

Muestreo. Problemas.

Diseños Bi-etápicas

Problema 1

Se quiere conocer el tiempo medio por semana que pasan delante de su televisor los clientes de una cadena de televisión por cable. Las entrevistas, que comportan otras preguntas, se harán cara a cara.

Hay 100000 abonados. Interesa un tipo de muestreo que ocasione pocos desplazamientos del entrevistador.

Se decide entrevistar a 100 personas, seleccionadas en dos etapas. Se forma 1000 UP según la zona geográfica, de forma que cada UP tenga los mismos efectivos. Se decide seleccionar 10 UP en la primera etapa y emplear una tasa de muestreo constante en la segunda etapa.

1. ¿cuál es el tamaño de cada submuestra?

En la muestra, se observan los resultados recogidos en la tabla 1.

UP	1	2	3	4	5	6	7	8	9
	20	30	20	20	10	30	10	10	20
S^2	12	16	10	10	2	20	4	4	4

2. Estimar el tiempo medio por semana y el tiempo total frente al televisor, por punto y por intervalo

Problema 2

El disco duro de un PC contiene 400 archivos y cada uno de dichos archivos contiene exactamente 50 registros. Para estimar el número medio de caracteres por registro, se decide extraer al azar 80 archivos y, después, 5 registros de cada archivo seleccionado.

Al observar la muestra así diseñada, se obtiene:

- $s_T^2 = 905000$
- La media de las m dispersiones de y en las m UP seleccionadas (media de los s_i^2) es igual a 805

Preguntas:

1. Cómo estimar sin sesgo la media de caracteres por registro
2. Cómo estimar sin sesgo la precisión del estimador empleado en 1.
3. Calculen un intervalo de confianza (95%) para la media de caracteres por registro.

Problema 3

El departamento de Sanidad de una región metropolitana quiere conocer el número de personas mayores de edad que tienen alguna discapacidad. Por esta razón, dicho departamento decide hacer una encuesta por muestreo e interrogar a 2000 personas mayores de edad escogidas al azar.

El registro de sanidad tiene censados 800 000 personas mayores de edad repartidos en 200 “zonas”, que se corresponden con áreas geográficas. Se decide efectuar el muestreo en dos etapas, aprovechando esta repartición en zonas. Se decide seleccionar 10 zonas, mediante una extracción aleatoria simple sin reposición entre las 200 zonas. A continuación, se seleccionan 200 mayores de edad en cada zona geográfica, mediante un muestreo aleatorio simple.

Se obtienen en la muestra los siguientes resultados:

Zonas seleccionadas	1	2	3	4	5	6	7	8	9	10
Número total de mayores de 70 años residentes en la zona N_i	6000	8000	6000	8000	10000	5000	5000	10000	10000	8000
Tamaño de la muestra de la segunda etapa en la correspondiente zona	200	200	200	200	200	200	200	200	200	200
Número de mayores de edad que tienen alguna discapacidad en la muestra	20	40	10	30	10	20	20	10	30	40

1. ¿Cuál es la variable de interés y cuáles son los valores que puede tomar?
 2. Estimen por punto y por intervalo el número total de personas mayores de edad con alguna discapacidad en la región metropolitana estudiada.
 3. Estimen por punto y por intervalo la proporción p de personas mayores de edad con alguna discapacidad en la región metropolitana estudiada.
- En los dos casos, se aconseja emplear un nivel de confianza del 95%.
4. ¿Cuál es el punto débil de este método de extracción de la muestra? ¿Qué modificación se podría hacer en la extracción para evitar el deterioro de la precisión?

Problema 4

Cierto fabricante de lavadoras quiere conocer el equipamiento de sus clientes en otros aparatos, en particular en lavaplatos. Tiene 100 000 clientes en su fichero de clientes. Dicho fichero viene organizado en 2000 "grupos" de 50 clientes que viven en zonas geográficas suficientemente contiguas. El fabricante pide que se haga un sondeo a 1000 clientes. Es interesante, por razones económicas, efectuar dicho sondeo por conglomerados, aprovechando la organización del fichero. Por lo tanto, se decide extraer 20 grupos de clientes, mediante una extracción aleatoria simple sin reposición entre los 2000 grupos. A continuación, se entrevistan todos los clientes de los grupos seleccionados.

