

# UdeA

## **Probabilidad y Estadística** **Proyecto 1: Estadística Descriptiva.** **2025-1**

Para evaluar el beneficio de implementar prácticas estandarizadas para la medición de la temperatura corporal elevada (EBT) con termógrafos infrarrojos (IRTs), se realizó un estudio clínico con más de mil sujetos. Se midieron las temperaturas orales de los sujetos y se capturaron imágenes térmicas faciales con dos IRTs evaluados. A partir de las imágenes térmicas, se extrajeron temperaturas de diferentes ubicaciones en el rostro, y se construyó una base de datos abierta disponible en este [enlace](#). Toda la información en estos archivos ha sido anonimizada.

La información de este estudio será utilizada en este proyecto para realizar diferentes análisis estadísticos.

Se solicita ingresar al enlace para leer la descripción de la base de datos y entender su organización y el significado de las variables medidas.

El estudio se realizó con dos temperaturas ambientales diferentes, y con dos termógrafos, sin embargo, el análisis se hará para un solo termógrafo. El archivo correspondiente se anexa. Adicionalmente, se cuenta con dos archivos PDF donde se aclara la convención del nombre de las variables. Asimismo, se tiene una figura en la que se ilustra las zonas donde se midió la temperatura facial con los termógrafos.

Se solicita realizar los siguiente con la base de datos:

### **Inspección del archivo y carga en Python**

1. Realizar una inspección de la base de datos (abrir el archivo en Excel) para identificar cómo está organizada la base de datos, para esto identifique las variables con ayuda de las convenciones aportados por lo autores del estudio.
2. Se deberá cargar el archivo en Python como un DataFrame con ayuda de la librería Panda.

# UdeA

## Análisis Exploratorio de Datos

Se realizará el análisis de las siguientes variables, las demás variables serán ignoradas en el análisis:

- aveOralM
- Gender
- Age
- Ethnicity
- T\_atm
- Humidity
- Cosmetics
- 

Las siguiente variable fue tomada en 4 ocasiones, promédielas para obtener solo 1 variable promedio.

- Max1R13\_

1. Construya un DataFrame de la siguiente forma: (Para las variables categóricas llene el valor con nan, excepto la moda) Imprima el DataFrame en Consola **(10%)**

	aveOralM	Gender	Age	Ethnicity	T_atm	...
Media						
Mediana						
Moda						
SD						
MAD						
Varianza						
IQR						
CV						
CVM						

Analice el comportamiento de las métricas calculadas. Compare en cada caso la SD y la MAD, ¿son parecidas? ¿Espera que haya datos atípicos según ese resultado? Analice si los datos son muy dispersos o si no lo son según los coeficientes de variación. **(15%)**

2. Determine usando el método del rango intercuartil, si existen datos atípicos y guárdelos en una variable. Imprímalos en consola **(5%)**

# UdeA

3. Realice un histograma para cada variable y analícelo, hable acerca de la distribución de los datos, ¿es simétrica? ¿Parece normal? **(5%)**
4. Realice gráficos de cajas y bigotes para las variables numéricas, analice los gráficos. **(5%)**
5. Realice gráficos de barras para las variables categóricas. Analice **(5%)**
6. Analice si existe correlaciones entre las variables analizadas, se sugiere construir una matriz de correlación. Analice los resultados, ¿alguna de estas variables podría ser usada para describir otra? **(5%)** (en este análisis debe omitir las variables categóricas).

## Muestreo de los datos.

Asuma que los datos del termógrafo son una población completa y no una muestra.

1. Realice un muestreo simple aleatorio a la población con  $n=500$ . Y guarde la muestra en un DataFrame, calcule las métricas del primer punto, ¿difieren mucho? Explique (Imprímalo en consola) **(7%)**
2. Basado en el análisis exploratorio, determine cuál podría ser una forma de estratificación y realice un muestreo estratificado para obtener una muestra de  $n=500$ . Explique el procedimiento y utilice gráficos para ilustrar lo que hizo. Además, calcule las métricas del primer punto ¿difieren mucho? Explique. **(8%)**

## Entregables

1. Código funcional y completo sin errores **(15%)**. Puede entregar un archivo .py o un .ipynb. En el caso de entregar un archivo .py utilice secciones para separar los bloques de código (`#%`). En ambos casos el código debe estar comentado. En el caso del .ipynb el informe del que habla el siguiente punto, debe estar en el mismo notebook, no aparte.
2. En caso de entregar un .py, hacer un informe escrito en PDF donde se analice cada punto y se dé respuesta a las preguntas hechas. **(20%)**