Muestreo estratificado

1. Para un estudio socio-económico se agrupó en estratos a todas las localidades de una región, incluyendo las deshabitadas, utilizando como criterios la altitud sobre el nivel del mar y la densidad de la población. En cada estrato se seleccionó mediante un m.a.s. a 10 unidades. Los datos sobre el número de familias residentes en las localidades observadas fueron:

Estrato	# localidades	1	2	3	4	5	6	7	8	9	10
I	1411	43	84	98	0	10	44	0	124	13	0
II	4705	50	147	62	87	84	158	170	104	56	160
III	2558	228	262	110	232	139	178	334	0	63	220
IV	14997	17	34	25	34	36	0	25	7	15	31

   Se pide:

   a) Obtener una estimación del número total de familias y su error de muestreo.

   b) Estimar la ganancia debida al uso de estratificación en términos de la varianza.

   c) Comparar la eficiencia de la afijación considerada con la óptima manteniendo constante el tamaño de la muestra.

2. Los siguientes datos muestran la estratificación de las granjas de un pais por su tamaño, incluyendo el número de granjas de cada estrato, N_h, el promedio de acres dedicados al cultivo de maíz por granja, $\bar{Y}_h$, para cada estrato, así como la cuasidesviación típica S_h en cada uno de ellos. Para un tamaño muestral de 100 granjas, calcular los tamaños muestrales de cada estrato bajo afijación proporcional y óptima. Comparar las precisiones de estos dos métodos con un muestreo aleatorio simple.

Tamaño	N_h	$\bar{Y}_h$	S_h
0-40	394	5.4	8.3
41-80	461	16.3	13.3
81-120	391	24.3	15.1
121-160	334	34.5	19.8
161-200	169	42.1	24.5
201-240	113	50.1	26
más de 240	148	63.8	35.2

3. En una población finita se consideran dos estratos de tamaños $N_1 = 400$ y $N_2 = 600$ siendo las correspondientes cuasidesviaciones típicas $S_1 = 12$ y $S_2 = 20$. Se pide:

   a) Determinar la varianza del estimador de la media, siendo el tamaño muestral $n = 200$ y la afijación proporcional.

b) Determinar para el mismo tamaño muestral anterior la afijación óptima y la varianza que tendría el estimador en ese caso.

c) Si los costes por unidad muestral son $c_1 = 225$ y $c_2 = 169$, y suponiendo que se dispone de un presupuesto total de 30000 euros y se desea afijación óptima, ¿cuál será el tamaño muestral? Determinar la varianza del estimador en esta situación.

4. Se quiere estimar la media de una población estratificada en tres estratos, de tamaños $N_1 = 3000$, $N_2 = 2000$ y $N_3 = 5000$, y cuasivarianzas $S_1^2 = 100$, $S_2^2 = 400$ y $S_3^2 = 900$, por medio de una muestra aleatoria de tamaño 120. Determinar el error de muestreo que se comete si se utiliza:

   a) afijación igual;

   b) afijación proporcional;

   c) afijación óptima.

5. Se desea estimar el número medio de activos financieros por familia de una población. Para ello se estratificó dicha población en dos estratos, el correspondiente a rentas altas, de tamaño $N_1 = 4000$ familias, y el correspondiente a rentas bajas, de $N_2 = 20000$ familias. Se piensa que una familia de renta alta posee, aproximadamente, nueve veces ms activos financieros que una de la clase baja, y que la cuasidesviación típica de cada estrato es proporcional a la raíz cuadrada de la media del estrato: $S_h = c\sqrt{\bar{X}_h}$.

   a) Con objeto de obtener la máxima precisión en la estimación, ¿cómo distribuirías entre los dos estratos una muestra de 1000 familias?

   b) ¿Y si se quisiera estimar la diferencia entre el número medio de activos financieros de los dos estratos?

6. Sea una población de tamaño N, con estratos de tamaño N_h.

   a) Si la función de coste es de la forma $C = c_0 + \sum t_h \sqrt{n_h}$, donde c_0 y t_h son valores conocidos, demostrar que, para minimizar $var(\bar{x}_{st})$ para un coste total fijo C, n_h debe ser proporcional a $\left(\frac{N_h^2 S_h^2}{t_h}\right)^{2/3}$.

   b) Encontrar los valores de n_h para una muestra de tamaño 1000, siendo $N = 100000$ y

Estrato	I	II	III
N_h	40000	30000	30000
S_h	4	5	6
t_h	1	2	4

   c) Si $C = 500$ y $c_0 = 300$, determinar el tamaño muestral y las afijaciones halladas en el apartado 6*a*).

   d) Si $n = 2968$, determinar la afijación proporcional y comparar en términos de la varianza los resultados obtenidos con los del apartado anterior.

7. En un muestreo estratificado se ha obtenido para cada estrato una muestra aleatoria simple de n_h elementos, con $n = \sum n_h$. Para estimar la media poblacional, se quiere utilizar un estimador del tipo $\hat{\bar{X}} = \sum c_h \bar{x}_h$, con $\sum c_h = 1$ y $\bar{x}_h$ la media muestral en el estrato h. Se pide:

   a) determinar el sesgo de $\hat{\bar{X}}$. ¿Qué condición deberán cumplir los c_h para que sea insesgado?

   b) calcular los c_h que minimizan la varianza del estimador.