

## Taller 3: Tipos de muestreo

1. En una zona se desea estimar la producción total de papa un productor que cuenta con tres parcelas, las cuales están divididas en tres ranchos con diferentes condiciones climáticas, por lo que se plantea un muestreo estratificado. Del primer rancho se muestrearán 10 plantas de 900, en el segundo rancho 12 plantas de 1100 y en el tercero 12 de 1050. Con los datos que se muestran a continuación calcular.

a) El total de la producción del consumidor, límite para el error y el intervalo de confianza.

b) Si se quiere error máximo de 350 kilos. ¿Cual debe ser el tamaño de muestra adecuado?, para que los pesos ( $w_i$ ) son totales con respecto al total de plantas

	Rancho 1	Rancho 2	Rancho 3
1	2	3	2
2	2,5	3,5	2,5
3	2	4	3
4	2,5	4	3,5
5	3	3,5	2,5
6	2	3,5	3
7	3	4	2,5
8	3	2,5	3,5
9	2,5	3	2
10	3	3,5	3
11		3,5	3
12		4	3,5

## Solución

Obtenemos los datos preliminares:

Tamaño de estratos	Valor
N_1	900
N_2	1100
N_3	1050
N	3050

Tamaño de muestra	Valor
n_1	10
n_2	12
n_3	12

## Ejercicio 1-A

Una vez tenemos los datos preliminares, se puede calcular algunos datos necesarios para el punto A, para ello podemos determinar las formulas que se han de usar:

$$\tau = N * \bar{Y}_{st} = \sum_{i=1}^L N_i * \bar{y}_i$$
$$\hat{V}(\bar{Y}_{st}) = N^2 \hat{V}(\bar{Y}_{st}) = \sum_{i=1}^L N_i^2 * \left( \frac{N_i - n_i}{N_i} \right) * \left( \frac{S_{\perp}^2}{n_i} \right)$$
$$2 * \sqrt{\hat{V}(N\bar{Y}_{st})} = 2 * \sqrt{\sum_{i=1}^L N_i^2 * \left( \frac{N_i - n_i}{N_i} \right) * \left( \frac{S_{\perp}^2}{n_i} \right)}$$
$$E = \sqrt{V(\tau)} * 2$$

Podemos usar los datos anteriores y las formulas para facilmente calcular los resultado con excel

	Punto A		
Media	2,55	3,5	2,83333333
Varianza (S)	0,19166667	0,22727273	0,28787879
N_i	900	1100	1050
n_i	10	12	12
N_i^2	810000	1210000	1102500
N_i - n_i / N_i	0,98888889	0,98909091	0,98857143
S_i^2 / n_i	0,01916667	0,01893939	0,02398899

Ahora bien, con los estos datos podemos determinar el total de la producción del consumidor ( $\tau$ ), límite para el error ( $\epsilon$ ) y el intervalo de confianza

$\tau$	9120
V( $\tau$ )	64165,75758
Error	506,6192163
Intervalo	8613,380784
	9626,619216

## Ejercicio 1-B

Ahora para poder calcular el tamaño de muestra adecuado se pueden usar los datos anteriores y la formula:

$$n = \frac{\sum_{i=1}^L N_i^2 * \frac{\sigma_i^2}{w_i}}{N^2 * D + \sum_{i=1}^L N_i * \sigma_i^2}$$

Recordar que el valor de D será:

$$D = \frac{B^2}{4 * N^2}$$

	Punto B		
Varianza (P)	0,1725	0,20833333	0,26388889
W_i	0,29508197	0,36065574	0,3442623
N_i^2*( $\sigma_i^2/w_i$ )	473512,5	698958,333	845104,167
N_i * $\sigma_i^2$	155,25	229,166667	277,083333

