

# Inferencia Estadística: Tarea 1

## Análisis Descriptivo

Fecha de entrega: 5 de septiembre

1. (2 puntos) Considera la base de datos *AlumnosInferencia.xlsx* y realiza lo siguiente:
  - (a) Para cada una de las columnas especifica el tipo de variable y la escala de medición.
  - (b) Realiza un análisis descriptivo de la variable *promedio*. Para ello deberás presentar y analizar las medidas numéricas y gráficas adecuadas.
  - (c) ¿Se puede pensar que existe diferencia entre los promedios de los dos turnos? Justifica adecuadamente tu respuesta.
  - (d) ¿Se puede pensar que existe alguna diferencia de promedios entre géneros? Justifica adecuadamente tu respuesta.
  - (e) ¿Se puede pensar que existe alguna diferencia entre promedios de las diferentes carreras? Justifica adecuadamente tu respuesta.
2. (2 puntos) En el archivo *cucos.txt* se presentan 50 mediciones (en mm) de huevos de cucos. El objetivo es realizar un análisis descriptivo de los datos. Para esto se requiere:
  - (a) Obtener la media, la desviación estándar y cualquier otro tipo de medida numérica que consideres oportuna (e.g. curtosis, coeficiente de asimetría, cuantiles, rango, etc.)
  - (b) Obtén el histograma y el diagrama de caja y bigotes.
  - (c) ¿Qué puedes decir de estos datos?
3. (2 puntos) Considera la base de datos *pesos.txt*, la cual contiene los pesos registrados de 985 varones de Ciudad Universitaria. Realiza un análisis descriptivo de los datos. ¿Consideras que los pesos se pueden asumir que vengan de una distribución normal?
4. (2 puntos) En un ensayo clínico se seleccionaron 1000 voluntarios mexicanos y se registró su grupo sanguíneo. Los datos están contenidos en el archivo *tipoSangre.txt*, donde los valores de la variable *tipo* corresponden a los siguientes grupos sanguíneos:  $1 = O^+$ ,  $2 = A^+$ ,  $3 = B^+$ ,  $4 = AB^+$ ,  $5 = O^-$ ,  $6 = A^-$ ,  $7 = B^-$  y  $8 = AB^-$ . Considerando la variable *tipo* como una variable categórica realice lo siguiente:

- (a) Calcule la tabla de frecuencias absolutas y relativas.
  - (b) Construya la gráfica de barras.
  - (c) Obtenga la gráfica de pie/pastel/dona.
  - (d) ¿Qué se puede decir sobre la distribución de los tipos de sangre? ¿Puede dar una aproximación para la proporción de la población mexicana que tiene grupo sanguíneo  $B^+$ ?
5. (2 puntos) Para reducir los costos del proceso productivo en una empresa de manufactura de gimnasios se midieron los tiempos de operación de 120 empleados. La información se encuentra en el archivo *tiemposOperacion.txt*. Realiza un análisis descriptivos completo sobre estos datos, considerando medidas numéricas así como gráficas de interés Haz especial énfasis en la forma de la distribución y escribe tus conclusiones y realiza posibles hipótesis de los datos.

## Actividades de DataCamp

Realizar los siguientes cursos de manera individual

1. *Introduction to R*
2. *Introduction to the Tidyverse*
3. *Introduction to Importing Data in R*