EJERCICIOS RESUELTOS TEMA 4: Muestreo y técnicas de encuesta

Ejercicio 1

En una determinada ciudad deseamos conocer la proporción actual de hogares con más de un automóvil. Por los datos de un estudio anterior sabemos que del total de 30.521 hogares, 12.530 tenían más de un vehículo. Estableciendo un nivel de confianza del 95,45% y un error e = 4%, calcular:

a) El tamaño de la muestra requerido para realizar la estimación.
b) El tamaño de la muestra si desconociéramos los datos anteriores sobre el número coches por hogar

Solución

a) A partir de los datos del enunciado podemos conocer la proporción de hogares con más de un vehículo:

$$p=\frac{12530}{30521}=0,41 \qquad \text{Nótese que } q = 1- p = 0,59$$

Dado que se trata de una población finita (N<100.000), para calcular el tamaño muestral utilizaremos la siguiente fórmula, que incorpora el tamaño poblacional a través de la corrección para poblaciones finitas:

$$n=\frac{Z^2PQN}{e^2(N-1)+Z^2PQ}=\frac{2^2\cdot 0,41\cdot 0,59\cdot 30521}{0,04^2\cdot 30520+2^2\cdot 0,41\cdot 0,59}=\frac{29532,1196}{49,7996}=593,02$$

El tamaño muestral es ligeramente superior a 593. En el cálculo del tamaño muestral los redondeos son por exceso, por lo que la muestra debe comprender 594 hogares.

b) Si desconociéramos la proporción de hogares con más de un coche, consideraríamos el caso más desfavorable, es decir p=q=0,5, que es el que maximiza la varianza. Por tanto, el tamaño muestral se calcula de forma análoga al caso anterior, como sigue:

$$n=\frac{Z^2PQN}{e^2(N-1)+Z^2PQ}=\frac{2^2\cdot 0,5\cdot 0,5\cdot 30521}{0,04^2\cdot 30520+2^2\cdot 0,5\cdot 0,5}=\frac{30521}{49.832}=612,48$$

El tamaño muestral es de 613 hogares. Como puede observarse, el desconocimiento de la proporción de hogares con más de un vehículo y la utilización, por tanto, del supuesto de máxima varianza, implica la necesidad de un tamaño muestral mayor, para el mismo nivel de confianza y error.

Ejercicio 2

Se desea estimar la percepción media (en una escala del 0 al 10) sobre la incidencia del acoso escolar en una Comunidad Autónoma. Sabiendo por estudios previos que la desviación típica es 3,6, se necesita saber el tamaño muestral necesario con un error admitido de ±0,5 puntos y un nivel de confianza del 95%

Solución

El enunciado nos aporta los siguientes datos:

- Estimación de la media (μ)
- N > 100.000
- e= ±0,5
- Nc= 95%   Z= ± 1,96
- σ=3,6

Debemos utilizar la fórmula del cálculo de *n* para la estimación de medias con población infinita:

$$n = \frac{Z^2 \cdot \sigma^2}{e^2} = \frac{1{,}96^2 \cdot 3{,}6^2}{0{,}5^2} = \frac{3{,}8416 \cdot 12{,}96}{0{,}25} = 199{,}1 \approx 200 \text{ individuos}$$

Ejercicio 3

Calcule el tamaño muestral para un estudio sobre la percepción de la seguridad ciudadana en un municipio de 60.500 habitantes. Considere P=Q, error del 4% y Nc=98%

Solución

Nuestros datos son:

- Estimación proporciones
- Población finita: N<100.000
- e=0,04
- Nc=98%
- P=Q supuesto de máxima varianza al desconocerse P

