

Tarea Individual III

Ciencia de Datos con R

Introducción

En esta actividad vamos a trabajar con datos provenientes de una muestra aleatoria compleja: la Encuesta Continua de Hogares (ECH) que realiza el Instituto Nacional de Estadística (INE) de Uruguay. El objetivo principal es realizar estimaciones de distintos parámetros poblacionales para el interior del país (es decir, excluyendo Montevideo), ya sea a nivel de toda la población o para distintos subconjuntos o dominios de interés (por ejemplo, por departamento). El análisis se realizará utilizando el paquete `srvyr`, que permite especificar diseños muestrales complejos de forma sencilla y realizar estimaciones válidas considerando los estratos, conglomerados (UPMs) y los ponderadores provistos en la encuesta. Además de estimar los parámetros, deberán incluir sus respectivas medidas de calidad (como errores estándar, coeficientes de variación, intervalos de confianza y efecto de diseño).

Archivos de Datos

- `P_2019_Terceros.sav`: microdatos individuales de la ECH 2019.
- `ESTRATO_UPM_ECH2019.sav`: contiene la información del diseño muestral.

Preparación inicial

```
library(tidyverse)
library(srvyr)

library(haven)
library(dplyr)
library(srvyr)
s = read_sav('P_2019_Terceros.sav')
s = s %>%
  filter(dpto != 1) %>% # Excluye Montevideo
```

```

transmute(
  numero,
  nper,
  sexo = e26,
  edad = e27,
  ingreso = PT1,
  ocupado = case_when(pobpcoac == 2 ~ 1, TRUE ~ 0),
  desocupado = case_when(pobpcoac >= 3 & pobpcoac <= 5 ~ 1, TRUE ~ 0),
  pet = case_when(edad >= 14 ~ 1, TRUE ~ 0),
  activo = ocupado + desocupado,
  pobreza = pobre06,
  w = pesoano
)

info_design = read_sav('ESTRATO_UPM_ECH2019.sav')

s = s |> left_join(info_design)

```

Parte 1 Definir el diseño muestral

```

design_srvyr <- s %>%
  as_survey_design(
    ids = upm_fic,
    strata = estrato,
    weights = w,
    nest = TRUE
  )

```

Parte 2: Tasa de Desempleo

- Estimar la tasa de desempleo para el total del interior.
- Incluir IC al 95% y CV.
- Repetir la estimación segmentada por departamento.
- Interpretar resultados generales y departamentales.

Parte 3: Proporción de personas bajo la línea de pobreza

- Estimar proporción total y por departamento.
- Incluir IC y CV.

- Comentar diferencias relevantes.

Parte 4: Ingreso promedio por departamento (mujeres 30–45 años)

- Calcular ingreso medio por departamento.
- Incluir IC y CV.
- Evaluar confiabilidad según criterios (por ejemplo, $CV < 20\%$).

Parte 5: Proporción de desocupados por tramo de edad

- Definir tramos: 14–24, 25–34, 35–49, 50+. Tener en cuenta que la edad de trabajar en Uruguay es 14 años o más.
- Estimar proporción en cada tramo.
- Visualizar con `ggplot2` y `geom_errorbar`

Parte 6: Conteo de personas de 18 años o más por departamento

- Usar `survey_count()` para obtener totales por dpto.
- Interpretar resultados para al menos tres departamentos.

Nota: se espera que acompañen cada resultado con comentarios interpretativos breves, considerando tanto el valor estimado como su precisión estadística (IC, CV, deff).

Entrega

- **La fecha limite de entrega es el 14 de julio de 2025.**
- Las respuestas deben estar en este mismo archivo `.qmd`, el contenido deber ser completamente reproducible, es decir, cada `chunk` debe de funcionar sin errores para poder replicar los resultados.
- No se aceptan archivos `.Rmd` o `.R` para la entrega. Solamente subir al repositorio el archivo `.qmd` con las respuestas.
- Cada respuesta del ejercicio debe estar en el chunk correspondiente, no borrar la etiqueta del chunk `#| label: ejercicio_XX`.
- Puede realizar pasos intermedios los que sean necesarios dentro del chunk pero debe de respetar el nombre del objeto final en el caso que se indique.

- Los gráficos deben ser guardados en objetos y luego impresos en el caso que se indique que lo almacenen en un objeto. En el caso que no se indique, pueden ser impresos directamente.
- Para comenzar la tarea deben de ir al siguiente link: [GitHub Classroom](#). Una vez allí les va a pedir que indiquen su cuenta de GitHub y luego les va a crear un repositorio en su cuenta. Una vez creado el repositorio, deben de clonar el repositorio en su computadora y abrirlo con RStudio