

Proyecto final

Muestreo II

noviembre

Objetivos

El objetivo de este trabajo es crear un sistema de ponderadores w_i para todas las unidades elegibles respondientes, siguiendo los criterios vistos en la asignatura.

En base a dicho sistema de ponderadores w_i se computarán estimaciones para distintos parámetros θ junto con una medida de la calidad de las mismas. Dichos parámetros pueden ser calculados, ya sea, a nivel de toda la población, como para distintos dominios/áreas de estimación.

Sobre el trabajo

El informe final puede ser entregado como un Rmarkdown o en un documento (word, latex, etc.) junto con el script en R que replica las medidas de resumen y las visualizaciones del documento.

La entrega del trabajo se realizará por el EVA y deberá ser entregado por parte de un solo integrante del grupo.

El trabajo tiene un peso del 40% en la nota total del curso.

Introducción

- **Condición de elegibilidad:** son elegibles para participar de la encuesta todas las personas que residen en viviendas particulares en cualquier parte del país.
- **Diseño muestral:** La muestra de personas, fue seleccionada bajo un diseño muestral aleatorio, estratificado, por conglomerados y en un etapa de selección. Los conglomerados son los hogares y todas las personas dentro del hogar son censadas (a priori). Los estratos de muestreo son los siguientes: 1 [Montevideo Bajo], 2 [Montevideo Medio Bajo], 3 [Montevideo Medio], 4 [Montevideo Medio Alto], 5 [Montevideo Alto], 6 [Zona Metropolitana], 7 [Interior Norte], 8 [Costa Este], 9 [Litoral Norte], 10 [Litoral Sur], 11 [Centro Norte] y 12 [Centro Sur]

Descripción de las variables

- **id_hogar** = variable indicadora del hogar
- **estrato** = estrato del diseño
- **dpto** = departamento de residencia del hogar
- **sexo** = sexo de la persona, donde 1 = Hombre y 2 = Mujer
- **edad** = edad de la persona en años cumplidos
- **ingresos** = ingreso de la personas por todo concepto
- **ocupado** = variable indicadora si la persona es ocupada/empleada
- **desocupado** = variable indicadora si la persona esta desocupada

- **pet**= variable indicadora si la persona pertenece a la población en edad de trabajar (14 años o más)
- **activo**= variable indicadora si la persona se encuentra activa (empleada o desocupada)
- **pobreza**= variable indicadora si la persona se encuentra por debajo de la Línea de pobreza. Recordar que una persona es pobre o no, dependiendo si su hogar lo es o no.
- **w0**= ponderador original o basado en el diseño, es decir, $w_{0i} = \pi_i^{-1}$
- **R**= variable indicadora si la persona respondió o no.

Naturaleza de las variables

- las variables **id_hogar**, **estrato** y **dpto** pertenecen al marco muestral F
- las variables **sexo** y **edad** son conocidas para todas las personas incluidas en la muestra original s
- las variables de interés son las de mercado laboral (**ocupado**, **desocupado**, **activo** y **pet**), **ingresos** y **pobreza**.

Parte 1

- Calcular las estimaciones puntuales (junto con el SE, CV, y el deff) de la tasa de desempleo, proporción de personas pobres y el ingreso promedio, utilizando los ponderadores originales. Bajo que esquema de no respuesta la estrategia para el cómputo de dichas estimaciones son correctas?

Explicar los resultados obtenidos.

- Realizar un breve análisis de la tasa de respuesta (ponderada y no ponderada). Explicar brevemente los resultados obtenidos.

Parte 2

Asumiendo que la no respuesta es MCAR se pide:

- Realizar un ajuste por no respuesta por medio de clases o post-estratos de no respuesta (NR) utilizando **únicamente información del marco muestral**.

Explique de forma detallada el enfoque utilizado, la construcción de las clases de NR y calcule el **deffK** a nivel global.

Compute los ponderadores ajustados por no respuesta y realice las estimaciones de los parámetros del punto 1. Compare los resultados obtenidos.

- Estime las propensiones simples de responder $\hat{\phi}_i$ utilizando uno de los siguientes algoritmos: logit, árbol de clasificación o random forest; utilizando información del marco muestral y de las covariables x conocidas para los respondientes y no respondientes.

Calcule el **deffK** a nivel global.

Compute los ponderadores ajustados por no respuesta utilizando las propensiones de NR a nivel de unidad y realice las estimaciones de los parámetros del punto 1. Compare los resultados obtenidos.

- Realice un ajuste por no respuesta creando clases o post-estratos de no respuesta utilizando las propensiones de respuesta computadas en el punto anterior.

Explique brevemente la construcción de las clases de NR y el ajuste elegido a nivel de dichas clases.

Calcule nuevamente el **deffK** a nivel global y realice las estimaciones de los parámetros del punto 1. Compare los resultados obtenidos.

Parte 3

En base a los resultados obtenidos en la parte 2, seleccione **únicamente** uno de los tres ajustes realizados para el computo de los ponderadores ajustados por no respuesta $w_i^{nr} = w_{i0} \times nr_i$.

Utilizando los ponderadores w_i^{nr} , calibre los ponderadores a las proyecciones de población utilizando como variables auxiliares la edad, sexo y departamento; y utilice la post-estratificación completa o incompleta (con o sin truncamiento) de los factores de ajuste g_i .

Observación: en el caso de tener ponderadores influyentes, considere si es necesario o no realizar un trimming de los ponderadores finales $w_i = w_{i0} \times nr_i \times g_i$

Parte 4

Una vez determinado los ponderadores finales realice las estimaciones de los parámetros del punto 1 y compare los resultados obtenidos. Comentar brevemente las diferencias obtenidas respecto a las estimaciones realizadas en el punto 1

Calcular los SEs, CV e IC al 95 % de la proporción de personas pobres, ingreso promedio (para las personas empleadas mayores de 25 años) y la tasa de desempleo. Dichas estimaciones, se debe computar para el total país y a nivel departamental.

El computo de los SEs debe ser llevado a cabo utilizando el método del ultimo conglomerado y utilizando Bootstrap Rao-Wu, fijando como mínimo 500 réplicas. Compare los SEs obtenidos bajo los distintos métodos utilizados para la estimación de los SEs.

Observación: es necesario que para el computo de los SEs, independientemente del método utilizado, se tenga en cuenta el ajuste por calibración.