La fórmula a aplicar es la correspondiente al cálculo de n para estimar proporciones en poblaciones finitas:

$$n = \frac{Z^2 PQN}{e^2(N-1) + Z^2 PQ}$$

Antes de nada hemos de buscar en la tabla de la curva normal estándar el Z correspondiente a un nivel de confianza de 0,98. Como las tablas proporcionan la probabilidad que existe entre μ=0 y un valor de Z cualquiera (es decir, de la "mitad" de la curva), y sabiendo que la curva normal es totalmente simétrica y comprende en su totalidad un área (probabilidad) igual a 1, entonces, debemos buscar en la tabla la casilla correspondiente a $\frac{Nc}{2}$:

$\frac{Nc}{2}$: $\frac{0{,}98}{2} = 0{,}4900$ Buscamos la celda 0,4900 y comprobamos a qué Z corresponde

En la tabla de la curva normal estándar vemos que el Z que más se aproxima a esa probabilidad es Z= ±2,33. Aplicamos la fórmula:

$$n = \frac{Z^2PQN}{e^2(N-1)+Z^2PQ} = \frac{2{,}33^2 \cdot 0{,}5 \cdot 0{,}5 \cdot 60500}{0{,}04^2(60500-1)+2{,}33^2 \cdot 0{,}5 \cdot 0{,}5} = \frac{82112{,}113}{96{,}7984+1{,}357225} = 836{,}55$$

n ≈ 837 entrevistas

Ejercicio 4

En un municipio de 1.500.000 habitantes, se sabe que el 60% suele realizar sus compras en grandes almacenes. Se ha realizado una encuesta para valorar la posibilidad de mantener abiertos dichos establecimientos en horario nocturno, con una muestra de 900 personas y un nivel de confianza del 95,45%.

a) ¿Qué error máximo se ha admitido?
b) ¿Qué tamaño debería tener la muestra para que, con el mismo nivel de confianza, el error admitido fuera del 2%?
c) ¿Qué ocurriría con el tamaño de la muestra si deseáramos aplicar un nivel de confianza del 99,5%, con e=2%? Explique las ventajas e inconvenientes de la ampliación y reducción del nivel de confianza.

Solución

a) Nuestros datos son:

- Población infinita
- P=0,6 Q=0,4
- n= 900
- Nc= 0,9545 por lo que Z= ±2

Utilizando la fórmula del tamaño muestral para poblaciones "infinitas", puede despejarse el valor del error:

$$e = \sqrt{\frac{Z^2PQ}{n}} = \sqrt{\frac{2^2 \cdot 0{,}6 \cdot 0{,}4}{900}} = \sqrt{\frac{0{,}96}{900}} = 0{,}0327$$

Una forma análoga es sustituir directamente los datos en la fórmula del tamaño muestral y despejar el error al final:

$$n = \frac{Z^2PQ}{e^2} \quad ; \quad 900 = \frac{2^2 \cdot 0{,}6 \cdot 0{,}4}{e^2} \quad ; \quad 900 = \frac{0{,}96}{e^2}$$

$$900 \cdot e^2 = 0{,}96 \quad ; \quad e^2 = \frac{0{,}96}{900} \quad ; \quad e = \sqrt{\frac{0{,}96}{900}} = 0{,}0327$$

Por tanto, el máximo error admitido es del 3,27%

b) Ahora e= 0,02. Aplicamos la fórmula del tamaño muestral:

$$n = \frac{Z^2PQ}{e^2} = \frac{2^2 \cdot 0{,}6 \cdot 0{,}4}{0{,}02^2} = 2.400 \text{ entrevistas}$$

Como podemos observar, al reducir el error admisible (siempre que mantengamos el mismo nivel de confianza) el tamaño muestral necesario aumenta.

c) Para un Nc= 99,5% tenemos que dividir 0,995 entre 2 para localizar el Z correspondiente.

$\frac{0{,}995}{2} = 0{,}4975$ Este es el área bajo la curva normal que debemos buscar en las tablas, pues marcará el Z que buscamos: Z=± 2,81.

