

EJERCICIO PRÁCTICO 10: REMUESTREO Y MÉTODOS ROBUSTOS

CONTEXTO

Hasta ahora hemos discutido dos alternativas que nos permiten trabajar con datos que no necesariamente cumplen con las condiciones que impone la teoría de la inferencia estadística: la transformación de datos y los métodos no paramétricos. Sin embargo, ambas tienen algunas limitaciones, por lo que no siempre son las más adecuadas, por lo que el remuestreo y los métodos robustos son otras buenas opciones a tener en cuenta.

OBJETIVO DE APRENDIZAJE

Aplicar remuestreo y métodos robustos para lidiar con datos problemáticos utilizando el ambiente R.

ÉXITO DE LA ACTIVIDAD

1. El equipo es capaz de plantear preguntas de investigación interesantes que pueden ser respondidas con los datos de la encuesta Casen 2017.
2. El equipo es capaz de responder las preguntas de investigación utilizando los métodos estudiados para abordar datos problemáticos.

ACTIVIDAD

1. El equipo obtiene los datos (Datos-Casen-v2.xls) y revisan las columnas disponibles según la descripción en el libro de códigos de la encuesta. Es importante notar que en esta encuesta hay datos de **carácter colectivo** sobre “el hogar” del entrevistado, pero también hay datos de **carácter individual**, que se refieren “al jefe o la jefa de hogar” (no del entrevistado).
2. El equipo propone preguntas de investigación **originales** que involucren:
 - a. La comparación de las medias de dos grupos independientes
 - b. La comparación de más de dos medias independientes.Algunos ejemplos (que no pueden ser ocupados) son:
 - a. El ingreso per cápita (ytotcorh / numper) en la Región Metropolitana (region) es el mismo entre hombres y mujeres (sexo) no heterosexuales (r23).
 - b. El ingreso *per cápita* es similar en las cuatro macro zonas (norte grande, norte chico, central, sur y austral).
 - c. La situación en que las familias chilenas ocupan su vivienda (v13) tiene relación con el nivel educacional (educ) del jefe o la jefa del hogar.
3. El equipo fija una semilla y realiza un muestreo aleatorio de los datos ($250 < n < 500$).
4. El equipo analiza los datos para ambas preguntas utilizando remuestreo.
5. El equipo analiza los datos para ambas preguntas utilizando métodos robustos.