

Facultad de Ingeniería  
Departamento de Ingeniería Industrial  
Probabilidad e Inferencia Estadística  
Trabajo II

Nombre: \_\_\_\_\_ Cédula: \_\_\_\_\_

Nombre: \_\_\_\_\_ Cédula: \_\_\_\_\_

Antes de iniciar con el desarrollo del trabajo **LEA ATENTAMENTE:**

- La calificación de cada punto desarrollado estará conformado por tres puntos:
    1. El planteamiento del ejercicio, que abarca la complejidad en la creación del enunciado, el correcto planteamiento del juego de hipótesis.
    2. El desarrollo del ejercicio planteado junto a los códigos, en donde se debe evidenciar la correcta evaluación de las condiciones que conducen al empleo de uno u otro estadístico.
    3. Las interpretaciones resultantes en el contexto de los datos y del enunciado planteado.
  - Si se presenta una situación en donde al momento de la calificación, surge alguna duda o inconsistencia con los códigos usados durante el desarrollo del trabajo, el docente solicitará sustentación de los mismos por parte de los estudiantes que generaron el código y desarrollo del punto >:c
  - Enviar al correo **jivan.perez@udea.edu.co**, los archivos correspondientes a los enunciados, códigos e interpretaciones, una vez finalice con su desarrollo.
  - Si el trabajo lo desarrollan en parejas, recuerden que deben juntas la base de datos de cada uno, ya sea en R o en Excel.
1. **(5 puntos)** El Ministerio de Salud de Colombia ha decidido realizar un estudio epidemiológico a nivel nacional para evaluar la prevalencia de enfermedades cardiovasculares en la población, y por ello, ha decidido recopilar la información de una muestra representativa personas de diferentes regiones, edades, género y condiciones socioeconómicas.

La base de datos contiene los siguientes campos:

- **Region:** Región del paciente, donde 'Andina', 'Caribe', 'Pacífica', 'Orinoquía' y 'Amazonía', representan la región de Colombia donde vive el paciente.
- **Edad:** Edad del paciente en años.
- **Altura:** Altura del paciente en centímetros.
- **Peso:** Peso del paciente en kilogramos.
- **Genero:** Género del paciente, donde 'Hombre' representa un paciente de género masculino y 'Mujer' representa un paciente de género femenino.
- **Niv\_Soc:** Nivel socioeconómico del paciente, donde 'Bajo', 'Medio' y 'Alto' son los niveles bajo los cuales se cataloga la capacidad adquisitiva y estado de la vivienda donde vive el paciente.
- **Pre\_S:** Presión arterial sistólica del paciente en mmHg.
- **Pre\_D:** Presión arterial diastólica del paciente en mmHg.
- **Col:** Nivel de colesterol del paciente, donde 'normal' representa un nivel normal de colesterol, 'por encima de lo normal' representa un nivel de colesterol por encima de lo normal y 'muy por encima de lo normal' representa un nivel de colesterol muy por encima de lo normal.
- **Col\_sangre:** Nivel de colesterol del paciente en la sangre en mg/dL.

- **Glu:** Nivel de glucosa del paciente, donde ‘normal’ representa un nivel normal de glucosa, ‘por encima de lo normal’ representa un nivel de glucosa por encima de lo normal y ‘muy por encima de lo normal’ representa un nivel de glucosa muy por encima de lo normal.
- **Az\_sangre:** Nivel de glucosa del paciente en forma numérica en mg/dL.
- **Fumador:** Si el paciente es fumador o no, donde ‘No’ representa que el paciente no es fumador y ‘Sí’ representa que el paciente es fumador.
- **Con\_Alc:** Si el paciente consume alcohol o no, donde ‘No’ representa que el paciente no consume alcohol y ‘Sí’ representa que el paciente consume alcohol.
- **Enf\_Car:** Si el paciente tiene enfermedad cardiovascular o no, donde ‘No’ representa que el paciente no tiene enfermedad cardiovascular y ‘Sí’ representa que el paciente tiene enfermedad cardiovascular.
- **IMC:** Índice de masa corporal del paciente, calculado como el peso en kilogramos dividido por el cuadrado de la altura en metros.
- **Niv\_Estres:** Nivel de estrés del paciente, donde ‘Bajo’ representa un nivel bajo de estrés, ‘Moderado’ representa un nivel moderado de estrés y ‘Alto’ representa un nivel alto de estrés.
- **Con\_Tabaco:** Frecuencia de consumo de tabaco del paciente, donde ‘Nunca’ representa que el paciente nunca consume tabaco, ‘Ocasional’ representa que el paciente consume tabaco ocasionalmente y ‘Diario’ representa que el paciente consume tabaco a diario.
- **Dieta:** Tipo de dieta del paciente, donde ‘Saludable’ representa una dieta saludable y ‘No saludable’ representa una dieta no saludable.
- **Act\_Social:** Nivel de actividad social del paciente, donde ‘Baja’ representa un nivel bajo de actividad social, ‘Moderada’ representa un nivel moderado de actividad social y ‘Alta’ representa un nivel alto de actividad social.
- **Niv\_Act:** Nivel de actividad física del paciente, donde ‘Bajo’ representa un nivel bajo de actividad física, ‘Moderado’ representa un nivel moderado de actividad física y ‘Alto’ representa un nivel alto de actividad física.
- **Con\_Sal:** Cantidad en mg de sal del paciente.
- **Con\_Azucar:** Cantidad en mg de consumo de azúcar del paciente.
- **Frec\_Card:** Frecuencia cardíaca del paciente en latidos por minuto.
- **Con\_FV:** Número de veces que consume de frutas y verduras al día.
- **Horas\_Sueño:** Horas de sueño del paciente al día.
- **Horas\_Ejercicio:** Horas de ejercicio que realiza el paciente al día.
- **Con\_Agua:** Consumo de agua del paciente en litros diarios.

El objetivo del estudio es analizar los factores de riesgo asociados con las enfermedades cardiovasculares, como la edad, el género, el estilo de vida (fumador, obesidad), los niveles de presión arterial, colesterol y diabetes. Además, se busca evaluar la prevalencia de estas enfermedades en diferentes regiones y grupos socioeconómicos.

Basado en la información contenida en la base de datos suministrada,

**NOTA: DEBE APLICAR AL MENOS UN FILTRO DIFERENTE EN CADA PUNTO >:c**

**NOTA: SI, DIJE FILTRO DIFERENTE, PARA QUE NO HAGAN TODOS LOS PUNTOS CON EL MISMO >:c**

- a) **(1 punto)** Plantee una prueba de hipótesis del 11 % para la diferencia de dos proporciones.
- b) **(1 punto)** Plantee una prueba de hipótesis para diferencia de medias empleando un nivel de significancia del 14 %.
- c) **(0.5 puntos)** Plantee una prueba de hipótesis para una media empleando un nivel de significancia del 11 % .

- d)* **(0.5 puntos)** Basado en el inciso c), plantee una hipótesis alternativa específica para calcular el Error Tipo II.
- e)* **(1 punto)** Plantee una prueba de bondad de ajuste para una variable, y pruebe 3 distribuciones diferentes. Use un nivel de significancia del 11 %.
- f)* **(1 punto)** Plantee una prueba de hipótesis del 14 % para razón de varianzas, en donde el objetivo es saber si hay o no diferencias significativas entre las varianzas.