2. Muestreo Aleatorio Estratificado

2.1 Selección de una muestra aleatoria estratificada.
2.2 Estimación de la media, proporción y total poblacionales.
2.3 Determinación del tamaño muestral.
2.4 Asignación de la muestra.
  2.4.1 Asignación Óptima.
  2.4.2 Asignación de Neyman.
  2.4.3 Asignación Proporcional.
2.5 Estratificación después de seleccionar la muestra.

2. Muestreo Aleatorio Estratificado

2.1 Selección de una muestra aleatoria estratificada. Notación.

$L =$ número de estratos

$N =$ tamaño de la población

$n =$ tamaño de la muestra

$N_i =$ tamaño del estrato

$n_i =$ tamaño de la muestra del estrato i

$$N = \sum_{i=1}^{L} N_i \qquad n = \sum_{i=1}^{L} n_i$$

$\mu_i =$ media poblacional del estrato i

$\overline{y}_i =$ media muestral del estrato i

$\tau_i =$ total poblacional del estrato i

$\sigma_i^2 =$ varianza poblacional del estrato i

$S_i^2 =$ Cuasivarianza muestral del estrato i

$p_i =$ proporción poblacional del estrato i

$\hat{p}_i =$ proporción muestral del estrato i

$c_i =$ coste de una observación del estrato i

2. Muestreo Aleatorio Estratificado

2.2 Estimación de la media, proporción y total poblacionales.

$$N_i \overline{y}_i = \hat{\tau}_i \qquad \tau = \sum_{i=1}^{L} \tau_i$$

$$\hat{\tau}_{st} = \sum_{i=1}^{L} N_i \, \overline{y}_i$$

$$\hat{V}(\hat{\tau}_{st}) = \sum_{i=1}^{L} N_i^2 \hat{V}(\overline{y}_i) = \sum_{i=1}^{L} N_i^2 \frac{S_i^2}{n_i} \frac{N_i - n_i}{N_i}$$

$$\overline{y}_{st} = \frac{\hat{\tau}_{st}}{N} = \frac{1}{N} \sum_{i=1}^{L} N_i \overline{y}_i \qquad \hat{V}(\overline{y}_{st}) = \frac{1}{N^2} \sum_{i=1}^{L} N_i^2 \hat{V}(\overline{y}_i) = \frac{1}{N^2} \sum_{i=1}^{L} N_i^2 \frac{S_i^2}{n_i} \frac{N_i - n_i}{N_i}$$

$$\overline{y}_{st} \neq \overline{y} \qquad \hat{\tau}_{st} \neq \hat{\tau} = N\overline{y}$$

3

3

2. Muestreo Aleatorio Estratificado

2.2 Estimación de la media, proporción y total poblacionales.

$$\frac{(1-7)^2+(2-7)^2+(3-7)^2+(11-7)^2+(12-7)^2+(13-7)^2}{6} = 25,6667 \qquad \frac{(1-2)^2+(2-2)^2+(3-2)^2}{3} = 0,6667 \quad (\#)$$

①②③ ⓫⓬⓭	*Población finita*	*Media poblacional,* μ	*Varianza poblacional,* σ^2
	N=6	$\mu = \frac{1+2+3+11+12+13}{6} = 7$	25,6667 (#)
①②③	*Estrato 1*	*Media poblacional,* μ_1	*Varianza poblacional,* σ_1^2
	N_1=3	$\mu_1 = \frac{1+2+3}{3} = 2$	0,6667 (#)
⓫⓬⓭	*Estrato 2*	*Media poblacional,* μ_2	*Varianza poblacional,* σ_2^2
	N_2=3	$\mu_2 = \frac{11+12+13}{3} = 12$	0,6667

4

4

2. Muestreo Aleatorio Estratificado

2.2 Estimación de la media, proporción y total poblacionales.

MUESTRAS ***n*=4** **(n_1=2, n_2=2)**		$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_{st}$
②③	⓬⓭	$\frac{2+3}{2}=2,5$	$\frac{12+13}{2}=12,5$	$\frac{(3\times 2,5)+(3\times 12,5)}{6}=7,5$
②③	⓫⓭	2,5	12	7,25
②③	⓫⓬	2,5	11,5	7
①③	⓬⓭	2	12,5	7,25
①③	⓫⓭	2	12	7
①③	⓫⓬	2	11,5	6,75
①②	⓬⓭	1,5	12,5	7
①②	⓫⓭	1,5	12	6,75
①②	⓫⓬	1,5	11,5	6,5
	TOTAL:	18	108	63
	MEDIA:	$E\left[\bar{y}_1\right]=\frac{18}{9}=2$	$E\left[\bar{y}_2\right]=\frac{108}{9}=12$	$E\left[\bar{y}_{st}\right]=\frac{63}{9}=7$
	VARIANZA:	$V\left[\bar{y}_1\right]=0,1667$	$V\left[\bar{y}_2\right]=0,1667$	$V\left[\bar{y}_{st}\right]=0,0833$

