma_h^2}$$

donde, las w_h-son la fracción de observaciones afijadas al estrato h, σ_h^2-es la varianza poblacional del estrato h.

EJEMPLO: Una empresa publicitaria está interesada en determinar cuánto debe enfatizar la publicidad televisiva en una determinada región, y decide realizar una encuesta por muestreo para estimar el número promedio de horas por semana que se ve la televisión en los hogares de la región. La región comprende dos pueblos, pueblo A y pueblo B, y un área rural. El pueblo A está ubicado en torno a una fábrica, y la mayoría de los hogares son de trabajadores industriales con niños de edad escolar. El pueblo B es un suburbio exclusivo de una ciudad vecina y consta de habitantes más viejos con pocos niños en casa. Existen 155 hogares en el pueblo A, 62 hogares en el pueblo B y 93 hogares en el área rural.

Una encuesta anterior a los hogares de dicha región, sugiere que las varianzas de cada uno de los estratos (Pueblo-A, Pueblo-B y Área Rural) son aproximadamente: $\sigma_1^2 \approx 25, \sigma_2^2 \approx 225$ y $\sigma_3^2 \approx 100$.

Si se desea estimar la media poblacional mediante $\overline{y}_{est}$. Cuál será el tamaño de muestra adecuado para obtener un límite para el error de estimación de 2-horas, si las fracciones fijadas en cada estrato son: $a_1 = 1/3 = a_2 = a_3 = 1/H$, es decir, se tomará un número igual de observaciones por estrato.

Solución: En este caso se tiene que:

$$B = 2\sqrt{Var[\overline{y}_{est}]} = 2\ , \quad \text{o bien} \quad Var[\overline{y}_{est}] = 1 = D$$

Para hallar a n se usará la expresión:

$$n = \frac{\sum_{h=1}^{3}\left(N_h^2\sigma_h^2/w_h\right)}{N^2D + \sum_{h=1}^{3} N_h\sigma_h^2},$$

con $N_1 = 155, N_2 = 62, N_3 = 93,\ \sigma_1^2 = 25, \sigma_2^2 = 225, \sigma_3^2 = 100,$ $w_1 = 1/3 = w_2 = w_3,\ D = \frac{B^2}{4} = \frac{2^2}{4} = 1$ y $N = 310,$

luego, se procede como sigue:

$$\begin{aligned}
\sum_{h=1}^{3}\left(N_h^2\sigma_h^2/w_h\right) &= \frac{N_1^2\sigma_1^2}{w_1} + \frac{N_2^2\sigma_2^2}{w_2} + \frac{N_3^2\sigma_3^2}{w_3} \\
&= \frac{(155)^2(25)}{1/3} + \frac{(62)^2(225)}{1/3} + \frac{(93)^2(100)}{1/3} \\
&= 6991275,
\end{aligned}$$

$$\sum_{h=1}^{3} N_h\sigma_h^2 = N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2$$
$$= 155(25) + 62(225) + 93(100)$$
$$= 27125, \quad \text{y}$$

$$N^2D = (310)^2(1) = 96100, \ \text{de donde}$$

$$n = \frac{\sum_{h=1}^{3}\left(N_h^2\sigma_h^2/w_h\right)}{N^2D + \sum_{h=1}^{3} N_h\sigma_h^2} = \frac{6991275}{96100 + 27125} = 56{,}7 \approx 57$$

Es decir, que el experimentador debería tomar $n = 57$-observaciones.

La asignación de este tamaño muestral a los diferentes estratos es:

$$n_1 = nw_1 = 57(1/3) = 19, \quad n_2 = nw_2 = 57(1/3) = 19$$

$$\text{y} \quad n_3 = nw_3 = 57(1/3) = 19$$

Ahora, si lo que se desea estimar es el total τ-de horas semanales que dedican a ver televisión todos los hogares de esa región, con un límite para el error de estimación de 400-horas, entonces se tendría lo siguiente:

$$D = \frac{B^2}{4N^2} = \frac{(400)^2}{4(310)^2} = \frac{40000}{(310)^2} = 0{,}4162$$

luego,

$$n = \frac{\sum_{h=1}^{3}\left(N_h^2\sigma_h^2/w_h\right)}{N^2D + \sum_{h=1}^{3} N_h\sigma_h^2} = \frac{6991275}{(310)^2(0{,}4162) + 27125} = \frac{6991275}{40000 + 27125} = 104{,}2$$

Es decir, que se necesitarían $n = 105$-observaciones y con una asignación por estratos dada por:

$$n_1 = n_2 = n_3 = \frac{n}{3} = \frac{105}{3} = 35$$

Tarea 2 fecha Entrega: HASTA el 16 de Abril al inicio de Clase

1. Considere el Caso de Estudio de Este Capítulo, diapositiva 4. Responda las preguntas de la diapositiva 6 sobre este caso de estudio.

2. Del Taller 2 de MAE, resuelva el problema asignado. Para ello, en la siguiente tabla aparece, según el número asignado en la lista de estudiantes adjunta (**NLista) el problema a resolver asignado (Pr. As.) ALEATORIAMENTE.**

NLista	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Pr. As.	12	18	15	8	4	6	5	11	20	2	13	19	17	21	22

NLista	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Pr. As.	28	27	25	26	24	23	16	3	5	1	20	2	13	19	14

NLista	31	32	33	34	35	36	37	38	39	40
Pr. As.	9	10	12	3	10	15	20	7	19	6

	APELLIDOS Y NOMBRE		APELLIDOS Y NOMBRE
1	Agudelo Carmona, Valentina	21	Murillo Anzola, Christian Camilo
2	Alvarez Morales, Guillermo	22	Naranjo Garcia, Yised Katerine
3	Aristizabal Echeverri, Genaro Alfonso	23	Nieto Morales, Alexis Andrés
4	Arzuaga Gonzalez, Maria Isabel	24	Pabón Palacio, Antonio
5	Bula Charry, Valentina	25	Palacios Duque, Sara
6	Bula Isaza, Juan Daniel	26	Pico Arredondo, Daniela
7	Carvajal Torres, Santiago	27	Rios Castro, Kaline Andrea
8	Duque Calle, David	28	Rios Garcia, Jhon Alexander
9	Franco Valencia, Santiago	29	Rojas Bolaños, Carlos Arturo
10	Galeano Arenas, Juan José	30	Salazar Mejía, Alejandro
11	Galeano Muñoz, Simón Pedro	31	Serrano Santos, Andrea
12	Garcia Muñoz, Jhonatan Smith	32	Suarez Ledesma, Jose Luis
13	Gaviria Sanchez, Sebastian	33	Tous Diaz, Cleidy Jimena
14	Gómez Valencia, Beatriz Valentina	34	Vanegas Castaño, Valentina
15	Granada Alvarez, Santiago	35	Vasco Ruiz, Daniela
16	Henao Vargas, Luisa Fernanda	36	Vásquez Gómez, Kleider Stiven
17	Hoyos Arias, John Daniel	37	Velez Rivera, Vanessa
18	Hoyos Peña, Laura Daniela	38	Yepes Pareja, Maria Fernanda
19	Jaramillo Calle, David	39	Zabaleta Cardeño, Carmen Daniela
20	Martínez Echavarría, Juan Pablo	40	Zuluaga Ayala, Santiago