

Tarea Breve – Teoría de muestreo

Indicaciones Generales:

1. Formato de trabajo:

- Esta tarea se realizará en grupos armados por los propios estudiantes.
- La tarea se entregará únicamente en un archivo con los códigos utilizados para resolver las preguntas (puede ser un `.do` para Stata o un `.Rmd` para R). Los códigos deben ser claros, ordenados y suficientemente comentados para que otra persona pueda reproducirlos fácilmente.
- Los comentarios deben indicar claramente cada subsección, así como explicar los pasos seguidos.
- Solo un miembro del equipo deberá enviar por correo el “Contrato de grupo de trabajo” (disponible en Canvas) y el archivo de solución de esta tarea.

2. Evaluación:

- Se valorará la claridad, replicabilidad y orden del código.
- El informe debe ser claro y bien estructurado, con respuestas concisas.

3. Materiales de referencia:

- Se deberán utilizar los comandos revisados (por ejemplo, `sample`, `slice_sample`, `svydesign` en R; `sample` y `svyset` en STATA).

4. Fecha de entrega:

- Límite: Domingo 12/01 a las 11:59 p.m. La demora en el envío se penalizarán con cinco puntos por cada cuatro horas de tardanza. Tomen sus precauciones.

Parte I: Fundamentos del Muestreo

- 1. Asociación entre características y enfoques (2 puntos):** Relacione cada una de las siguientes características con el enfoque correspondiente (Diseño Muestral o Modelo Estadístico):
 - Asume que la población está completamente definida y conocida.
 - Los datos muestrales se modelan bajo supuestos específicos.
 - Depende únicamente del método de selección de las unidades.
 - Introduce supuestos probabilísticos sobre la variable de interés.
- 2. Evaluación de afirmaciones (2 puntos):** Indique si las siguientes afirmaciones son verdaderas o falsas. Justifique su respuesta en no más de 50 palabras:
 - En un MAS, al duplicar el tamaño de la muestra, la varianza de la media muestral se reduce a la mitad.
 - En un MAE, el uso de pesos mal especificados puede sesgar la estimación de la media poblacional.
 - En un MAE, estratos de tamaño desigual siempre conducen a estimadores menos eficientes.
 - El efecto de diseño (DEFF) solo depende del tamaño muestral, independientemente del diseño muestral empleado.

3. **Preguntas abiertas** (6 puntos): Responda brevemente (máximo 80 palabras por respuesta):
- ¿Cuál es la diferencia clave entre un diseño probabilístico y uno no probabilístico?
 - ¿Qué es un marco muestral y qué implicaciones tiene su incorrecta definición?
 - Explique cómo el muestreo por conglomerados puede reducir los costos pero aumentar la varianza.
 - ¿Por qué el muestreo estratificado puede mejorar la precisión de las estimaciones en comparación con el MAS?
-

Parte II: Ejercicio de juguete

1. **Simulación de un muestreo aleatorio simple (MAS)** (1 punto):
- Genere un conjunto de datos ficticio con las siguientes características:
 - Tamaño de población: 10,000 individuos.
 - Establezca una semilla con el número 12345.
 - Variables:
 - ingresos: distribución normal con media 2,000 y desviación estándar 500
 - edad: uniforme entre 18 y 65 años
 - estrato: tres estratos aleatorios con proporciones: 30%, 50%, 20%
 - sexo: variable categórica con valores “Hombre” y “Mujer”
 - Extrae una muestra del 5% de la población utilizando un MAS
 - Calcule la media y el error estándar de las variables `ingresos` y `edad`.
2. **Muestreo estratificado (MAE)** (2 puntos):
- Extrae una muestra estratificada del 5% de la población, asegurando que la proporción de cada estrato en la muestra sea la misma que en la población.
 - Compare la media y el error estándar de los `ingresos` obtenidos en el MAS y el MAE. Explique las diferencias.
 - Explique cómo el diseño muestral afecta la precisión y el costo de las estimaciones.
3. **Exceso de estratos** (2 puntos):
- Genere la variable `grupo_edad` considerando los siguientes umbrales: 30, 40, 50. En total, la variable incluirá cuatro categorías.
 - Considere que los estratos se construirán a partir de las variables `estrato`, `sexo` y `grupo_edad`. Si lo prefiere, puede crear una nueva variable que resuma todos los estratos extendidos, pero no es necesario.
 - Calcule el número de observaciones en cada estrato extendido y reporte si hay estratos con menos de 30 observaciones.
 - Extrae una muestra estratificada del 5% basada en los nuevos estratos.
 - Compare la media y el SE de los `ingresos` obtenidos con MAS y MAE.
 - Explique cómo un exceso de estratos puede afectar el diseño muestral, la precisión de las estimaciones y la viabilidad del muestreo.
-

Parte II: Ejercicio Aplicado

Utilice la base de datos del CENAGRO, que está disponible en Canvas.

1. Preparación (1 punto):
 - Establece una semilla con el número 12345
 - Genere las siguientes variables:
 - `rango_edad`: variable categórica que identifica los siguientes grupos etarios: de 12 a 24 inclusive, de 25 a 64 inclusive, de 65 a más.
 - `mujer`: variable dicotómica que identifique a las Unidades Agropecuarias (UAs) que son dirigidas por mujeres
2. Aplicando un MAS (2 puntos):
 - Extrae una MAS de 11,234 observaciones de la población total (base completa).
 - Calcule el intervalo de confianza de la variable `sup_total`.
 - Estime el factor de expansión de cada observación. ¿A qué corresponde la sumatoria de esta variable? ¿Por qué?
3. Mejorando eficiencia con MAE (2 puntos):
 - Extraer una MAE de 11,234 observaciones utilizando la variable `rango_edad` como estrato.
 - Declare el diseño muestral y obtenga el promedio de la variable `mujer` para toda la muestra utilizando el diseño muestral.