

Trabajo 1 – Teoría de muestreo

Indicaciones Generales:

1. Formato de trabajo:

- Este trabajo se realizará en grupos armados por los propios estudiantes.
- El trabajo 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 (vfuentes@pucp.edu.pe) el “Contrato de grupo de trabajo” (disponible en Canvas) y el archivo de solución del trabajo.

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 19/01 a las 11:59 p.m. La demora en el envío se penalizarán con cinco puntos por cada cuatro horas de tardanza.

Parte I (5 puntos)

Responda en máximo 80 palabras **dos** de las siguientes preguntas

1. Explique en sus propias palabras las diferencias entre dominio y estrato. Utilice ejemplos no discutidos durante las clases teóricas.
2. ¿Qué esperaríamos 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?
3. Explique, con sus propias palabras, por qué se espera que el efecto diseño (*deff*) en un MAE sea menor o igual que 1.
4. ¿En qué caso el muestreo aleatorio por conglomerados tiende a aumentar la precisión de las estimaciones?

Parte II (5 puntos)

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.

4. Extraiga una muestra estratificada y por conglomerados que contenga entre el 3% y 5% de la población (toda la base). Justifique la elección de las variables para los estratos y conglomerados. Discuta también, de ser el caso, qué otras variables que podrían haberse utilizado para los estratos o conglomerados.
5. Declare el diseño muestral completo y calcule las siguientes estadísticas (deben ser representativas):
 - a. El porcentaje de personas que cometió “Delitos contra la seguridad pública”
 - b. El total de personas que cometió “Delitos contra la seguridad pública”

Parte III (5 puntos)

Explique el objetivo, la población objetivo, macro muestral, el diseño muestral de la encuesta y las unidades de muestreo. Luego, declare el diseño muestral (**completo**). Explique los niveles de inferencia, construcción de pesos y tasa de no respuesta. Para esto, se recomienda revisar la ficha técnica de la encuesta.

Grupo	Encuesta	Link de acceso
1	Encuesta del Impacto del COVID-19 en las Empresas de Lima Metropolitana 2020	https://www.gob.pe/institucion/inei/informes-publicaciones/3353970-encuesta-del-impacto-del-covid-19-en-las-empresas-de-lima-metropolitana-2020
2	Encuesta sobre Transición de la Escuela al Trabajo 2012	https://www.gob.pe/institucion/inei/informes-publicaciones/3364206-encuesta-sobre-transicion-de-la-escuela-al-trabajo-2012
3	Encuesta Mensual de Servicios 2024	https://www.gob.pe/institucion/inei/informes-publicaciones/3364713-encuesta-mensual-de-servicios-2024
4	Encuesta Nacional de Satisfacción de Usuarios en Salud 2016	https://www.gob.pe/institucion/inei/informes-publicaciones/3355216-encuesta-nacional-de-satisfaccion-de-usuarios-en-salud-2016
5	Encuesta Permanente de Empleo Nacional – EPEN 2024	https://www.gob.pe/institucion/inei/informes-publicaciones/3880733-encuesta-permanente-de-empleo-nacional-epen-2024
6	Segunda Encuesta Mundial a la Comunidad Peruana en el Exterior 2021*	https://www.inei.gob.pe/media/MenuRecursivo/publicaciones-digitales/Est/Lib1868/libro.pdf
7	Encuesta Nacional Sobre Relaciones Sociales - ENARES 2024	https://www.gob.pe/institucion/inei/informes-publicaciones/3354732-encuesta-nacional-sobre-relaciones-sociales-enares-2024
8	Encuesta Habilidades al Trabajo en el Perú 2017	https://www.gob.pe/institucion/inei/informes-publicaciones/3355072-encuesta-habilidades-al-trabajo-en-el-peru-2017
9	Encuesta Nacional de Programas Presupuestales - ENAPRES 2024	https://www.gob.pe/institucion/inei/informes-publicaciones/3364568-encuesta-nacional-de-programas-presupuestales-enapres-2024
10	Encuesta Nacional de Lectura - ENL 2022	https://www.gob.pe/institucion/inei/informes-publicaciones/3364245-encuesta-nacional-de-lectura-enl-2022
11	Encuesta Nacional Continua de Victimización a Empresas - 2018	https://www.gob.pe/institucion/inei/informes-publicaciones/3354961-encuesta-nacional-continua-de-victimizacion-a-empresas-2018
12	Encuesta Nacional Especializada sobre Discapacidad 2012	https://www.gob.pe/institucion/inei/informes-publicaciones/3364195-encuesta-nacional-especializada-sobre-discapacidad-2012
13	Encuesta Económica Anual - EEA 2024	https://www.gob.pe/institucion/inei/informes-publicaciones/3881658-encuesta-economica-anual-eea-2024
14	Encuesta Nacional de Uso del Tiempo – ENUT 2024	https://www.gob.pe/institucion/inei/informes-publicaciones/5920191-encuesta-nacional-de-uso-del-tiempo-enut-2024