Con ello se puede calcular el tamaño de muestra adecuado para esto será necesario determinar las sumatorias:

$$\sum_{i=1}^L N_i^2 * \left( \frac{\sigma_i^2}{w_i} \right)$$

Cuyo valor será: 2017575

$$\sum_{i=1}^L N_i * \sigma_i^2$$

Cuyo valor será: 661.5

B	350
D	0,003292126
Total de la suma	2017575
Total de la suma	661,5
Valor de n	64,48707909

Por tanto, se puede afirmar estadísticamente hablando que se requiere de 64.487 plantas (Aproximado a 65 plantas), esto para que el error maximo sea de 350 kilos

2. La policía de carreteras está interesada en la proporción de automovilistas que portan su licencia. Se instala un puesto de verificación en una carretera principal, y se detiene un vehículo cada siete. Use los datos de la tabla de la derecha para estimar la proporción de conductores que portan su licencia.

a) Establezca un límite para el error de estimación y presente el correspondiente intervalo de confianza. Suponga que N = 2800 autos pasan por el puesto de verificación durante el periodo de muestreo.

b) Calcular n si se tiene un error del 0,015 con el p estimado

c) Calcular n maximizando el tamaño de muestra

## Solución

## Ejercicio 2 - A

Necesitamos calcular algunos datos preliminares con ayuda de las formulas:

$$\hat{p} = \frac{\sum Y_i}{n}$$
$$\hat{q} = 1 - \hat{p}$$
$$\hat{V}(\hat{p}) = \frac{\hat{p} * \hat{q}}{n - 1} * \frac{N - n}{N}$$

Una vez se tengan estos datos se podrá calcular el intervalo de confianza y el limite para el error

N	2800	Punto A-1
n	400	
$\epsilon_{y_i}$	324	
p_hat	0,81	
q_hat	0,19	
V(p)	0,00033061	

Ya para determinar el limite para el error podemos usar la formula:

$$2 * \sqrt{\hat{V}(\hat{p})}$$

Para el intervalo se puede usar:

$$int_{1,2} = \hat{p} \pm \epsilon$$

Cota	0,03636549	Punto A-2
Intervalo de confianza	0,77363451	
	0,84636549	

Ejercicio 2 - B

Para el calculo de n con los datos que nos proporciona el ejercicio podemos usar la formula:

$$n = \frac{N * \hat{p} * \hat{q}}{(N - 1) * D + \hat{p} * \hat{q}}$$

Para este caso de uso, se puede determinar D usando:

$$D = \frac{B^2}{4}$$

B	0,015	Punto B
D	0,00005625	
n	1384,06504	

∴ Estadísticamente hablando se puede decir que el tamaño de muestra necesario será de  $n = 1384$

Ejercicio 2 - C

Recordar el punto maximo está representado por  $\hat{p} = 0.5$  y  $\hat{q} = 0.5$

Para este caso podemos usar las formulas anteriores:

$$n = \frac{N * \hat{p} * \hat{q}}{(N - 1) * D + \hat{p} * \hat{q}}$$

Para este caso de uso, se puede determinar D usando:

$$D = \frac{B^2}{4}$$

N	2800	Punto C
p_hat	0,5	
q_hat	0,5	
B	0,05	
D	0,000625	
n	350,109409	

∴ Estadísticamente hablando podemos decir que el tamaño de muestra maximizado será de  $n = 350$

3. En un estudio sociológico realizado en una pequeña ciudad, se hicieron llamadas telefónicas para estimar la proporción de hogares donde habita cuando menos una persona mayor de 65 años de edad. La ciudad tiene 750 hogares, se tomó una muestra aleatoria de 85 hogares y se obtuvo la información de que en 21 de ellos habita por lo menos un adulto mayor a 65 años

a) Estime la proporción de hogares donde habita por lo menos una persona mayor de 65 años de edad y establezca un límite para el error de estimación.

b) Partiendo del supuesto de que no se ha realizado estudios similares en esta ciudad y se desea realizar una estadística que maneje una cota de error de 3 puntos y un nivel de significancia de 5%, ¿De cuánto debe ser el tamaño de la muestra?

