Parcial 3 de Mostratge, 27 de mayo 2015
Duración: 1h15. Publicación de notas: martes 2 de junio
Revisión de notas: lunes 8 de junio 15h, despacho C5209

Por favor, Hojas distintas por teoría, prácticas y cada problema.

Teoría (en una hoja como mucho) (1 punto)

Expliquen, en una hoja máximo, el método de recomposición llamado de cociente. Indiquen claramente la información auxiliar requerida. Comparen con la extracción con probabilidades desiguales.

Prácticas (3 puntos)

En la práctica de diseño en conglomerados, indiquen:

La variable que ha servido a formar las UP

El tamaño de la muestra de UP escogida y el tamaño de la muestra de US

La probabilidad de inclusión de una UP. Digan por qué se obtiene dicha probabilidad.

La probabilidad de inclusión de una US. Digan por qué se obtiene dicha probabilidad.

Solución

La variable que se ha utilizado para formar las UP es GruD.

El tamaño de la muestra de UP ha sido una decisión, en general m=9 (o m=8). Se tiene que saber la decisión tomada y el porqué.

El tamaño de la muestra de US depende de la muestra. Se tiene que saber cuál es el n obtenido, diferente para cada uno. Se sabe consultando los resultados de la ejecución.

La probabilidad de una inclusión de una UP es m/M, es decir 9/37. Hay M=37 UP y se selecciona una muestra ASSR de tamaño m=9

La probabilidad de una inclusión de una US es también 9/37, dado que una vez que se ha seleccionado una UP, se seleccionan a todas las US de esta UP. Por lo tanto la probabilidad de que una US pertenezca a la muestra es igual a la probabilidad de que la UP a la cual pertenece esté seleccionada en la muestra de UP.

Problema 2 (4 puntos)

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 600 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 3 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
Número total de mayores de edad residentes en la zona N_i	6000	8000	6000
Tamaño de la muestra de la segunda etapa en la correspondiente zona	200	200	200
Número de mayores de edad que tienen alguna discapacidad en la muestra	20	40	10

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. Comenten los resultados y opinen sobre la precisión de la estimación ¿Cómo se podría mejorar dicha precisión?

Solución

N=800000 M=200

m=3 n_i=200 n=600 N_i varia según la UP (ver tabla)

1.

La variable de interés es una variable dicotómica que toma el valor 1 si el individuo interrogado presenta discapacidad, 0 si no presenta.

2.

Para facilitar los cálculos, Se construye la siguiente tabla:

zona	N_i	$\hat{p}_i$	$\hat{T}_i$	$1-n_i/N_i$	$\hat{Z}_i$
1	6000	0,10	600	0,966667	15738,7
2	8000	0,20	1600	0,975000	50170,9
3	6000	0,05	300	0,966667	8306,5
total			2500		

$$\hat{T}(Y) = \frac{M}{m}\sum_{i=1}^{m}\hat{T}_i(Y) = 166667$$

$$\hat{V}\left(\hat{T}(Y)\right) = \frac{M^2}{m}\left(1-\frac{m}{M}\right)s_T^2 + \frac{M}{m}\sum_{i=1}^{m}\hat{Z}_i$$

con $$S_T^2 = \frac{1}{m-1}\sum_i\left(\hat{T}-\hat{\bar{T}}\right)^2 = 463333$$

$$\hat{V}\left(\hat{T}(Y)\right) = 6085111111 + 4947740 = 6090058851$$
$$= (78038{,}8)^2$$

$$\textbf{IC para T:}\quad [10589{,}4\ \ 322745]$$

3. Estimación de la proporción de personas con discapacidad

$$\hat{p} = \frac{\hat{T}}{N} = \frac{166667}{800000} = 0{,}21$$

$$\textbf{IC para P:}\quad [\,0{,}013;\ 0{,}403431]$$

4. El resultado no es nada satisfactoria. Se han seleccionado pocas UP: Se debería, imperativamente, aumentar el tamaño de la muestra de UP.

Problema 1 (2 puntos)

En una ciudad grande, se desea conocer la proporción de adultos que distinguen el sabor del agua mineral del sabor del agua del grifo. Para esto, se organiza una sesión de cata a ciegas con una muestra de 100 adultos. Se supone que se puede considerar la muestra como una muestra aleatoria simple sin reposición extraída de la población de adultos de la ciudad. Además, no se han producido no-respuestas.
En la muestra, se establece, mediante una prueba adecuada, que 72 individuos discriminan realmente entre los dos tipos de agua por el sabor.

1. Estimen la proporción de consumidores que distinguen los dos tipos de agua por punto y por intervalo.

2.. Se sabe que la edad tiene una influencia sobre la capacidad a discriminar los sabores. Se tiene la información de que, entre los adultos, la proporción de menores 60 años es de 74% del total de los adultos. Como la información sobre la edad es disponible para los individuos de la muestra, se les reparte en menores de 60 años, por una parte, y con 60 años o más, por otra parte y se construye la siguiente tabla:

	< 60 años	>= 60 años
Discriminan	59	13
No discriminan	19	9
Total	78	22

Utilizar esta información para recomponer la estimación y proponer una nueva estimación **puntual** de la verdadera proporción de adultos de la ciudad que discriminan los dos tipos de agua por el sabor. Indiquen cómo se llama el método de recomposición utilizado.

Solución

1. El estimador de Horwitz-Thompson es, en este caso :

$$\hat{p} = \frac{72}{100} \cdot = 0.72$$

La estimador de la varianza de es

$$\hat{V}(\hat{p}) = \frac{1-f}{n} \cdot \frac{\hat{p}(1-\hat{p})}{n-1} =$$

Se asume que *f=n(N* es nulo

Así
$$\hat{V}(\hat{p}) = \frac{0.72(1-0.72)}{99} = 0.002036364 = (0.045)^2$$

y

$$P \in [0.63, 0.81]$$

2. El estimador post-estratificado es:

$$\hat{p}_{post} = \sum_{h=1}^{k} \frac{N_h}{N} \hat{p}_h = 0.74 \cdot \frac{59}{78} + 0.26 \frac{13}{22} \cdot = 0.71$$