Aplicamos la fórmula del tamaño muestral:

$$n = \frac{Z^2 PQ}{e^2} = \frac{2{,}81^2 \cdot 0{,}6 \cdot 0{,}4}{0{,}02^2} = 4.737{,}66 \approx 4.738 \text{ entrevistas}$$

Vemos que el tamaño de la muestra aumenta de forma sensible al incrementar el nivel de confianza. Aumentar el nivel de confianza implica incrementar el tamaño muestral porque deseamos aumentar la probabilidad de que la proporción poblacional se encuentre dentro del margen de error establecido.

¿Qué significa un nivel de confianza del 99,5%? Significa que un 99,5% de todas las muestras posibles que se podrían extraer de la población tendrían un valor de p dentro del margen de error establecido (p±0,02). Solo se quedarían fuera de ese margen de error el 0,05% restante de las muestras que pudiéramos extraer. Sin embargo, la desventaja es que, al aumentar el tamaño de la muestra, aumentan los costes de la investigación.

Como hemos visto en el apartado anterior, también aumentará el tamaño de la muestra si, para un mismo nivel de confianza, quisiéramos reducir el error admitido.

Ejercicio 5

Se desea realizar un estudio sobre la prevalencia del consumo de cannabis en las Comunidades Autónomas de Madrid, Castilla-la Mancha y Comunidad Valenciana. Según datos del Ministerio de Sanidad de 2013, la proporción de consumidores en cada una de esas regiones era la siguiente:

Región	Población (N_i)	Consumo de cannabis (%)
Com. Madrid	6.466.996	9,1
Castilla-La Mancha	2.041.631	6,4
Com. Valenciana	4.959.968	10,8
Total 3 regiones	13.468.595	

¿Qué tamaño muestral sería necesario para un estudio actualizado teniendo en cuenta un nivel de confianza del 99% y un error máximo admitido del 3%? Considere cada región como un estrato diferente, teniendo en cuenta su tamaño para el cálculo de n.

Solución

Nos encontramos ante un problema de muestreo estratificado proporcional al tamaño de cada estrato, siendo cada Comunidad Autónoma un estrato diferente. Debemos calcular el tamaño muestral (n_i) para cada estrato.

En este diseño, la afijación (el reparto de la muestra entre los distintos estratos) debe considerar el tamaño de cada estrato respecto al total. Por eso, es necesario calcular el peso de cada uno de ellos: $W_i = \frac{N_i}{N}$

La fórmula del tamaño del tamaño muestral tiene en cuenta la varianza conjunta de los estratos considerando su peso relativo:

Varianza total: $(pq)_{ST}$: $\sum p_i q_i W_i$

Así:

$$n = \frac{Z^2 (pq)_{ST}}{e^2}$$

A partir de la tabla anterior, hacemos los cálculos necesarios:

Región	Población (N_i)	Peso del estrato en la población (W_i)	p_i	q_i	$p_i\,q_i\,W_i$
Com. Madrid	6.466.996	0,48	0,091	0,909	0,0397051
Castilla-La Mancha	2.041.631	0,15	0,064	0,936	0,0089856
Com. Valenciana	4.959.968	0,37	0,108	0,892	0,0356443
Total 3 regiones	13.468.595	1			0,084335

$$n = \frac{Z^2 (pq)_{st}}{e^2} = \frac{2{,}58^2 \cdot 0{,}084335}{0{,}03^2} = \frac{0{,}5613674}{0{,}0009} = 623{,}7 \approx 624 \text{ individuos en total}$$

Seguidamente repartimos n entre los estratos según su peso en la población: $n_i = n \cdot W_i$