Se denota T_i el número de poseedores de un lavaplatos en el grupo i . Se obtienen en la muestra los siguientes resultados:

$$\sum T_i = 124$$

$$\sum T_i^2 = 1818$$

1. Estimar por punto y por intervalo el número total de poseedores de lavaplatos entre los clientes de la marca de lavadoras.
2. Estimar por punto y por intervalo la proporción p de poseedores de lavaplatos entre los clientes. En los dos casos, emplear un nivel de confianza del 95%.

Problema 5

Un cierto periódico se distribuye en 5000 quioscos. El servicio se hace con motoristas. Cada motorista atiende a 50 quioscos, relativamente cercanos. Se desea conocer el número de quiosqueros partidarios de abrir los domingos por la tarde. La dirección del periódico encarga un estudio destinado a estimar dicho número.

Se decide observar 1000 quioscos, en 20 conglomerados, aprovechando el reagrupamiento geográfico establecido.

En los 20 conglomerados seleccionados se observa los efectivos de partidarios de abrir los domingos por la tarde

nr	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	16	18	19	20	suma
efec	10	9	12	8	34	2	7	22	10	12	8	11	10	10	32	4	5	24	8	14	252

1. Estimar el número total de quiosqueros partidarios de abrir los domingos por la tarde
2. Sea P la proporción de quiosqueros partidarios de abrir los domingos por la tarde. Estimar P .
3. Opinen sobre la buena/mala calidad de la estimación y de las razones
4. ¿Cuál habría sido la precisión de la estimación con una muestreo aleatorio simple? Calculen $\hat{V}(\hat{p})$ y el efecto del diseño DEFF.

Problema 6

Un banco tiene 39800 clientes-empresas en sus ficheros informáticos repartidos en 3980 agentes. Cada agente gestiona exactamente 10 clientes. Se desea estimar la proporción de clientes a los cuales el banco ha hecho un préstamo.

Al azar (MAS), se seleccionan 40 agentes. Para cada agente seleccionado, se observa el número de clientes que han pedido un prestamos. Se obtiene:

nr	1		40	
efec	185
Efec ²	1263

Estimar la proporción de clientes a los cuales el banco ha hecho un préstamo.

Problema 7

Un club de excursionistas cuenta con 40000 socios. Se desea conocer el número de socios que son abonados a una determinada revista.

Existe una ficha para cada socio del club, y el fichero viene clasificado por orden geográfico de manera que las zonas geográficas se siguen las unas a las otras.

Se decide hacer una encuesta a 800 socios, tomados en conglomerados de 10. Así, se ahorran gastos de transporte, dado que un encuestador puede realizar 10 entrevistas al día, por término medio, si los entrevistados viven en una área geográfica cercana.

Se pueden considerar las $N=40000$ fichas como una base compuesta por $M=4000$ conglomerados de 10 unidades cada uno.

Se selecciona, con un procedimiento ASSR, 80 conglomerados. En cada conglomerado, se observa T_i , el número de socios abonados a la revista ($0 \leq T_i \leq 10$). Se obtiene:

$$\sum T_i = 370$$

$$\sum T_i^2 = 2536$$

- Calcular un intervalo de confianza al 95% para el número total de abonados a la revista entre los socios del club
- Sea P , la proporción de socios abonados a la revista. Estimar P , por punto y por intervalo.
- ¿Cuál sería la precisión obtenida en la estimación de la proporción con un diseño ASSR?
- ¿Cómo explica la diferencia entre las amplitudes de los intervalos de confianza calculados en las anteriores preguntas?
- ¿Qué sugerencia se podría formular para mejorar la precisión de la selección bietápica empleada en la selección de la muestra?