Muestreo aleatorio simple en poblaciones finitas

	(no escojo)	**MUESTRA** **(*n*=4)**	Proporción muestral, $\hat{p}$ (de negras)	Media muestral, $\bar{y}$	S^2_{n-1}
1	①②	③⓫⓬⓭	$\frac{3}{4}=0,75$	$\frac{3+11+12+13}{4}=9,75$	20,92
2	①③	②⓫⓬⓭	$\frac{3}{4}=0,75$	$\frac{2+11+12+13}{4}=9,5$	25,67
3	①⓫	②③⓬⓭	$\frac{2}{4}=0,50$	$\frac{2+3+12+13}{4}=7,5$	33,67
4	①⓬	②③⓫⓭	$\frac{2}{4}=0,50$	$\frac{2+3+11+13}{4}=7,25$	30,92
5	①⓭	②③⓫⓬	$\frac{2}{4}=0,50$	$\frac{2+3+11+12}{4}=7$	27,33
6	②③	①⓫⓬⓭	$\frac{3}{4}=0,75$	$\frac{1+11+12+13}{4}=9,25$	30,92
7	②⓫	①③⓬⓭	$\frac{2}{4}=0,50$	$\frac{1+3+12+13}{4}=7,25$	37,58
8	②⓬	①③⓫⓭	$\frac{2}{4}=0,50$	$\frac{1+3+11+13}{4}=7$	34,67
9	②⓭	①③⓫⓬	$\frac{2}{4}=0,50$	$\frac{1+3+11+12}{4}=6,75$	30,92
10	③⓫	①②⓬⓭	$\frac{2}{4}=0,50$	$\frac{1+2+12+13}{4}=7$	40,67
11	③⓬	①②⓫⓭	$\frac{2}{4}=0,50$	$\frac{1+2+11+13}{4}=6,75$	37,58
12	③⓭	①②⓫⓬	$\frac{2}{4}=0,50$	$\frac{1+2+11+12}{4}=6,5$	33,67
13	⓫⓬	①②③⓭	$\frac{1}{4}=0,25$	$\frac{1+2+3+13}{4}=4,75$	30,92
14	⓫⓭	①②③⓬	$\frac{1}{4}=0,25$	$\frac{1+2+3+12}{4}=4,5$	25,67
15	⓬⓭	①②③⓫	$\frac{1}{4}=0,25$	$\frac{1+2+3+11}{4}=4,25$	20,92
		TOTAL:	7,50	105	462

Muestreo aleatorio simple en poblaciones finitas

MUESTRAS (*n*=4)	Proporción muestral, $\hat{p}$ (de negras)	Media muestral, $\bar{y}$
TOTAL:	7,50	105
MEDIA:	$E\left[\hat{p}\right]=\frac{7,5}{15}=0,50$	$E\left[\bar{y}\right]=\frac{105}{15}=7$
VARIANZA:	$V\left[\hat{p}\right]=0,025$	$V\left[\bar{y}\right]=2,5667$

MUESTRA (*n*=5)	Proporción muestral, $\hat{p}$ (*de negras*)	Media muestral, $\bar{y}$
MEDIA:	$E\left[\hat{p}\right]=\frac{3}{6}=0,50$	$E\left[\bar{y}\right]=\frac{42}{6}=7$
VARIANZA:	$V\left[\hat{p}\right]=0,0100$	$V\left[\bar{y}\right]=1,0267$

2. Muestreo Aleatorio Estratificado

2.2 Estimación de la media, proporción y total poblacionales.

El muestreo aleatorio estratificado no conduce a mejores estimadores si los estratos no son más homogéneos que el conjunto de la población y/o no hay claras diferencias entre los elementos de uno y otro estrato.

