

# Actividad de Seguimiento 2

## Contexto

Los datos anexos sirven como un recurso valioso para explorar las complejas dinámicas de la salud del corazón y sus predictores. Los ataques cardíacos, o infartos de miocardio, siguen siendo un problema importante de salud a nivel mundial, lo que hace necesario comprender mejor sus precursores y posibles factores de mitigación. Este conjunto de datos incluye diferentes variables, como la edad, los niveles de colesterol, la presión arterial, los hábitos de tabaquismo, los patrones de ejercicio, las preferencias dietéticas y más, con el objetivo de esclarecer la compleja interacción de estas variables en la determinación de la probabilidad de un ataque cardíaco.

Mediante el uso de analítica predictiva y aprendizaje automático en este conjunto de datos, investigadores y profesionales de la salud pueden trabajar en el desarrollo de estrategias proactivas para la prevención y el manejo de las enfermedades cardíacas. El conjunto de datos se erige como un testimonio de los esfuerzos colectivos por mejorar nuestra comprensión de la salud cardiovascular y allanar el camino hacia un futuro más saludable.

El conjunto de datos, compuesto por **8.763 registros de pacientes de todo el mundo**, culmina en una característica de **clasificación binaria** que indica la presencia o ausencia de ataque cardíaco, proporcionando así un recurso integral para el análisis predictivo y la investigación en salud cardiovascular.

Variables:

- Patient ID – Identificador único de cada paciente
- Age – Edad del paciente
- Sex – Género del paciente (Masculino/Femenino)
- Cholesterol – Niveles de colesterol del paciente
- Blood Pressure – Presión arterial del paciente (sistólica/diastólica)
- Heart Rate – Frecuencia cardíaca del paciente
- Diabetes – Si el paciente tiene diabetes (Sí/No)
- Family History – Antecedentes familiares de problemas cardíacos (1: Sí, 0: No)
- Smoking – Estado de tabaquismo del paciente (1: Fumador, 0: No fumador)
- Obesity – Estado de obesidad del paciente (1: Obeso, 0: No obeso)
- Alcohol Consumption – Nivel de consumo de alcohol del paciente (Nulo/Leve/Moderado/Alto)
- Exercise Hours Per Week – Número de horas de ejercicio por semana
- Diet – Hábitos alimenticios del paciente (Saludable/Promedio/Poco saludable)
- Previous Heart Problems – Problemas cardíacos previos del paciente (1: Sí, 0: No)
- Medication Use – Uso de medicamentos por parte del paciente (1: Sí, 0: No)

- Stress Level – Nivel de estrés reportado por el paciente (1-10)
- Sedentary Hours Per Day – Horas de actividad sedentaria por día
- Income – Nivel de ingresos del paciente
- BMI – Índice de Masa Corporal (IMC) del paciente
- Triglycerides – Niveles de triglicéridos del paciente
- Physical Activity Days Per Week – Días de actividad física por semana
- Sleep Hours Per Day – Horas de sueño por día
- Country – País del paciente
- Continent – Continente donde reside el paciente
- Hemisphere – Hemisferio donde reside el paciente
- Heart Attack Risk – Ataque cardiaco (1: Sí, 0: No)

## Actividad

Responda a las siguientes preguntas. Para esto deberá hacer uso de pruebas de hipótesis. En todos los casos se deberá plantear las hipótesis a verificar, escrita en términos del contexto del análisis. Mencione el Valor\_p obtenido y su interpretación. Previamente corrobore los supuestos necesarios, cuando no sea posible alcanzar los supuestos, recuerde que debe intentar transformaciones y cuando no se posible de ninguna forma lograr los supuestos, se hará uso de una prueba no paramétrica.

1. ¿El valor promedio de colesterol de la muestra, correspondiente a pacientes que sufrieron infarto, es menor 200mg/dL? Y en los pacientes que no sufrieron infarto, ¿es menor 200mg/dL?
2. ¿Hay diferencias entre el nivel de colesterol entre pacientes con menos de 50 años y más de 50 años de edad?
3. ¿Hay diferencia en el tiempo que dedican a hacer ejercicio entre las personas que sufrieron infarto y las que no?
4. ¿Hay diferencia en el nivel de triglicéridos entre las personas entre las personas obesas y las que no lo son?
5. Se sabe que aproximadamente el 30% de las personas con diabetes tienen enfermedades cardiacas a nivel mundial. Las personas del estudio que tuvieron ataques cardiacos guardan la misma proporción de la población mundial respecto de la prevalencia de diabetes?

## Entregable

Deberá entregar un archivo tipo notebook (.ipybn) con el desarrollo de la actividad. **Sea muy organizado**, haga uso de celdas de texto y de código de forma separada para mostrar paso a paso lo que hace y lo que significa cada resultado para finalmente llegar a la respuesta de cada pregunta.

## **Rúbrica**

Cada numeral tiene igual valor. En cada uno se tendrá en cuenta lo siguiente, y cada cosa tendrá el mismo peso.

- Menciona cual prueba o pruebas paramétricas podría usar para dar respuesta a la pregunta
- Plantea las hipótesis que quiere corroborar para dar respuesta a la pregunta
- Corroborar los supuestos de la prueba paramétrica seleccionada. Implica plantear las hipótesis de las pruebas con las que corrobora supuesta, e interpretar cada resultado.
- Se realizan transformaciones a las variables de ser necesario
- Se selecciona una prueba no paramétrica de no ser posible usar una paramétrica
- Se aplica la prueba seleccionada, y se interpreta y responde a la pregunta hecha.