Universidad Nacional de Colombia-Sede Medellín

Pacial-2 Muestreo Estadístico. **Fecha: 2021-07-30**

Nombre Completo: ______________________________

Firma: ____________________ **C.C** ____________________

1. **fun**

2. **fun**

1. **Ejercicio**
   Una empresa publicitaria está interesada en determinar que tanto debe enfatizar la publicidad televisiva en una parte del Bajo Cauca Antioqueño. La empresa decide realizar una encuesta por muestreo con el objetivo de estimar el número promedio de horas por semana, μ, que se ve televisión en los hogares de dicha región y el Número Total, A, de hogares de dicha región que tienen mas de 2 niños por hogar.

   Debido a la falta de presupuesto y tiempo, la empresa decide dividir la parte considerada en la región de estudio en tres Zonas para llevar a cabo un Muestreo Aleatorio Estratificado por Zonas. Las zonas son denominadas: **A**, **B** y **C**. Existen 98 hogares en la Zona-A, 65 en la Zona-B y 152 en la Zona-C, es decir existen un total de: $N = N_1 + N_2 + N_3$=98+65+152 =315-hogares en la Región de Estudio.

   Suponga que se selecciona una MAS en cada una de las Zonas de tamaños: n_1=21, n_2=16 y n_3=13 para cada una de las Zonas **A**, **B** y **C** respectivamente, para un total de $n = n_1 + n_2 + n_3$=21+16+13=50-hogares en la muestra.

   En la siguiente tabla se tienen el resúmen descriptivo de cada una de las muestras seleccionada de cada una de las Zonas de estudio.

Zona	N_h	n_h	$\bar{y}_h$	a_h	$p_h = a_h/n_h$	S_h
A	N_1 =98	n_1 =21	$\bar{y}_1$=24	a_1=17	p_1=0.8095238	S_1=9
B	N_2 =65	n_2 =16	$\bar{y}_2$=32	a_2=9	p_1=0.5625	S_2=7
C	N_3 =152	n_3 =13	$\bar{y}_3$=27	a_3=8	p_1=0.6153846	S_3=12

   (a) Estime el número promedio de horas por semana que cada hogar de la región de estudio dedica a ver TV. Halle el respectivo LLE=B con un nivel de confianza de **aproximadamente** 95%. Interprete Resultados.

   (b) Estime el Porcentaje o Proporción de Hogares con mas de 2 Niños por hogar. Halle el respectivo LLE=B con un nivel de confianza de **aproximadamente** 95%. Interprete Resultados.

   (c) Suponga que la muestra anterior es una muestra aletoria piloto. Hallar el tamaño de muestra necesario y la asignación correspondiente de dicha muestra, para estimar el número promedio de horas por semana que cada hogar de la región de estudio dedica a ver TV con un LEE=B=3-horas y un nivel de confianza de **aproximadamente el** 95%.

Solución
(a) Para hallar la estimación de μ-se utiliza la expresión dada por:

$$\hat{\mu} = \bar{y}_{ets} = \frac{1}{N}\sum_{h=1}^{k} N_k\bar{y}_k, \quad \text{es decir que:}$$

$$\hat{\mu} = \frac{1}{N}\sum_{h=1}^{3} N_k\bar{y}_k = \frac{1}{N}\Big(N_1\bar{y}_1 + N_2\bar{y}_2 + N_3\bar{y}_3\Big)$$

$$= \frac{1}{315}\Big[(98)(24) + (65)(32) + (152)(27)\Big] = \frac{1}{315}\Big[2352 + 2080 + 4104\Big]$$

$$= \frac{1}{315}(8536)$$

$$\bar{y}_{ets} = 27.09841$$

Ahora para hallar el $LEE = B$ con nivel de confianza de **aproximadamente el** 95%, se procede como sigue:

$$B = 2\sqrt{\widehat{Var}[\bar{y}_{est}]},$$

$$\text{con:} \quad \widehat{Var}[\bar{y}_{est}] = \widehat{Var}\left[\frac{1}{N}\sum_{h=1}^{3} N_k\bar{y}_k\right]$$