①②③ ❶❷❸	Población finita	Media poblacional, μ	Varianza poblacional, σ^2
	N=6	$\mu=\frac{1+2+3+1+2+3}{6}=2$	0,6667

	MUESTRAS (*n*=4)	Media muestral, $\bar{y}$
1	③❶❷❸	$\frac{3+1+2+3}{4}=2,25$
2	②❶❷❸	2
3	②③❷❸	2,5
4	②③❶❸	2,25
5	②③❶❷	2
6	①❶❷❸	1,75
7	①③❷❸	2,25
8	①③❶❸	2
9	①③❶❷	1,75
10	①②❷❸	2
11	①②❶❸	1,75
12	①②❶❷	1,5
13	①②③❸	2,25
14	①②③❷	2
15	①②③❶	1,75
	TOTAL:	30

MUESTRAS (*n*=4)	Media muestral, $\bar{y}$
MEDIA:	$E\left[\bar{y}\right]=\frac{30}{15}=2$
VARIANZA:	$V\left[\bar{y}\right]=0,0667$

2. Muestreo Aleatorio Estratificado

2.2 Estimación de la media, proporción y total poblacionales.

①②③ ❶❷❸	*Población finita*	*Media poblacional,* μ	*Varianza poblacional,* σ^2
	N=6	$\mu = \frac{1+2+3+1+2+3}{6} = 2$	0,6667
①②③	*Estrato 1*	*Media poblacional,* μ_1	*Varianza poblacional,* σ_1^2
	N_1=3	$\mu_1 = \frac{1+2+3}{3} = 2$	0,6667
❶❷❸	*Estrato 2*	*Media poblacional,* μ_2	*Varianza poblacional,* σ_2^2
	N_2=3	$\mu_2 = \frac{1+2+3}{3} = 2$	0,6667

MUESTRAS *n*=4 (*n₁*=2, *n₂*=2)		$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_{st}$
②③	❷❸	$\frac{2+3}{2} = 2,5$	$\frac{2+3}{2} = 2,5$	$\frac{(3\times 2,5)+(3\times 2,5)}{6} = 2,5$
②③	❶❸	2,5	2	2,25
②③	❶❷	2,5	1,5	2,00
①③	❷❸	2	2,5	2,25
①③	❶❸	2	2	2,00
①③	❶❷	2	1,5	1,75
①②	❷❸	1,5	2,5	2,00
①②	❶❸	1,5	2	1,75
①②	❶❷	1,5	1,5	1,50
TOTAL:		18	18	18
MEDIA:		$E[\bar{y}_1] = \frac{18}{9} = 2$	$E[\bar{y}_2] = \frac{18}{9} = 2$	$E[\bar{y}_{st}] = \frac{18}{9} = 2$
VARIANZA:		$V[\bar{y}_1] = 0,1667$	$V[\bar{y}_2] = 0,1667$	$V[\bar{y}_{st}] = 0,0833$

9

9

Ejemplo 2.5 Un distribuidor de productos de limpieza desea conocer el consumo por hogar durante un año de un determinado producto en una comarca formada por cuatro municipios. Para estimar de paso también el consumo en cada municipio decide usar muestreo estratificado tomando cada municipio como un estrato. Se sabe que el 20% de la población de la comarca vive en el municipio 1, el 30% en el municipio 2, el 25% en el municipio 3 y el 25% restante en el municipio 4. El distribuidor tiene medios suficientes para controlar y obtener datos sobre el consumo anual de 20 hogares.

Dado que no tiene información previa respecto a las varianzas de los estratos y porque el coste del muestreo es el mismo en cada municipio, aplica asignación proporcional, la cual conduce a

$$n_1 = n\frac{N_1}{N} = 20 \times 0,20 = 4 \quad \text{de forma similar} \quad n_2 = 6 \quad n_3 = 5 \quad n_4 = 5\,.$$

Obteniendo los resultados de la tabla siguiente (consumo expresado en valor en euros).

Estrato 1	Estrato 2	Estrato 3	Estrato 4
470	490	540	450
510	500	480	560
500	470	500	460
550	520	470	440
	550	470	580
	500		
$\bar{y}_1 = 507,5 \quad S_1^2 = 1091,67$	$\bar{y}_2 = 505 \quad S_2^2 = 750$	$\bar{y}_3 = 492 \quad S_3^2 = 870$	$\bar{y}_4 = 498 \quad S_4^2 = 4420$

Estime el consumo anual medio por hogar y fije un límite para el error de estimación.