*/ No encontré la documentación, pero la información se encuentra cubierta a lo largo del documento

Parte IV (5 puntos)

Responder uno de los siguientes bloques de preguntas (elija entre Bloque A o Bloque B). Prestar atención al límite de palabras de cada pregunta.

BLOQUE A

Pregunta 1 (Palabras: 200)

Suponga que usted trabaja en el Área de Estudios del Ministerio de Educación. En su área quieren medir el nivel de bienestar psicológico en estudiantes de educación secundaria en Lima. Para ello, el psicólogo especialista sugiere utilizar como instrumento la Escala de Bienestar Psicológico de Ryff. A usted le piden encargarse del diseño muestral de la encuesta a levantar. Explique cuál sería el marco muestral que usaría y por qué. Comente la calidad del marco muestral elegido (incluya los conceptos de población objetivo, población encuestada). Además, explique cuál sería el diseño muestral que usted sugeriría para la encuesta y por qué. Comente, de manera breve, las ventajas y desventajas del diseño elegido frente a los otros. Finalmente, de manera sencilla e intuitiva, comente el tamaño de muestra que elegiría (n) y a qué factores responde el tamaño propuesto.

Pregunta 2 (Palabras: 300)

En el Ministerio de Desarrollo e Inclusión Social (Midis), le han encargado a Ud. como muestrista principal diseñar una encuesta probabilística para medir la prevalencia de vulnerabilidad a la pobreza (variable dicotómica) en todas los departamentos del Perú.

Su supervisor, quien es un muy buen *policymaker* pero sin ninguna formación cuantitativa ni de muestreo, se enteró del libro de Dillman (2014) y de las tablas de cálculo de tamaño de muestra. Él está solicitándole que justifique sus decisiones metodológicas pues argumenta que “como máximo se necesitaría una muestra de tamaño $n = 1067$ como se observa en la figura 3.5 de libro”.

Ud. ha planteado que se levante un muestreo complejo (estratificado y biétapico) para lograr el objetivo. Explique en **castellano sencillo y de forma intuitiva** por qué necesitaría (o no) una muestra de mayor tamaño. Incluya en su respuesta el concepto de tamaño de muestra efectivo, correlación intracluster y efecto de diseño. Recuerde que su supervisor no maneja estos temas.

BLOQUE B

Pregunta 1 (Palabras: 300)

Suponga que usted es ascendido como muestrista principal en el Ministerio de Salud (Minsa). Su primera tarea es terminar de definir el diseño de una encuesta probabilística para medir el nivel de desnutrición en niños de educación primaria. El anterior muestrista había sugerido un diseño por conglomerados donde los conglomerados sean los distritos. Discuta si esta opción le parece la más idónea. ¿Qué beneficios y/o perjuicios traería hacerlo de dicho modo? Además, si usted considera que hay otra opción, explique cuál sería y por qué. Incluya los conceptos de UPM y USC.

Suponga, además, que su superior le dice que han aumentado el presupuesto para el estudio, por lo cual pueden incrementar el tamaño de muestra calculado por el anterior

muestrista. En ese sentido, le consulta si resulta más conveniente aumentar los conglomerados o más bien el número de estudiantes dentro de cada conglomerado. Argumente su respuesta.

Discuta las ventajas y/o desventajas del diseño planteado por el anterior muestrista (diseño por conglomerados) frente a un diseño por estratos. De optar por un diseño por estratos, cuáles serían las ventajas y desventajas frente al diseño inicialmente planteado. Incluya conceptos como homogeneidad, heterogeneidad, efecto de diseño.

Pregunta 2 (Palabras: 200)

Un hacedor de políticas públicas que cuenta con conocimientos intermedios de estadística le pide que utilice la Encuesta Nacional Agropecuaria más reciente para obtener estadísticas distritales de acceso a asistencia técnica en agricultores de Cajamarca. Estos datos serán utilizados en el diseño de un programa público de apoyo a los productores. Explíquelo al funcionario si esto es posible y por qué. En caso no sea posible, propóngale qué otra fuente de datos podría usarse y cuáles serían sus ventajas y desventajas.