Com. Madrid: $n_i = 624 \cdot 0{,}48 = 299{,}5 \approx 299$ entrevistas

Castilla-La Mancha: $n_i = 624 \cdot 0{,}15 = 93{,}6 \approx 94$ entrevistas

Com. Valenciana: $n_i = 624 \cdot 0{,}37 = 230{,}9 \approx 231$ entrevistas

Ejercicio 6

Una empresa de publicidad quiere conocer la proporción de hogares en los que se escucha una determinada emisora de radio en una región. Para ello, dicha región se divide en tres estratos: Municipio A, Municipio B y Área Rural con N_1=620, N_2=1.210 y N_3=340 hogares respectivamente. Las proporciones p_i se aproximan por estimaciones de un estudio anterior: p_1=0,40, p_2=0,45 y p_3=0,32. Calcule el tamaño de la muestra para estimar la proporción de oyentes en el conjunto de la región con un error máximo del 5% y un nivel de confianza del 95,45%. Considerando la diferencia de varianza entre estratos, aplique una afijación de Neyman.

Solución

En primer lugar calculamos el peso de los estratos y la varianza total de la población:

	N_i	W_i	p_i	q_i	$p_i\,q_i\,W_i$
Municipio A	620	0,28571429	0,40	0,60	0,06857143
Municipio B	1210	0,55760369	0,45	0,55	0,13800691
Área Rural	340	0,15668203	0,32	0,68	0,03409401
Total	2170	1			0,24067235

La varianza total de la población es: $(pq)_{st} = \sum p_i q_i W_i = 0{,}24$

Al tratarse de una población finita (N total < 100.000), debemos aplicar la siguiente fórmula para hallar el tamaño total de la muestra:

$$n = \frac{Z^2 N (pq)_{st}}{e^2(N-1)+Z^2(pq)_{st}} = \frac{2^2 \cdot 2170 \cdot 0{,}24}{0{,}05^2 \cdot 2169 + 2^2 \cdot 0{,}24} = \frac{2083{,}2}{5{,}4225+0{,}96} = 326{,}39 \approx 327 \text{ hogares}$$

Ahora debemos repartir el tamaño muestral total entre los distintos estratos según la afijación de Neyman:

$$n_i = n \frac{N_i\sqrt{p_i q_i}}{\sum N_i\sqrt{p_i q_i}}$$

Realizamos los cálculos necesarios en la siguiente tabla:

	N_i	p_i	q_i	$N_i\sqrt{p_i q_i}$
Municipio A	620	0,40	0,60	303,736728
Municipio B	1210	0,45	0,55	601,967399
Área rural	340	0,32	0,68	158,601892
Total				1064,30602

Finalmente, aplicando la fórmula repartimos n entre los distintos estratos:

Municipio A: $n_1 = 327 \frac{303{,}736728}{1064{,}30602} = 93{,}32 \approx 93$ hogares

Municipio B: $n_2 = 327 \frac{601{,}967399}{1064{,}30602} = 184{,}95 \approx 185$ hogares

Área rural: $n_3 = 327 \frac{158{,}601892}{1064{,}30602} = 48{,}73 \approx 49$ hogares

Ejercicio 7

El Ayuntamiento de un municipio va a realizar una encuesta de opinión sobre la actuación de la policía municipal. Cada entrevista tiene un coste de 20€, siendo el presupuesto para el total de entrevistas de 15.000€. Conociendo, gracias al registro padronal, la composición de la población adulta del municipio, se decide establecer cuotas para componer la muestra. Los datos poblacionales según sexo y edad son los siguientes:

		Edad			
		18-34	35-64	65 y más	Total
Sexo	Hombres	25.453	67.950	29.236	
	Mujeres	26.415	70.487	37.429	
	Total				256.970

Solución

Para calcular las cuotas correspondientes a cada categoría de sexo y edad, vemos su peso en el conjunto de la población. Por ejemplo, la cuota correspondiente a los hombres de 18-34 años es:

$$\frac{25453}{256970} = 0{,}099$$

Así, para cada una de las categorías de la población, obtenemos los siguientes resultados:

		Edad			
		18-34	35-64	65 y más	Total
Sexo	Hombres	0,099	0,264	0,114	
	Mujeres	0,103	0,274	0,146	
	Total				1

