FACULTAD DE



**CIENCIAS SOCIALES**

**Tarea 1 – Teoría de Muestreo**

Indicaciones generales:

* Este trabajo se realizará en parejas que serán armadas por los propios estudiantes.
* Se entregará un archivo con los códigos utilizados (do-file o Rmd). Estos archivos deben estar ordenados y comentados. Los comentarios deben indicar claramente cada subsección de la Tarea, así como explicar los pasos seguidos.
* Se valorará en la calificación el orden y claridad del código. Este debe ser replicable y llegar a los mismos resultados que se reporten como respuestas a las preguntas.
* Junto con el trabajo se entregará el contrato de trabajo grupal que se encuentra al final del documento. Este debe ser firmado por todos los miembros del grupo.
* Para extraer muestras y declarar el diseño muestral, utilizar los comandos utilizados en las PDs (sample/slice\_sample y svydesign en R, sample y svyset en Stata)
* La fecha de entrega será **hasta el** **jueves 17/08 hasta las 06:00 p.m. (6 de la tarde)**. **A las entregas que se realicen entre las 06:00 p.m. y 07:00 p.m. se les descontará dos puntos.** **Luego de las 07:00 p.m. ya no se recibirán más trabajos y serán calificados con cero.**

Indicaciones de la Tarea 1

**Parte I**

1. Indique la veracidad o falsedad de las siguientes afirmaciones:
   1. En un MAS con reemplazo, si tomo muestras repetidas del mismo tamaño, la media muestral es siempre la misma.
   2. En un MAS, cuando disminuye el tamaño de la muestra, el error estándar de la media muestral también disminuye.
   3. En un MAS, es posible que la media de una muestra pequeña se acerque más a la media poblacional que la de una muestra grande.
   4. En un muestreo aleatorio por conglomerados, para un tamaño de muestra dado, existe un *trade-off* entre el número de conglomerados y el número de unidades dentro de cada conglomerado.
2. Indique a qué enfoque (diseño o modelo) corresponde cada una de las siguientes características:
   1. Asume una población finita
   2. Asume que la variable objetivo es aleatoria
   3. Asume un proceso generador de datos
   4. Asume que el componente aleatorio viene de cómo seleccionamos las observaciones
3. Responda **tres** de las siguientes preguntas. La respuesta a cada pregunta no debe pasar de 75 palabras.
   1. ¿En qué se diferencian un dominio y un estrato?
   2. ¿Qué es el factor de expansión y cómo se construye?
   3. ¿Qué esperaría que suceda con el error estándar de la media de una variable cuando en vez de realizar un MAS se realiza un MAE? ¿Por qué?
   4. ¿Qué sucede con la varianza de una variable de interés cuando se realiza un muestreo por conglomerados respecto a cuando se hace un muestreo aleatorio simple? ¿Por qué?
   5. Explique en sus propias palabras que es el efecto de diseño (DEFF)
   6. Explique en sus propias palabras en qué se diferencia un muestreo aleatorio estratificado de uno por conglomerados

**Parte II**

Utilizando la base de datos del CENAGRO que ha sido colgada en PAIDEIA, realizar los siguientes ejercicios.

1. Calcular las siguientes variables:
   1. Una variable categórica que identifique los siguientes grupos etarios: de 12 a 24 inclusive, de 25 a 64 inclusive, de 65 a más. Llamar a esta variable “rango\_edad”.
   2. Una variable dicotómica que identifique a las Unidades Agropecuarias (UAs) que son dirigidas por personas cuya lengua materna es una lengua nativa\*. Llamar a esta variable “pp\_lenguanativa”.

*\*Considerar como lenguas nativas: el quechua, el aymara, el ashaninka y a “otra lengua nativa”*

1. Establecer una semilla (“seed”) usando los dígitos del código PUCP de alguno de los miembros de la pareja.
2. Extraer una Muestra Aleatoria Simple (MAS) de 2247 observaciones de la población total (base completa). Con esta muestra, utilizar la variable “sup\_total” y obtener los siguientes estadísticos.
   1. La suma de todos los valores,
   2. La media,
   3. La fracción muestral,
   4. La corrección de población finita (fpc), , donde f es la fracción muestral
   5. La suma de desviaciones al cuadrado,
   6. La varianza de la muestra,
   7. La varianza muestral de la media,
   8. El error estándar de la media,
   9. El factor de expansión de cada observación. ¿A qué corresponde la sumatoria de esta variable? ¿Por qué?
3. Extraer una Muestra Aleatoria Estratificada (MAE) de 2247 observaciones utilizando la variable “rango\_edad” como estrato. Con esta muestra, utilizar la variable “pp\_lenguanativa” y:
   1. Obtener los mismos estadísticos de las preguntas 3.1-3.9 según la variable de estratificación (intra estrato).
   2. Declare el diseño muestral y obtenga el promedio de la variable “pp\_lenguanativa” para toda la muestra utilizando el diseño muestral
   3. Discuta como pueden utilizarse los resultados obtenidos en 4.1 para poder calcular el promedio de la variable “pp\_lenguanativa” para toda la muestra.

**Contrato de grupo de trabajo**

**Nombres completos de cada miembro:**

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Objetivos y fechas**

Rellenar con las tareas que consideren necesarias para completar el trabajo grupal. Asignar un encargado/a, fecha de entrega y si se cumplió con la tarea asignada.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tareas** | **Encargado/a** | **Fecha de entrega** | **¿Se cumplió?** |
| 1. |  |  |  |
| 2. |  |  |  |
| 3. |  |  |  |
| Añadir según sea necesario |  |  |  |

**Sobre la relación entre miembros del grupo**

Normas grupales: Consideramos importantes las siguientes actitudes y comportamientos y nos esforzaremos en mantenerlas a lo largo del trabajo grupal.

|  |
| --- |
|  |

Toma de decisiones: Tomaremos decisiones de la siguiente manera:

|  |
| --- |
|  |

**Reglas de comunicación**

Los siguientes medios serán utilizados para coordinar los aspectos relacionados al trabajo grupal. Si el grupo tendrá reuniones de grupo (presenciales o virtuales), especificar las fechas de reunión.

|  |
| --- |
|  |

**Resolución de conflictos**

Especificar cómo el grupo lidiará con los conflictos que pueda haber o situaciones en los que no se cumpla este contrato.

|  |
| --- |
|  |

**Firmas**

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Al escribir nuestros nombres completos, aceptamos la responsabilidad de completar las tareas del trabajo grupal.