$$= \frac{1}{N^2}\left[\sum_{h=1}^{3} N_k^2\ \widehat{Var}[\bar{y}_k]\right]$$

$$= \frac{1}{N^2}\Big(N_1^2\ \widehat{Var}[\bar{y}_1] + N_2^2\ \widehat{Var}[\bar{y}_2] + N_3^2\ \widehat{Var}[\bar{y}_3]\Big)$$

$$= \frac{1}{N^2}\left[N_1^2\left(\frac{N_1-n_1}{N_1}\right)\frac{S_1^2}{n_1} + N_2^2\left(\frac{N_2-n_2}{N_2}\right)\frac{S_2^2}{n_2} + N_3^2\left(\frac{N_3-n_3}{N_3}\right)\frac{S_3^2}{n_3}\right]$$

$$= \frac{1}{(315)^2}\left[(98)^2\left(\frac{98-21}{98}\right)\frac{(9)^2}{21} + (65)^2\left(\frac{65-16}{65}\right)\frac{(7)^2}{16} + (152)^2\left(\frac{152-13}{152}\right)\frac{(12)^2}{13}\right]$$

$$= \frac{1}{(315)^2}\Big[29106 + 9754.062 + 234033.2\Big]$$

$$= \frac{1}{(315)^2}\Big[272893.3\Big]$$

$$\widehat{Var}[\bar{y}_{est}] = 2.750247$$

Luego,

$$B = 2\sqrt{\widehat{Var}[\bar{y}_{est}]} = 2\sqrt{2.750247} = 3.316774$$

de donde, un IC estimado de **aproximadamente el** 95% para μ es:

$$\hat{\mu} \;\pm\; B \iff \bar{y}_{est} \;\pm\; B \iff 27.09841 \;\pm\; 3.316774 \iff \Big(23.7816388 \;,\; 30.4151866 \Big)$$

(b) Para hallar la estimación de P-se utiliza la expresión dada por:

$$\hat{P} = \hat{p}_{ets} = \frac{1}{N}\sum_{h=1}^{k} N_k \hat{p}_k, \quad \text{es decir que:}$$

$$\begin{aligned}
\hat{p}_{est} = \frac{1}{N}\sum_{h=1}^{3} N_k \hat{p}_k &= \frac{1}{N}\Big[N_1\hat{p}_1 + N_2\hat{p}_2 + N_3\hat{p}_3\Big] \\
&= \frac{1}{N}\left[N_1\left(\frac{a_1}{n_1}\right) + N_2\left(\frac{a_2}{n_2}\right) + N_3\left(\frac{a_3}{n_3}\right)\right] \\
&= \frac{1}{315}\left[(98)\left(\frac{17}{21}\right) + (65)\left(\frac{9}{16}\right) + (152)\left(\frac{8}{13}\right)\right] \\
&= \frac{1}{315}\Big[(98)\,(0.8095238) + (65)\,(0.5625) + (152)\,(0.6153846)\Big] \\
&= \frac{1}{315}\Big[79.3333333 + 36.5625 + 93.5384615\Big] \\
&= \frac{1}{315}(209.4342949) \\
\hat{p}_{ets} &= 0.6648708
\end{aligned}$$

Ahora para hallar el $LEE = B$ con nivel de confianza de **aproximadamente el** 95%, se procede como sigue:

$$B = 2\sqrt{\widehat{Var}[\hat{p}_{est}]},$$

con: $$\widehat{Var}[\hat{p}_{est}] = \widehat{Var}\left[\frac{1}{N}\sum_{h=1}^{3} N_k \hat{p}_k\right] = \frac{1}{N^2}\left[\sum_{h=1}^{3} N_k^2\ \widehat{Var}[\hat{p}_k]\right]$$

$$= \frac{1}{N^2}\left[N_1^2\ \widehat{Var}[\hat{p}_1] + N_2^2\ \widehat{Var}[\hat{p}_2] + N_3^2\ \widehat{Var}[\hat{p}_3]\right]$$

