

David Montaña Castro

Tarea 5: Muestreo Estratificado con afijación igual en cada estrato.

Muestro

Se dispone de información sin algún contexto inicial. Se cuentan con 625 registros, de los cuales 200 perteneces al estrato 1; al estrato 2, 275; al estrato 3, 150.

N1	200	Aleatorio	N2	275	Aleatorio	N3	150	Aleatorio
1	207.00	0.63	1.00	535.00	0.89	1.00	151.00	0.30
2	287.00	0.20	2.00	336.00	0.33	2.00	524.00	0.34
3	200.00	0.15	3.00	322.00	0.12	3.00	546.00	0.70
4	214.00	0.06	4.00	494.00	0.01	4.00	280.00	0.28
5	272.00	0.46	5.00	267.00	0.89	5.00	603.00	0.48
6	246.00	0.12	6.00	359.00	0.32	6.00	738.00	0.35
7	209.00	0.94	7.00	435.00	0.73	7.00	387.00	0.31
8	136.00	0.27	8.00	334.00	0.57	8.00	734.00	0.31
9	328.00	0.21	9.00	258.00	0.46	9.00	597.00	0.13
10	250.00	0.59	10.00	257.00	0.95	10.00	611.00	0.65

Se pretende estimar un intervalo de confianza para la media global; es decir, la media de los 3 estratos.

MUESTRA PILOTO

Se plantea una muestra piloto de **8 registros**. Para asegurar en todo momento la aleatoriedad, se realizó primero un muestreo simple con ayuda de la función **.rand()**

MUESTRAS PILOTO		
ESTRATO 1	ESTRATO 2	ESTRATO 3
321.00	290.00	286.00
179.00	257.00	399.00
223.00	360.00	442.00
322.00	482.00	603.00
249.00	339.00	260.00
333.00	341.00	625.00
214.00	326.00	413.00
336.00	394.00	519.00

Con esta muestra se planea encontrar el tamaño de la muestra final. Para ello, primero se necesitan hacer algunos cálculos auxiliares:

Estrato	N_h	w_h	\hat{S}_h^2
1	200	0.32	3,951.55
2	275	0.44	4,638.84
3	150	0.24	17,924.84
	625	1.00	

David Montaña Castro

Tarea 5: Muestreo Estratificado con afijación igual en cada estrato.

Muestro

Los pesos (W_h) fueron calculados dividiendo el número de registros en el estrato entre el total de registros:

Estrato	N_h	W_h
1	200	=C6/\$C\$9
2	275	0.44
3	150	0.24
	625	1.00

La varianza fue calculada usando la función **.VAR.S()**

Después, se calculan los siguientes elementos necesarios para calcular el tamaño de la muestra final:

Muestra Piloto					
Estrato	N_h	W_h	\hat{S}_h^2	$N_h^2 \hat{S}_h^2$	$N_h \hat{S}_h^2$
1	200	0.32	3,951.55	158,062,142.9	790,310.71
2	275	0.44	4,638.84	350,812,221.0	1,275,680.80
3	150	0.24	17,924.84	403,308,883.9	2,688,725.89
SUMA	625	1.00		912,183,247.77	4,754,717.41

Además, se necesitan los siguientes elementos. Desde el principio, el nivel de confianza y el error admisible fueron pactados por el profesor.

$$D^2 = \left(\frac{d}{Z_{95}}\right)^2$$

L (Estratos)	3
Z95%	1.96
d	16.00
D ²	66.6

CÁLCULO DEL TAMAÑO DE LA MUESTRA POR ESTRATO

$$n = \frac{L \sum_{h=1}^L N_h^2 S_h^2}{N^2 D^2 + \sum_{h=1}^L N_h S_h^2}$$

David Montaña Castro

Tarea 5: Muestreo Estratificado con afijación igual en cada estrato.

Muestro

Estrato	N_h	W_h	\hat{S}_h^2	$N_h^2 \hat{S}_h^2$	$N_h \hat{S}_h^2$
1	200	0.32	3,951.55	158,062,142.9	790,310.71
2	275	0.44	4,638.84	350,812,221.0	1,275,680.80
3	150	0.24	17,924.84	403,308,883.9	2,688,725.89
SUMA	625	1.00		912,183,247.77	4,754,717.41
L (Estratos)	3	ESTRATOS			
Z95%	1.96				
d	16.00				
D ²	66.6				
			$L \sum_{h=1}^L N_h^2 S_h^2$		
n =	=CEILING.MATH(C11*F9/(C9^2*C14+G9))				

NOTA: se utilizó la función .CEILING.MATH() para asegurar que el tamaño de la muestra se redondeara al entero mayor más cercano.

