



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE GOBIERNO
GOB3032 - ANÁLISIS DE DATOS I

Tarea 3

Análisis de Datos I

13 de mayo 2024

Instrucciones

Se espera que entregue sus respuestas en un archivo pdf en conjunto con el código (script) utilizado en el buzón de canvas designado hasta el día **19 de mayo a las 23:59**.

Considere un extracto de las respuestas de la Encuesta de Caracterización Socioeconómica CASEN de 2017 adjunta en el archivo "casen.csv" con las siguientes variables

Variables	Descripción
region	Número de la región
sexo	1 (Hombre), 2 (Mujer)
edad	Edad en años
e_civil	Estado civil
o1	Trabajó al menos una hora
o2	Actividad de al menos una hora
o3	Ausencia temporal al empleo
o6	Búsqueda de trabajo
o25a_hr	Horas que tarda en viaje al trabajo
o25a_min	Minutos que tarda en viaje al trabajo
o25b	Veces a la semana en que realiza el viaje

Tabla 1. Variables a utilizar de la encuesta CASEN 2017.

Pregunta 1

- a. Presente en una tabla de calidad los estadísticos descriptivos del tiempo de viaje al trabajo. Ahora filtre la base de datos para la región Metropolitana y presente los mismos estadísticos. Compare y comente al respecto.

Hint 1: El tiempo de viaje al trabajo corresponde a la suma del registro de horas y minutos.

Hint 2: Una tabla de calidad presenta título, numeración y fuente de los datos.

- b. ¿Cuál es la probabilidad de que siendo mujer y realizando el viaje al menos 5 veces a la semana, la persona encuestada se demore una hora o más en llegar al trabajo?

Pregunta 2

- a. Presente una tabla de calidad de frecuencia relativa de las personas ocupadas y desocupadas de la muestra.

Hint 1: Las personas ocupadas y desocupadas se estiman en base a las respuestas de 01, 02, 03 y 06 según el libro de códigos.

- b. Estime los estadísticos descriptivos de edad distinguiendo si son ocupados o desocupados y reportelos en una tabla de calidad.

Bonus

- a. Presente en un histograma la distribución de tiempos de viaje de las personas ocupadas de la muestra.
- b. Estime la edad sobre la que se encuentra el 10% superior de la muestra en la distribución de tiempos de viaje de a).