Una vez establecidas las cuotas, distribuimos proporcionalmente la muestra entre las distintas categorías de la población. Sabiendo los datos del presupuesto municipal y el coste de cada entrevista, hallamos el número de entrevistas a realizar (tamaño muestral):

$$n = \frac{15000}{20} = 750 \text{ entrevistas}$$

Para calcular el número de entrevistas para los hombres entre 18-34 años hay que ponderar el número total de entrevistas por la cuota correspondiente a esta categoría (redondeamos los resultados para ajustarnos al *n* total presupuestado):

$n_i = n \cdot w_i = 750 \cdot 0{,}099 = 74{,}25 \approx 74$ entrevistas

Así sucesivamente para cada categoría poblacional, la distribución de entrevistas es la siguiente:

		Edad			
		18-34	35-64	65 y más	Total
Sexo	Hombres	74	198	86	
	Mujeres	77	205	110	
	Total entrevistas				750

Como puede observarse, no hemos utilizado para este supuesto ninguna fórmula probabilística para el cálculo del tamaño muestral, ya que el muestreo por cuotas, aunque muy utilizado en la fase final de muchos estudios mediante encuesta, no es un muestro probabilístico. En la práctica, para la elaboración de encuetas se utilizan diseños mixtos con técnicas probabilísticas en la primera parte del diseño muestral y cuotas para la selección de individuos solo en la fase final.

****

DESARROLLO EXPLICADO DE LOS EJERCICIOS RESUELTOS DEL *CAPÍTULO VIII* (*Las muestras estadísticas: teoría y diseño*) del manual de la asignatura:

Estadística para la Investigación Social (2ª edición), de Camarero, L. et al.

El enunciado de los ejercicios se encuentra en el capítulo correspondiente.

Ejercicio 2. Solución

a) Calculamos la media de la población:

$$\mu_{x_i} = \frac{\sum x_i n_i}{N} = \frac{12+8+6+4+10+18+16+14}{8} = 11$$

b) Para construir la distribución muestral de las medias para muestras de tamaño 2, vamos a calcular primero cuántas muestras de n=2 podemos obtener con el total de elementos de la población. Tratándose de una población de N=8, hallamos las combinaciones de 8 elementos, tomados de 2 en 2:

$$C_{8,2} = \binom{8}{2} = \frac{8!}{2!\cdot 6!} = 28 \text{ muestras}$$

Podemos construir 28 muestras de tamaño 2. Ahora desarrollamos todas las muestras posibles de n=2 y calculamos el valor de la media muestral $\bar{x}$ para cada una de ellas:

Muestra	Valor del elemento 1	Valor del elemento 2	$\bar{x}$
a, b	12	8	10
a, c	12	6	9
a, d	12	4	8
a, e	12	10	11
a, f	12	18	15
a, g	12	16	14

