

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
MAESTRÍA EN ESTADÍSTICA

TÉCNICAS DE MUESTREO

Exámen final

(Primer semestre 2017)

Prof. Luis Valdivieso

1.- a) Si se toma un MASc de tamaño 30 de una población de tamaño 500 para estimar un total de una variable y ¿cuál sería el efecto de este diseño? (2.0 puntos)

b) Indique, justificando, si es verdadera o falsa la siguiente afirmación: Si en un muestreo por conglomerados de tres etapas para una variable estadística y en una población con N unidades primarias, M_i unidades secundarias en la unidad primaria i y L_{ij} unidades terciarias en la unidad secundaria j de la unidad primaria i , se extraen n unidades primarias, m_i unidades secundarias de la unidad primaria elegida y finalmente l_{ij} unidades terciarias de la unidad secundaria seleccionada, entonces el estimador del total $\hat{\tau} = \sum_{i=1}^N \sum_{j=1}^{M_i} \sum_{k=1}^{L_{ij}} \omega_{ijk} y_{ijk} \delta_{ijk}$ es insesgado, donde $\omega_{ijk} = \frac{NM_i L_{ij}}{nm_i l_{ij}}$ y δ_{ijk} es una variable aleatoria que vale 1 si la observación de y en la i -ésima unidad primaria, j -ésima unidad secundaria y k -ésima unidad terciaria es seleccionada y 0 en caso contrario. (2.0 puntos)

2.- Una empresa textil posee 90 plantas localizadas en todo el país y desea estimar el número promedio de horas que sus máquinas de coser estuvieron paradas por reparaciones en los últimos meses. Dado que las plantas están distribuidas de manera dispersa por todo el país, ellos deciden utilizar un muestreo aleatorio por conglomerados bietápico, seleccionando primero mediante un MASs a 7 de las plantas y dentro de cada planta a aproximadamente el 20% de las máquinas de cada planta. Los datos se muestran a continuación, donde M_i es el número de máquinas por planta y m_i el número de máquinas seleccionadas por planta. Para su ayuda se muestran también la media y varianza muestral del número de horas que las máquinas estuvieron de para durante el periodo de estudio.

Planta	M_i	m_i	Tiempo de para (en horas)	Media	Varianza
1	50	10	5, 7, 9, 0, 11, 2, 8, 4, 3, 5	5.40	11.38
2	65	13	4, 3, 7, 2, 11, 0, 1, 9, 4, 3, 2, 1, 5	4.00	10.67
3	45	9	5, 6, 4, 11, 12, 0, 1, 8, 4	5.67	16.75
4	48	10	6, 4, 0, 1, 0, 9, 8, 4, 6, 10	4.80	13.29
5	52	10	11, 4, 3, 1, 0, 2, 8, 6, 5, 3	4.30	11.12
6	58	12	12, 11, 3, 4, 2, 0, 0, 1, 4, 3, 2, 4	3.83	14.88
7	42	8	3, 7, 6, 7, 8, 4, 3, 2	5.00	5.14

a) Halle el coeficiente de determinación ajustado de este diseño. Interprete. (2.0 puntos)

b) Estime el parámetro de interés tanto en forma insesgada como por el estimador de razón. Para el primero considere que la empresa posee un total de 4,500 máquinas de coser. (2.0 puntos)

c) Estime el error estándar de estimación para ambos estimadores e indique, justificando adecuadamente, ¿cuál de las dos estimaciones reportaría para este estudio? (2.0 puntos)

3.- En una ciudad hay 720 fábricas, de las cuales 10, 20 y 8 pertenecen respectivamente a los consorcios A,B y C. El ministerio de trabajo desea hacer un estudio de salud ocupacional en las fábricas de la ciudad. Dado que muchos de los indicadores a estudiarse son proporciones, el ministerio desea tomar un MASs de tamaño n de tal manera que pueda estimar cualquier proporción con un margen de error no mayor a 0.1 y un nivel de confianza del 95%

- a) ¿Cuál debería el tamaño de muestra a tomarse? (1.0 punto)
- b) Halle la probabilidad de que se seleccionen en la muestra del tamaño tomado en a) a sólo una fábrica de cada consorcio. (2.0 puntos)
- c) Si se duplicase el tamaño de muestra anterior n ¿cuál sería el máximo error de estimación que se estaría cometiendo en la estimación de cualquier proporción? (1.0 punto)

4.- Con el fin de recabar información acerca de la calidad de los servicios que brindan los hospitales públicos de una ciudad, se ha planificado tomar una muestra sin reemplazamiento de 4 hospitales con probabilidades proporcionales a la capacidad de atención (número de camas) de cada hospital. De particular interés es estimar el gasto total en suministros médicos que semanalmente demandan estos hospitales. Dada la siguiente información (en la que el gasto se conocerá sólo después de la inspección por muestreo)

Hospital	Capacidad de atención	Gasto semanal (en soles)
A	542	7,183.74
B	400	2,159.92
C	258	3,938.31
D	623	7,202.94
E	102	794.22
F	320	3,929.31
G	990	9,366.96
H	159	1,840.18

- a) Halle las probabilidades de inclusión de primer orden para cada hospital. (1.0 punto)
- b) Si mediante algún esquema ppt resultaron seleccionados los hospitales B, D, F y G ¿cuál sería la estimación pedida? (1.0 punto)
- c) Estime, en base a la muestra en b) el cuantil 0.8 de la distribución de gastos semanales en suministros médicos de los hospitales públicos de la ciudad. (1.0 punto)
- d) Utilizando el esquema de Sampford tome la muestra y obtenga la estimación pedida. (2.0 puntos)
- e) Tome ahora una muestra ppt de tamaño 4 con reemplazamiento y obtenga la estimación pedida junto con una estimación de su error estándar de estimación. (2.0 puntos)