Trabajo 1 – Teoría de muestreo

Parte I (5 puntos) Responda en máximo 80 palabras

1. Explique en sus propias palabras las diferencias entre dominio y estrato. Utilice ejemplos no discutidos durante las clases teóricas.

Si estamos analizando los datos de una encuesta de satisfacción del cliente sobre una página web o tienda online, el dominio corresponde a comercio electrónico en atención al área de aplicación y/o tipo de datos a tratar. Por otra parte, el estrato podría comprender a los clientes con compras mayores de una cantidad X, es decir se trataría de una subcategoría dentro de los datos del dominio.

En síntesis, el dominio es el campo específico de interés, mientras el estrato comprende una partición o grupo con características similares.

2. ¿Qué esperaría que ocurra con la varianza de una variable cuando se aplique un muestreo aleatorio clusterizado (MAC) en lugar de un muestreo aleatorio estratificado (MAE)? ¿Por qué?

Primero, en el MAE, la población de la muestra se divide en grupos homogéneos llamados "estratos".

Luego, en el MAC, los clusters suelen ser heterogéneos y las observaciones en un mismo cluster suelen ser similares entre sí, por lo cual, aumenta la varianza.

3. Explique, con sus propias palabras, por qué se espera que el efecto diseño (Def) en un MAE sea menor o igual que 1.

Primero, el efecto diseño está dado por la relación entre la varianza de la estimación bajo el diseño complejo y la varianza bajo el muestreo aleatorio simple. Además, se infiere que el MAE presenta grupos homogéneos la variabilidad se reduce y las estimaciones tienden a ser más precisas, por cuanto la varianza tiende a reducirse. En otros términos, se presume que la MAE es más eficiente que el muestreo aleatorio simple.

4. ¿En qué caso el muestreo aleatorio por conglomerados tiende a aumentar la precisión de las estimaciones?

Se tendrán que presentar dos condiciones, que los conglomerados sean muy heterogéneos entre sí y que los elementos de los conglomerados sean relativamente homogéneos. Puesto que si seleccionamos un número de conglomerados representativos de una muestra, y la diferencia entre estos, es probable que las estimaciones sean muy precisas.

Parte II

1. Descargue el Censo Nacional de Población Penitenciaria 2016 de la página del INEI (https://proyectos.inei.gob.pe/microdatos/). Puede descargar solo la carátula o también más módulos de dicho censo.

2. Genere una variable dicotómica con el valor de 1 para quienes cometen “Delitos contra la seguridad pública”, y 0 para el resto de casos. Llamar a esta variable “pp\_dcsp”

3. Establezca una semilla (“seed”) usando los dígitos 12345.