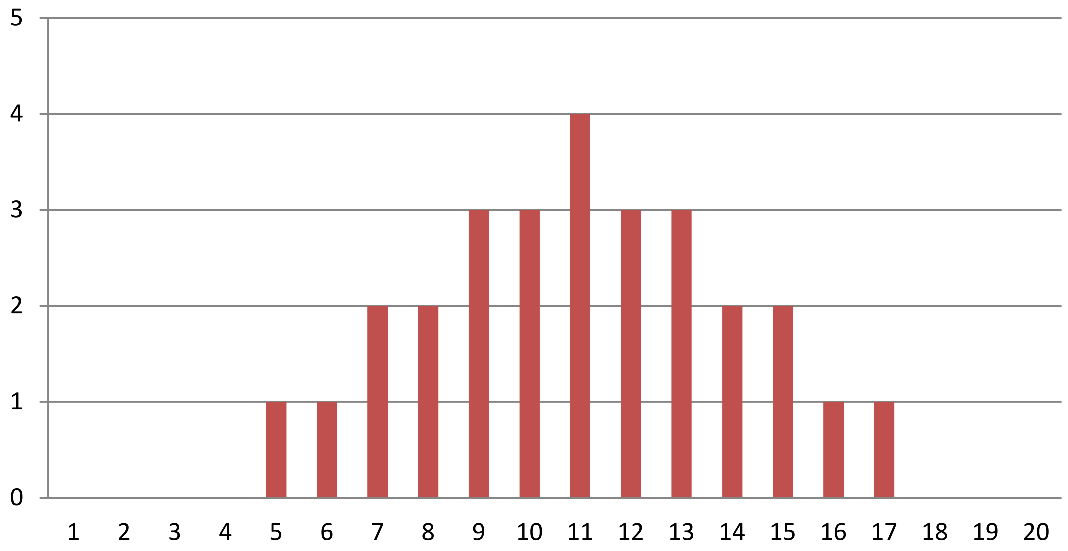
a, h	12	14	13
b, c	8	6	7
b, d	8	4	6
b, e	8	10	9
b, f	8	18	13
b, g	8	16	12
b, h	8	14	11
c, d	6	4	5
c, e	6	10	8
c, f	6	18	12
c, g	6	16	11
c, h	6	14	10
d, e	4	10	7
d, f	4	18	11
d, g	4	16	10
d, h	4	14	9
e, f	10	18	14
e, g	10	16	13
e, h	10	14	12
f, g	18	16	17
f, h	18	14	16
g, h	16	14	15

Ahora vamos a observar la distribución de frecuencias de las medias muestrales, es decir, cuántas veces se repite cada una de las medias o, lo que es lo mismo, en cuántas muestras aparece cada valor de la media. Para ello construimos una tabla de frecuencias donde el valor de las medias se ordena de menor a mayor (tal como hacemos en cualquier otra distribución de frecuencias con una variable cuantitativa):

$\bar{x}$	Frecuencia
5	1
6	1
7	2
8	2
9	3
10	3
11	4
12	3
13	3
14	2
15	2
16	1
17	1

A simple vista observamos que el valor que más repite de esta distribución (la moda) es $\bar{x}$ = 11. Algo que también podemos comprobar con la representación gráfica de los datos.

c) Tomando los valores de $\bar{x}$ como discretos, podemos construir un diagrama de barras:



En el gráfico se aprecia que las medias de las distintas muestras convergen hacia el valor de la media poblacional que hallamos en el apartado a): $\mu = 11$

Pero también podemos comprobarlo al calcular la media de las medias muestrales:

$$\mu_{\bar{x}} = \frac{\sum \bar{x}_i\, n_i}{N} = \frac{5 \cdot 1+6 \cdot 1+7 \cdot 2+8 \cdot 2+9 \cdot 3+10 \cdot 3+11 \cdot 4+12 \cdot 3+13 \cdot 3+14 \cdot 2+15 \cdot 2+16 \cdot 1+17 \cdot 1}{28} = 11$$

$$\mu_{\bar{x}} = 11$$

Podemos comprobar a través de este ejercicio que el parámetro poblacional coincide con la media de los estadísticos muestrales. Es decir, que la media poblacional coincide con la media de todas las medias muestrales:

$$\mu_{x_i} = \mu_{\bar{x}}$$

Ejercicio 3. Solución

Calculamos el tamaño de la muestra. Tenemos estos datos:

- Estimación proporciones
- Población infinita: N>100.000
- e= 5% = 0,05
- Nc=95% entonces Z= ±1,96
- P=Q=0,5 supuesto de máxima varianza al desconocerse P

$$n = \frac{Z^2PQ}{e^2} = \frac{1{,}96^2 \cdot 0{,}5 \cdot 0{,}5}{0{,}05^2} = 384{,}16 \approx 385 \text{ entrevistas}$$