10

$$\frac{N_1}{N}=0,20 \qquad \frac{N_2}{N}=0,30 \qquad \frac{N_3}{N}=0,25 \qquad \frac{N_4}{N}=0,25$$

$$\bar{y}_{st}=\frac{1}{N}\sum_{i=1}^{4}N_i\bar{y}_i=\sum_{i=1}^{4}\frac{N_i}{N}\bar{y}_i=(0,20\times 507,5)+(0,30\times 505)+(0,25\times 492)+(0,25\times 498)=500,5€$$

$$\hat{V}(\bar{y}_{st})=\frac{1}{N^2}\sum_{i=1}^{4}N_i^2\frac{S_i^2}{n_i}\frac{N_i-n_i}{N_i}=\sum_{i=1}^{4}\frac{N_i^2}{N^2}\frac{S_i^2}{n_i}=\sum_{i=1}^{2}\left(\frac{N_i}{N}\right)^2\frac{S_i^2}{n_i}=$$

$$=\left(0,20^2\frac{1091,67}{4}\right)+\left(0,30^2\frac{750}{6}\right)+\left(0,25^2\frac{870}{5}\right)+\left(0,25^2\frac{4420}{5}\right)=88,29$$

$$2\sqrt{\hat{V}(\bar{y}_{st})}=18,79\text{ €}$$

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{20}y_i=500,5\text{ €} \qquad S_{n-1}^2=1520,79$$

$$\hat{V}(\bar{y})=\frac{S_{n-1}^2}{n}\frac{N-n}{N}=\frac{1520,79}{20}=76,04 \qquad 2\sqrt{\hat{V}(\bar{y})}=17,44\text{ €}$$

11

2. Muestreo Aleatorio Estratificado

2.3 Determinación del tamaño muestral.

$$2\sqrt{V(\bar{y}_{st})}=B \qquad V(\bar{y}_{st})=\frac{1}{N^2}\sum_{i=1}^{L}N_i^2\frac{\sigma_i^2}{n_i}\frac{N_i-n_i}{N_i-1} \qquad n_i=n\omega_i$$

$$n=\frac{\sum_{i=1}^{L}\frac{N_i^2\sigma_i^2}{\omega_i}}{N^2D+\sum_{i=1}^{L}N_i\sigma_i^2} \qquad D=\frac{B^2}{4}\ (media) \qquad D=\frac{B^2}{4N^2}\ (total)$$

$$\sigma_i^2=p_iq_i \qquad n=\frac{\sum_{i=1}^{L}\frac{N_i^2p_iq_i}{\omega_i}}{N^2D+\sum_{i=1}^{L}N_ip_iq_i} \qquad D=\frac{B^2}{4}\ (proporción) \qquad D=\frac{B^2}{4N^2}\ (total)$$

12

12

2. Muestreo Aleatorio Estratificado

2.4.1 Asignación óptima. $c_1n_1 + c_2n_2 + c_3n_3 = C$

Minimiza el coste de obtención de la muestra para un límite del error de estimación fijado.

$$\omega_j = \frac{\dfrac{N_j\sigma_j}{\sqrt{c_j}}}{\sum_{i=1}^{L}\dfrac{N_i\sigma_i}{\sqrt{c_i}}} \qquad \omega_j = \frac{N_j\sqrt{\dfrac{p_jq_j}{c_j}}}{\sum_{i=1}^{L}N_i\sqrt{\dfrac{p_iq_i}{c_i}}}$$

$$n = \frac{\sum_{i=1}^{L}N_i\sigma_i\sqrt{c_i}\ \sum_{i=1}^{L}\dfrac{N_i\sigma_i}{\sqrt{c_i}}}{N^2D + \sum_{i=1}^{L}N_i\sigma_i^2} \qquad n = \frac{\sum_{i=1}^{L}N_i\sqrt{p_iq_ic_i}\ \sum_{i=1}^{L}N_i\sqrt{\dfrac{p_iq_i}{c_i}}}{N^2D + \sum_{i=1}^{L}N_ip_iq_i}$$

Minimiza el límite del error de estimación para un coste de obtención de la muestra fijo.

$$c_1n_1 + c_2n_2 + c_3n_3 = C$$

$$c_1\omega_1n + c_2\omega_2n + c_3\omega_3n = C$$

$$n = \frac{C}{c_1\omega_1 + c_2\omega_2 + c_3\omega_3}$$

$$n = \frac{C\sum_{i=1}^{L}\dfrac{N_i\sigma_i}{\sqrt{c_i}}}{\sum_{i=1}^{L}N_i\sigma_i\sqrt{c_i}} \qquad n = \frac{C\sum_{i=1}^{L}N_i\sqrt{\dfrac{p_iq_i}{c_i}}}{\sum_{i=1}^{L}N_i\sqrt{p_iq_ic_i}}$$

5

13

2. Muestreo Aleatorio Estratificado

2.4.2 Asignación de Neyman.

Caso numérico:

$$\omega_j = \frac{N_j\sigma_j}{\sum_{i=1}^{L}N_i\sigma_i} \qquad n = \frac{\left(\sum_{i=1}^{L}N_i\sigma_i\right)^2}{N^2D + \sum_{i=1}^{L}N_i\sigma_i^2}$$

