

Indicaciones Proyecto 4

Rodrigo Zepeda

Noviembre 2021

Objetivo

El objetivo de este proyecto es analizar y presentar resultados, a través de los estimadores vistos en clase una encuesta aleatorizada por clusters, multietápica de carácter público. Para ello se sugiere analizar una encuesta [del INEGI](#), [del DANE](#), [del INE español](#) o [Statistics Canada](#).

NOTA Se busca analizar una encuesta no un censo. En la nota metodológica de la ENCUESTA debe decir que es aleatorizada por clusters y multietápica. Casi todas las *Nacionales* del INEGI lo son.

Fecha de entrega

5 diciembre

Indicaciones

1. El proyecto debe ser entregado en pdf.
2. El proyecto requiere realizar al menos 3 gráficas distintas donde se describa la muestra y 3 gráficas (puede ser el mismo tipo que las otras) donde se use la muestra para describir la población y colocarlas en un pdf. Ejemplo: puede usarse una gráfica de barras para describir la proporción de hombres/mujeres en la muestra y *exactamente el mismo tipo de gráfica* (incluso puede ponerse dentro de la misma) una que describa la proporción de hombres/mujeres que se deducen en la población a partir de la muestra. Esta última debe estar ponderada por los π_k correspondientes.
3. El pdf debe explicar claramente cómo se realizó el muestreo explicando las definiciones *a alguien que no sabe estadística*. Por ejemplo si la nota metodológica de la encuesta dice que es un panel multietápico es necesario definir *panel* (como una encuesta que se repite de manera periódica) y *multietápico* que, en este caso, puede ser que fueron hasta las casas seleccionadas en calles aleatorias (primera etapa) y ahí muestrearon dentro de casa a personas (segunda etapa). Específicamente los conceptos que uses en tu pdf debes poder explicarlos.
4. El proyecto puede escribirse en R, Python, Julia, Matlab, Octave, C, C++, Java, Javascript, LISP ó Prolog. En caso de no hacerse en R especificar la versión (ej Python 2 vs Python 3) y el compilador (ej cmake vs gcc).
5. Editar el código usando las [guías de estilo de Google](#) o bien la de [Matlab](#) para el caso de Matlab/Octave, [la de Prolog](#) ó la de [Julia](#) en sus respectivos casos.
6. El proyecto debe incluir también el análisis de 3 estimadores *poblacionales* distintos con sus intervalos de confianza al % que tú decidas. Debes tomar en cuenta el diseño muestral. En particular puedes hacer cualquiera de estos:
 - a. Calcular e interpretar un total con su intervalo.
 - b. Calcular e interpretar una media con su intervalo.
 - c. Calcular e interpretar una varianza con su intervalo.

- d. Calcular e interpretar una proporción con su intervalo.
 - e. Calcular e interpretar un cuantil (incluyendo mediana) con su intervalo.
 - f. Calcular e interpretar cualquiera de los estadísticos vistos en la primera fase del curso (como el MAD, el riesgo relativo, un OR, etc) con su intervalo.
7. Puedes usar referencias o código de cualquier parte de Internet. Sólo ¡CÍTALO! y asegúrate de explicar lo que hace hasta donde lo entiendes. Ejemplo: `este código entiendo que accede a Twitter y descarga Tweets elegidos por número de Tweet de manera uniforme a partir del 2020.`
8. Se sugiere usar el paquete [survey](#) de R