Ejercicio 4. Solución

Solo cambian, respecto al ejercicio anterior, los siguientes datos:

e= 2% = 0,02
Nc=99% entonces Z= ±2,58

Nótese que en comparación con el ejercicio anterior, vamos a calcular el tamaño muestral con unas condiciones más exigentes: es menor el error (e) que estamos dispuestos a admitir, y es mayor el nivel de confianza (Nc) que exigimos a la estimación a partir de un tamaño muestral (n)

$$n = \frac{Z^2PQ}{e^2} = \frac{2{,}58^2 \cdot 0{,}5 \cdot 0{,}5}{0{,}02^2} = 4.160{,}25 \approx 4.161 \text{ entrevistas}$$

Al establecer unas condiciones más exigentes para la estimación del resultado del referéndum, el número de individuos que debe formar parte de la muestra ha aumentado considerablemente.

Ejercicio 5. Solución

N= 100.000
e= 0,05
Nc= 95,45% de forma que Z= ±2
P=Q=0,5 supuesto de máxima varianza

a) Considerando población finita:

$$n = \frac{Z^2PQN}{e^2(N-1)+Z^2PQ} = \frac{2^2 \cdot 0{,}5 \cdot 0{,}5 \cdot 100.000}{0{,}05^2(100.000-1)+2^2 \cdot 0{,}5 \cdot 0{,}5} = \frac{100.000}{249{,}9975+1} = 398{,}41 \approx 399 \text{ entrevistas}$$

b) Considerando población infinita:

$$n = \frac{Z^2PQ}{e^2} = \frac{2^2 \cdot 0{,}5 \cdot 0{,}5}{0{,}05^2} = 400 \text{ individuos}$$

Como puede comprobarse, para N = 100.000 el tamaño de la muestra (n) resulta prácticamente idéntico aplicando o no el factor de corrección para poblaciones finitas. Por eso, no es necesario su uso cuando el tamaño poblacional es igual o mayor a 100.000 individuos.

Ejercicio 6. Solución

N ≥ 100.000
e = 2,5% = 0,025

Nc= 95% $\rightarrow$ Z= ±1,96
P= 0,12 Q=0,88

$$n = \frac{Z^2 PQ}{e^2} = \frac{1{,}96^2 \cdot 0{,}12 \cdot 0{,}88}{0{,}025^2} = 649{,}08 \approx 650 \text{ individuos}$$

Ejercicio 7. Solución

N ≥ 100.000
e = 3
Nc= 95,45%
σ = 4

Debemos hallar el tamaño muestral para la estimación de una media:

$$n = \frac{Z^2 \cdot \sigma^2}{e^2} = \frac{2^2 \cdot 4^2}{3^2} = \frac{64}{9} = 7{,}1 \approx 8 \text{ individuos}$$

Con una muestra de únicamente 8 personas podemos estimar el tiempo medio en que los parados encuentran empleo. Este reducido *n* se debe fundamentalmente a que una desviación típica de tan solo 4 meses en relación con el recorrido de la variable implica que existe bastante homogeneidad en el tiempo en que los parados encuentran normalmente empleo. Por eso, solo hace falta conocer muy pocos casos para estimar el comportamiento del conjunto total.

Se propone al estudiante calcular el tamaño muestral suponiendo que la dispersión es mayor, por ejemplo, una desviación típica de un año (12 meses) exigiendo mayor precisión a la estimación: reduciendo el error a 1 mes y/o trabajando con un nivel de confianza del 98%. Compare los resultados con el tamaño muestral anterior.

Con ello podremos comprobar que el tamaño de la muestra no depende tanto del tamaño de la población como de la varianza, del error admitido y del nivel de confianza.

Ejercicio 8. Solución

Estamos ante un ejercicio de muestreo estratificado uniforme. A partir del enunciado sabemos que en cada una de las 5 provincias de Castilla-La Mancha se han realizado 968 entrevistas (*n* es igual para todas las provincias, de ahí que sea un muestreo estratificado uniforme). Consideramos cada provincia como un estrato diferente.