Caso dicotómico:

$$\omega_j = \frac{N_j\sqrt{p_jq_j}}{\sum_{i=1}^{L}N_i\sqrt{p_iq_i}} \qquad n = \frac{\left(\sum_{i=1}^{L}N_i\sqrt{p_iq_i}\right)^2}{N^2D + \sum_{i=1}^{L}N_ip_iq_i}$$

Si $c_1 = c_2 = \ldots = c_i = \ldots = c_L$ *coincide con la Optima*

14

14

2. Muestreo Aleatorio Estratificado

2.4.3 Asignación proporcional.

Caso numérico:

$$\omega_j = \frac{N_j}{N} \qquad n = \frac{\sum_{i=1}^{L} N_i \sigma_i^2}{ND + \frac{1}{N}\sum_{i=1}^{L} N_i \sigma_i^2}$$

Caso dicotómico:

$$\omega_j = \frac{N_j}{N} \qquad n = \frac{\sum_{i=1}^{L} N_i p_i q_i}{ND + \frac{1}{N}\sum_{i=1}^{L} N_i p_i q_i}$$

Si $c_1 = c_2 = \ldots = c_i = \ldots = c_L \qquad \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_i^2 = \ldots = \sigma_L^2$

coincide con la Optima

Ventajas:

$$\overline{y}_{st} = \overline{y} \qquad \hat{p}_{st} = \hat{p} \qquad \hat{\tau}_{st} = \hat{\tau}$$

Resuelve complicaciones en la asignación para varias mediciones muestrales 15

15

2. Muestreo Aleatorio Estratificado

2.4.3 Asignación proporcional.

Determinación del tamaño muestral y asignación para varias estimaciones

Asignación óptima/Neyman

1ª estimación: $n = 100 \qquad \omega_1 = 0,10 \Rightarrow n_1 = 10 \qquad \omega_2 = 0,90 \Rightarrow n_2 = 90$

2ª estimación: $n = 40 \qquad \omega_1 = 0,50 \Rightarrow n_1 = 20 \qquad \omega_2 = 0,50 \Rightarrow n_2 = 20$

Asignación proporcional

1ª estimación: $n = 100 \qquad \omega_1 = 0,30 \Rightarrow n_1 = 30 \qquad \omega_2 = 0,70 \Rightarrow n_2 = 70$

2ª estimación: $n = 40 \qquad \omega_1 = 0,30 \Rightarrow n_1 = 12 \qquad \omega_2 = 0,70 \Rightarrow n_2 = 28$

16

16

2. Muestreo Aleatorio Estratificado

2.5 Estratificación después de seleccionar la muestra.

Ejemplo 2.6 En una ciudad se sabe que el 30% de los hogares tienen calefacción eléctrica. Al realizar una encuesta sobre el consumo de energía (valor en euros de la factura bimensual) se obtuvieron los siguientes resultados:

Tipo Calefacción	Nº casas	Valor total de las facturas	Cuasidesviación típica
Eléctrica	60	5730	200
No electrica	40	2080	90

Obtenga una estimación del valor medio de la factura de electricidad en la ciudad. Dé un límite para el error de estimación.

17

2. Muestreo Aleatorio Estratificado

2.5 Estratificación después de seleccionar la muestra.

$$\bar{y} = \frac{5730+2080}{60+40} = \frac{7810}{100} = 78{,}10€$$

Solución:

$$\bar{y}_1 = \frac{5730}{60} = 95{,}5€ \qquad \bar{y}_2 = \frac{2080}{40} = 52€$$

$$\bar{y}_{st} = \frac{1}{N}\sum_{i=1}^{2} N_i \bar{y}_i = \sum_{i=1}^{2} \frac{N_i}{N}\bar{y}_i = (0{,}30 \times 95{,}5) + (0{,}70 \times 52) = 65{,}05€$$

$$\hat{V}(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{2} N_i^2 \frac{S_i^2}{n_i}\frac{N_i - n_i}{N_i} = \sum_{i=1}^{2} \frac{N_i^2}{N^2}\frac{S_i^2}{n_i}\frac{N_i - n_i}{N_i} =$$

$$= \sum_{i=1}^{2} \frac{N_i^2}{N^2}\frac{S_i^2}{n_i} = \sum_{i=1}^{2} \left(\frac{N_i}{N}\right)^2 \frac{S_i^2}{n_i} = \left(0{,}30^2 \frac{200^2}{60}\right) + \left(0{,}70^2 \frac{90^2}{40}\right) = 159{,}225$$

$$2\sqrt{\hat{V}(\bar{y}_{st})} = 25{,}24€$$

18