## Solución

## Ejercicio 3 - A

Una vez se identificando los datos preliminares se pueden usar algunas formulas para facilitar el calculos del limite de error y de la proporción de hogares habitados por mayores de 65 años:

$$\hat{p} = \frac{\sum Y_i}{n}$$
$$\hat{q} = 1 - \hat{p}$$
$$\hat{V}(\hat{p}) = \frac{\hat{p} * \hat{q}}{n - 1} * \frac{N - n}{N}$$

N	750	Punto A
n	85	
Y_i	21	
p_hat	0,247058824	
q_hat	0,752941176	
V(p)	0,001963552	
Cota	0,08862398	
Intervalo de confianza	0,158434843	
	0,3355682804	

## Ejercicio 3 - B

Para determinar el tamaño de esa población se puede usar la formula:

$$\frac{Z_{\frac{\alpha}{2}} * \hat{p} * \hat{q}}{\epsilon^2}$$

Z_alpha/2	1,9599	Punto B
$\epsilon$	0,03	
n	405,0912111	

∴ Estadísticamente hablando el tamaño de muestra que cumple con esos constrains será de aproximadamente  $n = 405$

4. Las granjas de una cierta región se dividen en cuatro categorías según su superficie. El número de granjas en cada categoría es 72, 37, 50 y 11. Un estudio para estimar el total de vacas productoras de leche en la región produce una muestra estratificada de 28 granjas. El total de vacas productoras de leche en estas 28 granjas viene dado en la siguiente tabla.

Categoría	Total de vacas										
Categoría I	61	47	44	70	28	39	51	52	101	49	54
Categoría II	160	148	89	139	142	93					
Categoría III	26	19	21	34	28	15	20	24			
Categoría IV	17	11									

a) Estimar el total de vacas productoras de leche, así como el límite para el error de estimación

## Solución

Primero, se re ordena la tabla

		Categoría I	Categoría II	Categoría III	Categoría IV
Total de Vacas	1	61	160	26	17
	2	47	148	19	11
	3	44	89	21	
	4	70	139	34	
	5	28	142	28	
	6	39	93	15	
	7	51		20	
	8	52		24	
	9	101			
	10	49			
	11	54			
	12	71			

Luego se pueden determinar algunos datos

Tamaño de estratos	N_i	72	37	50	11
Tamaño de muestras	n_i	12	6	8	2

Con los datos anteriores se puede usar excel para calcular algunos datos para poder reemplazar en las formulas que nos permitirán determinar la solución al problema:

$$\tau = N * \bar{Y}_{st} = \sum_{i=1}^L N_i * \bar{y}_i$$
$$\hat{V}(\bar{Y}_{st}) = N^2 \hat{V}(\bar{Y}_{st}) = \sum_{i=1}^L N_i^2 * \left( \frac{N_i - n_i}{N_i} \right) * \left( \frac{S_{\perp}^2}{n_i} \right)$$
$$2 * \sqrt{\hat{V}(N\bar{Y}_{st})} = 2 * \sqrt{\sum_{i=1}^L N_i^2 * \left( \frac{N_i - n_i}{N_i} \right) * \left( \frac{S_{\perp}^2}{n_i} \right)}$$
$$E = \sqrt{V(\tau)} * 2$$

Media	55,5833333	128,5	23,375	14
Varianza	350,992424	837,1	35,4107143	18
N_i^2	5184	1369	2500	121
N_i - n_i / N_i	0,83333333	0,83783784	0,84	0,81818182
S_i^2 / n_i	29,2493687	149,516667	4,42633929	9

Con estos datos podemos determinar los datos necesarios para responder el punto a

N (Total)	170
$\tau$	10079,25
V( $\tau$ )	308039,202
Error	1110,02559
Intervalo	8969,22441
	11189,2756