$$= \frac{1}{N^2}\left[N_1^2\left(\frac{N_1-n_1}{N_1}\right)\frac{p_1q_1}{n_1-1} + N_2^2\left(\frac{N_2-n_2}{N_2}\right)\frac{p_2q_2}{n_2-1} + N_3^2\left(\frac{N_3-n_3}{N_3}\right)\frac{p_3q_3}{n_3-1}\right]$$

$$= \frac{1}{(315)^2}\left[(98)^2\left(\frac{98-21}{98}\right)\frac{\left(\frac{17}{21}\right)\left(1-\frac{17}{21}\right)}{21-1} + (65)^2\left(\frac{65-16}{65}\right)\frac{\left(\frac{9}{16}\right)\left(1-\frac{9}{16}\right)}{16-1} + (152)^2\left(\frac{152-13}{152}\right)\frac{\left(\frac{8}{13}\right)\left(1-\frac{8}{13}\right)}{13-1}\right]$$

$$= \frac{1}{(315)^2}\Big[58.17778 + 52.25391 + 416.7258\Big]$$

$$= \frac{1}{(315)^2}\Big[527.1575\Big]$$

$$\widehat{Var}[\hat{p}_{est}] = 0.005312749$$

Luego,

$$B = 2\sqrt{\widehat{Var}[\hat{p}_{est}]} = 2\sqrt{0.005312749} = 0.1457772$$

de donde, un IC estimado de **aproximadamente el** 95% para P es:

$$\hat{P} \ \pm\ B \iff \hat{p}_{est}\ \pm\ B \iff 0.6648708\ \pm\ 0.1457772 \iff \Big(0.5190936\ ,\ 0.810648\Big)$$

(c) Para hallar el tamaño de muestra n, se utiliza la siguiente expresión:

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

$$\text{con:}\quad D = \frac{B^2}{Z^2} = \frac{B^2}{4} = \frac{(3)^2}{4} = 2.25,$$

$Z^2 = 4$, debido a que el Nivel de confianza es de **aproximadamente el** 95%.

Ahora hallamos las fracciones w_h para cada estrato utilizando la asignación correspondiente a **costos iguales, debido a que éstos costos son desconocidos**, es decir se utiliza:

$$w_h = \frac{N_h\sigma_h}{\sum_{h=1}^{H} N_h\sigma_h}$$

De la muestra piloto se utilizan las varianzas muestrales como estimaciones de las varianzas poblacionales, es decir:

$$\hat{\sigma}_1^2 = S_1^2\ ,\quad \hat{\sigma}_2^2 = S_2^2 \quad \text{y} \quad \hat{\sigma}_3^2 = S_3^2.$$

Luego,

$$w_1 = \frac{N_1 S_1}{\sum_{h=1}^{3} N_h \hat{\sigma}_h} = \frac{N_1 S_1}{N_1 S_1 + N_2 S_2 + N_3 S_3} = \frac{(98)(9)}{(98)(9) + (65)(7) + (152)(12)} = \frac{882}{3161} = 0.2790256$$

$$w_2 = \frac{N_2 S_2}{\sum_{h=1}^{3} N_h \hat{\sigma}_h} = \frac{N_2 S_2}{N_1 S_1 + N_2 S_2 + N_3 S_3} = \frac{(65)(7)}{(98)(9) + (65)(7) + (152)(12)} = \frac{455}{3161} = 0.1439418$$

$$w_3 = \frac{N_3 S_3}{\sum_{h=1}^{3} N_h \hat{\sigma}_h} = \frac{N_3 S_3}{N_1 S_1 + N_2 S_2 + N_3 S_3} = \frac{(152)(12)}{(98)(9) + (65)(7) + (152)(12)} = \frac{1824}{3161} = 0.5770326$$

Ahora, reemplazmos estas cantidades en:

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

pero, primero hallemos las sumas involucradas en la expresión anterior:

$$(I) \qquad \sum_{h=1}^{3} N_h^2 \sigma_h^2 / w_h = \frac{N_1^2 S_1^2}{w_1} + \frac{N_2^2 S_2^2}{w_2} + \frac{N_3^2 S_3^2}{w_3}$$

$$= \frac{(98)^2 (9)^2}{0.2790256} + \frac{(65)^2 (7)^2}{0.1439418} + \frac{(152)^2 (12)^2}{0.5770326}$$

$$= 2788002 + 1438255 + 5765664$$

$$\sum_{h=1}^{3} N_h^2 \sigma_h^2 / w_h = 9991921 \qquad (I)$$

$$(II) \qquad \sum_{h=1}^{H} N_h \sigma_h^2 = N_1 S_1^2 + N_2 S_2^2 + N_3 S_3^2$$

$$= 98(9)^2 + 65(7)^2 + 152(12)^2$$

$$= 7938 + 3185 + 21888$$

$$\sum_{h=1}^{H} N_h \sigma_h^2 = 33011 \qquad (II)$$

y Ahora, hallamos n:

$$n = \frac{\sum_{h=1}^{H} \left(N_h^2 \sigma_h^2 / w_h \right)}{N^2 D + \sum_{h=1}^{H} N_h \sigma_h^2} = \frac{9991921}{(315)^2 (2.25) + 33011} = 38.99024 \approx 39$$

por último, hallemos las asignaciones $n_1 = n\ w_1$, $n_2 = n\ w_2$ y $n_3 = n\ w_3$, es decir:

$$n_1 = n\ w_1 = 39(0.2790256) = 10.8792754 \approx 11$$

$$n_2 = n\ w_2 = 39(0.1439418) = 5.6123246 \approx 6$$

$$n_3 = n\ w_3 = 39(0.5770326) = 22.4986377 \approx 23$$

2. **Ejercicio**
   Se desea estimar el valor promedio pagado por un almuerzo en una de las cafeterías de la Universidad Nacional de Colombia, Sede Medellín. Para ello se realizó una encuesta a un total de N=820 estudiantes que frecuentan la cafetería objeto de estudio.

   Se selecciona una muestra aleatoria mediante Muestreo Sistemático-Estándar de tamaño n=40 estudiantes, es decir se muestrea 1 de cada $k = \frac{N}{n} = \frac{820}{40}$=20.5=20 estudiantes, y la muestra sistemática obtenida arrojó una media muestral y una desviación estándar muestral dadas por: **$\bar{y}$=4800 pesos, S = 1045 pesos, respectivamente**.

   Estimar el valor promedio pagado por almuerzo en dicha cafetería utilizando **Muestreo Sistemático-Estándar** y establezca el LEE=B de dicha estimación con un nivel de confianza de **Aproximadmente** el 95%.

Solución

(a) Para hallar la estimación del valor promedio pagado por almuerzo en dicha cafetería μ, utilizando **Muestreo Sistemático-Estándar** se utiliza la expresión dada por:

$$\hat{\mu} = \bar{y}_{sis} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{40} y_i}{40} = 4800 \text{ pesos.}$$

Ahora para hallar el $LEE = B$ con un nivel de confianza de **aproximadamente el** 95%, se procede como sigue:

$$B = 2\sqrt{\widehat{Var}[\bar{y}_{sis}]},$$

$$\text{con:} \quad \widehat{Var}[\bar{y}_{sis}] = \left(\frac{N-n}{N}\right)\frac{S_{sis}^2}{n}$$

$$= \left(\frac{820-40}{820}\right)\frac{(1045)^2}{40}$$

$$\widehat{Var}[\bar{y}_{sis}] = 25968.89$$

$$\textbf{Luego}, \quad B = 2\sqrt{\widehat{Var}[\bar{y}_{sis}]} = 2\sqrt{25968.89} = 322.2973$$

de donde, un IC estimado de **aproximadamente el** 95% para μ es:

$$\hat{\mu} \ \pm \ B \iff \bar{y}_{sis} \ \pm \ B \iff 4800 \ \pm \ 322.2973 \iff \Big(\ 4477.7027013 \ , \ 5122.2972987\ \Big)$$