Ahora, se tiene que dividir el resultado entre en número de estratos para saber cuántos registros por estrato se requieren:

$$nh = \frac{n}{L}$$

n_h
30
30
30
90.00

MUESTRA DEFINITIVA

Por medio de un proceso de aleatorización con la función .RAND(), se adquiere una muestra de cada estrado de **tamaño 30**.

MUESTRAS FINAL		
ESTRATO 1	ESTRATO 2	ESTRATO 3
250.00	304.00	732.00
284.00	424.00	603.00
322.00	448.00	527.00
309.00	333.00	583.00
319.00	402.00	533.00
266.00	366.00	561.00
268.00	375.00	565.00

David Montaña Castro

Tarea 5: Muestreo Estratificado con afijación igual en cada estrato.

Muestro

317.00	301.00	669.00
220.00	354.00	537.00
356.00	242.00	603.00
188.00	318.00	549.00
272.00	482.00	602.00
277.00	257.00	602.00
256.00	354.00	601.00
301.00	528.00	669.00
227.00	307.00	739.00
181.00	391.00	522.00
101.00	452.00	536.00
380.00	257.00	563.00
207.00	334.00	649.00
228.00	330.00	619.00
264.00	422.00	624.00
373.00	386.00	616.00
195.00	391.00	473.00
342.00	345.00	399.00
218.00	332.00	593.00
253.00	348.00	516.00
328.00	394.00	455.00
203.00	437.00	442.00
106.00	234.00	864.00

Con tal información, se procede a calcular los siguientes resultados:

Muestra Final					
n_h	Yh med	$W_h YhMed$	\hat{S}_h^2	$N_h^2 \hat{S}_h^2$	$N_h \hat{S}_h^2$
30	260.37	83.32	4,822.65	192,906,160.92	964,530.80
30	361.60	159.10	4,987.08	377,147,612.07	1,371,445.86
30	584.87	140.37	8,734.81	196,533,206.90	1,310,221.38
90.00		382.79		766,586,979.89	3,646,198.05

La media global estimada es de **382.79** y fue calculada de la siguiente manera:

$$y_{st} = \sum_{h=1}^L W_h y_h$$

David Montaña Castro

Tarea 5: Muestreo Estratificado con afijación igual en cada estrato.

Muestro

INTERVALOS DE CONFIANZA PARA LA MEDIA GLOBAL

El cálculo de la varianza de realiza con:

$$V(\bar{y}_{st}) = \frac{L}{N^2} \sum_{h=1}^L \frac{N h^2 S_h^2}{n} - \frac{1}{N^2} \sum_{h=1}^L N h S_h^2$$

Muestra Piloto						Muestra Final					
Estrato	N_h	W_h	S_h^2	$N_h^2 S_h^2$	$N_h S_h^2$	n_h	Yh med	W_h YhMed	S_h^2	$N_h^2 S_h^2$	$N_h S_h^2$
1	200	0.32	3,951.55	158,062,142.9	790,310.71	30	260.37	83.32	4,822.65	192,906,160.92	964,530.80
2	275	0.44	4,638.84	350,812,221.0	1,275,680.80	30	361.60	159.10	4,987.08	377,147,612.07	1,371,445.86
3	150	0.24	17,924.84	403,308,883.9	2,688,725.89	30	584.87	140.37	8,734.81	196,533,206.90	1,310,221.38
SUMA	625	1.00		912,183,247.77	4,754,717.41	90.00		382.79		766,586,979.89	3,646,198.05
L (Estratos)	3	ESTRATOS									
Z95%	1.96										
d	16.00										
D ²	66.6										
n =	89	n =									
$n_h = \frac{n}{L}$		$N^2 D^2 + \sum_{h=1}^L N h S_h^2$									
						$y_{st} = \sum_{h=1}^L W_h y_h$					
						$V(\bar{y}_{st}) = \frac{L}{N^2} \sum_{h=1}^L \frac{N h^2 S_h^2}{n} - \frac{1}{N^2} \sum_{h=1}^L N h S_h^2$					
						Var media = (C11/C9^2)*L9/H9 - M9/C9^2					

Aún más importante es el Error Estándar (la raíz de la varianza):

Var media	56.08
EE media	7.489

Finalmente, se obtiene el siguiente intervalo de confianza al 95%:

Lim Inf 95%	368.11
Lim Sup 95%	397.47

Que son calculados de la misma forma en la que se han venido calculando:

$$\bar{y}_{st} \pm Z_{95} EE$$

El intervalo de confianza, de acuerdo a la muestra seleccionada, Sí contiene a la estimación de la media global.

Más aún, también contiene a la verdadera media global, calculada a partir de todos los datos:

MEDIA REAL	372.72
------------	--------