En total, tenemos 968 x 5 = 4.840 entrevistas.

a) El coeficiente de ponderación u_i es la relación entre:

- El peso de cada estrato poblacional sobre el total de la población (W_i)
- El peso de cada estrato muestral entre el nº total de elementos de la muestra (w_i)

$$u_i = \frac{W_i}{w_i}$$

Calculamos W_i de cada provincia a partir de la tabla de los datos poblacionales:

$$W_i = \frac{N_i}{N}$$

		W_i
Albacete	55.845	0,2178789
Ciudad Real	62.992	0,2457629
Cuenca	26.289	0,1025664
Guadalajara	29.633	0,115313
Toledo	81.553	0,3181786
Total (N)	256.312	

Ahora calculamos el w_i de cada estrato. Dado que se trata de una estratificación uniforme, todos los w_i serán idénticos, pues el tamaño muestral de cada estrato es idéntico. Si observamos la tabla de datos muestrales del enunciado, basta con que hagamos el cálculo para un estrato (una provincia):

$$w_i = \frac{n_i}{n} = \frac{968}{4.840} = 0{,}2$$

Calculamos los **coeficientes de ponderación u_i**

$$u_i = \frac{W_i}{w_i}$$

		W_i	w_i	u_i
Albacete	55.845	0,217878991	0,2	1,08939496
Ciudad Real	62.992	0,245762976	0,2	1,22881488
Cuenca	26.289	0,102566403	0,2	0,51283202
Guadalajara	29.633	0,115613003	0,2	0,57806501
Toledo	81.553	0,318178626	0,2	1,59089313
Total (N)	256.312			

Los coeficientes > 1 indican que el estrato (en este caso, las provincias de Albacete, Ciudad Real y Toledo) han sido infrarrepresentadas en la muestra, mientras que los estratos cuyo u_i <1 han sido sobrerrepresentados en la muestra.

b) Ahora debemos estimar la proporción total de activas en Castilla –La Mancha a partir de los datos de la muestra, teniendo en cuenta que se trata de un muestreo estratificado uniforme.

Para ello, obtenemos la cantidad de activas en cada una de las provincias ajustando esta frecuencia con el coeficiente de ponderación u_i. Por ejemplo, para Albacete:

Frecuencia de activas obtenido en la muestra: 447
Coeficiente u_i = 1,08939496
Frecuencia ponderada = 447 x 1,08939496 = 486,95957 activas en Albacete

Haciendo la misma operación, obtenemos la frecuencia ponderada de activas para todas las provincias. La frecuencia ponderada total de activas de Castilla-La Mancha será la suma de las frecuencias ponderadas de cada estrato:

Datos de la muestra

	Activas	u_i	Frecuencia ponderada: nº Activas x u_i
Albacete	447	1,089395	486,95957
Ciudad Real	402	1,228815	493,98363
Cuenca	406	0,512832	208,20979
Guadalajara	487	0,578065	281,51766
Toledo	445	1,590893	707,94739
Total (N)	2.187		**2178,618**

El total ponderado de activas en Castilla-La Mancha es 2178,618 (obtenemos decimales porque se trata de una ponderación).

La proporción total de activas en Castilla-La Mancha, que es lo que nos pide el ejercicio, es el resultado de dividir el total ponderado de activas entre el número de entrevistas realizadas, es decir, entre el tamaño muestral (n=4.840):

Proporción total de activas = $\frac{2178,618}{4840} = \mathbf{0,45\ \ (45\%)}$

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivar 4.0 Internacional.
La autoría de este trabajo corresponde a los siguientes profesores del Departamento de Sociología I de la UNED: Beatriz Mañas Ramírez, Alejandro Almazán Llorente y Luis Alfonso Camarero Rioja.

http://www2.uned.es/socioestadistica/Crim/Ejercicios_resueltos_Tema4_Muestreo_Tecnicas_encuesta.pdf